Vector-valued concentration inequalities on discrete spaces

Miriam Gordin

A Dissertation
Presented to the Faculty
of Princeton University
in Candidacy for the Degree
of Doctor of Philosophy

Recommended for Acceptance
by the Program in
Applied and Computational Mathematics
Adviser: Ramon van Handel

September 2025

Abstract

Existing concentration inequalities for functions that take values in a general Banach space, such as the classical results of Pisier (1986) and the more recent contribution of Ivanisvili, van Handel, and Volberg (2020), are known only in very special settings, such as the Gaussian measure on $\mathbb{R}^n$ and the uniform measure on the discrete cube $\{-1, 1\}^n$.

In this thesis, we prove such vector-valued concentration inequalities for more general probability measures on discrete spaces. In the first chapter, we present such an inequality for the biased product measure on the discrete cube with an optimal dependence on the bias parameter and the Rademacher type of the target Banach space. This result yields scaling limits to the product of Poisson measures, as well as lower bounds on the average distortion of embeddings of the discrete cube into Banach spaces of nontrivial type, implying average nonembeddability.

Moreover, we present a novel vector-valued concentration inequality for the uniform measure on the symmetric group, which is the first to go beyond the setting of product measures. Our inequality attains optimal dimensional dependency for Banach space of Rademacher type $p \in [1, 2)$, which implies average nonembeddability of the symmetric group into Banach spaces of nontrivial Rademacher type. Our approach enriches the Markov semigroup interpolation argument used by Ivanisvili, van Handel, and Volberg. In particular, we further techniques for capturing the concentration of random coefficients arising from the semigroup method.

Acknowledgements

Thank you to my advisor, Ramon van Handel, for his guidance in completing the research in this thesis and in general during my years in the PhD. It has been an inspiration to witness his breadth of mathematical knowledge and deep intuition. His discerning taste in problems and high standards for writing will continue to shape me as a mathematician for years to come. I am deeply grateful for all he has taught me.

Thank you to Assaf Naor and Grigoris Paouris for serving in my thesis committee. I am grateful to Alexandros Eskenazis for undertaking the duty of serving as a reader, as well as for many interesting and fruitful mathematical discussions, supportive guidance, and for being a generous host and organizer.

Many thanks to Evita Nestoridi for all you have taught me, our fascinating mathematical conversations, for hosting me numerous times at Stony Brook, and for your exemplary support as a mentor and friend. Thank you also to Dan Mikulincer, Max Fathi, Allan Sly, and Peter Sarnak for your support and interesting mathematical discussions during my time in graduate school.

The support and relentless kindness and efficiency of the PACM staff was indispensable in facilitating my time at Princeton: in particular, I would like to thank Audrey Mainzer, Bernadeta Wysocka, Victoria Beltra, and Katherine Lamos. I am also grateful to Jill LeClair in the Mathematics Department. Finally, I am grateful to the NSF for its financial support.

I would like to thank many fellow graduate students and postdocs for enriching my time at Princeton: Laura Shou, Katy Woo, Cosmas Kravaris, Kunal Chawla, Alan Chang, Seung-Yeon Ryoo, Eyob Tsegaye, Sergio Cristancho Sanchez, Anna Skorobogatova, Graeme Baker, and Hezekiah Grayer.

Completing the PhD would have been immeasurably more challenging and less rewarding without the support of many dear friends: thank you to everyone in the Princeton tango community, in particular, Danielle, Barbara, Suying, Juan, Alena, Angela, and Elif. I am grateful to Iris Hauser whose patience and thoughtfulness as a teacher will inspire me forever, both in my own learning and in pedagogy. Thank you to Anaïs, Felipe, Carolyn, Derek, Hugo, Suren, Seba, Tere, Michael, Tyler, Karl, Sharada, Eu, Gio, and Ollie for many lovely shared moments. Thank you to Álvaro and Ale for your relentless optimism and always making me feel at home. I am especially grateful to Cali for being the best writing companion I could have dreamed of for the writing of Chapter 3. I am thankful to Miloš and Amelia for their deep understanding and support in the moments when I needed it most. Thank you to Jackie, Sean, and Sophia for always being there for me through and through; your generous care will always be an inspiration to me and I will continue to carry forward your values as teachers. Minsoo and Ying, thank you for always cheering me on and believing in me; I am grateful that our friendship continues to overcome space and time. Tanya, I could not imagine going through the PhD without you, thank you for always being there for me as a mathematical sister and dear friend. Edi, you were there through it all, thank you for your unwavering belief in me and all of your encouragement, support, and love.

And to my parents, for everything.

Contents

CHAPTER 1

Introduction

Functional inequalities capturing concentration phenomena have important consequences in algorithmic and statistical settings, as well as problems in analysis and geometry. They have been characterized extensively through scalar-valued functional inequalities, such as the classical Poincaré [**25**] or log-Sobolev inequalities [**11**]; less is known for *vector-valued* functions. Decades of work have focused on proving vector-valued inequalities for target Banach spaces satisfying certain special conditions, such as UMD Banach spaces [**22**].

There are few known vector-valued concentration inequalities for functions $f : \mathbb{R}^n \to X$ which hold for any arbitrary Banach space $(X, \|\cdot\|)$. The classical work of Pisier [**24**] introduced the two main known examples of vector-valued concentration inequalities for target Banach spaces with no restrictions: one for the standard Gaussian measure in $\mathbb{R}^n$ and a second for the uniform measure on the discrete cube $\{-1, 1\}^n$ [**24**]. The former theorem states that for any locally Lipschitz function $f : \mathbb{R}^n \to X$, $G, G' \sim N(0, I_n)$ independent standard Gaussian random vectors in $\mathbb{R}^n$, and $1 \leq p < \infty$,

$$\mathbb{E}\,\|f(G) - \mathbb{E}f(G)\|^p \leq \left(\frac{\pi}{2}\right)^p \mathbb{E}\left\|\sum_{j=1}^n G'_j \frac{\partial f}{\partial x_j}(G)\right\|^p .$$

This Sobolev-type inequality captures a general phenomenon inherent to many high-dimensional settings – the worst-case fluctuations of a (potentially nonlinear) random quantity are captured by a linear quantity. Notably, the constant on the right-hand side of the inequality does not depend on n, meaning that this phenomenon is characterized *independently* of the dimension n. Pisier's elegant proof relied heavily on special properties of the standard Gaussian measure on $\mathbb{R}^n$, which makes it challenging to obtain similar dimension-free results for other measures.

In particular, the same paper [**24**] contains the proof of an analogous inequality on the discrete cube $\{-1,1\}^n$, stating that for any Banach space $(X, \|\cdot\|)$, a function $f : \{-1,1\}^n \to X$, and ε, δ independent random vectors uniformly distributed on the discrete cube $\{-1,1\}^n$, and $p \geq 1$,

$$\mathbb{E}\|f(\varepsilon) - \mathbb{E}f(\varepsilon)\|^p \leq C(n)^p \mathbb{E}\left\|\sum_{j=1}^n \delta_j D_j f(\varepsilon)\right\|^p, \tag{1}$$

where

$$D_i f(x) = f(x) - f(x_1, \ldots, -x_i, \ldots, x_n).$$

The theorem holds with a constant $C(n) \sim \log n$; Talagrand showed in [**28**] this is sharp.

If (1) were true with a dimension-free constant, this would settle a classical conjecture of Enflo in Banach space geometry which concerns the equivalence of Enflo and Rademacher type [**8**]. Enflo type is a nonlinear metric invariant of a Banach space given by the following condition: we say a Banach space has Enflo type $p \geq 0$ if there exists a $C \in (0,\infty)$ such that for any function $f : \{-1,1\}^n \to (X, \|\cdot\|)$,

$$\mathbb{E}\,\|f(\varepsilon) - f(-\varepsilon)\| \leq C^p \sum_{j=1}^n \mathbb{E}\|D_j f(\varepsilon)\|^p. \tag{2}$$

In contrast, the notion of Rademacher type describes the linear structure of the Banach space. We say that Banach space $(X, \|\cdot\|)$ has Rademacher type $p \in [1,2]$ if there exists a $C \in (0,\infty)$ so that for all $n \geq 1$ and $x_1, \ldots, x_n \in X$, and $\varepsilon_1, \ldots, \varepsilon_n$ i.i.d. symmetric Rademacher random variables,

$$\mathbb{E}\left\|\sum_{j=1}^n \varepsilon_j x_j\right\|^p \leq C^p \sum_{j=1}^n \|x_j\|^p. \tag{3}$$

We use $T_p(X)$ to denote the smallest possible constant in the inequality. By choosing the function $f(x) = \sum_{i=1}^n \varepsilon_i x_i$ in (2), one readily obtains the property (3). The opposite direction – to go from a linear to nonlinear property – requires a deeper insight.

After more than 40 years, the conjecture of Enflo was settled in the affirmative in [**13**]. This result is a consequence of a Pisier-like inequality with a key difference: rather than considering random coefficients δ which have the same distribution as the uniformly distributed random vectors ε on the discrete cube, the authors introduce *biased* random coefficients $\delta(t)$, which arise naturally from a Markov semigroup interpolation argument. More precisely, they show that for any function $f : \{-1,1\}^n \to X$, $\varepsilon \sim \mathrm{Unif}(\{-1,1\}^n)$, and $p \geq 1$,

$$(\mathbb{E}\|f(\varepsilon) - \mathbb{E}f(\varepsilon)\|^p)^{\frac{1}{p}} \leq \int_0^\infty \left(\mathbb{E}\left\|\sum_{j=1}^n \delta_j(t) D_j f(\varepsilon)\right\|^p\right)^{\frac{1}{p}} dt \tag{4}$$

where

$$\delta_j(t) := \frac{\xi_j(t) - e^{-t}}{e^t - e^{-t}}$$

and $\{\xi_j(t)\}_{j=1}^n$ are i.i.d. biased Rademacher random variables, independent of ε, with

$$\mathbb{P}(\xi_j(t) = 1) = \frac{1+e^{-t}}{2} \quad \text{and } \mathbb{P}(\xi_j(t) = -1) = \frac{1-e^{-t}}{2}.$$

(We modify the notation of their result to correspond to the notation in this exposition.) In particular, for a Rademacher space of type $p \in [1,2]$, (4) implies that for $f : \{-1,1\}^n \to X$, and $\varepsilon \sim \mathrm{Unif}(\{-1,1\}^n)$,

$$\mathbb{E}\|f(\varepsilon) - \mathbb{E}f(\varepsilon)\|^p \leq C^p \sum_{j=1}^n \mathbb{E}\|D_j f(\varepsilon)\|^p, \tag{5}$$

where C does not depend on n. In contrast, the original inequality of Pisier (1) yields $C \sim \log n$. Thus, the approach of [**13**] introducing appropriately biased random coefficients in the inequality of Pisier yields a functional inequality (4) that yields a vector-valued Poincaré inequality (5) with dimension-free constant.

In fact the constants of such generalized Poincaré inequalities are notable in their consequences for the metric geometry of Banach spaces. Just as the constants in scalar-valued Poincaré inequalities capture geometric properties of the underlying measure space, the study of non-linear spectral gaps, i.e. generalizations for the classical spectral gap for other geometries than the Euclidean geometry, captures geometric properties in these more general settings. The work of Mendel and Naor [**19**] introduced the systematic study of on nonlinear spectral gaps in particular with regard to consequences for *metric embeddings*. Thus, proving vector-valued analogs of Poincaré inequalities yields estimates on the distortion of such embeddings of metric spaces of Banach spaces. In particular, vector-valued Poincaré inequalities, such as those presented in this thesis, yield statements of *average* nonembeddability which are much stronger than worst-case embeddability. For example, if an n-point metric space embeds into a Banach space with constant distortion, one could "plant" a set of $\log n$ points which does not embed well. Then the worst-case distortion will pick up the planted set immediately, whereas on average, when choosing two points from the space uniformly on average, one is unlikely to choose two of the bad points.

The inequality (4) sets the stage for the results presented in this thesis, which enrich our understanding of the semigroup technique for proving Pisier-like inequalities in other settings than the uniform measure on the discrete cube. In Chapter 2, we introduce the preliminaries for the general framework of proving inequalities in the spirit of (4) on a general weighted graph, as well as making precise the notions of metric embeddings which we will study in this thesis. In Chapter 3, we present a Pisier-like inequality for the biased product measure on the discrete cube, which generalizes the inequality (4) of [**13**] to a wider class of product measures on the discrete cube. In particular, allowing the bias parameter to decay yields scaling limits which give useful intuition about the behavior of the random coefficients $\delta(t)$. In Chapter 4, we present a vector-valued Poincaré inequality on the symmetric group and on the so-called k-slice of the discrete cube. We obtain optimal dimensional constants in the latter case, and for the symmetric group for all Banach spaces p of type $[1, 2)$. Furthermore, for $p \in [1, 2]$, we obtain optimal constants up to a factor of $\log n$ which already yield novel embedding applications. These are the first results of this kind to go beyond the setting of product measures and requires us to introduce a variety of new techniques to the problem. Finally, in Chapter 5, we obtain Poisson-like concentration on the random coefficients $\delta(t)$ under a curvature-dimension condition on a discrete analog of Ricci curvature and discuss challenges in furthering our understanding of their concentration in order to obtain vector-valued concentration inequalities.

CHAPTER 2

General preliminaries

2.1. Markov semigroups on discrete spaces

Let G denote the complete graph on N vertices with associated edge weights $\lambda_{xy} \geq 0$ for $x, y \in G$ and $x \neq y$. Set $\lambda_{xx} = 0$ for convenience. We denote the discrete (weighted) Laplacian on G as

$$\Delta f(x) = -\sum_{y \in G} \lambda_{xy} D_y f(x) \tag{6}$$

where

$$D_y f(x) := f(x) - f(y).$$

Denote by $P_t := e^{t\Delta}$ the heat semigroup on G.

Let the weights $\lambda = \{\lambda_{xy}\}_{x,y \in G}$ be such that the Markov process $X = \{X_t\}_{t \geq 0}$ on G with transition rates given by λ is irreducible and aperiodic; thus, X converges to a unique stationary measure μ on G. Furthermore, suppose the transition rates satisfy the detailed balance condition

$$\mu(x)\lambda_{xy} = \mu(y)\lambda_{yx}.$$

Thus, X is a reversible Markov process on G.

Throughout this thesis, we will denote $\mathbb{E}_x[\,\cdot\,] := \mathbb{E}[\,\cdot \mid X_0 = x]$ and $\mathbb{E}_\mu[\,\cdot\,] := \mathbb{E}[\,\cdot \mid X_0 \sim \mu]$

The carré du champ (sometimes called Gamma operator) associated to P_t, applied to functions $f, g : G \to \mathbb{R}$, is given by

$$\Gamma(f,g)(x) = \frac{1}{2} \sum_{y \in G} \lambda_{xy} D_y f(x) D_y g(x) \tag{7}$$

for $x \in G$. In fact Δ is self-adjoint on G with Dirichlet form given by

$$\mathcal{E}(f,g) = \mathbb{E}_\mu[\Gamma(f,g)] = -\mathbb{E}_\mu[f \Delta g] \tag{8}$$

2.2. General framework for Pisier's inequality

Let p_t denote the heat kernel of P_t with respect to the stationary measure μ, i.e. for all $x, y \in G$,

$$p_t(x, y) := \frac{1}{\mu(y)} \mathbb{P}(X_t = y \mid X_0 = x), \tag{9}$$

such that

$$P_t f(x) = \sum_{y \in G} p_t(x, y)\mu(y) f(y).$$

Then we define random coefficients $\{\delta_y(t)\}_{y \in G}$ by

$$\delta_y(t) := \frac{p_t(X_0, X_t) - p_t(y, X_t)}{p_t(x, X_t)} = \frac{D_y p_t(X_0, X_t)}{p_t(X_0, X_t)}, \tag{10}$$

where the discrete derivative D_y is applied to the first coordinate of $p_t(\cdot, \cdot)$.

The key identity connecting the random coefficients $\delta(t)$ with discrete derivatives of the semigroup is given in the following lemma.

LEMMA 2.1. *For all $t \geq 0$ and $x, y \in G$,*

$$D_y P_t f(x) = \mathbb{E}_x[f(X_t)\delta_y(t)].$$

PROOF. By definition,

$$\begin{aligned}
D_y P_t f(x) &= P_t f(x) - P_t f(y) \\
&= \sum_{k \in G} (p_t(x, k) - p_t(y, k))\mu(k) f(k) \\
&= \sum_{k \in G} \frac{p_t(x, k) - p_t(y, k)}{p_t(x, k)} p_t(x, k)\mu(k) f(k) \\
&= \mathbb{E}_x[f(X_t)\delta_y(t)].
\end{aligned}$$

□

REMARK 2.2. Taking $f \equiv 1$ in Lemma 2.1, we have as an immediate consequence that

$$\mathbb{E}_x[\delta_y(t)] = 0$$

uniformly for $x \in G$ and $t \geq 0$, which will be a useful fact in later computations.

The following lemma generalizes the proof of (4) to the setting of a general weighted graph; the proof is almost identical to the one in [**13**].

LEMMA 2.3. *For any Banach space $(X, \|\cdot\|)$, function $f : G \to X$, and $1 \leq p < \infty$, we have*

$$(\mathbb{E}_\mu \|f - \mathbb{E}_\mu f\|^p)^{\frac{1}{p}} \leq \frac{1}{2} \int_0^\infty \left(\mathbb{E}_\mu \left\| \sum_{y \in G} \lambda_{X_0 y} \delta_y(t) D_y f(X_0) \right\|^p \right)^{\frac{1}{p}} dt.$$

PROOF. By Proposition 1.3.1 of [**12**],

$$(\mathbb{E}_\mu \|f - \mathbb{E}_\mu f\|^p)^{\frac{1}{p}} = \sup_{\mathbb{E}_\mu \|g\|^{p'} \leq 1} \mathbb{E}_\mu [\langle g, f - \mathbb{E}_\mu f \rangle]$$

where $\frac{1}{p} + \frac{1}{p'} = 1$. Fix any g such that $\mathbb{E}_\mu \|g\|^{p'} \leq 1$. By the fundamental theorem of calculus, and using the ergodicity of the Markov semigroup P_t, which implies that $P_0 f = f$ and $\lim_{t\to\infty} P_t f = \mathbb{E}_\mu f$, we obtain

$$\begin{aligned}
\mathbb{E}_\mu [\langle g, f - \mathbb{E}_\mu f \rangle] &= -\int_0^\infty \mathbb{E}_\mu \left[\langle g, \frac{d}{dt} P_t f \rangle \right] dt \\
&= -\int_0^\infty \mathbb{E}_\mu [\langle g, \Delta P_t f \rangle] \, dt \\
&= -\int_0^\infty \mathbb{E}_\mu [\langle P_t g, \Delta f \rangle] \, dt \\
&= \frac{1}{2} \int_0^\infty \sum_{y \in G} \mathbb{E}_\mu [\langle \lambda_{X_0 y} D_y P_t g(X_0), D_y f(X_0) \rangle] \, dt,
\end{aligned}$$

where in the third equality we have used the reversibility of the semigroup and the fact that Δ and P_t commute and in the last equality we have applied the representation of the Dirichlet form shown in (8) extended to X-valued functions. By Lemma 2.1, we have that

$$\begin{aligned}
\mathbb{E}_\mu [\langle \lambda_{X_0 y} D_y P_t g(X_0), D_y f(X_0) \rangle] &= \mathbb{E}_\mu [\langle \lambda_{X_0 y} \delta_y(t) g(X_t), D_y f(X_0) \rangle] \\
&= \mathbb{E}_\mu [\langle g(X_t), \lambda_{X_0 y} \delta_y(t) D_y f(X_0) \rangle] \, .
\end{aligned}$$

Therefore,

$$\mathbb{E}_\mu[\langle g, f-\mathbb{E}_\mu f\rangle] = \frac{1}{2}\int_0^\infty \mathbb{E}_\mu\left[\left\langle g(X_t), \sum_{y\in G}\lambda_{X_0 y}\delta_y(t)D_y f(X_0)\right\rangle\right] dt$$
$$\leq \frac{1}{2}\int_0^\infty \left(\mathbb{E}_\mu\|g(X_t)\|^{p'}\right)^{\frac{1}{p}'}\left(\mathbb{E}_\mu\left\|\sum_{y\in G}\lambda_{X_0 y}\delta_y(t)D_y f(X_0)\right\|^p\right)^{\frac{1}{p}} dt.$$

By the stationarity of P_t, X_t is equal in distribution to X_0 so by our assumption on g,

$$\left(\mathbb{E}_\mu\|g(X_t)\|^{p'}\right)^{\frac{1}{p}'} \leq 1$$

and we obtain the desired result. $\square$

2.3. Rademacher type and metric embeddings

We say that a function $f:(M,d)\to(X,\|\cdot\|)$ is a bi-Lipschitz embedding of a metric space (M,d) into a Banach space $(X,\|\cdot\|)$ if there exists a $D\geq 1$ and $s>0$ such that

$$sd(x,y)\leq \|f(x)-f(y)\|\leq Dsd(x,y). \tag{11}$$

We call D the distortion of f, i.e. $D=\|f\|_{\mathrm{Lip}}\cdot\|f^{-1}\|_{\mathrm{Lip}}$. In general, we denote as

$$c_X(M) := \inf\{D : \exists f:(M,d)\to(X,\|\cdot\|) \text{ with bi-Lipschitz distortion } D\}$$

the bi-Lipschitz distortion of M into X.

The bi-Lipschitz distortion quantifies the worst-case multiplicative distortion when embedding a metric space. However, in many cases, one cares only about distortion *on average* with respect to a probability measure. The study of average distortion was introduced by Rabinovich in [**26**] and has led to new insights in the Ribe program [**21**] and combinatorics [**16**]. Let ν be a probability measure on a finite metric space (M,d). We say a mapping $f:(M,d)\to(X,\|\cdot\|)$ has ν-average distortion $D\geq 1$ if f is D-Lipschitz and

$$\mathbb{E}_{\nu\otimes\nu}\|f(\varepsilon)-f(\varepsilon')\|\geq \mathbb{E}_{\nu\otimes\nu}|d(\varepsilon,\varepsilon')|.$$

Let

$$c_X^\nu(M) := \inf\{D : \exists f : (M,d) \to (X, \|\cdot\|) \text{ with } \nu\text{-average distortion } D\}.$$

LEMMA 2.4. *For any finite metric space (M,d), Banach space $(X,\|\cdot\|)$, and probability measure ν on M,*

$$c_X(M) \geq c_X^\nu(M).$$

PROOF. Suppose that a function f has bi-Lipschitz distortion D with scaling factor $s > 0$. Then the rescaled function $\frac{1}{s}f$ also achieves bi-Lipschitz distortion D and is D-Lipschitz, so without loss of generality, we will take f to be D-Lipschitz. Furthermore, we have that

$$\sum_{x,y\in M} \nu(x)\nu(y)\|f(x)-f(y)\| \geq \sum_{x,y\in M} \nu(x)\nu(y)d(x,y).$$

Thus, f also satisfies ν-average distortion D. □

CHAPTER 3

Vector-valued concentration on the biased discrete cube

This chapter is based on the work [**10**].

3.1. Introduction

One elegant consequence of the optimal dimensional dependence obtained in main inequality (4) of [**13**] for the uniform measure on the discrete cube is that Pisier's Gaussian inequality can be derived from it via the central limit theorem.

Thus, obtaining optimal constants for vector-valued inequalities on the high-dimensional spaces can yield important results on other spaces through appropriate scaling limits.

In this chapter, we present a Pisier-like concentration inequality for the *biased* product measure on the n-dimensional hypercube $\{-1,1\}^n$ with parameter $\alpha \in (0,1)$ given by

$$\mu = \bigotimes_{i=1}^{n} \mu_i$$

where

$$\mu_i(+1) = \alpha \qquad \text{and} \qquad \mu_i(-1) = 1-\alpha.$$

For the main result, we make no assumption on the target Banach space $(X, \|\cdot\|)$ in bounding the concentration of a function $f : \{-1,1\}^n \to (X, \|\cdot\|)$. Furthermore, we obtain an optimal dependence on the bias parameter and the Rademacher type of the Banach space by a special observation about the Dirichlet form on the biased cube. This result is an extension of the inequality (4) for the uniform measure on the hypercube of [**13**] which can be recovered by setting $\alpha = \frac{1}{2}$. If one takes $X = \mathbb{R}$ endowed with the Euclidean norm, we capture the same dependence on the bias as the scalar Poincaré inequality. Thus, in terms of the dependence of the bias, one loses nothing in going from the scalar to the vector-valued Poincaré inequality.

The constant also improves the scalar-valued Poincaré inequality on the biased hypercube in [**28**].

Our inequality on the biased hypercube allows us to prove a novel vector-valued concentration inequality on the space $\mathbb{N}^n$ endowed with the measure given by a product of Poisson distributions. The optimal dependence on the bias parameter α is essential to our proof technique, which obtains the inequality as a result of an appropriate scaling limit of the hypercube inequality.

As described in the introduction, one can derive vector-valued Poincaré inequalities of the form of (5) for spaces of Rademacher type $p \in [1, 2]$ as a consequence of Pisier-like inequalities. The corollary of the Pisier-like inequality presented in this chapter also achieves optimal dependence on α and n. In [**9**], Eskenazis independently obtained the same vector-valued Poincaré inequality. However, the form of Pisier-like inequality in [**9**] is not sufficient in order to obtain the scaling limit of the inequality in this chapter. See Section 3.2.2 for a more detailed comparison of the results.

An important motivation for proving such vector-valued functional inequalities is applications in metric geometry. The result (4) of [**13**] implies optimal lower bounds for the distortion of bi-Lipschitz embeddings from the discrete hypercube into an arbitrary Banach space of Rademacher type p. In [**9**], Eskenazis refines this for *finite-dimensional* target Banach spaces.

In fact, (4) has deeper consequences than this for metric embeddings, since it implies that the non-embeddability phenomenon occurs not only in the worst case but also on average with respect to the uniform measure on the hypercube. This poses the natural question of whether average non-embeddability holds for other measures on the hypercube such as the biased product measure. We show that non-embeddability occurs for all bias parameters α depending on n decaying slower than $\frac{1}{n}$.

3.1.1. Markov heat semigroup on the biased discrete cube. For a function $f : \{-1,1\}^n \to X$, we define the discrete derivative on the biased cube as

$$D_i^\alpha f(x) := f(x) - \mathbb{E}_i f(x),$$

where $\mathbb{E}_i f(x) = \mathbb{E} f(x_1, \dots, x_{i-1}, \varepsilon_i, x_{i+1}, \dots, x_n)$ where $\varepsilon_i \sim \mu_i$.

Remark 3.1. This definition is consistent with [**9**]. We can also see it as a mixture of two other notions of discrete derivative. Let

$$D_i f(x) := \frac{1}{2}(f(x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_n) - f(x_1, \dots, x_{i-1}, -x_i, x_{i+1}, \dots, x_n))$$

$$\partial_i f(x) := \frac{1}{2}(f(x_1, \dots, x_{i-1}, 1, x_{i+1}, \dots, x_n) - f(x_1, \dots, x_{i-1}, -1, x_{i+1}, \dots, x_n)).$$

Then we can express

$$D_i^\alpha f(\varepsilon) = (1-2\alpha)\partial_i f(\varepsilon) + D_i f(\varepsilon) = (1 - 2\alpha + \varepsilon_i)\partial_i f(\varepsilon). \tag{12}$$

Notice that $\partial_i f(\varepsilon)$ does not depend on the value of ε_i, so by the independence of the coordinates of ε, we have that the random quantities $\partial_i f(\varepsilon)$ and ε_i are in fact independent! This will be a key observation in our further analysis.

The generator of the random walk on the hypercube with stationary measure μ is given by

$$\Delta = -\sum_{i=1}^n D_i^\alpha.$$

Remark 3.2. We can express Δ in the general form of generator defined on a weighted graph in (6) by writing

$$\Delta f(x) = -\sum_{y \in G} \lambda_{xy} D_y f(x),$$

where $\lambda_{xy} = 0$ if x and y differ at more than one coordinate and otherwise, if x and y differ at coordinate i, $\lambda_{xy} = \frac{1}{2}(1 - 2\alpha + \varepsilon_i)$. To streamline the notation in this section, we index the discrete derivatives D_i by indices of the coordinates rather than by elements $y \in \{-1,1\}^n$.

Recall that the Dirichlet form associated to Δ and μ is given by

$$\mathcal{E}(f,g) := -\mathbb{E}_\mu[f(\varepsilon)\Delta g(\varepsilon)].$$

The following representation of the Dirichlet form is critical to obtaining the optimal dependence on α in the constant in Theorem 3.6.

Lemma 3.3. *The Dirichlet form associated with Δ is given by*

$$\mathcal{E}(f,g) = 4\alpha(1-\alpha)\sum_{i=1}^{n}\mathbb{E}_\mu[D_i f(\varepsilon)D_i g(\varepsilon)]$$

Proof. Notice that by the identity (12),

$$\mathbb{E}_\mu[f(\varepsilon)D_i^\alpha g(\varepsilon)] = \mathbb{E}_\mu[(1-2\alpha+\varepsilon_i)f(\varepsilon)\partial_i g(\varepsilon)]$$

Recalling that $\partial_i g(\varepsilon)$ is independent of ε_i, we compute that

$$\begin{aligned}\mathbb{E}_i[(1-2\alpha+\varepsilon_i)f(\varepsilon)\partial_i g(\varepsilon)] &= \alpha(2-2\alpha)f(1)\partial_i g(\varepsilon) + (1-\alpha)(-2\alpha)f(-1)\partial_i g(\varepsilon)\\ &= 4\alpha(1-\alpha)\partial_i f(\varepsilon)\partial_i g(\varepsilon).\end{aligned}$$

Finally, using the fact that $\varepsilon_i\partial_i f(\varepsilon) = D_i f(\varepsilon)$ and $\varepsilon_i^2 = 1$, we have that

$$\mathbb{E}_\mu[f(\varepsilon)D_i^\alpha g(\varepsilon)] = 4\alpha(1-\alpha)\mathbb{E}[D_i f(\varepsilon)D_i g(\varepsilon)],$$

so

$$-\mathbb{E}[f(\varepsilon)\Delta g(\varepsilon)] = \sum_{i=1}^{n}\mathbb{E}_\mu[f(\varepsilon)D_i^\alpha g(\varepsilon)] = 4\alpha(1-\alpha)\sum_{i=1}^{n}\mathbb{E}[D_i f(\varepsilon)D_i g(\varepsilon)].$$

□

We will consider the heat semigroup on the biased hypercube given by $P_t := e^{t\Delta}$. It will be useful to work with the corresponding continuous-time random walk on the biased hypercube denoted by $\{X(t)\}_{t\geq 0}$.

For the purpose of explict computations, we may construct the random walk for $n = 1$ as follows: let $Z_0, Z_1, Z_2, \ldots$ be an i.i.d. sequence of biased Rademacher variables such that $Z_i \sim \mu$. Let $\{N_t\}$ be a rate 1 Poisson process independent of the Z_is. Then the continuous-time random walk on the 1-dimensional biased hypercube with initial condition $X(0) = Z_0$ is given by

$$X(t) = Z_{N_t} \text{ for } t \geq 0.$$

Then for $n > 1$, the random walk is given simply by n one-dimensional random walks independently on each coordinate. It can be readily checked (e.g. see Section 2.3.2 of [**29**]) that for $t \geq 0$,

$$P_t f(x) = \mathbb{E}[f(X(t)) \mid X(0) = x].$$

We denote the heat kernel of the semigroup by $p_t(\cdot, \cdot)$ so that

$$\mathbb{P}(X(t) = y \mid X(0) = x) =: p_t(x, y) = \prod_{i=1}^{n} p_t(x_i, y_i),$$

where it will be clear from context whether we are considering the heat kernel on a single coordinate or on the n-dimensional discrete hypercube. Note that for this section, we define $p_t(x, \cdot)$ to be a probability measure, rather than considering the heat kernel with respect to the stationary measure as in (9).

We can write the heat kernel explicitly as

$$p_t(x_i, y_i) = \frac{1 - e^{-t}}{2}(2\alpha - 1)y_i + \frac{1 + e^{-t} x_i y_i}{2}.$$

Remark 3.4 (on notation). Unless otherwise specified, throughout the chapter ε will denote a μ-distributed random vector and all probabilities and expectations will be taken on the probability space on which the stationary random walk $\{X(t)\}_{t \geq 0}$ is defined. Furthermore, we will take $X(0) = \varepsilon$.

We define the random vector $\delta(t)$ by

$$\delta_i(t) := \frac{D_i\, p_t(\varepsilon, X(t))}{p_t(\varepsilon, X(t))},$$

where D_i is acting on the first coordinate of $p_t(\cdot,\cdot)$. We can explicitly write $\delta(t)$ as

$$\delta_i(t) = \frac{e^{-t}\varepsilon_i X_i(t)}{(1-e^{-t})(2\alpha-1)X_i(t) + 1 + e^{-t}\varepsilon_i X_i(t)}.$$

REMARK 3.5. In this work, we focus on the biased product measure which has the same bias α for each coordinate $i = 1, \ldots, n$. The methods of proof of the results on the biased cube extend without issue to considering a different bias α_i for each coordinate. For clarity of notation and due to the focus on obtaining scaling limits as α becomes small, we have not expressed the results in this generality.

3.2. Results

3.2.1. Pisier-like inequality on the biased hypercube. The main result is the following Pisier-like inequality on the biased hypercube, which is a generalization of the inequality (4) for uniform measure on the hypercube in [**13**]. Setting the bias $\alpha = \frac{1}{2}$ retrieves their result.

THEOREM 3.6. *For any Banach space $(X, \|\cdot\|)$, function $f : \{-1,1\}^n \to X$, and $p \geq 1$, we have*

$$(\mathbb{E}\|f(\varepsilon) - \mathbb{E}f(\varepsilon)\|^p)^{\frac{1}{p}} \leq \int_0^\infty 4\alpha(1-\alpha)\left(\mathbb{E}\left\|\sum_{i=1}^n \delta_i(t) D_i f(\varepsilon)\right\|^p\right)^{\frac{1}{p}} dt.$$

We have precise estimates on the coefficients $\delta(t)$. First we observe that the coefficients are independent and identically distributed across indices $i = 1, \ldots, n$. Furthermore, recall that by Remark 2.2, for any $t > 0$ and for any $x_i \in \{-1, 1\}$,

$$\mathbb{E}\left[\,\delta_i(t) \mid X(0) = x\right] = 0.$$

We have explicit bounds on the L^p norms of the $\delta_i(t)$, which are critical to obtain Corollary 3.8.

LEMMA 3.7. *For $\alpha < \frac{1}{2}$ and $p \geq 1$,*

$$\sup_{x \in \{-1,1\}^n} \mathbb{E}\Big[\big|\delta_i(t)\big|^p \mid X(0) = x\Big] \leq (2\alpha)^{1-p} e^{-tp}(1-e^{-t})^{1-p}.$$

Applying a standard symmetrization argument, we obtain the following vector-valued biased Poincaré inequality.

COROLLARY 3.8. *For any Banach space $(X, \|\cdot\|)$ of Rademacher type $p \in [1,2]$, function $f : \{-1,1\}^n \to X$, and $\alpha < \frac{1}{2}$, we have that*

$$(\mathbb{E}\|f(\varepsilon) - \mathbb{E}f(\varepsilon)\|^p)^{\frac{1}{p}} \leq 32 T_p(X) \alpha^{\frac{1}{p}} \left(\sum_{i=1}^n \mathbb{E}\|D_i f(\varepsilon)\|^p \right)^{\frac{1}{p}}.$$

REMARK 3.9. In order capture the optimal dependence on α, we impose the assumption that $\alpha < \frac{1}{2}$. By symmetry, the case $\alpha \geq \frac{1}{2}$ can be obtained by replacing α by $1-\alpha$.

Through a simple example, one can see that these inequalities are optimal in the sense of capturing the dependence on α and the type p of X.

EXAMPLE 3.10. Let $(X, \|\cdot\|) = (\mathbb{R}^n, \|\cdot\|_{\ell^p})$ and

$$f(\varepsilon) := n^{-\frac{1}{p}} \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}.$$

Suppose without loss of generality that $\alpha < \frac{1}{2}$. Then by the law of large numbers,

$$\|f(\varepsilon) - \mathbb{E}f(\varepsilon)\| = \left(\frac{1}{n} \sum_{i=1}^n |\varepsilon_i - \mathbb{E}\varepsilon_i|^p \right)^{\frac{1}{p}} \xrightarrow[n\to\infty]{} (\mathbb{E}|\varepsilon_1 - \mathbb{E}\varepsilon_1|^p)^{\frac{1}{p}} \sim \alpha^{\frac{1}{p}}.$$

Meanwhile, notice also that for this choice of f, we can bound the right-hand side of Corollary 3.8 by $\alpha^{\frac{1}{p}}$. Thus, we have shown that Theorem 3.6 and Corollary 3.8 attain the optimal dependence in terms of α and n, up to universal constants. In fact, even the inequality

$$\mathbb{E}\|f(\varepsilon)-\mathbb{E}f(\varepsilon)\| \leq \int_0^\infty 4\alpha(1-\alpha)\left(\mathbb{E}\left\|\sum_{i=1}^n \delta_i(t) D_i f(\varepsilon)\right\|^p\right)^{\frac{1}{p}} dt$$

which is weaker than that of Theorem 3.6 attains the optimal dependence on α, p, and n.

3.2.2. Comparison with inequality obtained in [9]. As noted in the introduction, in [**9**], Eskenazis obtains a vector-valued Poincaré inequality on the biased hypercube. In the notation of this paper, Theorem 14 of [**9**] states that for any function $f:\{-1,1\}^n \to (X,\|\cdot\|)$ where X is a Banach space of type $p \in [1,2]$,

$$\left(\mathbb{E}\|f(\varepsilon)-\mathbb{E}f(\varepsilon)\|^p\right)^{\frac{1}{p}} \leq 2\pi T_p(X)\left(\sum_{i=1}^n \mathbb{E}\|D_i^\alpha f(\varepsilon)\|^p\right)^{\frac{1}{p}}.$$

Recall by the identity (12) and the fact that $\partial_i f(\varepsilon)$ is independent of ε_i (see Remark 3.1),

$$\mathbb{E}\|D_i^\alpha f(\varepsilon)\|^p = \mathbb{E}[(1-2\alpha+\varepsilon_i)^p]\,\mathbb{E}\left\|\partial_i f(\varepsilon)\right\|^p.$$

As computed in the example above, when $\alpha < \frac{1}{2}$,

$$\mathbb{E}[(1-2\alpha+\varepsilon_i)^p] \sim \alpha.$$

Thus, up to universal constants, Theorem 14 of [**9**] and Corollary 3.8 are equivalent, but identity (12) is necessary in order to obtain the explicit dependence on the bias parameter.

As an intermediate step in the proof of Theorem 14 of [**9**], Eskenazis also states a Pisier-like inequality similar to that of Theorem 3.6; however, this result does not suffice in order to obtain the scaling limit to prove the Pisier inequality for the product of Poisson distributions present in Theorem 3.13.

3.2.3. Average non-embeddability of the biased discrete hypercube. Consider the discrete cube $\{-1,1\}^n$ with the Hamming metric which we will denote as $d(\cdot,\cdot)$, where for $x,y \in \{-1,1\}^n$,

$$d(x,y) = \sum_{i=1}^n \mathbb{1}_{\{x_i \neq y_i\}}$$

Corollary 3.8 implies the following lower bound on the μ-average distortion for embeddings of the hypercube into a Banach space of type p.

Corollary 3.11. *Let $(X, \|\cdot\|)$ be a Banach space of Rademacher type p and μ the α-biased product measure on $\{-1,1\}^n$ as defined above. Then for any $f : \{-1,1\}^n \to (X, \|\cdot\|)$ with f having μ-average distortion D,*

$$D \gtrsim_X (\alpha n)^{1-\frac{1}{p}},$$

where the result holds up to universal constants depending on $T_p(X)$.

Example 3.12. Let α_n be the bias parameter of the product measure μ as defined above such that $\lim_{n\to\infty} n\alpha_n \to \infty$. Suppose $f : \{-1,1\}^n \to (X, \|\cdot\|)$ has μ-average distortion D. Then by Corollary 3.11

$$D \gtrsim (n\alpha_n)^{1-\frac{1}{p}} \to \infty.$$

Thus, even for very sparse random vector distributions on the hypercube (in terms of the number of entries equal to 1), the distortion of D must still grow with n for any Banach space X of nontrivial type, i.e. $p > 1$.

In fact, this condition on α_n is optimal: take $(X, \|\cdot\|)$ to be $(\mathbb{R}^n, |\cdot|)$ and consider the identity map $f : \{-1,1\}^n \to \mathbb{R}^n$, which has Lipschitz constant 2. If $\alpha_n \sim \frac{1}{n}$, then

$$\mathbb{E}_{\mu\otimes\mu}|f(\varepsilon) - f(\varepsilon')| = 2\mathbb{E}_{\mu\otimes\mu}[d(\varepsilon,\varepsilon')^{\frac{1}{2}}] \geq 2\mathbb{E}_{\mu\otimes\mu}[\min\{d(\varepsilon,\varepsilon'),1\}].$$

Notice that $\min\{d(\varepsilon,\varepsilon'),1\} = 0$ with probability $(1-2\alpha_n(1-\alpha_n))^n$. Therefore,

$$\mathbb{E}_{\mu\otimes\mu}|f(\varepsilon) - f(\varepsilon')| \geq 2(1-(1-2\alpha_n(1-\alpha_n))^n) \sim 1$$

On the other hand, $\mathbb{E}[d(\varepsilon,\varepsilon')] = 2n\alpha_n(1-\alpha_n) \sim 1$. Therefore, when $\alpha_n \lesssim \frac{1}{n}$, constant average distortion can be achieved.

3.2.4. Pisier-like inequality for product Poisson measure. Let $N = \{N_i\}_{i=1,\dots,m}$ be a vector of i.i.d. Poisson(1) random variables.

THEOREM 3.13. *For any Banach space* $(X, \|\cdot\|)$*, bounded function* $f : \mathbb{N}^m \to X$*,* $1 \leq p < \infty$*, and ,*

$$(\mathbb{E}\|f(N) - \mathbb{E}f(N)\|^p)^{\frac{1}{p}} \leq \int_0^\infty \left(\mathbb{E}\left\|\sum_{i=1}^m \tilde{\eta}_i(t) D_i^{\mathbb{Z}} f(N)\right\|^p\right)^{\frac{1}{p}} dt,$$

where

$$D_i^{\mathbb{Z}} := f(x_1, \ldots, x_i + 1, \ldots, x_m) - f(x_1, \ldots, x_i, \ldots, x_m)$$

and

$$\tilde{\eta}_i(t) := e^{-t} - \frac{e^{-t}}{(1 - e^{-t})}\eta_i(t),$$

for $\eta_i(t)$ *independent* $\mathrm{Poisson}(1 - e^{-t})$ *random variables over all* $i = 1, \ldots, m$*, also independent of* N*.*

COROLLARY 3.14. *For any Banach space* $(X, \|\cdot\|)$ *of Rademacher type* $p \in [1, 2]$ *with optimal constant* $T_p(X)$*, bounded function* $f : \mathbb{N}^m \to X$*,*

$$(\mathbb{E}\|f(N) - \mathbb{E}f(N)\|^p)^{\frac{1}{p}} \leq 4\, T_p(X) \left(\sum_{i=1}^m \mathbb{E}\left\|D_i^{\mathbb{Z}} f(N)\right\|^p\right)^{\frac{1}{p}}.$$

3.3. Proofs

3.3.1. Proofs of Theorem 3.6 and Corollary 3.8. The following proof is almost identical to that of Lemma 2.3 except for the key step of using the Dirichlet form derived in Lemma 3.3.

PROOF OF THEOREM 3.6. By Proposition 1.3.1 of [**12**]

$$(\mathbb{E}\|f(\varepsilon) - \mathbb{E}f(\varepsilon)\|^p)^{\frac{1}{p}} = \sup_{\mathbb{E}\|g(\varepsilon)\|^{p'} \leq 1} \mathbb{E}[\langle g(\varepsilon), f(\varepsilon) - \mathbb{E}f(\varepsilon)\rangle]$$

where $\frac{1}{p} + \frac{1}{p'} = 1$.

By the fundamental theorem of calculus, and using the stationarity of the Markov semigroup P_t, which implies that $P_0 f = f$ and $\lim_{t\to\infty} P_t f = \mathbb{E}f(\varepsilon)$, we obtain

$$\begin{aligned}\mathbb{E}[\langle g(\varepsilon), f(\varepsilon) - \mathbb{E}f(\varepsilon)\rangle] &= -\int_0^\infty \mathbb{E}\left[\langle g(\varepsilon), \frac{d}{dt}P_t f(\varepsilon)\rangle\right] dt \\ &= -\int_0^\infty \mathbb{E}\left[\langle g(\varepsilon), \Delta P_t f(\varepsilon)\rangle\right] dt \\ &= \int_0^\infty \sum_{i=1}^n 4\alpha(1-\alpha)\mathbb{E}[\langle D_i g(\varepsilon), D_i P_t f(\varepsilon)\rangle]\, dt,\end{aligned}$$

where in the second to last line we have used the key representation of the Dirichlet form derived in Lemma 3.3. By Lemma 2.1, we have that

$$\mathbb{E}[\langle D_i g(\varepsilon), D_i P_t f(\varepsilon)\rangle] = \mathbb{E}[\langle g(X(t)), \delta_i(t) D_i f(\varepsilon)\rangle]\,.$$

Therefore,

$$\begin{aligned}\mathbb{E}[\langle g(\varepsilon), f(\varepsilon) - \mathbb{E}f(\varepsilon)\rangle] &= \int_0^\infty 4\alpha(1-\alpha)\mathbb{E}\left[\langle g(X(t)), \sum_{i=1}^n \delta_i(t) D_i f(\varepsilon)\rangle\right] dt \\ &\le \int_0^\infty 4\alpha(1-\alpha)\left(\mathbb{E}\|g(X(t))\|^{p'}\right)^{\frac{1}{p}'}\left(\mathbb{E}\left\|\sum_{i=1}^n \delta_i(t) D_i f(\varepsilon)\right\|^p\right)^{\frac{1}{p}} dt.\end{aligned}$$

By the reversibility of P_t, $X(t)$ is equal in distribution to ε so that we obtain:

$$(\mathbb{E}\|f(\varepsilon) - \mathbb{E}f(\varepsilon)\|^p)^{\frac{1}{p}} \le \int_0^\infty 4\alpha(1-\alpha)\left(\mathbb{E}\left\|\sum_{i=1}^n \delta_i(t) D_i f(\varepsilon)\right\|^p\right)^{\frac{1}{p}} dt.$$

□

The proof of Corollary 3.8 requires an estimate on the moments of $\delta(t)$ which is captured in Lemma 3.7. It is particularly important that we obtain an upper bound which is uniform over all initial conditions of the underlying Markov process. We postpone the proof of this lemma until after the proof of Corollary 3.8.

Proof of Corollary 3.8. We begin by performing a routine symmetrization argument on the expectation on the right hand side of Theorem 3.6. Recall that

$$\mathbb{E}\left[\,\delta_i(t) \mid X(0)=\varepsilon\right]=0.$$

Let $\zeta_1,\dots,\zeta_n$ be i.i.d. *unbiased* Rachemacher random variables independent of all other randomness. Then by Lemma 7.3 of [**29**], applied conditionally on ε, we have that

$$\mathbb{E}\left\|\sum_{i=1}^n \delta_i(t) D_i f(\varepsilon)\right\|^p \le 2^p \mathbb{E}\left\|\sum_{i=1}^n \zeta_i\delta_i(t) D_i f(\varepsilon)\right\|^p$$

Then by the definition of Rademacher type and our assumption on $(X,\|\cdot\|)$, applied conditionally on $\delta(t)$ and ε,

$$\mathbb{E}\left\|\sum_{i=1}^n \delta_i(t) D_i f(\varepsilon)\right\|^p \le T_p(X)^p 2^p \sum_{i=1}^n \mathbb{E}\|\delta_i(t) D_i f(\varepsilon)\|^p\,.$$

Now applying the uniform bound on $\mathbb{E}_x|\delta_i(t)|^p$ obtained in Lemma 3.7, conditionally on ε, we obtain

$$\mathbb{E}\|\delta_i(t) D_i f(\varepsilon)\|^p \le (2\alpha)^{1-p} e^{-tp}(1-e^{-t})^{1-p}\mathbb{E}\|D_i f(\varepsilon)\|^p$$

Now by Theorem 3.6, we have that

$$(\mathbb{E}\|f(\varepsilon)-\mathbb{E}f(\varepsilon)\|^p)^{\frac{1}{p}} \le T_p(X) 8\alpha(1-\alpha)\left(\sum_{i=1}^n \mathbb{E}\|D_i f(\varepsilon)\|^p\right)^{\frac{1}{p}} \int_0^\infty (2\alpha)^{\frac{1}{p}-1} e^{-t}(1-e^{-t})^{\frac{1}{p}-1}\,dt.$$

Evaluating the integral by a change of variables, we have that

$$\int_0^\infty (2\alpha)^{\frac{1}{p}-1} e^{-t}(1-e^{-t})^{\frac{1}{p}-1}\,dt = (2\alpha)^{\frac{1}{p}-1}\int_0^1 u^{\frac{1}{p}-1}\,du = p(2\alpha)^{\frac{1}{p}-1},$$

which yields the final result up to universal constants, using the assumption that $\alpha<\frac{1}{2}$ and the fact that $p\in[1,2]$. $\square$

Proof of Lemma 3.7. Fix any $x\in\{-1,1\}^n$. By definition of $\delta(t)$,

$$\mathbb{E}_x\left[\left|\delta_i(t)\right|^p\right] = \sum_{y\in\{-1,1\}}\left|\frac{e^{-t}}{2p_t(x_i,y)}\right|^p p_t(x_i,y).$$

Explicitly,

$$\mathbb{E}_x\left[\left|\delta_i(t)\right|^p\right] = \frac{e^{-tp}}{2}\big(|(1-2\alpha+x_i)e^{-t}+2\alpha|^{1-p} + |(2\alpha-1-x_i)e^{-t}+2-2\alpha|^{1-p}\big).$$

Denote

$$f_1(x_i,\alpha,t) := (1-2\alpha+x_i)e^{-t}+2\alpha$$

$$f_2(x_i,\alpha,t) := (2\alpha-1-x_i)e^{-t}+2-2\alpha.$$

Notice that by construction $f_1, f_2 \geq 0$, since they arise from the heat kernel p_t. Through a case-by-case analysis based on the initial condition x, we lower bound the terms f_1 and f_2 in order to obtain an upper bound on $\mathbb{E}|\delta_i(t)|^p$.

Case I: $x_i = -1$. We have that

$$f_2(-1,\alpha,t) = 2\alpha e^{-t}+2-2\alpha.$$

Therefore, using $\alpha < \frac{1}{2}$,

$$f_1(-1,\alpha,t) = 2\alpha(1-e^{-t}) \leq 2\alpha e^{-t}+2-2\alpha = f_2(-1,\alpha,t)$$

for all $t \geq 0$. Thus,

$$\mathbb{E}\left[\left|\delta_i(t)\right|^p \,\middle|\, X_i(0) = -1\right] \leq e^{-tp} f_1(-1,\alpha,t)^{1-p} = e^{-tp}(2\alpha)^{1-p}(1-e^{-t})^{1-p},$$

using that $2\max\{a,b\} \geq a+b$.

Case II: $x_i = +1$. We have that

$$f_1(1,\alpha,t) = (2-2\alpha)e^{-t}+2\alpha \geq 2\alpha(1-e^{-t}).$$

Also, using the assumption that $\alpha < \frac{1}{2}$, we have that

$$f_2(1,\alpha,t) = (2\alpha-2)e^{-t}+2-2\alpha = (2-2\alpha)(1-e^{-t}) \geq 2\alpha(1-e^{-t}).$$

Combining the analysis of the two cases, we obtain that

$$\mathbb{E}\left[\left|\delta_i(t)\right|^p \,\middle|\, X_i(0) = x_i\right] \leq e^{-tp}(2\alpha)^{1-p}(1 - e^{-t})^{1-p}$$

uniformly in the value of x. □

3.3.2. Proof of Corollary 3.11.

Proof of Corollary 3.11. Observe that since f is D-Lipschitz, we have that

$$\|D_i f\| \leq D.$$

Therefore, we have by Corollary 3.8 that

$$T_p(X)(\alpha n)^{\frac{1}{p}} D \gtrsim (\mathbb{E}\|f(\varepsilon) - \mathbb{E}f(\varepsilon)\|^p)^{\frac{1}{p}} .$$

By Jensen's inequality,

$$(\mathbb{E}\|f(\varepsilon) - \mathbb{E}f(\varepsilon)\|^p)^{\frac{1}{p}} \geq \mathbb{E}\|f(\varepsilon) - \mathbb{E}f(\varepsilon)\|.$$

Let ε' be an independent identically distributed copy of ε. Notice that

$$\mathbb{E}\|f(\varepsilon) - \mathbb{E}f(\varepsilon)\| = \frac{1}{2}\mathbb{E}\|f(\varepsilon) - \mathbb{E}f(\varepsilon)\| + \frac{1}{2}\mathbb{E}\|f(\varepsilon') - \mathbb{E}f(\varepsilon')\| \geq \frac{1}{2}\mathbb{E}\|f(\varepsilon) - f(\varepsilon')\|.$$

By the assumption that f has μ-average distortion D, we have that

$$\mathbb{E}\|f(\varepsilon) - f(\varepsilon')\| \geq \mathbb{E}|d(\varepsilon, \varepsilon')| \sim \alpha n.$$

□

3.3.3. Proofs of Theorem 3.13 and Corollary 3.14. The main idea of this section is to obtain the concentration inequality for Poisson random variables as the limit of appropriately scaled biased Rademachers. This would not be possible without the optimal dependence on α obtained in Theorem 3.6. We will work on the $m \times n$-dimensional discrete hypercube $\{-1, 1\}^{m \times n}$, where we fix m and send n to infinity. The following definitions are analogous to those given in Section 3.1.1, but with more convenient indexing for this setting.

We consider the biased product measure on $\{-1,1\}^{m\times n}$ with parameter $\alpha = \frac{1}{n}$,

$$\mu = \bigotimes_{i=1}^{m} \bigotimes_{j=1}^{n} \mu_{ij}$$

where

$$\mu_{ij}(+1) = \frac{1}{n} \qquad \text{and} \qquad \mu_{ij}(-1) = 1 - \frac{1}{n}.$$

We will consider the heat semigroup P_t associated with the random walk

$$X(t) = \{X_{ij}(t)\}_{i=1,\dots,m;\, j=1,\dots,n}$$

on the with bias $\alpha = \frac{1}{n}$.

We denote

$$D_{ij} f(x) := \frac{1}{2}(f(x) - f(x_{11}, \dots, x_{i-1\,j}, -x_{ij},\, x_{i+1\,j} \dots, x_{mn})),$$

where we take the difference with respect to a single coordinate x_{ij}. Let ε be a μ-distributed random vector and denote as $Y \in \{0,1\}^{m\times n}$ the rescaled version of ε given by

$$Y_{ij} := \frac{\varepsilon_{ij}+1}{2} \in \{0,1\}.$$

We will apply our results on the biased hypercube to a "structured" function g of the form $g : \{-1,1\}^{m\times n} \to X$ given by

$$g(\varepsilon_{11}, \dots, \varepsilon_{1n}, \dots, \varepsilon_{mn}) = f\left(\sum_{j=1}^{n} Y_{1j}, \sum_{j=1}^{n} Y_{2j}, \dots, \sum_{j=1}^{n} Y_{mj}\right), \tag{13}$$

for some function $f : \mathbb{N}^m \to X$. The following observation will be useful.

Lemma 3.15. *The discrete partial derivative of g can be written as*

$$D_{ij} g(\varepsilon) = \frac{1}{2}\left(f\left(\dots, \sum_{j'=1}^{n} Y_{ij'}, \dots\right) - f\left(\dots, -\varepsilon_{ij} + \sum_{j'=1}^{n} Y_{ij'}, \dots\right)\right).$$

Proof. Using the definition of g given in (13), we have that

$$D_{ij}g(\varepsilon) = \frac{1}{2}\left(f\left(\dots, \sum_{j'=1}^{n} Y_{ij'}, \dots \right) - f\left(\dots, \frac{-\varepsilon_{ij}+1}{2} + \sum_{\substack{j'=1\\ j'\neq j}}^{n} Y_{ij'}, \dots \right)\right).$$

Noticing that

$$\frac{-\varepsilon_{ij}+1}{2} + \sum_{\substack{j'=1\\ j'\neq j}}^{n} Y_{ij'} = \frac{-\varepsilon_{ij}+1}{2} + \varepsilon_{ij} - \varepsilon_{ij} + \sum_{\substack{j'=1\\ j'\neq j}}^{n} Y_{ij'} = -\varepsilon_{ij} + \sum_{j'=1}^{n} Y_{ij'}$$

completes the proof. □

We will denote the heat kernel of P_t as $p_t^{(n)}(\cdot,\cdot)$ to emphasize the dependence on n in the bias parameter. Explicitly, for any $x_{ij}, y_{ij} \in \{-1,1\}$,

$$p_t^{(n)}(x_{ij}, y_{ij}) = \frac{1-e^{-t}}{2}\left(\frac{2}{n}-1\right) y_{ij} + \frac{1+e^{-t}y_{ij}x_{ij}}{2}.$$

Let us denote

$$\delta_{ij}(t) := \frac{e^{-t}X_{ij}(t)\varepsilon_{ij}}{(1-e^{-t})(\frac{2}{n}-1)X_{ij}(t) + 1 + e^{-t}\varepsilon_{ij}X_{ij}(t)}.$$

where $\varepsilon = X(0) \in \{-1,1\}^{m\times n}$.

Lemma 3.16. *Conditional on the event that $\sum_{j'=1}^{n} Y_{ij'} = k$, define independent binomial random variables*

$$B_i^{(n)} \sim \text{Binomial}\left(n-k, p_t^{(n)}(-1,1)\right), \quad i = 1,\dots,m.$$

Then for $i = 1,\dots,m$, we have the following joint convergence in distribution,

$$\left(\sum_{j'=1}^{n} Y_{ij'}, B_i^{(n)}\right) \xrightarrow{d} (N_i, \eta_i(t)) \quad \text{as } n\to\infty,$$

where $N_i \sim \text{Poisson}(1)$ and $\eta_i(t) \sim \text{Poisson}(1-e^{-t})$, with N_i independent of $\eta_i(t)$ and all random variables independent over i.

PROOF. By the law of rare events, for each $i = 1, \dots, m$,

$$\sum_{j'=1}^{n} Y_{ij'} \xrightarrow{d} N_i \text{ as } n \to \infty.$$

Notice that

$$n \cdot p_t^{(n)}(-1, 1) = 1 - e^{-t},$$

so that, conditional on $\sum_{j'=1}^{n} Y_{ij'}$, by the law of rare events,

$$B_i^{(n)} \xrightarrow{d} \eta_i(t) \text{ as } n \to \infty,$$

independent of N_i. □

Now we have the necessary elements for the proof of the main theorem in the Poisson setting.

PROOF OF THEOREM 3.13. Let g be a function of the form given in (13). By Theorem 3.6, we have that

$$(\mathbb{E}\|g(\varepsilon) - \mathbb{E}g(\varepsilon)\|^p)^{\frac{1}{p}} \leq 4\left(1 - \frac{1}{n}\right) \int_0^\infty \left(\mathbb{E}\left\|\frac{1}{n}\sum_{i=1}^{m}\sum_{j=1}^{n} \delta_{ij}(t) D_{ij} g(\varepsilon)\right\|^p\right)^{\frac{1}{p}} dt. \tag{14}$$

We decompose the sum based on the value of ε_{ij} as follows

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{m}\sum_{j=1}^{n} \delta_{ij}(t) D_{ij} g(\varepsilon) = {} & \frac{1}{n}\sum_{i=1}^{m}\sum_{j=1}^{n} \mathbb{1}_{\{\varepsilon_{ij}=1\}} \delta_{ij}(t) D_{ij} g(\varepsilon) \\ & + \frac{1}{n}\sum_{i=1}^{m}\sum_{j=1}^{n} \mathbb{1}_{\{\varepsilon_{ij}=-1\}} \delta_{ij}(t) D_{ij} g(\varepsilon).\end{aligned} \tag{15}$$

Also, for all ε,

$$\|D_{ij} g(\varepsilon)\| \leq 2\|g\|_\infty = 2\|f\|_\infty,$$

which is finite by assumption. Therefore, by the triangle inequality,

$$\lim_{n\to\infty}\left\|\frac{1}{n}\sum_{i=1}^{m}\sum_{j=1}^{n} \mathbb{1}_{\{\varepsilon_{ij}=1\}} \delta_{ij}(t) D_{ij} g(\varepsilon)\right\| \leq 2\|f\|_\infty \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{m}\sum_{j=1}^{n} \mathbb{1}_{\{\varepsilon_{ij}=1\}} |\delta_{ij}(t)| = 0,$$

where by using the explicit expression of $\delta_{ij}(t)$ conditioned on $\varepsilon_{ij} = 1$, we see that

$$\mathbb{1}_{\{\varepsilon_{ij}=1\}}|\delta_{ij}(t)| \leq \max\left\{\left|\frac{e^{-t}}{\frac{2}{n}(1-e^{-t})+2e^{-t}}\right|, \left|\frac{e^{-t}}{\frac{2}{n}(1-e^{-t})-2+2e^{-t}}\right|\right\}$$

is bounded as $n \to \infty$. Now we consider the second sum in (15):

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{m}\sum_{j=1}^{n}\mathbb{1}_{\{\varepsilon_{ij}=-1\}}\delta_{ij}(t)D_{ij}g(\varepsilon)\\ &\quad= \frac{1}{n}\sum_{i=1}^{m}\frac{1}{2}\left(f\left(\dots,\sum_{j'=1}^{n}Y_{ij'},\dots\right)-f\left(\dots,\sum_{j'=1}^{n}Y_{ij'}+1,\dots\right)\right)\sum_{j=1}^{n}\mathbb{1}_{\{\varepsilon_{ij}=-1\}}\delta_{ij}(t).\end{aligned}$$

Let $B_i^{(n)}$ be as in Lemma 3.16. Then, for any $i = 1,\dots,n$, we obtain

$$\frac{1}{n}\sum_{j=1}^{n}\mathbb{1}_{\{\varepsilon_{ij}=-1\}}\delta_{ij}(t) = -B_i^{(n)}\frac{e^{-t}}{2(1-e^{-t})} + \frac{n-\sum_{j'=1}^{n}Y_{ij'}-B_i^{(n)}}{n}\cdot\frac{e^{-t}}{2-\frac{2}{n}(1-e^{-t})}.$$

Therefore, by Lemma 3.16, we have that

$$-B_i^{(n)}\frac{e^{-t}}{2(1-e^{-t})} \xrightarrow{d} -\frac{e^{-t}}{2(1-e^{-t})}\eta_i(t) \text{ as } n \to \infty,$$

and

$$\frac{n-\sum_{j'=1}^{n}Y_{ij'}-B_i^{(n)}}{n}\cdot\frac{e^{-t}}{2-\frac{2}{n}(1-e^{-t})} \xrightarrow{d} \frac{e^{-t}}{2} \text{ as } n \to \infty.$$

Finally, we conclude that

$$\frac{1}{n}\sum_{i=1}^{m}\sum_{j=1}^{n}\mathbb{1}_{\{\varepsilon_{ij}=-1\}}\delta_{ij}(t)D_{ij}g(\varepsilon)$$

converges in distribution to

$$\frac{1}{4}\sum_{i=1}^{m}(f(N_1,\dots,N_m)-f(N_1,\dots,N_i+1,\dots,N_m))\left(e^{-t}-\frac{e^{-t}}{(1-e^{-t})}\eta_i(t)\right).$$

Thus, using the assumption that f is bounded, we achieve the desired result by taking $n \to \infty$ in (14). $\square$

Proof of Corollary 3.14. By a similar symmetrization argument to that in the proof of Corollary 3.8, observing that $\mathbb{E}\tilde{\eta}_i(t) = 0$, we have that

$$\mathbb{E}\left\|\sum_{i=1}^m \tilde{\eta}_i(t) D_i^{\mathbb{Z}} f(N)\right\|^p \le 2^p T_p(X)^p \sum_{i=1}^m \mathbb{E}\left\|\tilde{\eta}_i(t) D_i^{\mathbb{Z}} f(N)\right\|^p .$$

By the fact that the $\tilde{\eta}_i(t)$ are i.i.d. and independent of $D_i^{\mathbb{Z}} f(N)$, we have by Theorem 3.13 that

$$(\mathbb{E}\|f(N) - \mathbb{E}f(N)\|^p)^{\frac{1}{p}} \le 2T_p(X) \left(\int_0^\infty (\mathbb{E}|\tilde{\eta}_1(t)|^p)^{\frac{1}{p}}\ dt\right) \left(\sum_{i=1}^m \mathbb{E}\left\|D_i^{\mathbb{Z}} f(N)\right\|^p\right)^{\frac{1}{p}} .$$

Notice that

$$(\mathbb{E}|\tilde{\eta}_1(t)|^p)^{\frac{1}{p}} \le \left(\mathbb{E}|\tilde{\eta}_1(t)|^2\right)^{\frac{1}{2}} = \frac{e^{-t}}{\sqrt{1-e^{-t}}}$$

and

$$\int_0^\infty \frac{e^{-t}}{\sqrt{1-e^{-t}}}\, dt = 2,$$

which concludes our proof. $\square$

CHAPTER 4

Vector-valued concentration on the symmetric group

This chapter is based on joint work with Ramon van Handel.

4.1. Introduction

The settings of the classical results of Pisier for the uniform measure on the discrete cube and the Gaussian measure in $\mathbb{R}^n$, the inequality (4) of [**13**] for the uniform measure on the discrete cube, and the results presented in the previous chapter all have in common that the measures considered are examples of *product measures*. The main aim of the project presented in this chapter is to go beyond the product setting, which we achieve through our main result on the symmetric group.

Capturing vector-valued concentration on the symmetric group is desirable due the immediate consequences for our understanding of the metric geometry Banach spaces of nontrivial Rademacher type. Our inequality yields novel (to our knowledge) nonembeddability results of the symmetric group endowed with the word metric given by transpositions in terms of bi-Lipschitz and average distortion for all embeddings into Banach spaces of nontrivial type. This enriches previous understanding of this nonembeddability phenomenon which follows from the study of nonembeddability of n-length metric spaces by Khot and Naor [**14**] in the setting of p-smooth Banach spaces.

In general, the symmetric group is an object which has been well-studied in the areas of discrete probability, analysis, and combinatorics, notably in the classical work of Diaconis and Shahshahani [**6**], which established through the use of representation theory a precise understanding of the phase transition known as cutoff in the mixing of the random walk on the symmetric group generated by all transpositions. More recently, [**4**] uses the notion of

Ollivier-Ricci curvature in order to understand the fine-scale ergodic behavior of conjugacy-invariant random walks on the symmetric group.

4.2. Main results

4.2.1. Preliminaries for Markov semigroups on finite groups. We work within the same framework for Pisier-like inequalities on a general graph described in Section 2. However, there are a few aspects of the group setting which streamline the notation and allow some useful simplifications, so we will summarize those here.

Let (G, S) be a finite group with a symmetric set of generators S. Let Y_n be a discrete-time random walk on the group initialized at $x \in G$ given by

$$Y_n = \gamma_n \gamma_{n-1} \dots \gamma_1 x$$

for $\gamma_1, \dots, \gamma_n \sim \nu =: \mathrm{Unif}(S)$ i.i.d. and $x \in G$. We consider the continuous-time version of the random walk: let $\{N_t\}$ be a rate $|S|$ Poisson process independent of the discrete-time random walk. Then let

$$X_t := Y_{N_t} = \gamma_{N_t} \gamma_{N_t - 1} \dots \gamma_1 x.$$

Let $\mu = \mathrm{Unif}(G)$ denote the stationary distribution of the random walk on G.

For intuition, it is useful to have in mind the Cayley graph associated to (G, S): since S is symmetric, the graph is an undirected graph with vertices indexed by the elements of G where two vertices x, y are connected by an edge if there exists an $s \in S$ such that $x = sy$. Then the random walk X_t can be analogously defined as a random walk on this graph, and the word metric of G corresponds to the standard graph metric on the Cayley graph.

For $x, y \in G$ and $t \geq 0$, the transition probability is given by

$$\mathbb{P}_x(X_t = y) = \sum_{k=0}^{\infty} \frac{(|S|t)^k e^{-|S|t}}{k!} \nu^{*k}(yx^{-1})$$

where ν^{*k} is the k-fold convolution of ν with itself.

Therefore, we have that the heat kernel in terms of its density with respect to the stationary measure is given by

$$p_t(x,y) = \frac{1}{\mu(y)} \sum_{k=0}^{\infty} \frac{(|S|t)^k e^{-|S|t}}{k!} \nu^{*k}(yx^{-1}) =: q_t(yx^{-1}), \tag{16}$$

i.e. we observe that the heat kernel depends in a univariate manner on the quantity yx^{-1}.

We denote the Laplacian on G by

$$\Delta f(x) = -\sum_{s \in S} D_s f(x)$$

where

$$D_s f(x) = f(x) - f(sx).$$

Finally, $P_t := e^{t\Delta}$ denotes the standard heat semigroup on G.

We define the random coefficients $\delta_g(t)$, for any element $g \in G$ of the group, analogously to the general discrete setting (see (10)) as

$$\delta_g(t) := \frac{p_t(x, X_t) - p_t(gx, X_t)}{p_t(x, X_t)} = \frac{D_g p_t(x, X_t)}{p_t(x, X_t)} \tag{17}$$

In the group setting, they can be written in terms of the univariate heat kernel as

$$\delta_g(t) = \frac{q_t(X_t x^{-1}) - q_t(X_t x^{-1} g^{-1})}{q_t(X_t x^{-1})},$$

which will prove useful in the proofs of the following lemmas.

Remark 4.1. Notice that we can write

$$\delta_g(t) = \frac{q_t(\gamma_{N_t}\gamma_{N_t-1}\dots\gamma_1) - q_t(\gamma_{N_t}\gamma_{N_t-1}\dots\gamma_1 g^{-1})}{q_t(\gamma_{N_t}\gamma_{N_t-1}\dots\gamma_1)},$$

so in fact $\delta_g(t)$ does not depend on the initial condition x of the random walk X_t.

4.2.2. Vector-valued Poincaré inequalities on the symmetric group. The results in this section establish vector-valued Poincaré inequalities for the uniform measure on the symmetric group. As described in Chapter 1 and in Corollaries 3.8 and 3.14, we derive these inequalities from an inequality in the spirit of Pisier's original inequality (1), but the form of inequality is more technical compared to the previous results on product spaces, which is why we do not state it as a main result.

We consider the symmetric group of permutations on n elements generated by the set of all transpositions, which we will denote

$$S = \{(ij) : i \neq j = 1, \dots, n\}.$$

Let $s_{ij} := (ij)$ for $i \neq j = 1, \dots, n$. For simplicity in this section, we will denote

$$\delta_{ij}(t) := \delta_{s_{ij}}(t)$$

and

$$D_{ij} f(x) := D_{s_{ij}} f(x) = f(x) - f(s_{ij} x).$$

Let μ denote the uniform measure on S_n.

THEOREM 4.2. *Let $(X, \|\cdot\|)$ be a Banach space of type $p \in [1,2)$, with optimal constant $T_p(X)$, $f : (S_n, S) \to (X, \|\cdot\|)$, and μ the uniform measure on S_n. Then up to universal constants, for $n \geq 2$,*

$$(\mathbb{E}_\mu \|f - \mathbb{E}_\mu f\|^p)^{\frac{1}{p}} \lesssim \frac{1}{\sqrt{2-p}} T_p(X)^2 \left(\frac{1}{n}\right)^{\frac{1}{p}} \left(\sum_{\substack{i,j=1, \\ i<j}}^{n} \mathbb{E}_\mu \|D_{ij} f\|^p \right)^{\frac{1}{p}}.$$

THEOREM 4.3. *Let $(X, \|\cdot\|)$ be a Banach space of type $p \in [1,2]$, with optimal constant $T_p(X)$, $f : (S_n, S) \to (X, \|\cdot\|)$, and μ the uniform measure on S_n. Then for $n \geq 2$,*

$$(\mathbb{E}_\mu \|f - \mathbb{E}_\mu f\|^p)^{\frac{1}{p}} \lesssim T_2(X)^2 \left(\frac{\log n}{n}\right)^{\frac{1}{p}} \left(\sum_{\substack{i,j=1, \\ i<j}}^{n} \mathbb{E}_\mu \|D_{ij} f\|^p \right)^{\frac{1}{p}}.$$

4.2.2.1. *Optimality of dimensional constant in Theorem 4.2.* Even simple examples show that the inequality of Theorem 4.2 captures the optimal dependence on the dimension n and the type p of the target Banach space $(X, \|\cdot\|)$ for $p \in [1,2)$. Take $(X, \|\cdot\|) = (\mathbb{R}^{n \times n}, \|\cdot\|_{\ell^p})$, where for $M = (m_{ij})_{i,j=1}^n \in \mathbb{R}^{n\times n}$, we define

$$\|M\|_{\ell^p} = \left(\sum_{i,j=1}^n |m_{ij}|^p\right)^{\frac{1}{p}}.$$

Let $f : S_n \to X$ to be the function that maps a permutation σ to its representation as a permutation matrix, i.e. for any $\sigma \in S_n$, we set $f(\sigma)_{ij} = \mathbb{1}_{\{\sigma: i\mapsto j\}}$. We can see that

$$(\mathbb{E}_\mu\|f - \mathbb{E}_\mu f\|^p)^{\frac{1}{p}} = \left(n\left(\frac{n-1}{n}\right)^p + n(n-1)\left(\frac{1}{n}\right)^p\right)^{\frac{1}{p}} \sim n^{\frac{1}{p}}.$$

On the other hand, notice that $\|D_{ij}f(\sigma)\|_{\ell^p} = 2^{\frac{1}{p}}$ for all σ, so we have that

$$\left(\frac{1}{n}\right)^{\frac{1}{p}}\left(\sum_{\substack{i,j=1,\\ i<j}}^n \mathbb{E}_\mu\|D_{ij}f\|^p\right)^{\frac{1}{p}} \sim n^{\frac{1}{p}}.$$

Thus, for this choice of f, both sides of the inequality in Theorem 4.2 are of the same order in terms of n and p.

Recall that for any function $f : S_n \to \mathbb{R}$, the classical Poincaré inequality on S_n is given by

$$\mathrm{Var}_\mu(f) \leq \frac{2}{n}\sum_{\substack{i,j=1,\\ i<j}}^n \mathbb{E}_\mu|D_{ij}f|^2. \tag{18}$$

where the optimal constant is given by the spectral gap computed in [**6**]. Notice that when $(X, \|\cdot\|) = (\mathbb{R}, |\cdot|)$, which has type 2, up to a factor of $(\log n)^{\frac{1}{p}}$ and universal constants, Theorem 4.3 retrieves (18). We conjecture that this logarithmic term is unnecessary.

4.2.3. Vector-valued Poincaré inequality on the k-slice of the discrete cube. The utility of Theorems 4.2 and 4.3 is that they capture the behavior of any function $f : S_n \to X$ whenever we have good control on the discrete derivates $D_{ij}f$. In fact, many discrete spaces of interest can be seen through the action of the symmetric group on another space, and thus our analysis on S_n yields immediate consequences in these other settings. In this section we focus on one such space, the so-called k-slice of the discrete cube, which is defined as follows.

Let k be a fixed value with $1 \leq k < n$ and consider the set

$$H_n^k := \left\{ x \in \{0,1\}^n : \sum_{i=1}^n x_i = k \right\}.$$

We call this set the k-slice of the discrete cube. We will consider the action of the symmetric group S_n on H_n^k as follows: for any element $x = (x_i)_{i=1}^n \in H_n^k$ and $g \in S_n$, we define

$$(gx)_i = x_{g(i)}.$$

Then gx is also an element of H_n^k. Then for any function $f : H_n^k \to X$ and $x \in H_n^k$, let

$$D_{ij} f(x) = f(x) - f((ij)x).$$

Then through a simple modification of the proof of Theorem 4.2, we obtain the following result on the k-slice.

THEOREM 4.4. *Suppose $(X, \|\cdot\|)$ is a Banach space of type $p \in [1,2]$, with optimal constant $T_p(X)$, and $f : H_n^k \to X$. Then for $n \geq 2$,*

$$(\mathbb{E}_\mu \| f - \mathbb{E}_\mu f \|^p)^{\frac{1}{p}} \lesssim T_p(X)^2 n^{-\frac{1}{p}} \left(\sum_{\substack{i,j=1, \\ i<j}}^n \mathbb{E} \| D_{ij} f(X_0) \|^p \right)^{\frac{1}{p}},$$

up to universal constants.

REMARK 4.5. For $p \in [1,2)$, Theorem 4.4 is a direct consequence of Theorem 4.2. To see this, fix any element $z \in H_n^k$ and for any function $f : H_n^k \to X$, apply Theorem 4.2 to the function $g : S_n \to X$ given by $g(\sigma) = f(\sigma(z))$. Due to the better Bakry-Émery curvature properties of the k-slice compared to S_n generated by transpositions, we are able to obtain a bounded constant in p for all $p \in [1,2]$; we conjecture that this should be possible for S_n as well, but our current techniques fall short of showing this.

4.2.3.1. *Optimality of dimensional constant in Theorem 4.4.* Let $(X, \|\cdot\|) = (\mathbb{R}^n, \|\cdot\|_{\ell^p})$ and $f : H_n^k \to R^n$ be given by $f(x) = x$. Then for μ the uniform measure on H_n^k,

$$\begin{aligned}
\mathbb{E}_\mu \|f - \mathbb{E}_\mu f\|_{\ell^p} &= \mathbb{E}\left[\left(\sum_{i=1}^n \left|\varepsilon_i - \tfrac{k}{n}\right|^p\right)^{\frac{1}{p}}\right] \\
&= \left(k\left(\frac{n-k}{n}\right)^p + (n-k)\left(\frac{k}{n}\right)^p\right)^{\frac{1}{p}} \\
&= n^{-1}\big(k(n-k)^p + (n-k)k^p\big)^{\frac{1}{p}} \\
&\gtrsim \min\{k, n-k\}^{\frac{1}{p}},
\end{aligned}$$

since $\max\{k, n-k\} \sim n$. Thus, since for all $p \in [1, 2]$, $\mathbb{E}_\mu\|f - \mathbb{E}_\mu f\| \leq (\mathbb{E}_\mu\|f - \mathbb{E}_\mu f\|^p)^{\frac{1}{p}}$, we have that the left-hand side of of the inequality of Theorem 4.4 is bounded from below by $\min\{k, n-k\}^{\frac{1}{p}}$, up to universal constants. On the other hand,

$$\sum_{\substack{i,j=1, \\ i<j}}^n \mathbb{E}\|D_{ij} f(X_0)\|^p \leq k(n-k)2^p,$$

since at most $k(n-k)$ terms in the sum are nonzero and $\|D_{ij} f(X_0)\| \leq 2$. Therefore, the right-hand side of the inequality of Theorem 4.4 is on the order of $n^{-\frac{1}{p}}\big(k(n-k)\big)^{\frac{1}{p}} \sim \min\{k, n-k\}^{\frac{1}{p}}$. Thus, the dependence on n and k in Theorem 4.4 is optimal.

4.2.4. Bi-Lipschitz and average nonembeddability. Theorem4.2 yield immediate consequences showing the bi-Lipschitz and average nonembeddability of the symmetric group generated by transpositions,

COROLLARY 4.6 (of Theorem 4.2). *Let (S_n, S) denote the symmetric group with word metric induced by the generating set of all transpositions and μ the uniform measure on S_n. Then for $(X, \|\cdot\|)$ of type $p \in [1, 2)$ up to constants depending only on p and $T_p(X)$,*

$$c_X(S_n, S) \geq c_X^\mu(S_n, S) \gtrsim n^{1-\frac{1}{p}}.$$

For $(X, \|\cdot\|)$ of type 2, we have that

$$c_X^\mu(S_n, S) \gtrsim n^{1-\frac{1}{2}}(\log n)^{-\frac{1}{2}}.$$

REMARK 4.7. If one cares only about the *bi-Lipschitz* nonembeddability of the symmetric group, the lower bound

$$c_X(S_n, S) \gtrsim n^{1-\frac{1}{p}}$$

can be seen directly as consequence of the nonembeddability of the Hamming cube, since the Hamming cube embeds isometrically into (S_n, S). (See Appendix A.) Our result captures *average* nonembeddability, a much stronger property. For $p \in [1, 2)$, Corollary 4.6 implies that for (S_n, S) *average* distortion is just as bad as the worst-case bi-Lipschitz distortion. For $p = 2$, we conjecture that $c_X^\mu(S_n, S) \gtrsim n^{1-\frac{1}{2}}$, but cannot capture this lower bound on the average distortion with our techniques (see Chapter 5 for further discussion). Nevertheless, our result still achieves a novel nearly optimal lower bound on the average distortion for $p = 2$, up to a logarithmic factor.

Corollary 4.6 already yields a nontrivial lower bound on the bi-Lipschitz and average distortions of embeddings of the k-slice of the cube into $(X, \|\cdot\|)$. However, careful application of Theorem 4.4 yields a stronger lower bound which captures the dependence on k. We consider the natural metric on H_n^k given by

$$d(x, y) = \frac{1}{2}\sum_{i=1}^{n} |x_i - y_i|. \tag{19}$$

for any $x, y \in H_n^k$. In particular, two elements $x, y \in H_n^k$ are distance one apart if $x \neq y$ and one can obtain y by transposing two entries of x.

COROLLARY 4.8 (of Theorem 4.4). *For $(X, \|\cdot\|)$ of type $p \in [1, 2]$, up to universal constants depending only on $T_p(X)$,*

$$c_X(H_n^k) \geq C_X^\mu(H_n^k) \gtrsim (\min\{k, n-k\})^{1-\frac{1}{p}}.$$

By the argument presented in Section 4.2.3.1, this lower bound is optimal up to universal constants.

4.2.5. Overview of the proof method. Applying the semigroup method of [**13**] to obtain Pisier-like inequalities with meaningful dimensional constants poses considerable challenges for measures which are *not* product measures.

Recall that the key to obtaining optimal constants in these inequalities came down to understanding the distribution of the random coefficients $\delta(t)$ as defined in (10). In the case when the stationary measure is a product measure across some indices $i = 1, \ldots, n$, each coordinate $\delta_i(t)$ is *independent* of the others. This independence makes it very natural to apply the property of Rademacher type through standard symmetrization arguments (see e.g. the proof of Corollary 3.8).

Moreover, this consequence of the product structure means that it is sufficient to understand the behavior of a single coordinate to understand the behavior of $\delta(t)$. In the case of the discrete cube, as in (4) and in Chapter 3 of this thesis, the distribution is simple enough to work with the heat kernel explicitly. Thus, one challenge in going beyond the product setting is understanding the distribution of $\delta(t)$ as a random vector with dependencies between coordinates.

The following two sections summarize informally how our approach overcomes these challenges.

4.2.5.1. *Exchangeability and symmetrization in the group setting.* The main advantage of the group setting is that we can exploit the symmetries of the underlying group structure in order to derive advantageous probabilistic properties to replace independence. In particular, by choosing a conjugacy-invariant generating set of the symmetric group, Lemma 4.12 ensures that the coordinates of $\delta(t)$ are *exchangeable*. Thus, using a slightly more involved form of symmetrization for exchangeable (rather than independent) random variables introduced in [**7**], we are able to apply Rademacher type (see Lemma 4.17). This reduces the problem to understanding the distribution of $\delta(t)$.

4.2.5.2. *Concentration of* $\delta(t)$. In this section, we derive concentration estimates on the random coefficients $\delta(t)$ using the notion of Bakry-Émery curvature, a geometric condition on Markov semigroups proposed as an analog to Ricci curvature by [**2**]. Numerous recent works in the literature have studied the utility of this condition, as well as modifications of it, in the the discrete setting. In particular, [**15**] gives several important consequences of Bakry-Émery curvature on discrete spaces and computes the curvature constant for several examples, including the symmetric group generated by transpositions and the k-slice of the discrete cube. In Section, 4.3.1, we derive concentration of $\delta(t)$ under this curvature condition.

Unfortunately, these curvature estimates alone are not enough to capture the dimensional constants in Theorems 4.2 and 4.3. Therefore, we derive further concentration estimates of $\delta(t)$ relying on the mixing properties of the transition random walk on the symmetric group. In order to obtain the optimal dimensional constant in Theorem 4.2 for $p \in [1, 2)$, we adapt precise bounds obtained on the mixing of the random walk shown in [**4**]; informally, we use the fact that two discrete time random walks starting at neighboring vertices of the Cayley graph of (S_n, S) converge in total variation distance in on the order of n steps. Interpolating between these bounds and the Bakry-Émery estimates, we are able to capture the optimal dimensional constant for $p \in [1, 2)$. However, this interpolation cannot capture the $p = 2$ endpoint of the range; in this regime, we use a more crude mixing estimate: the time it takes the two initially-neighboring random walks to converge is bounded by the L^2 mixing time to stationarity, which is on the order of $n \log n$ steps by [**27**]. This results in the suboptimal factor $\log n$ in Theorem 4.3, which we believe is an artifact of the proof.

4.3. Proofs

4.3.1. Bakry-Émery curvature. For simplicity of notation, we restrict ourselves to the setting of a finite group (G, S) as given in Section 4.2.1; however, the following two lemmas can readily be generalized to the setting of weighted graphs introduced in Section 2.

Recall the definition of the carré-du-champ operator Γ given in (7), which takes the following form on (G, S). For any functions $f, g : G \to \mathbb{R}$,

$$\Gamma(f,g) = \frac{1}{2} \sum_{s \in S} D_s f D_s g.$$

Then the iterated carré du champ is given by

$$\Gamma_2(f,g) := \frac{1}{2} \left(\Delta\Gamma(f,g) - \Gamma(f,\Delta g) - \Gamma(\Delta f, g) \right).$$

We will denote

$$\Gamma(f) := \Gamma(f,f) \quad \text{and} \quad \Gamma_2(f) := \Gamma_2(f,f)$$

The Bakry-Émery curvature of G is given by the largest constant $K \in \mathbb{R}$ such that for any $x \in G$ and $f : G \to \mathbb{R}$,

$$\Gamma_2(f)(x) \geq K\Gamma(f)(x). \tag{20}$$

The Bakry-Émery curvature of several discrete spaces of interest has been computed in [**15**]: for instance, the Bakry-Émery curvature of S_n generated by all transpositions is 2 and that of the k-slice of the discrete cube is $\frac{n}{2} + 1$. The fact that the k-slice not only has strictly positive curvature, but positive and increasing with n allows us to achieve the optimal constant for all $p \in [1, 2]$ in Theorem 4.4, as Lemma 4.9 already yields bounds on the L^2 norm of $\delta(t)$ on the order of e^{-nt}. In the case of S_n, we must combine the Bakry-Émery estimates of this section with further bounds derived in Sections 4.3.4 and 4.3.5 which leverage known results about the mixing time of the transposition walk in order to achieve bounds of this order.

Lemma 4.9. *Suppose the Bakry-Émery condition* (20) *with constant K is satisfied for all $f : G \to \mathbb{R}$. Then we have that for any $a \in \mathbb{R}^{|S|}$ and $x \in G$*

$$\left(\mathbb{E}_x \left| \sum_{s \in S} a_s \delta_s(t) \right|^2 \right)^{\frac{1}{2}} \leq \left(\frac{2K}{e^{2Kt} - 1} \right)^{\frac{1}{2}} \|a\|_2.$$

PROOF. Notice that

$$\frac{1}{2}\sum_{s\in S}|D_sP_tf(x)|^2=\Gamma(P_tf)(x).$$

By the Bakry-Émery condition, we have the local Poincaré inequality (see, for example, equation (4.1) of [**15**]):

$$\Gamma(P_tf)(x)\leq\frac{K}{e^{2Kt}-1}P_tf^2(x).$$

Then by Lemma 2.1, we have that

$$\frac{1}{2}\sum_{s\in S}\Big(\mathbb{E}_x[f(X_t)\delta_s(t)]\Big)^2\leq\frac{K}{e^{2Kt}-1}P_tf^2(x). \tag{21}$$

By scaling, we can assume $\|a\|_2=1$. By Cauchy-Schwarz, we have that

$$\sum_{s\in S}\Big(\mathbb{E}_x[f(X_t)\delta_s(t)]\Big)^2\geq\left(\mathbb{E}_x\left[f(X_t)\sum_{s\in S}a_s\delta_s(t)\right]\right)^2. \tag{22}$$

Set

$$f(X_t):=\sum_{s\in S}a_s\delta_s(t).$$

Then combining (21) and (22) for this value of f, we have

$$\frac{1}{2}\left(\mathbb{E}_x\left[\left(\sum_{s\in S}a_s\delta_s(t)\right)^2\right]\right)^2\leq\frac{K}{e^{2Kt}-1}\mathbb{E}_x\left[\left(\sum_{s\in S}a_s\delta_s(t)\right)^2\right].$$

To obtain the desired result, we divide by $P_tf^2(x)$ on both sides of the inequality. ☐

PROPOSITION 4.10. *Suppose the Bakry-Émery condition* (20) *with constant K is satisfied for all $f:G\to\mathbb{R}$. Then we have that for any $a\in\mathbb{R}^G$, $x\in G$, and $p\in[1,2]$,*

$$\left(\mathbb{E}_x\left|\sum_{s\in S}a_s\delta_s(t)\right|^p\right)^{\frac{1}{p}}\leq 2\left(\frac{4K}{e^{2Kt}-1}\right)^{1-\frac{1}{p}}\|a\|_p. \tag{23}$$

PROOF. We prove the result by interpolation over $p\in[1,2]$. Consider the linear operator $T:(\mathbb{C}^S,\|\cdot\|_p)\to L^p(\Omega,\mathbb{P}_x)$ given by

$$T:a\mapsto\sum_{s\in S}a_s\delta_s(t).$$

Observe that for any $a = a_1 + ia_2 \in \mathbb{C}^S$, by applying Lemma 4.9 twice, we have

$$\|Ta\|_{L^2} = \|Ta_1\|_{L^2} + \|Ta_2\|_{L^2} \leq \left(\frac{2K}{e^{2Kt}-1}\right)^{\frac{1}{2}} (\|a_1\|_2 + \|a_2\|_2).$$

Furthermore,

$$(\|a_1\|_2 + \|a_2\|_2)^2 \leq 2(\|a_1\|_2^2 + \|a_2\|_2^2) = 2\|a\|_2^2.$$

Thus, T is a bounded operator from $\ell^2(\mathbb{C}^S)$ to $L^2(\Omega, \mathbb{P}_x)$, with

$$\|T\|_{\ell^2 \to L^2} \leq \sqrt{2}\left(\frac{2K}{e^{2Kt}-1}\right)^{\frac{1}{2}}.$$

Furthermore, we notice that by definition of $\delta(t)$,

$$\mathbb{E}_x|\delta_s(t)| = \mathbb{E}_x\left|\frac{p_t(x,X_t) - p_t(sx,X_t)}{p_t(x,X_t)}\right| = \sum_{y\in G} |p_t(x,y) - p_t(sx,y)| \leq 2.$$

Therefore, we have that for any $a \in \mathbb{C}^S$,

$$\mathbb{E}_x\left|\sum_{s\in S} a_s\delta_s(t)\right| \leq 2\|a\|_1,$$

so T is also a bounded operator from $\ell^1(\mathbb{C}^S)$ to $L^1(\Omega, \mathbb{P}_x)$. Then by the Riesz-Thorin interpolation theorem, we have that for $\theta \in (0,1)$ and $p_\theta := \frac{2}{2-\theta}$,

$$\|T\|_{\ell^{p_\theta} \to L^{p_\theta}} \leq 2^{1-\theta}\left(\frac{4K}{e^{2Kt}-1}\right)^{\frac{\theta}{2}}.$$

Now setting $\theta = 2 - \frac{2}{p}$, we obtain the desired result. □

4.3.2. Exchangeability of $\delta(t)$. Let $\mathrm{N}_G(S) := \{g \in G : gSg^{-1} = S\}$ denote the normalizer of (G,S).

LEMMA 4.11. *For any $y \in G$ and $g \in \mathrm{N}_G(S)$,*

$$q_t(y) = q_t(gyg^{-1}) = q_t(g^{-1}yg).$$

Proof. We first observe that

$$g\gamma_{N_t}\gamma_{N_t-1}\dots\gamma_1 g^{-1} = g\gamma_{N_t}g^{-1}g\gamma_{N_t-1}\dots g^{-1}g\gamma_1 g^{-1} \overset{d}{=} \gamma_{N_t}\gamma_{N_t-1}\dots\gamma_1,$$

since the γ_i are independent and chosen uniformly from S and the mapping $s \mapsto gsg^{-1}$ is a bijection on S, because we assumed that $g \in \mathrm{N}_G(S)$. Then we have that for any $x \in G$,

$$\begin{aligned} q_t(y) &= |G|\mathbb{P}(\gamma_{N_t}\gamma_{N_t-1}\dots\gamma_1 = y) \\ &= |G|\mathbb{P}(g\gamma_{N_t}\gamma_{N_t-1}\dots\gamma_1 g^{-1} = y) \\ &= |G|\mathbb{P}(\gamma_{N_t}\gamma_{N_t-1}\dots\gamma_1 = g^{-1}yg) \\ &= q_t(g^{-1}yg). \end{aligned}$$

Since $g \in \mathrm{N}_G(S)$ implies that $g^{-1} \in \mathrm{N}_G(S)$, we can replace g by g^{-1} to obtain the other equality. □

Lemma 4.12. *Let $g \in \mathrm{N}_G(S)$ and $(s_1, \dots, s_n)$ be an ordered subset of S.*

$$(\delta_{s_1}(t), \dots, \delta_{s_n}(t)) \overset{d}{=} (\delta_{gs_1g^{-1}}(t), \dots, \delta_{gs_ng^{-1}}(t)) \tag{24}$$

Proof. We have that

$$\begin{aligned} &(\delta_{s_1}(t), \dots, \delta_{s_n}(t)) \\ &\quad = \left(1 - \frac{q_t(\gamma_{N_t}\gamma_{N_t-1}\dots\gamma_1 s_1^{-1})}{q_t(\gamma_{N_t}\gamma_{N_t-1}\dots\gamma_1)}, \dots, 1 - \frac{q_t(\gamma_{N_t}\gamma_{N_t-1}\dots\gamma_1 s_n^{-1})}{q_t(\gamma_{N_t}\gamma_{N_t-1}\dots\gamma_1)}\right) \\ &\quad = \left(1 - \frac{q_t(g\gamma_{N_t}\gamma_{N_t-1}\dots\gamma_1 s_1^{-1}g^{-1})}{q_t(g\gamma_{N_t}\gamma_{N_t-1}\dots\gamma_1 g^{-1})}, \dots, 1 - \frac{q_t(g\gamma_{N_t}\gamma_{N_t-1}\dots\gamma_1 s_n^{-1}g^{-1})}{q_t(g\gamma_{N_t}\gamma_{N_t-1}\dots\gamma_1 g^{-1})}\right) \\ &\quad = \left(1 - \frac{q_t(g\gamma_{N_t}\gamma_{N_t-1}\dots\gamma_1 g^{-1}gs_1^{-1}g^{-1})}{q_t(g\gamma_{N_t}\gamma_{N_t-1}\dots\gamma_1 g^{-1})}, \dots, 1 - \frac{q_t(g\gamma_{N_t}\gamma_{N_t-1}\dots\gamma_1 g^{-1}gs_n^{-1}g^{-1})}{q_t(g\gamma_{N_t}\gamma_{N_t-1}\dots\gamma_1 g^{-1})}\right) \end{aligned}$$

by applying Lemma 4.11. Finally, we observe that

$$g\gamma_{N_t}\gamma_{N_t-1}\dots\gamma_1 g^{-1} \overset{d}{=} \gamma_{N_t}\gamma_{N_t-1}\dots\gamma_1,$$

which completes the proof. □

4.3.3. Symmetrization and decoupling on the symmetric group. Now let us restrict ourselves to the setting of the symmetric group S_n generated by the set of all transpositions, which we denote as S. Let I denote a random subset of $\{1, \ldots, n\}$, chosen uniformly among all subsets of size $\lceil \frac{n}{2} \rceil$. Then the following lemma allows us to decouple the indices contained in I with those in I^c.

LEMMA 4.13. *Let I denote a random subset of $\{1, \ldots, n\}$, chosen uniformly among all subsets of size $\lceil \frac{n}{2} \rceil$. For any $n \geq 2$, $f : (S_n, S) \to (X, \| \cdot \|)$, and $p \geq 1$,*

$$\mathbb{E}_\mu \left\| \sum_{\substack{i,j=1, \\ i<j}}^{n} \delta_{ij}(t) D_{ij} f(X_0) \right\|^p \leq 2^p \mathbb{E}_\mu \left\| \sum_{i \in I} \sum_{j \in I^c} \delta_{ij}(t) D_{ij} f(X_0) \right\|^p .$$

PROOF. Let $Y_i = \mathbb{1}_{\{i \in I\}}$. We have that when n is even and $i \neq j$,

$$\mathbb{E}[Y_i(1 - Y_j)] = \frac{\frac{n}{2}}{n} \cdot \frac{\frac{n}{2}}{n-1} = \frac{1}{4} \left(\frac{n}{n-1} \right) \geq \frac{1}{4},$$

and when n is odd,

$$\mathbb{E}[Y_i(1 - Y_j)] = \frac{\lceil \frac{n}{2} \rceil}{n} \cdot \frac{\lfloor \frac{n}{2} \rfloor}{n-1} = \frac{\frac{n+1}{2}}{n} \cdot \frac{\frac{n-1}{2}}{n-1} \geq \frac{1}{4}.$$

Therefore,

$$\begin{aligned}
\mathbb{E}_\mu \left\| \sum_{\substack{i,j=1, \\ i<j}}^{n} \delta_{ij}(t) D_{ij} f(X_0) \right\|^p &= \mathbb{E}_\mu \left\| \frac{1}{2} \sum_{\substack{i,j=1, \\ i \neq j}}^{n} \delta_{ij}(t) D_{ij} f(X_0) \right\|^p \\
&\leq 2^p \mathbb{E}_\mu \left\| \sum_{\substack{i,j=1, \\ i \neq j}}^{n} \mathbb{E}[Y_i(1 - Y_j)] \delta_{ij}(t) D_{ij} f(X_0) \right\|^p \\
&\leq 2^p \mathbb{E}_\mu \left\| \sum_{\substack{i,j=1, \\ i \neq j}}^{n} Y_i(1 - Y_j) \delta_{ij}(t) D_{ij} f(X_0) \right\|^p ,
\end{aligned}$$

where the inequality follows by Jensen's inequality. ∎

For any element $g \in S_n$, let $\sigma_g(i)$ denote the coordinate to which the permutation g sends the coordinate $i \in [n]$.

LEMMA 4.14. *Let I be defined as in Lemma 4.13. Let σ denote a uniform random permutation on I and σ' an independent random permutation on I^c. Then*

$$\left(\delta_{ij}\right)_{i\in I, j\in I^c} \stackrel{d}{=} \left(\delta_{\sigma(i)\sigma'(j)}\right)_{i\in I, j\in I^c}$$

PROOF. Since S is a conjugacy-invariant generating set, $G = \mathrm{N}_G(S)$. Then by Lemma 4.12, we have that for any $g \in G$,

$$\left(\delta_{s_{ij}}\right)_{i,j=1,\ldots,n} \stackrel{d}{=} \left(\delta_{g s_{ij} g^{-1}}\right)_{i,j=1,\ldots,n}.$$

Notice that we can write

$$g s_{ij} g^{-1} = s_{\sigma_g(i)\sigma_g(j)},$$

Then for any permutation $g \in S_n$, we obtain that

$$\left(\delta_{ij}\right)_{i\in I, j\in I^c} \stackrel{d}{=} \left(\delta_{\sigma_g(i)\sigma_g(j)}\right)_{i\in I, j\in I^c}.$$

For any realization of σ, σ', there exists a g such that $\sigma_g\big|_I = \sigma$ and $\sigma_g\big|_{I^c} = \sigma'$, so we obtain the final result. □

Now by the exchangeability property derived in Lemma 4.14, as well as the decoupling scheme of Lemma 4.13, we may perform a form of symmetrization introduced in [**7**]. We restate the main theorem of [**7**] in the form that will be useful for the following lemma:

THEOREM 4.15 (Theorem 1 of [**7**]). *Fix any Banach space $(X, \|x\|)$. Let $\Psi : X \to \mathbb{R}$ be given by $\Psi(x) = \psi(\|x\|)$ for a convex, nondecreasing function $\psi : [0,\infty) \to \mathbb{R}$. Let $A = (a_{ij})_{i,j=1}^n$ be a random matrix with independent entries and $\Pi \sim$ Uniform(S_n) an independently chosen random permutation on n elements. Then*

$$\mathbb{E}\Psi\left(\sum_{i=1}^n a_{i\Pi(i)} - \mathbb{E}\left[\sum_{i=1}^n a_{i\Pi(i)}\right]\right) \leq \mathbb{E}\left(\sum_{i=1}^n \varepsilon_i a_{i\Pi(i)}\right),$$

where ε_i are i.i.d. standard Rademacher signs.

PROPOSITION 4.16. *Let $\varepsilon, \varepsilon'$ denote independent symmetric Rademacher random vectors. Let I be defined as in Lemma 4.13. Then for any $n \geq 2$, $f : (S_n, S) \to (X, \|\cdot\|)$, and $p \geq 1$,*

$$\frac{1}{2}\left(\mathbb{E}_\mu\left\|\sum_{\substack{i,j=1,\\ i<j}}^{n}\delta_{ij}(t)D_{ij}f(X_0)\right\|^p\right)^{\frac{1}{p}} \leq 64^{\frac{1}{p}}\left(\mathbb{E}\left\|\sum_{i\in I}\sum_{j\in I^c}\varepsilon_i\varepsilon_j'\delta_{ij}(t)D_{ij}f(X_0)\right\|^p\right)^{\frac{1}{p}} \tag{25}$$

$$+\,8^{\frac{1}{p}}\left(\mathbb{E}\left\|\frac{1}{|I^c|}\sum_{i\in I}\varepsilon_i\left(\sum_{k\in I^c}\delta_{ik}(t)\sum_{j\in I^c}D_{ij}f(X_0)\right)\right\|^p\right)^{\frac{1}{p}} \tag{26}$$

$$+\,8^{\frac{1}{p}}\left(\mathbb{E}\left\|\frac{1}{|I|}\sum_{j\in I^c}\varepsilon_j'\left(\sum_{k\in I}\delta_{kj}(t)\sum_{i\in I}D_{ij}f(X_0)\right)\right\|^p\right)^{\frac{1}{p}} \tag{27}$$

$$+\left(\mathbb{E}\left\|\frac{1}{|I|}\frac{1}{|I^c|}\sum_{k\in I}\sum_{l\in I^c}\delta_{kl}(t)\sum_{i\in I}\sum_{j\in I^c}D_{ij}f(X_0)\right\|^p\right)^{\frac{1}{p}}. \tag{28}$$

PROOF. Let σ denote a uniform random permutation on I and σ' an independent random permutation on I^c. Recall that in the group setting, we have that $\delta(t)$ is independent of X_0 (see Remark 4.1). Therefore, we can apply Lemma 4.14 to the result of Lemma 4.13 to obtain

$$\mathbb{E}_\mu\left\|\sum_{\substack{i,j=1,\\ i<j}}^{n}\delta_{ij}(t)D_{ij}f(X_0)\right\|^p \leq 2^p\mathbb{E}_\mu\left\|\sum_{i\in I}\sum_{j\in I^c}\delta_{\sigma(i)\sigma'(j)}(t)D_{ij}f(X_0)\right\|^p.$$

Set

$$a_{ik} = \sum_{j\in I^c}\delta_{k\sigma'(j)}(t)D_{ij}f(X_0).$$

By Theorem 4.15 applied conditionally on all randomness except σ (since σ is independent of all other random quantities), we have that

$$\mathbb{E}\left\|\sum_{i\in I}a_{i\sigma(i)} - \mathbb{E}_\sigma\left[\sum_{i\in I}a_{i\sigma(i)}\right]\right\|^p = \mathbb{E}\left\|\sum_{i\in I}a_{i\sigma(i)} - \frac{1}{|I|}\sum_{i\in I}\sum_{k\in I}a_{ik}\right\|^p \leq 8\mathbb{E}\left\|\sum_{i\in I}\varepsilon_i a_{i\sigma(i)}\right\|^p.$$

Therefore,

$$\left(\mathbb{E}_\mu\left\|\sum_{i\in I}\sum_{j\in I^c}\delta_{\sigma(i)\sigma'(j)}(t)D_{ij}f(X_0)\right\|^p\right)^{\frac{1}{p}} \le 8^{\frac{1}{p}}\left(\mathbb{E}\left\|\sum_{i\in I}\varepsilon_i\sum_{j\in I^c}\delta_{\sigma(i)\sigma'(j)}(t)D_{ij}f(X_0)\right\|^p\right)^{\frac{1}{p}} + \left(\mathbb{E}\left\|\frac{1}{|I|}\sum_{i\in I}\sum_{k\in I}\sum_{j\in I^c}\delta_{k\sigma'(j)}(t)D_{ij}f(X_0)\right\|^p\right)^{\frac{1}{p}}. \tag{29}$$

Now set

$$b_{jk} := \sum_{i\in I}\varepsilon_i\delta_{\sigma(i)k}(t)D_{ij}f(X_0).$$

Again by Theorem 4.15 applied conditionally on all randomness except σ' (since σ' is independent of all other random quantities), we have that

$$\mathbb{E}\left\|\sum_{j\in I^c}b_{j\sigma'(j)} - \mathbb{E}_{\sigma'}\left[\sum_{j\in I^c}b_{j\sigma'(j)}\right]\right\|^p = \mathbb{E}\left\|\sum_{j\in I^c}b_{j\sigma'(j)} - \frac{1}{|I^c|}\sum_{j\in I^c}\sum_{k\in I^c}b_{jk}\right\|^p \le 8\mathbb{E}\left\|\sum_{j\in I^c}\varepsilon'_j b_{j\sigma'(j)}\right\|^p.$$

Therefore, plugging in the value of b_{jk}, we obtain the following bound on the first term on the right-hand side of (29)

$$\left(\mathbb{E}\left\|\sum_{i\in I}\sum_{j\in I^c}\varepsilon_i\delta_{\sigma(i)\sigma'(j)}(t)D_{ij}f(X_0)\right\|^p\right)^{\frac{1}{p}} \le 8^{\frac{1}{p}}\left(\mathbb{E}\left\|\sum_{i\in I}\sum_{j\in I^c}\varepsilon_i\varepsilon'_j\delta_{\sigma(i)\sigma'(j)}(t)D_{ij}f(X_0)\right\|^p\right)^{\frac{1}{p}} + \left(\mathbb{E}\left\|\frac{1}{|I^c|}\sum_{j\in I^c}\sum_{k\in I^c}\sum_{i\in I}\varepsilon_i\delta_{\sigma(i)k}(t)D_{ij}f(X_0)\right\|^p\right)^{\frac{1}{p}}.$$

Finally, we apply Theorem 4.15 again to the second term on the right-hand side of (29). Set

$$c_{jl} := \frac{1}{|I|}\sum_{i\in I}\sum_{k\in I}\delta_{kl}(t)D_{ij}f(X_0).$$

Then, applying the theorem conditionally on all randomness except for σ', as before,

$$\mathbb{E}\left\|\sum_{j\in I^c}c_{j\sigma'(j)} - \mathbb{E}_{\sigma'}\left[\sum_{j\in I^c}c_{j\sigma'(j)}\right]\right\|^p = \mathbb{E}\left\|\sum_{j\in I^c}c_{j\sigma'(j)} - \frac{1}{|I^c|}\sum_{j\in I^c}\sum_{l\in I^c}c_{jl}\right\|^p \le 8\mathbb{E}\left\|\sum_{j\in I^c}\varepsilon'_j c_{j\sigma'(j)}\right\|^p.$$

Therefore,

$$\left(\mathbb{E}\left\|\frac{1}{|I|}\sum_{i\in I}\sum_{k\in I}\sum_{j\in I^c}\delta_{k\sigma'(j)}(t)D_{ij}f(X_0)\right\|^p\right)^{\frac{1}{p}} \le 8^{\frac{1}{p}}\left(\mathbb{E}\left\|\frac{1}{|I|}\sum_{i\in I}\sum_{k\in I}\sum_{j\in I^c}\varepsilon'_j\delta_{k\sigma'(j)}(t)D_{ij}f(X_0)\right\|^p\right)^{\frac{1}{p}}$$
$$+\left(\mathbb{E}\left\|\frac{1}{|I|}\frac{1}{|I^c|}\sum_{j\in I^c}\sum_{k\in I}\sum_{i\in I}\sum_{l\in I^c}\delta_{kl}(t)D_{ij}f(X_0)\right\|^p\right)^{\frac{1}{p}}.$$

Finally, by independence of δ from X_0 (see Remark 4.1) and the exchangeability of δ with respect to σ, σ', which we have by Lemma 4.14, we may remove the σ, σ' from the final result. □

Now that we have introduced independent standard Rademacher signs through symmetrization, we may apply the Rademacher type property, yielding the following estimate.

LEMMA 4.17. *Suppose $(X, \|\cdot\|)$ is a Banach space of type $p \in [1,2]$, with type constant $T_p(X)$, and $f : (S_n, S) \to (X, \|\cdot\|)$. Let $C(t,n)$ be such that for all $t \ge 0$, $n \ge 2$, and $E \subset S$,*

$$\left(\mathbb{E}\left\|\sum_{s\in E}\delta_s(t)\right\|^p\right)^{\frac{1}{p}} \le C_p(t,n)|E|^{\frac{1}{p}}. \tag{30}$$

Then for $n \ge 2$,

$$(\mathbb{E}_\mu\|f-\mathbb{E}_\mu f\|^p)^{\frac{1}{p}} \lesssim T_p(X)^2\left(\int_0^\infty C_p(t,n)\,dt\right)\left(\sum_{\substack{i,j=1,\\ i<j}}^n \mathbb{E}_\mu\|D_{ij}f\|^p\right)^{\frac{1}{p}}.$$

PROOF. We apply the type p assumption to the first three terms on the right-hand side of Proposition 4.16, conditionally on all randomness except $\varepsilon, \varepsilon'$, using the independence of these random variables from all others present. Recall that I denotes a random subset of $\{1, \ldots, n\}$, chosen uniformly among all subsets of size $\lceil \frac{n}{2} \rceil$.

For the first term (25), we apply the type p assumption twice as follows:

$$\left(\mathbb{E}\left\|\sum_{i\in I}\sum_{j\in I^c}\varepsilon_i\varepsilon_j'\delta_{ij}(t)D_{ij}f(X_0)\right\|^p\right)^{\frac{1}{p}} \leq T_p(X)\left(\sum_{i\in I}\mathbb{E}\left\|\sum_{j\in I^c}\varepsilon_j'\delta_{ij}(t)D_{ij}f(X_0)\right\|^p\right)^{\frac{1}{p}}$$
$$\leq T_p(X)^2\left(\sum_{i\in I}\sum_{j\in I^c}\mathbb{E}\left\|\delta_{ij}(t)D_{ij}f(X_0)\right\|^p\right)^{\frac{1}{p}}.$$

Recall that by remark 4.1, X_0 and $\delta(t)$ are independent, and furthermore all entries $\delta_{ij}(t)$ are equal in distribution by Lemma 4.12. Thus, by (30),

$$\left(\sum_{i\in I}\sum_{j\in I^c}\mathbb{E}\left\|\delta_{ij}(t)D_{ij}f(X_0)\right\|^p\right)^{\frac{1}{p}} = C_p(t,n)\left(\sum_{\substack{i,j=1,\\ i<j}}^{n}\mathbb{E}\left\|D_{ij}f(X_0)\right\|^p\right)^{\frac{1}{p}}.$$

For the second term (26), we apply the type p assumption one time to obtain

$$\left(\mathbb{E}\left[\left\|\frac{1}{|I^c|}\sum_{i\in I}\varepsilon_i\left(\sum_{k\in I^c}\delta_{ik}(t)\sum_{j\in I^c}D_{ij}f(X_0)\right)\right\|^p\right]\right)^{\frac{1}{p}}$$
$$\leq T_p(X)\left(\mathbb{E}\left[\sum_{i\in I}\left\|\frac{1}{|I^c|}\sum_{k\in I^c}\delta_{ik}(t)\sum_{j\in I^c}D_{ij}f(X_0)\right\|^p\right]\right)^{\frac{1}{p}}.$$

Let us denote $\mathbb{E}^I[\,\cdot\,] := \mathbb{E}[\,\cdot \mid I]$. Therefore, again be independence of $\delta(t)$ and X_0, exchangeability of $\delta(t)$, and Hölder's inequality applied conditionally on I, we obtain that

$$(26) \leq 8^{\frac{1}{p}}T_p(X)\left(\mathbb{E}\left[\frac{1}{|I^c|}\sum_{i\in I}\mathbb{E}^I\left[\left|\sum_{k\in I^c}\delta_{ik}(t)\right|^p\right]\mathbb{E}^I\left[\sum_{j\in I^c}\left\|D_{ij}f(X_0)\right\|^p\right]\right]\right)^{\frac{1}{p}}$$

Then by (30), we have that

$$(26) \lesssim T_p(X)C_p(t,n)\left(\sum_{\substack{i,j=1,\\ i<j}}^{n}\mathbb{E}\left\|D_{ij}f(X_0)\right\|^p\right)^{\frac{1}{p}}.$$

Similarly, for the third term, we have that

$$(27) \lesssim T_p(X)C_p(t,n)\left(\sum_{\substack{i,j=1,\\ i<j}}^{n} \mathbb{E}\left\|D_{ij}f(X_0)\right\|^p\right)^{\frac{1}{p}}.$$

Lastly, for the fourth term, we see that

$$\left\|\frac{1}{|I|}\frac{1}{|I^c|}\sum_{k\in I}\sum_{l\in I^c}\delta_{kl}(t)\sum_{i\in I}\sum_{j\in I^c}D_{ij}f(X_0)\right\|^p = \left|\frac{1}{|I|}\frac{1}{|I^c|}\sum_{k\in I}\sum_{l\in I^c}\delta_{kl}(t)\right|^p\left\|\sum_{i\in I}\sum_{j\in I^c}D_{ij}f(X_0)\right\|^p$$

By applying Hölder's inequality and the assumption (30), we have that

$$(28) \lesssim C_p(t,n)\left(\sum_{\substack{i,j=1,\\ i<j}}^{n} \mathbb{E}\left\|D_{ij}f(X_0)\right\|^p\right)^{\frac{1}{p}}.$$

Combining all of these facts, we obtain the final result by Lemma 2.3. □

4.3.4. Bounds on $\|\delta_s(t)\|_1$ via Ollivier-Ricci curvature. For two random variables X, Y taking values in a metric space (M, d), the Wasserstein 1-distance (or L^1-Kantorovich distance) is given by

$$W_1(X,Y) := \inf \mathbb{E}[d(X,Y)], \tag{31}$$

where the infimum is taken over all couplings of X and Y. The Ollivier-Ricci curvature [**23**] of a discrete-time Markov chain $\{Y_m\}_{m\in\mathbb{N}}$ between two points $x, x' \in M$ is given by

$$\kappa_m(x,x') := 1 - \frac{W_1(Y_m^x, Y_m^{x'})}{d(x,x')},$$

where Y_m^x denotes the Markov chain initialized at $Y_0 = x$.

Now take $\{Y_m\}_{m\in\mathbb{N}}$ to be the random walk on the symmetric group S_n generated by all transpositions as described in Section 4.2.1 and let $d(\cdot,\cdot)$ denote the word metric on S_n. As in [**4**], for all $c > 0$, denote

$$\kappa_c^{(n)}(\sigma,\sigma') := 1 - \frac{W_1(Y^{\sigma}_{\lfloor cn/2\rfloor}, Y^{\sigma'}_{\lfloor cn/2\rfloor})}{d(\sigma,\sigma')}.$$

We define $\kappa_c(\sigma,\sigma)=1$. Now let

$$\kappa_c^{(n)} := \inf_{\sigma,\sigma'} \kappa_c^{(n)}(\sigma,\sigma'),$$

where the infimum is taken over σ and σ' being even distance apart in the transposition word metric.

Remark 4.18. Recall that the total variation distance between two probability measures μ and ν can be written as

$$\|\mu-\nu\|_{TV} = \inf \mathbb{E}[\mathbb{1}_{\{X\neq Y\}}]$$

where the infimum is taken over all couplings of X and Y such that $X\sim\mu$ and $Y\sim\nu$ (see e.g. Theorem 5.4 of [**17**]). Then for every metric $d(\cdot,\cdot)$ such that $x\neq y\in M$ implies that $d(x,y)\geq 1$, we have that by the definition (31) of Wasserstein 1-distance above,

$$\|\mu-\nu\|_{TV} = \inf \mathbb{E}[\mathbb{1}_{\{X\neq Y\}}] \leq \inf \mathbb{E}[d(X,Y)] = W_1(X,Y).$$

In particular, the word metric on any discrete group satisfies the necessary property. Furthermore, we have that

$$W_1(X,Y) \leq \operatorname{diam}(M)\|\mu-\nu\|_{TV}. \tag{32}$$

Finally, recall also that the total variation distance of two measures μ and ν on a discrete domain G can be expressed as one half times the L^1 norm between μ and ν:

$$\|\mu-\nu\|_{TV} = \frac{1}{2}\sum_{y\in G}|\mu(y)-\nu(y)|.$$

For the proof of the following lemma, we follow closely the argument in Section 2.2 of [**4**].

Lemma 4.19. *Let $Y=\{Y_m\}_{m\in\mathbb{N}}$ the random walk on the symmetric group S_n generated by all transpositions and let Y^x denote the Markov chain initialized at $Y_0^x=x$. Fix two distinct transpositions s_1,s_2. Then for all $m\in\mathbb{N}$ and a universal constant $C>0$,*

$$W_1(Y_m^x, Y_m^{s_2s_1x}) \lesssim e^{-C\frac{m}{n}}.$$

Proof. By Corollary 21 of [**23**], noting that Y_n^x and $Y_n^{s_2s_1x}$ are also at an even distance, we have that for all $l \in \mathbb{N}$,

$$W_1(Y^x_{l\lfloor cn/2\rfloor}, Y^{s_2s_1x}_{l\lfloor cn/2\rfloor}) \leq d(x, s_2s_1x)(1-\kappa_c^{(n)})^l \leq 2(1-\kappa_c^{(n)})^l.$$

Let $k = l\lfloor cn/2\rfloor$. Then for some universal constant $c_1 \geq 0$

$$W_1(Y_k^x, Y_k^{s_2s_1x}) \lesssim e^{c_1\frac{k}{n}\log(1-\kappa_c^{(n)})}. \tag{33}$$

Let $\theta(c)$ denote the largest solution in $[0,1]$ to the equation

$$\theta(c) = 1 - e^{-c\theta(c)}.$$

Equivalently, $\theta(c)$ is the survival probability of a Galton-Watson tree with offspring distribution given by Poisson(c). By Theorem 1.2 of [**4**], we have that

$$\liminf_{n\to\infty} \kappa_c^{(n)} \geq \theta(c)^4,$$

and $\theta(c) > 0$ for $c > 1$. Thus,

$$\limsup_{n\to\infty} \log(1-\kappa_c^{(n)}) \leq \log(1-\theta(c)^4) < 0.$$

Fix $c > 1$ independent of n. Then for some universal constant $C > 0$ and some fixed $n_0 \in \mathbb{N}$, for all $n \geq n_0$,

$$W_1(Y_k^x, Y_k^{s_2s_1x}) \lesssim e^{-C\frac{k}{n}}. \tag{34}$$

To see that the result holds for $n < n_0$, by (33) it is enough to show that $\kappa_c^{(n)} > 0$ for all $n \in \mathbb{N}$. To see this, fix any $\sigma \neq \sigma' \in S_n$ such that $d(\sigma, \sigma')$ are even distance apart. By definition,

$$W_1(Y^\sigma_{\lfloor cn/2\rfloor}, Y^{\sigma'}_{\lfloor cn/2\rfloor}) = \inf \mathbb{E}[d(Y^\sigma_{\lfloor cn/2\rfloor}, \tilde{Y}^{\sigma'}_{\lfloor cn/2\rfloor})]$$

where the infimum is over all couplings. In particular, we may choose the classic coupling of Aldous (see page 264 of [**1**]), which is constructed such that the expected distance between the random walks strictly decreases from their initial distance, since the fact that σ and σ'

are at an even distance means the random walk is aperiodic. Therefore,

$$W_1(Y^{\sigma}_{\lfloor cn/2 \rfloor}, Y^{\sigma'}_{\lfloor cn/2 \rfloor}) < d(\sigma, \sigma'),$$

which implies that $\kappa_c^{(n)} > 0$ for all $n \in \mathbb{N}$. Finally, recall that we chose $k = l\lfloor cn/2 \rfloor$. In order to see that the desired result holds for all $m \in \mathbb{N}$, we observe that for any fixed $x, y \in S_n$, $W_1(Y^x_m, Y^y_m)$ is a nonincreasing function in m. To see this, observe that

$$\begin{aligned} W_1(Y^x_m, Y^y_m) = \inf \mathbb{E}[d(Y^x_m, Y^y_m)] &\leq \inf \mathbb{E}[d(\gamma_m Y^x_{m-1}, \gamma_m Y^y_{m-1})] \\ &= \inf \mathbb{E}[d(Y^x_{m-1}, Y^y_{m-1})] = W_1(Y^x_{m-1}, Y^y_{m-1}), \end{aligned}$$

where we bound the first infimum over all couplings by choosing the coupling that applies the same transposition γ_m to both random walks at time m. This monotonicity property shows that the desired result holds for all $m \in \mathbb{N}$. □

LEMMA 4.20. *Fix $x \in S_n$ and s any transposition. Then for all $t \geq 0$ and a universal constant $C > 0$, we have that*

$$\|\delta_s(t)\|_1 \lesssim e^{-Cnt}$$

PROOF. Recall that in (16) we defined $p_t(x, y) = \frac{1}{\mu(y)} \mathbb{P}_x(X_t = y)$ to be the heat kernel normalized with respect to the stationary measure. Notice that by Remark 4.18,

$$\begin{aligned} \|\delta_s(t)\|_1 &= \sum_{y \in G} \frac{|\mu(y)p_t(x,y) - \mu(y)p_t(sx,y)|}{\mu(y)p_t(x,y)} \mu(y)p_t(x,y) \\ &= 2\|\mu(\cdot)p_t(x,\cdot) - \mu(\cdot)p_t(sx,\cdot)\|_{TV} \\ &\leq 2W_1(X^x_t, X^{sx}_t). \end{aligned}$$

By definition,

$$W_1(X^x_t, X^{sx}_t) = \inf \mathbb{E}[d(X^x_t, \tilde{X}^{sx}_t)],$$

where the infimum is taken over all couplings of X^x_t and $\tilde{X}^{sx}_t$. Consider the following coupling which ensures that the random walks will be at even distance after the first jump taken by either one of the walks.

Denote $\lambda := |S| = \binom{n}{2}$. Let N_t and $\tilde{N}_t$ be two Poisson processes, both with rate λ. We can write

$$X_t^x = Y_{N_t}^x \text{ and } \tilde{X}_t^{sx} = \tilde{Y}_{\tilde{N}_t}^{sx},$$

where Y and $\tilde{Y}$ are discrete-time transposition random walks on S_n. We couple N_t and $\tilde{N}_t$ as follows. Let N_t and $\tilde{N}_t$ start evolving independently from time $t = 0$. Denote

$$\tau := \inf\{t \geq 0 : N_t = 1 \text{ or } \tilde{N}_t = 1\}.$$

Notice that $\tau \sim \mathrm{Exp}(2\lambda)$, since it can be expressed as the minimum of two $\mathrm{Exp}(\lambda)$ random variables. By the strong Markov property, the processes $N_{t+\tau} - N_\tau$ and $\tilde{N}_{t+\tau} - \tilde{N}_\tau$ are also Poisson processes with rate λ, independent of the events prior to time τ. Let us set our coupling such that for all $t \geq \tau$, N_t and $\tilde{N}_t$ evolve according to the same exponential clocks, i.e. their jumps are synchronized.

Let γ_1 denote the random transposition chosen at time τ. Notice that $\mathbb{P}(d(X_\tau^x, \tilde{X}_\tau^{sx}) \in \{0, 2\}) = 1$. On the event that $d(X_\tau^x, \tilde{X}_\tau^{sx}) = 0$, for all $t \geq \tau$, we couple X and $\tilde{X}$ such that the random walks are equal. Furthermore, let $\sigma = \gamma_1 x$ and $\sigma' = sx$ if $N_\tau = 1$; otherwise, if $\tilde{N}_\tau = 1$, let $\sigma = x$ and $\sigma' = \gamma_1 sx$. Then, by the strong Markov property,

$$\begin{aligned} \mathbb{E}[d(X_t^x, \tilde{X}_t^{sx})] &= \mathbb{E}[d(X_{t-\tau}^\sigma, \tilde{X}_{t-\tau}^{\sigma'}) \mid \tau \leq t] \cdot \mathbb{P}(\tau \leq t) + \mathbb{P}(\tau > t) & (35) \\ &\leq \int_0^t \mathbb{E}[d(X_{t-r}^\sigma, \tilde{X}_{t-r}^{\sigma'})] 2\lambda e^{-2\lambda r}\, dr + e^{-2\lambda t}. & (36) \end{aligned}$$

Now by the definition of our coupling, we have that

$$\mathbb{E}[d(X_{t-r}^\sigma, \tilde{X}_{t-r}^{\sigma'})] \leq \mathbb{E}[d(Y_{N_{t-r}}^\sigma, \tilde{Y}_{N_{t-r}}^{\sigma'})].$$

By Lemma 4.19, we can find a coupling of Y and $\tilde{Y}$ such that

$$\mathbb{E}[d(Y_{N_{t-r}}^\sigma, \tilde{Y}_{N_{t-r}}^{\sigma'})] \leq \mathbb{E}[e^{-CN_{t-r}/n}].$$

Since $N_{t-r} \sim \text{Poisson}(\lambda(t-r))$, its moment generating function evaluated at $-\frac{C}{n}$ can be computed explicitly as

$$\mathbb{E}[e^{-CN_{t-r}/n}] = \exp\left(\lambda(t-r)(e^{-\frac{C}{n}} - 1)\right) \lesssim e^{-C\frac{\lambda}{n}(t-r)}.$$

Therefore, recalling that $\lambda = \binom{n}{2}$ and applying the bounds derived above to (36), we obtain

$$\begin{aligned}\mathbb{E}[d(X_t^x, \tilde{X}_t^{sx})] &\lesssim \int_0^t n^2 e^{-C(n-1)(t-r)} e^{-2n^2 r}\, dr + e^{-2n(n-1)t} \\ &= \frac{n^2}{-2n^2 + C(n-1)}(e^{-2n^2 t} - e^{C(n-1)t}) + e^{-2n(n-1)t} \\ &\lesssim e^{Ct} e^{-Cnt},\end{aligned}$$

which implies our desired result. □

4.3.5. Mixing time bounds on $\delta(t)$. In the previous section, we were able to derive bounds on the order of e^{-nt} on the L^1 norm of $\delta(t)$ for all $t \geq 0$. In this section, we present such bounds for the L^2 norm of $\delta(t)$ which hold for all $t \gtrsim \frac{\log n}{n}$. It would be desirable to obtain such bounds for all $t \geq 0$ and to develop methods that do not rely on specific mixing time bounds, but on easier-to-verify properties of the discrete space. We present a method to obtain such bounds in Chapter 5 which yield an alternate technique to obtain Theorem 4.3 than bounds in this section (see Section 5.4).

Recall that in (16) we defined the univariate heat kernel on a group (G, S) for $x, y \in G$ as

$$q_t(yx^{-1}) = p_t(x, y) = \frac{1}{\mu(y)}\mathbb{P}_x(X_t = y) = |G|\mathbb{P}_x(X_t = y).$$

For convenience in this section, we denote

$$\bar{q}_t(\cdot) := \frac{1}{|G|} q_t(\cdot).$$

Thus, $\bar{q}_t$ is a probability measure on G.

We say a stopping time T for a random walk $\{Y_m\}_{m\in\mathbb{N}}$ on G with stationary measure μ is a *strong stationary time* if for all $x, y \in G$ and times $m \in \mathbb{N}$,

$$\mathbb{P}_x(T = m, Y_m = y) = \mathbb{P}_x(T = m)\mu(y).$$

In other words, at the random time T, the random walk is distributed according to the stationary measure. (See e.g. Section 6.4 of [**17**]).

LEMMA 4.21. *For $t \geq 2\frac{\log n}{n}$, $n > 10$, and any $y \in S_n$,*

$$\bar{q}_t(y) \geq \frac{1}{2}\mu(y). \tag{37}$$

PROOF. Let T be the strong uniform time for the discrete-time random transposition random walk given in [**18**] (denoted in the paper as T_N).Recall that we denote the discrete-time random walk as $Y = \{Y_m\}_{m\in\mathbb{N}}$ and the continuous-time random walk is therefore given by $X_t := Y_{N_t}$. Then we may bound the separation distance of the chain by the tail probability of the strong uniform time (see e.g. Lemma 6.12 of [**17**]) as follows:

$$\begin{aligned}
\sup_{x\in G}\left(1 - \frac{\mathbb{P}_x(X_{N_t} = y)}{\pi(x)}\right) &\leq \mathbb{P}(T > N_t) \\
&= \mathbb{P}\left(T > N_t \middle| N_t > \frac{1}{2}n\log n + cn\right)\mathbb{P}\left(N_t > \frac{1}{2}n\log n + cn\right) \\
&\quad + \mathbb{P}\left(T > N_t \middle| N_t \leq \frac{1}{2}n\log n + cn\right)\mathbb{P}\left(N_t \leq \frac{1}{2}n\log n + cn\right) \\
&\leq \mathbb{P}\left(T > \frac{1}{2}n\log n + cn\right) + \mathbb{P}\left(N_t \leq \frac{1}{2}n\log n + cn\right).
\end{aligned}$$

By Equation (3.2) of [**18**], we have that for any $c \geq 0$

$$\mathbb{P}\left(T > \frac{1}{2}n\log n + cn\right) \leq 1 - e^{-(2e)e^{-2c}}[1 + o(1)] \leq \frac{1}{4}$$

for $n \geq 2$. For the second term, recalling that $N_t \sim \text{Poisson}(|S|t)$ where in our case

$$|S|t = \frac{n(n-1)}{2}t,$$

by the Chernoff bound we have that for $t, c \geq 0$ such that $\frac{1}{2}n\log n + cn < \frac{n(n-1)}{2}t$,

$$\mathbb{P}\left(N_t \leq \frac{1}{2}n\log n + cn\right) \leq \frac{\left(e\frac{n(n-1)}{2}t\right)^{\frac{1}{2}n\log n + cn} e^{-\frac{n(n-1)}{2}t}}{\left(\frac{1}{2}n\log n + cn\right)^{\frac{1}{2}n\log n + cn}}.$$

Then for any $t \geq 2\frac{\log n}{n}$ (setting $c = \frac{1}{10}$) and $n > 10$,

$$\mathbb{P}\left(N_t \leq \frac{1}{2}n\log n + cn\right) \leq \frac{(e(n\log n - \log n))^{\frac{1}{2}n\log n + cn} e^{-n\log n + \log n}}{\left(\frac{1}{2}n\log n + cn\right)^{\frac{1}{2}n\log n + cn}} \leq \frac{1}{4}$$

Therefore, we obtain that for all $n > 10$,

$$\sup_{x \in G}\left(1 - \frac{\bar{q}_t(y)}{\pi(x)}\right) \leq \frac{1}{4} + \frac{1}{4} = \frac{1}{2},$$

which gives us the desired result. $\square$

Definition 4.22. The χ^2-distance between two distributions μ, ν is given by

$$d_2(\nu, \mu) := \left(\sum_{x \in G} \frac{(\nu(x) - \mu(x))^2}{\mu(x)}\right)^{\frac{1}{2}}.$$

In the case when $\mu \equiv \frac{1}{|G|}$, we have that

$$d_2(\nu, \mu) := \left(|G| \sum_{x \in G} (\nu(x) - \mu(x))^2\right)^{\frac{1}{2}}.$$

Lemma 4.23. *For $t \gtrsim \frac{\log n}{n}$ and $i \neq j \in [n]$, we have that*

$$\left(\mathbb{E}|\delta_{ij}(t)|^2\right)^{\frac{1}{2}} \lesssim e^{-2nt}.$$

Proof. By the definition of $\delta(t)$ in (17), we have that

$$\mathbb{E}|\delta_{ij}(t)|^2 = \mathbb{E}\left[\left|\frac{p_t(x, X_t) - p_t(s_{ij}x, X_t)}{p_t(x, X_t)}\right|^2\right] = \frac{1}{|G|}\sum_{y \in G}\left|\frac{p_t(x, y) - p_t(s_{ij}x, y)}{p_t(x, y)}\right|^2 p_t(x, y)$$

$$= \frac{1}{|G|}\sum_{y \in G}\frac{|p_t(x, y) - p_t(s_{ij}x, y)|^2}{p_t(x, y)} = \frac{1}{|G|}\sum_{y \in G}\frac{\left|q_t(y) - q_t(ys_{ij}^{-1})\right|^2}{q_t(y)}.$$

Then we obtain

$$\mathbb{E}|\delta_{ij}(t)|^2 = \sum_{y\in G} \frac{\left|\bar{q}_t(y) - \bar{q}_t(ys_{ij}^{-1})\right|^2}{\bar{q}_t(y)}.$$

By Lemma 4.21, we have that

$$\mathbb{E}|\delta_{ij}(t)|^2 \le \sum_{y\in G} \frac{\left|\bar{q}_t(y) - \bar{q}_t(ys_{ij}^{-1})\right|^2}{\frac{1}{2}\mu(y)}.$$

Then we have that

$$\begin{aligned}\mathbb{E}|\delta_{ij}(t)|^2 &\le 4\left(\sum_{y\in G}\frac{|\bar{q}_t(y)-\mu(y)|^2}{\mu(y)} + \sum_{y\in G}\frac{\left|\mu(y)-\bar{q}_t(ys_{ij}^{-1})\right|^2}{\mu(y)}\right) \qquad (38)\\ &\le 8\sum_{y\in G}\frac{|\bar{q}_t(y)-\mu(y)|^2}{\mu(y)},\end{aligned}$$

where in the last line we have noticed that without loss of generality we may replace y by ys_{ij} in the second term on the right-hand side of (38). Therefore,

$$\mathbb{E}|\delta_{ij}(t)|^2 \le 8d_2(\bar{q}_t, \mu).$$

By Proposition 4.6 of [**27**] and noticing that in our definition of the continuous-time random walk we have rescaled time by $|S| = \frac{n(n-1)}{2}$, we have that for all $t \ge \frac{1}{|S|}\frac{n}{2}(\log n + c) = \frac{1}{n-1}(\log n + c)$ and $c \ge 2$,

$$d_2(\bar{q}_t, \mu) \le e^{-(c-2)}.$$

We obtain the final result by setting $c = 4nt$. □

Lemma 4.24. *For $t \gtrsim \frac{\log n}{n}$, $a \in \mathbb{R}^S$, and $p \in [1, 2]$, we have that*

$$\left\|\sum_{\substack{i,j=1,\\ i<j}}^{n} a_{ij}\delta_{ij}(t)\right\|_p \lesssim e^{-\frac{1}{2}nt}\|a\|_p.$$

PROOF. We first show the result for $q = 2$. Without loss of generality, let $a \in \mathbb{R}^n$ be such that $\|a\|_2 = 1$. By the Cauchy-Schwarz inequality,

$$\mathbb{E}\left[\left(\sum_{i<j} a_{ij}\delta_{ij}(t)\right)^2\right] \le \mathbb{E}\left[\left(\sum_{i<j} a_{ij}^2\right)\left(\sum_{i<j} \delta_{ij}(t)^2\right)\right] = \sum_{i<j} \mathbb{E}\left[\delta_{ij}(t)^2\right].$$

Then by Lemma 4.23, we have that

$$\mathbb{E}\left[\left(\sum_{i<j} a_{ij}\delta_{ij}(t)\right)^2\right] \lesssim \frac{n(n-1)}{2}e^{-4nt} \lesssim e^{2\log n}e^{-4nt} \lesssim e^{-2nt},$$

for $t \gtrsim \frac{\log n}{n}$, which gives us the result for $p = 2$. Finally by Jensen's inequality, we see that for $p \in [1, 2]$,

$$\mathbb{E}\left[\left(\sum_{i<j} a_{ij}\delta_{ij}(t)\right)^p\right]^{\frac{1}{p}} \le e^{-nt}\|a\|_2 \le e^{-nt}\|a\|_p.$$

□

4.3.6. Proofs of Theorems 4.2 and 4.3 on the symmetric group. To complete the proof of Theorem 4.2, we interpolate between the estimates obtained on the 1st moment of $\delta(t)$ in Section 4.3.4 via Ollivier-Ricci cuvature and the estimates on the 2nd moment obtained in Section 4.3.1 via Bakry-Émery curvature.

PROOF OF THEOREM 4.2. By Lemma 4.17, it suffices to bound

$$\int_0^\infty C_p(t, n)\, dt$$

where

$$\left(\mathbb{E}\left\|\sum_{s\in E} \delta_s(t)\right\|^p\right)^{\frac{1}{p}} \le C_p(t, n)|E|^{\frac{1}{p}}. \tag{39}$$

By Lemma 4.20, we have that by the triangle inequality,

$$\left\|\sum_{s\in E} \delta_s(t)\right\|_1 \lesssim |E|e^{-nt}.$$

By Theorem 2.10 of [**15**], we have that for S_n with all transpositions as generators, the Bakry-Émery criterion $\Gamma_2(f) \geq 2\Gamma(f)$ is satisfied. Therefore, by Lemma 4.9,

$$\left\|\sum_{s\in E}\delta_s(t)\right\|_2 \leq |E|^{\frac{1}{2}}\left(\frac{4}{e^{4t}-1}\right)^{\frac{1}{2}} \lesssim |E|^{\frac{1}{2}}t^{-\frac{1}{2}}. \tag{40}$$

Then for any $p \in [1,2]$, we have that by Hölder's inequality,

$$\left\|\sum_{s\in E}\delta_s(t)\right\|_p \leq \left\|\sum_{s\in E}\delta_s(t)\right\|_1^{1-\theta}\left\|\sum_{s\in E}\delta_s(t)\right\|_2^{\theta}, \tag{41}$$

for $\theta = 2\left(1-\frac{1}{p}\right)$. Therefore, using the L^1 bound derived in Lemma 4.20, we obtain that for a universal constant $C > 0$,

$$\left\|\sum_{s\in E}\delta_s(t)\right\|_p \lesssim |E|^{\frac{1}{p}}e^{-C\left(\frac{2}{p}-1\right)nt}t^{\frac{1}{p}-1} \tag{42}$$

By the change of variables $u := C\left(\frac{2}{p}-1\right)nt$, we compute that

$$\int_0^\infty C_p(t,n)\,dt \lesssim \left(\frac{2}{p}-1\right)^{-\frac{1}{p}}n^{-\frac{1}{p}}\int_0^\infty u^{\frac{1}{p}-1}e^{-u}\,dt = \Gamma\left(\frac{1}{p}\right)\left(\frac{2}{p}-1\right)^{-\frac{1}{p}}n^{-\frac{1}{p}},$$

where Γ is the generalized factorial function. For $p \in [1,2)$,

$$\Gamma\left(\frac{1}{p}\right)\left(\frac{2}{p}-1\right)^{-\frac{1}{p}} \lesssim \frac{1}{\sqrt{2-p}},$$

which yields the final result. $\square$

Since the proof method of Theorem 4.2 only works for p bounded away from 2, we use a different method for the case when $p = 2$.

Proof of Theorem 4.3. Again using the fact that S_n satisfies the Bakry-Émery condition with $K = 2$, we have that Proposition 4.10 yields

$$\left\|\sum_{s\in E}\delta_s(t)\right\|_p \lesssim |E|^{\frac{1}{p}}\left(\frac{1}{e^{4t}-1}\right)^{\frac{1}{p}} \lesssim |E|^{\frac{1}{p}}t^{\frac{1}{p}-1}, \tag{43}$$

By (43) and Lemma 4.24, we have that

$$C_p(t,n) \lesssim \begin{cases} t^{\frac{1}{p}-1} & \text{for } t \lesssim \frac{\log n}{n} \\ e^{-\frac{nt}{p}} & \text{for } t \gtrsim \frac{\log n}{n}. \end{cases}$$

for $C_p(t,n)$ as in (30). We compute that

$$\int_0^\infty C_2(t,n)\,dt \lesssim \int_0^{\frac{\log n}{n}} t^{\frac{1}{p}-1}\,dt + \int_{\frac{\log n}{n}}^\infty e^{-\frac{nt}{p}}\,dt \lesssim \left(\frac{\log n}{n}\right)^{\frac{1}{p}} + 2n^{-1-\frac{1}{p}} \lesssim \left(\frac{\log n}{n}\right)^{\frac{1}{p}}.$$

By Lemma 4.17, this yields the result of the theorem. □

Remark 4.25. See Section 5.4 for an alternative proof of Theorem 4.3.

4.3.7. Proof of Theorem 4.4 on the k-slice of the discrete cube. Due to the fact the the Bakry-Émery curvature of the k-slice of the cube is positive and on the order of n, rather than on the order of a constant, as is the case for S_n generated by all transpositions, we are able to obtain the optimal constant in Theorem 4.4 for all $p \in [1,2]$. For $p \in [1,2)$, the inequality of Theorem 4.4 can be derived as a consequence of Theorem 4.2 (see Remark 4.5 for a proof.)

Proof of Theorem 4.4. The key to the symmetrization argument on S_n generated by all transpositions was the exchangeability property shown in Lemma 4.14. By the action of the symmetric group on the k-slice H_n^k as described in Section 4.2.3, this same property holds for the random coefficients $\delta(t)$ induced by the random walk on the k-slice. Thus, the statements of Lemma 4.13, Proposition 4.16, and Lemma 4.17 hold for the k-slice in the same form as in the case of the symmetric group with random transpositions. Therefore, by the same reasoning as in the proof of Theorem 4.2, we have that for any function $f : H_n^k \to X$,

$$(\mathbb{E}_\mu\|f-\mathbb{E}_\mu f\|^p)^{\frac{1}{p}} \leq \left(128T_p(X)^2 + 32T_p(X) + 2\right)\left(\int_0^\infty C_p(t,n)\,dt\right)\left(\mathbb{E}\left[\sum_{\substack{i,j=1,\\ i<j}}^n \|D_{ij}f(X_0)\|^p\right]\right)^{\frac{1}{p}},$$

where as in (30),

$$\left(\mathbb{E}\left\|\sum_{s\in E}\delta_s(t)\right\|^p\right)^{\frac{1}{p}} \le C_p(t,n)|E|^{\frac{1}{p}}.$$

By Theorem 2.7 of [**15**], we have that for the k-slice of the hypercube, the Bakry-Émery condition $\Gamma_2(f) \ge \left(1+\frac{n}{2}\right)\Gamma(f)$ is satisfied. Then by Lemma 4.9, we have that

$$\left\|\sum_{s\in E}\delta_{ij}(t)\right\|_2 \lesssim \left(\frac{n}{e^{(n+2)t}-1}\right)^{1-\frac{1}{2}} |E|^{\frac{1}{2}}. \tag{44}$$

Notice that the random walk on H_n^k can be written as a deterministic function of the transposition random walk on S_n (for instance, for any $\sigma \in S_n$ take $h(\sigma)(i) = \mathbb{1}_{\{\sigma^{-1}(i)\le k\}}$). Therefore, Lemmas 4.19 and 4.20 also hold for the random walk on H_n^k. Therefore, we can interpolate between the L^1 bound of Lemma 4.20 and the L^2 bound given by (44) for $p=2$ as in (41):

$$\left\|\sum_{s\in E}\delta_s(t)\right\|_p \lesssim |E|^{\frac{1}{p}} e^{-\left(\frac{2}{p}-1\right)Cnt}\left(\frac{n}{e^{(2+n)t}-1}\right)^{1-\frac{1}{p}}.$$

Thus, for $p\in[1,2]$,

$$\begin{aligned}
\int_0^\infty C_p(n,t)\,dt &\lesssim \int_0^\infty e^{-\left(\frac{2}{p}-1\right)Cnt}\left(\frac{n}{e^{(2+n)t}-1}\right)^{1-\frac{1}{p}} dt \\
&\lesssim n^{1-\frac{1}{p}}\int_0^\infty e^{-\left(\frac{2}{p}-1\right)Cnt}\left(e^{nt}-1\right)^{\frac{1}{p}-1} dt \\
&\lesssim n^{-\frac{1}{p}}\int_0^\infty e^{-\left(\frac{2}{p}-1\right)u}\left(e^{u}-1\right)^{\frac{1}{p}-1} du \\
&\lesssim n^{-\frac{1}{p}},
\end{aligned}$$

which completes the proof. $\square$

4.3.8. Proof of Corollaries 4.6 and 4.8. The following lemma will be useful in finding lower bounds on the distortion. Denote as

$$\operatorname{diam}(G,S) := \sup_{x,y\in G} d_G(x,y).$$

LEMMA 4.26. *Let Y, Y' by independent uniform random elements of G. Then*

$$\frac{1}{2}\operatorname{diam}(G,S) \leq \mathbb{E}_{\mu\otimes\mu}[d_G(Y,Y')]$$

PROOF. Recall that for the graph distance on the Cayley graph of (G,S) with symmetric generating set S, we have that

$$d(gx,gy) = d(x,y)$$

for all $g,x,y \in G$.

Then

$$d(Y,Y') = d(e, Y^{-1}Y') = d(g, gY^{-1}Y').$$

We have that $gY^{-1}Y'$ has the same distribution as Y. So we have that

$$d(Y,Y') \stackrel{d}{=} d(g,Y).$$

Therefore,

$$d(x,y) \leq \mathbb{E}_\mu[d(x,Y)] + \mathbb{E}_\mu[d(y,Y)] = 2\mathbb{E}_\mu[d(Y,Y')].$$

We obtain

$$\operatorname{diam}(G,S) \leq 2\mathbb{E}_\mu[d_G(Y,Y')]$$

by maximizing over $x, y \in G$. □

Now equipped with this lemma, we are ready to prove the lower bound on the bi-Lipschitz distortion of S_n implied by Theorems 4.2 and 4.3.

PROOF OF COROLLARY 4.6. Fix $p \in [1,2)$. Recall that by Lemma 2.4, it is enough to show the lower bound on $c_X^\mu(S_n,S)$. Let $f : (S_n,S) \to (X, \|\cdot\|)$ have μ-average distortion $D \geq 1$. Then

$$\|D_{ij}f(X_0)\| \leq D,$$

so by Theorem 4.2,

$$\mathbb{E}_\mu\|f - \mathbb{E}_\mu f\| \lesssim n^{\frac{2}{p}} D \left(\frac{1}{n}\right)^{\frac{1}{p}}.$$

For any random variable Y and an independent copy of it $\tilde{Y}$, we have the following fact by the triangle inequality

$$\mathbb{E}\|f(Y)-f(\tilde{Y})\| \leq \mathbb{E}\|f(Y)-\mathbb{E}[f(Y)]\| + \mathbb{E}\|f(\tilde{Y})-\mathbb{E}[f(Y)]\| = 2\mathbb{E}\|f(Y)-\mathbb{E}[f(Y)]\|.$$

Thus, for X_0 and $\tilde{X}_0$ independent random variables uniformly distributed on G, we have that

$$\mathbb{E}_\mu\|f-\mathbb{E}_\mu f\| \geq \frac{1}{2}\mathbb{E}\|f(X_0)-f(\tilde{X}_0)\|.$$

By the definition of the average distortion, we have that

$$\mathbb{E}\|f(X_0)-f(\tilde{X}_0)\| \geq \mathbb{E}[d(X_0,\tilde{X}_0)].$$

Finally by Lemma 4.26, recalling that the diameter of the Cayley graph of of S_n with all transpositions as generators is $n-1$, we have that

$$n \lesssim \mathbb{E}_\mu\|f-\mathbb{E}_\mu f\| \lesssim n^{\frac{2}{p}} D\left(\frac{1}{n}\right)^{\frac{1}{p}} = Dn^{\frac{1}{p}}.$$

Thus,

$$D \gtrsim n^{1-\frac{1}{p}}.$$

For $p=2$, the proof is identical except that by Theorem 4.3, we have that

$$D \gtrsim n^{1-\frac{1}{p}}(\log n)^{-\frac{1}{p}}.$$

$\square$

Next we can prove an analogous result to Lemma 4.26 on the k-slice paying special attention to the dependence on the parameter k.

LEMMA 4.27. *Let Y, Y' by independent uniform random elements of H_n^k. Then*

$$\operatorname{diam}(H_n^k) = \min\{k, n-k\} \leq \mathbb{E}_{\mu\otimes\mu}[d_{H_n^k}(Y,Y')].$$

PROOF. Recall that the metric on H_n^k is given by

$$d(x,y) = \frac{1}{2}\sum_{i=1}^{n} |x_i - y_i|.$$

Then

$$\mathbb{E}_{\mu\otimes\mu}[d(Y,Y')] = \frac{1}{2}\sum_{i=1}^{n} \mathbb{E}|Y_i - Y_i'| = n\frac{k(n-k)}{n} = k(n-k).$$

On the other hand, $\operatorname{diam}(H_n^k) = \min\{k, n-k\}$. To see this, we can first consider $k \leq \frac{n}{2}$. For any two elements $x, y \in H_n^k$, we need to apply at most k transpositions to move from x to y (to change the position of at most k 1s). By symmetry, when $k > \frac{n}{2}$, we need to do at most $n-k$ moves. □

Now we can use this lemma in order to prove a lower bound on the bi-Lipschitz distortion of the k-slice of the discrete cube.

PROOF OF COROLLARY 4.8. Again, by Lemma 2.4, it suffices to show the lower bound on $c_X^{\mu}(H_n^k)$. Let $f : H_n^k \to (X, \|\cdot\|)$ have μ-average distortion $D \geq 1$. We have that for any $x \in H_n^k$,

$$\sum_{\substack{i,j=1,\\ i<j}}^{n} \|D_{ij}f(x)\|^p \leq D^p(n-k)k,$$

since by the definition of H_n^k, only at most $(n-k)k$ of the terms in the summation are nonzero, and we have that

$$\|D_{ij}f(x)\| \leq D$$

by the definition of the average distortion. Then by Theorem 4.4,

$$\mathbb{E}_{\mu}\|f - \mathbb{E}_{\mu}f\| \lesssim n^{-\frac{1}{p}} \left(D^p(n-k)k\right)^{\frac{1}{p}}$$

By the triangle inequality (as in the proof of Corollary 4.6), for Y and Y' independent random variables uniformly distributed on H_n^k, we have that

$$\mathbb{E}_{\mu}\|f - \mathbb{E}_{\mu}f\| \geq \frac{1}{2}\mathbb{E}\|f(Y) - f(Y')\| \geq \frac{1}{2}\mathbb{E}_{\mu\otimes\mu}[d(Y,Y')].$$

Then by Lemma 4.27, we have that

$$\min\{k, n-k\} \lesssim \mathbb{E}_\mu \|f - \mathbb{E}_\mu f\| \lesssim n^{-\frac{1}{p}} \left(D^p (n-k) k\right)^{\frac{1}{p}} .$$

In the case that $k \leq \frac{n}{2}$, this gives us that

$$D \gtrsim k^{1-\frac{1}{p}} .$$

By symmetry, in the other case that $k > \frac{n}{2}$, we have that

$$D \gtrsim (n-k)^{1-\frac{1}{p}} .$$

$\square$

CHAPTER 5

Bennett inequality for $\delta(t)$

Since in general it is not computationally tractable to work with the heat kernel of a random walk on a discrete space, as is possible e.g. for the biased measure on the discrete cube, it is necessary to derive more general estimates on the concentration of $\delta(t)$ to capture the desired dimensional constants. In the settings of the biased product measure on the discrete cube as well as the product of Poisson measures, we observe that the random coefficients in Theorems 3.6 and 3.13, respectively, arising from the heat kernel satisfy a form of Poissonian concentration; in fact, in the case of Theorem 3.13, they *are* Poisson random variables. This gives us reason to believe that this concentration will extend to other discrete settings.

In fact, this is confirmed in our analysis under a general discrete geometric condition: for a Markov heat semigroups on "flat" graphs, we obtain Poissonian concentration of $\delta(t)$ in the form of a Bennett inequality which is stated in Theorem 5.2. In particular, Cayley graphs of groups with conjugacy-invariant generating sets satisfy the necessary conditions. This allows us to present an alternative proof of Theorem 4.3 on S_n which attains an optimal constant up to a $\log n$ factor and does not use the L^2 mixing estimates obtained in Section 4.3.5.

5.1. Main results

DEFINITION 5.1. We say a centered random variable X satisfies a Bennett inequality [**3**] with constant η if it is sub-Poissonian in the sense that for all $\lambda \in \mathbb{R}$

$$\mathbb{E}[e^{\lambda X}] \leq \mathbb{E}[e^{\lambda(N-\mathbb{E}N)}]. \tag{45}$$

where $N \sim \text{Poisson}(\eta)$. More explicitly,

$$\mathbb{E}[e^{\lambda(N-\mathbb{E}N)}] = e^{\eta(e^{\lambda}-\lambda-1)}.$$

The main result of this chapter is a Bennett inequality for the random coefficients $\delta(t)$ associated to any Markov heat semigroup on a graph satisfying the Ricci-flat condition introduced by Chung and Yau in [**5**], which will be presented in Section 5.3.1. In particular, Cayley graphs of groups generated by conjugacy-invariant generating sets are Ricci-flat (see Remark 5.8).

Let $G = (V, E)$ be a finite graph and denote $w \sim v$ if $(w, v) \in E$. Then we consider the Laplacian on G applied to a function $f : G \to \mathbb{R}$ is given by

$$\Delta(f)(v) = \sum_{w \sim v} (f(w) - f(v)).$$

THEOREM 5.2. *Let Δ denote the Laplacian on a Ricci-flat graph G and $P_t := e^{t\Delta}$. Let $\delta(t)$ be as defined in* (10). *Fix $x \in G$ and let $|\partial x| := \{y \in G : y \sim x\}$. Then for any $E \subset |\partial x|$ and for all $t \geq 0$, we have that for any $\lambda \in \mathbb{R}^E$,*

$$\log \mathbb{E}_x \left[e^{-t \sum_{y \in E} \lambda_y \delta_y(t)} \right] \leq t \sum_{y \in E} (e^{\lambda_y} - \lambda_y - 1).$$

In particular, $-t \sum_{y \in E} \delta_y(t)$ satisfies a Bennett inequality (45) *with constant $t|E|$.*

Thus, Theorem 5.2 shows that the moment-generating function of the random variables $-t\delta_y(t)$ is bounded by that of a family of *independent* Poisson random variables. The main application of Theorem 5.2 will be an alternative proof of Theorem 4.3 presented in Section 5.4.

5.2. Local reverse log-Sobolev inequality under $CD\psi$ condition

In order to state the main result, we introduce the preliminaries of the Γ^ψ calculus and the $CD\psi$ condition introduced by Münch in [**20**]. To our knowledge, this condition has not previously been considered for our choice of the functional ψ, i.e. $\psi(x) = -x \log x$, which is the key to obtaining the discrete reverse log-Sobolev inequality presented in Theorem 5.4.

5.2.1. Preliminaries for Γ^ψ calculus and the $CD\psi$ condition. Let $|\partial x| := \{y \in G : y \sim x\}$.

Let $\psi : \mathbb{R} \to \mathbb{R}$. Then for any function $f : G \to (0, \infty)$ the operator Δ^ψ is given by

$$(\Delta^\psi f)(v) := \Delta\left[\psi\left(\frac{f}{f(v)}\right)\right](v).$$

In addition, Ω^ψ is given by

$$\Omega^\psi f(v) := \Delta\left[\psi'\left(\frac{f}{f(v)}\right) \cdot \frac{f}{f(v)}\left[\frac{\Delta f}{f} - \frac{(\Delta f)(v)}{f(v)}\right]\right](v).$$

Then the generalized iterated Γ operator Γ_2^ψ is defined as

$$2\Gamma_2^\psi(f) := \Omega^\psi f + \frac{\Delta f \Delta^\psi f}{f} - \frac{\Delta(f\Delta^\psi f)}{f}.$$

Then the $CD\psi(\infty, 0)$ condition is satisfied if for all $f : G \to (0, \infty)$,

$$\Gamma_2^\psi(f) \geq 0. \tag{46}$$

LEMMA 5.3. *Let $\psi(x) = -x \log x$. For any $t \geq r \geq 0$ and $f : G \to (0, \infty)$, let $\Lambda(r) := -P_r\psi(P_{t-r}f)(x)$. Then*

$$\Lambda''(r) = 2P_r(P_{t-r}f\Gamma_2^\psi(P_{t-r}f)).$$

PROOF. For $\psi(x) = -x \log x$, we compute that

$$\begin{aligned}
\Omega^\psi g(x) &= \Delta\left[\left(-\log\left(\frac{g}{g(x)}\right) - 1\right) \cdot \frac{g}{g(x)}\left(\frac{\Delta g}{g} - \frac{\Delta g(x)}{g(x)}\right)\right](x)\\
&= \frac{1}{g(x)}\Delta\left[(-\log g + \log g(x) - 1) \cdot \left(\Delta g - \frac{g\Delta g(x)}{g(x)}\right)\right](x)\\
&= \frac{1}{g(x)}\left[-\Delta(\Delta g \cdot \log g) + \frac{\Delta g}{g}\Delta(g \log g) + (\log g - 1)\left(\Delta\Delta g - \frac{(\Delta g)^2}{g}\right)\right](x).
\end{aligned}$$

Also,

$$\Delta^\psi g(x) = -\Delta\left[\frac{g}{g(x)}\log\left(\frac{g}{g(x)}\right)\right](x) = -\frac{1}{g}\left[\Delta(g \log g) - \log g \cdot \Delta g\right](x)$$

$$\Delta(g\Delta^\psi g)(x) = -\Delta[\Delta(g \log g) - \log g \cdot \Delta g](x) = -\Delta\Delta(g \log g)(x) + \Delta(\log g \cdot \Delta g)\,(x).$$

Therefore,

$$
\begin{aligned}
2g\Gamma_2^{\psi} g &= (g\Omega^{\psi} g) + \Delta g \Delta^{\psi} g - \Delta(g\Delta^{\psi} g) \\
&= -\Delta(\Delta g \cdot \log g) + \frac{\Delta g}{g}\Delta(g\log g) + \log g \cdot \Delta\Delta g - \log g \frac{(\Delta g)^2}{g} - \Delta\Delta g + \frac{(\Delta g)^2}{g} \\
&\qquad - \frac{\Delta g}{g} \cdot [\Delta(g\log g) - \log g \cdot \Delta g] + \Delta\Delta(g\log g) - \Delta(\log g \cdot \Delta g) \\
&= -2\Delta(\log g \cdot \Delta g) + \log g \cdot \Delta\Delta g - \Delta\Delta g + \frac{(\Delta g)^2}{g} + \Delta\Delta(g \log g). \qquad (47)
\end{aligned}
$$

We have that for $\Lambda(r) = -P_r\psi(P_{t-r}f)(x)$,

$$\Lambda'(r) = -P_r(\Delta\psi(P_{t-r}f) - \psi'(P_{t-r}f)\Delta P_{t-r}f)\,(x).$$

Let $g := P_{t-r}f$. Since $\psi'(x) = -\log x - 1$,

$$\Lambda'(r) = P_r(\Delta(g\log g) - (\log g + 1)\Delta g)\,.$$

Now we can see that,

$$
\begin{aligned}
\Lambda''(r) &= P_r(\Delta\Delta(g\log g) - 2\Delta(\log g \cdot \Delta g) - \Delta\Delta g + \frac{(\Delta g)^2}{g} + \log g \cdot \Delta\Delta g) \\
&= P_r(2g\Gamma_2^{\psi} g),
\end{aligned}
$$

where the last line follows from the expression for $2g\Gamma_2^{\psi} g$ derived in (47). $\square$

Recall that the entropy of a function $f : G \to (0, \infty)$ with respect to a probability measure ν on G is given by

$$\mathrm{Ent}_{\nu} f = \mathbb{E}_{\nu}[f \log f] - \mathbb{E}_{\nu}[f] \log \mathbb{E}_{\nu}[f].$$

The following lemma consists of a reverse local log-Sobolev inequality, i.e. a lower bound on the entropy of any function f with respect to the heat semigroup P_t.

THEOREM 5.4. *Let Δ denote the Laplacian on G and $P_t := e^{t\Delta}$. If P_t satisfies $CD\psi(\infty, 0)$ for $\psi(x) = -x \log x$, then for all $t \geq 0$ and $f : G \to (0, \infty)$,*

$$\mathrm{Ent}_{P_t} f \geq t(\Delta(P_t f \log P_t f) - (\log P_t f + 1)\Delta P_t f).$$

PROOF. Let $\Lambda(r) = -P_r\psi(P_{t-r}f)(x)$ for $\psi = -x \log x$ as in Lemma 5.3. By the fundamental theorem of calculus, we have that

$$\mathrm{Ent}_{P_t} f = \Lambda(t) - \Lambda(0) = \int_0^t \Lambda'(s)\, ds \geq t\Lambda'(0).$$

where the last line follows from the fact that $\Lambda''(r) = P_r(2g\Gamma_2^{\psi} g) \geq 0$ by the $CD\psi(\infty, 0)$ condition (46). $\square$

5.3. Bennett inequality for $t\delta(t)$

If X satisfies a Bennett inequality (45) with constant η, then by it satisfies the following Chernoff bound (see e.g. Theorem 2.9.2 of [**30**]): for any $r > 0$,

$$\mathbb{P}(X - \mathbb{E}X \geq r) \leq e^{-\eta\varphi(r/\eta)},$$

where for any $s > -1$,

$$\varphi(s) := (s + 1) \log (s + 1) - s \tag{48}$$

The function φ will prove useful in the next lemma in the form of the following dual formulation. Notice for any $s > -1$, the quantity

$$\alpha s - (e^{\alpha} - \alpha - 1)$$

is maximized when

$$\alpha = \log (s + 1) .$$

Therefore,

$$\varphi(s) = \sup_{\alpha\in\mathbb{R}} \left(\alpha s - (e^{\alpha} - \alpha - 1)\right). \tag{49}$$

This variational characterization of φ will be useful in the proof of Theorem 5.2, which proves a Bennett inequality which is dual to the reverse local log-Sobolev-like inequality derived in Theorem 5.4.

We will use the following form of Gibbs variational principle of Lemma 4.10 of [**29**].

LEMMA 5.5 (Gibbs variational principle). *For any random variable Z, we have that*

$$\sup_{\substack{X\geq 0,\\ \mathbb{E}X=1}} \Big\{ -\operatorname{Ent} X + \mathbb{E}[XZ]\Big\} = \log \mathbb{E}[e^{Z}].$$

Now we are ready for the proof of the main theorem.

PROOF OF THEOREM 5.2. Fix $x \in G$ and let us restrict ourselves to functions $f : G \to (0,\infty)$ such that $P_t f(x) = 1$. Then Theorem 5.4 implies that for any such f,

$$\operatorname{Ent}_{P_t} f(x) \geq t \sum_{w\sim x} \Big((P_t f \log P_t f)(w) + D_w P_t f(x)\Big). \tag{50}$$

Now notice that by our assumption that $P_t f(x) = 1$, we can write $P_t f(w) = 1-1+P_t f(w) = 1 - D_w P_t(f)$. Therefore, we can rewrite (50) as

$$\begin{aligned}\operatorname{Ent}_{P_t} f(x) &\geq t \sum_{w\sim x} \Big((1 - D_w P_t f(x)) \log(1 - D_w P_t f(x)) + D_w P_t f(x)\Big)\\ &= t \sum_{w\sim x} \varphi(-D_w P_t f(x))\end{aligned}$$

by the definition of φ given in (48). Recall that by Lemma 2.1,

$$D_w P_t f(x) = \mathbb{E}_x[f(X_t)\delta_w(t)].$$

Then combining this with the expression for φ in (49), we obtain

$$\begin{aligned}
0 &\geq \sup_{\substack{f\geq 0:\\ P_t f(x)=1}} -\operatorname{Ent}_{P_t} f(x) + t\sum_{w\sim x}\sup_{\beta\in\mathbb{R}}\big\{-\beta\mathbb{E}_x[f(X_t)\delta_w(t)] - (e^\beta-\beta-1)\big\}\\
&= \sup_{b\in\mathbb{R}^{|\partial x|}}\sup_{\substack{f\geq 0:\\ P_t f(x)=1}} -\operatorname{Ent}_{P_t} f(x) - t\sum_{w\sim x} b_w\mathbb{E}_x[f(X_t)\delta_w(t)] - \sum_{w\sim x} t(e^{b_w}-b_w-1)\\
&= \sup_{b\in\mathbb{R}^{|\partial x|}} \log\mathbb{E}_x\left[\exp\left(-t\sum_{w\sim x} b_w\delta_w(t)\right)\right] - \sum_{w\sim x} t(e^{b_w}-b_w-1),
\end{aligned}$$

where the last line follows from Lemma 5.5. We obtain the result by setting $b_w = \lambda_w \mathbb{1}_{\{w\in E\}}$.

□

The following lemmas show how to deduce moment bounds from Theorem 5.2. Such bounds are useful in applications. For example, the second lemma will be used in the alternative proof of Theorem 4.3.

Lemma 5.6. *Suppose Z satisfies a Bennett inequality* (45) *with constant η. Then for any even $p\in\mathbb{N}$ such that $\log\left(\frac{1}{\eta}\right)\geq p\geq 2$,*

$$\|Z\|_p \lesssim \eta^{\frac{1}{p}}.$$

Proof. By the Bennett inequality, we have that for any even $k\in\mathbb{N}$ and $\lambda\geq 0$,

$$\frac{\mathbb{E}[Z^k]\lambda^k}{k!} \leq \frac{1}{2}(E[e^{\lambda Z}] + E[e^{-\lambda Z}]) \leq e^{\eta(e^\lambda-\lambda-1)},$$

since $e^{\eta(e^\lambda-\lambda-1)}\geq e^{\eta(e^{-\lambda}+\lambda-1)}$, as $e^\lambda-\lambda\geq e^{-\lambda}+\lambda$ for $\lambda\geq 0$. Using $(k!)^{\frac{1}{k}}\sim k$, we obtain

$$\|Z\|_k \lesssim \frac{k}{\lambda} e^{\frac{\eta}{k}(e^\lambda-\lambda-1)}.$$

Set $\lambda = 1+\log\left(\frac{k}{\eta}\right)$. Then we have that

$$\|Z\|_k \lesssim \frac{k}{1+\log\left(\frac{k}{\eta}\right)}.$$

Now choosing an even $k \in \mathbb{N}$ such that $k \geq \lfloor \log(\frac{1}{\eta}) \rfloor$, we have that

$$\|Z\|_{\lfloor \log(\frac{1}{\eta}) \rfloor} \lesssim \frac{k}{1 + \log\left(\frac{k}{\eta}\right)}.$$

By our assumption that $\log\left(\frac{1}{\eta}\right) \geq 2$, we have that $\eta \leq e^{-2}$. Therefore, $\frac{\eta}{k} \lesssim 1$, and

$$\|Z\|_{\lfloor \log(\frac{1}{\eta}) \rfloor} \lesssim 1.$$

Therefore, by the Riesz convexity theorem for $\lfloor \log(\frac{1}{\eta}) \rfloor \geq p \geq 2$,

$$\|Z\|_p \leq \|Z\|_2^s \|Z\|^{1-s}_{\left\lfloor \log\left(\frac{1}{\eta}\right) \right\rfloor},$$

where

$$\frac{1}{p} = \frac{s}{2} + \frac{1-s}{\lfloor \log(\frac{1}{\eta}) \rfloor}.$$

Therefore, for $\lfloor \log(\frac{1}{\eta}) \rfloor \geq p \geq 2$,

$$\|Z\|_p \lesssim \|Z\|_2^s \lesssim \|Z\|_2^{2/p}. \tag{51}$$

Finally, let us denote $\psi_1(\lambda) := \mathbb{E}[e^{\lambda Z}]$ and $\psi_2(\lambda) := e^{\eta(e^\lambda - \lambda - 1)}$. Notice that

$$\psi_1(0) = \psi_2(0) = 1.$$

Furthermore, by our assumption that $\mathbb{E}Z = 0$,

$$\psi_1'(0) = 0,$$

and since $\psi_2'(\lambda) = e^{\eta(e^\lambda - \lambda - 1)}\eta(e^\lambda - 1)$, $\psi_2'(0) = 0$ as well. Therefore, we have that $\psi_1''(0) \leq \psi_2''(0)$, which implies that

$$\mathbb{E}[Z^2] \leq \left[e^{\eta(e^\lambda - \lambda - 1)}\eta^2(e^\lambda - 1)^2 + e^{\eta(e^\lambda - \lambda - 1)}\eta(e^\lambda) \right]_{\lambda = 0} = \eta.$$

Thus, $\|Z\|_2 = \eta^{\frac{1}{2}}$, which implies the final result by (51). $\square$

Lemma 5.7. *Suppose Z satisfies a Bennett inequality* (45) *with constant η. Then for any $\eta \gtrsim 1$,*

$$\|Z\|_4 \lesssim \eta^{1/2}.$$

Proof. Proceeding similarly to the preceeding lemma, we have that

$$\|Z\|_4 \lesssim \frac{4}{\lambda} e^{\frac{\eta}{4}(e^{\lambda}-\lambda-1)}.$$

Setting $\lambda = \frac{4}{\sqrt{\eta}}$, we have that

$$\|Z\|_4 \lesssim \sqrt{\eta} e^{\frac{\eta}{4}(e^{\frac{4}{\sqrt{\eta}}}-\frac{4}{\sqrt{\eta}}-1)}.$$

For $\eta \gtrsim 1$, we have that $e^{\frac{\eta}{4}(e^{\frac{4}{\sqrt{\eta}}}-\frac{4}{\sqrt{\eta}}-1)} \lesssim 1$, yielding the final result. □

5.3.1. Ricci flat graphs satisfy $CD\psi(\infty, 0)$. Since the $CD\psi$ condition can be computationally intensive to verify, it is convenient to work instead with the *Ricci flat* condition, which implies $CD\psi(\infty, 0)$.

We call a D-regular graph $G = (V, E)$ Ricci flat if for all $v \in V$ and $N(v) := \{v\} \cup \{w \in V : w \sim v\}$, there exist maps $\eta_1, \ldots, \eta_D : N(v) \to V$ such that for all $w \in N(v)$, and all $i, j \in \{1, \ldots, D\}$ with $i \neq j$, the following properties are satisfied:

(i) $\eta_i(w) \sim w$,

(ii) $\eta_i(w) \neq \eta_j(w)$, and

(iii) $\bigcup_k \eta_k(\eta_i(v)) = \bigcup_k \eta_i(\eta_k(v))$.

Remark 5.8. Cayley graphs of finite groups with conjugacy-invariant generating sets are Ricci flat. To see this, let (G, S) be a finite group with $S := \{\gamma_1, \ldots, \gamma_D\}$ a conjugacy-invariant generating set. For any $w \in G$ and $i = 1, \ldots, D$, let $\eta_i(w) := \gamma_i w$. It is clear that properties (i) and (ii) are satisfied. Furthermore, we have that

$$\eta_k(\eta_i(v)) = \gamma_k \gamma_i v \quad \text{and} \quad \eta_k(\eta_i(v)) = \gamma_i \gamma_k v.$$

Since S is conjugacy invariant, we have that $\gamma_i \gamma_k \gamma_i^{-1} \in S$. This implies that property (iii) holds as well, so the Cayley graph of (G, S) satisfies the Ricci-flat condition.

The following lemma follows from the proof of Theorem 6.6 in [**20**]. We reproduce the argument here for completeness.

LEMMA 5.9. *Ricci flat graphs satisfy* $CD\psi(\infty, 0)$ *for any concave, differentiable function* $\psi : (0, \infty) \to \mathbb{R}$.

PROOF. By the same argument as in equation (6.5) of [**20**], we have that for any $f : G \to \mathbb{R}$ and $v \in G$,

$$2\Gamma_2^{\psi}(f)(v) = \sum_{\substack{i,j=1 \\ i \neq j}}^{D} z_j \left[\psi'(z_i)(z_{ij} - z_i) - (\psi(z_{ij}) - \psi(z_i))\right]$$

for $z_i := f(\eta_i(v))/f(v)$ and $z_{ij} := f(\eta_j(\eta_i(v)))/f(\eta_i(v))$. Since ψ is concave, we have that for any $x, y \in \mathbb{R}^+$,

$$\psi'(x)(y - x) - (f(y) - f(x)) \geq 0.$$

Therefore,

$$\Gamma_2^{\psi}(f)(v) \geq 0,$$

which is our desired result. $\square$

5.4. Alternative proof of Theorem 4.3

The main application of the results in this chapter is the following alternative proof of Theorem 4.3.

PROOF OF THEOREM 4.3. Since the set of all transpositions is conjugacy-invariant, we have that the Cayley graph of S_n with transpositions satisfies the Ricci flat condition by Remark 5.8. Therefore, by Theorem 5.2, we have that $-t\sum_{y \in E} \delta_y(t)$ satisfies a Bennett inequality with constant $t|E|$. Then by Lemma 5.7, we have that for all $t \geq \frac{e^{-4}}{|E|}$,

$$\left\| \sum_{s \in E} \delta_s(t) \right\|_4 \lesssim |E|^{\frac{1}{2}} t^{-\frac{1}{2}} \tag{52}$$

Then interpolating between the result of Lemma 4.20 and (52), by Hölder's inequality, we have that for $p \in [1,2]$,

$$\left\| \sum_{s \in E} \delta_s(t) \right\|_p \lesssim |E|^{\frac{1}{3}\left(\frac{2}{p}+1\right)} t^{\frac{2}{3}(\frac{1}{p}-1)} e^{-C\frac{1}{3}\left(\frac{4}{p}-1\right)nt}.$$

Recall that we have $|E| \lesssim n^2$, so

$$\left\| \sum_{s \in E} \delta_s(t) \right\|_p \lesssim n^{\frac{2}{3}\left(\frac{2}{p}+1\right)} t^{\frac{2}{3}(\frac{1}{p}-1)} e^{-C\frac{1}{3}\left(\frac{4}{p}-1\right)nt}. \tag{53}$$

Therefore, we have that for $t \gtrsim \frac{\log n}{n}$, $p \in [1,2]$, and a universal constant $c > 0$,

$$\left\| \sum_{s \in E} \delta_s(t) \right\|_p \lesssim |E|^{\frac{1}{p}} e^{-cnt}.$$

By the fact that the symmetric group generated by transpositions has Bakry-Émery curvature $K = 2$ [**15**], by Proposition 4.10 for all $t \geq 0$,

$$|E|^{\frac{1}{p}} \left(\frac{1}{e^{4t}-1} \right)^{\frac{1}{p}} \lesssim |E|^{\frac{1}{p}} t^{\frac{1}{p}-1}, \tag{54}$$

Let

$$C_p(t,n) = \begin{cases} t^{1-\frac{1}{p}} & \text{for } t \lesssim \frac{\log n}{n} \\ e^{-cnt} & \text{for } t \gtrsim \frac{\log n}{n}. \end{cases}$$

Then

$$\int_0^\infty C_p(t,n)\, dt \lesssim \int_0^{\frac{\log n}{n}} t^{\frac{1}{p}-1}\, dt + \int_{\frac{\log n}{n}}^\infty e^{-cnt}\, dt \lesssim \left(\frac{\log n}{n}\right)^{\frac{1}{p}} + n^{-1} \lesssim \left(\frac{\log n}{n}\right)^{\frac{1}{p}}.$$

Applying this bound to the result of Lemma 4.17 concludes the proof. □

Appendix

A. Bi-lipschitz non-embeddability of S_n

Let H_n denote the subgroup of S_{2n} generated by disjoint transpositions

$$\{(12), (34), (56), \dots\}.$$

We can identify the elements of this subgroup with the elements of the n-dimensional Hamming cube $\{-1, +1\}^n$. To see this explicitly, fix $\sigma \in H_n$, i.e.

$$\sigma = \tau_m \circ \cdots \circ \tau_1$$

for some $\tau_1, \dots, \tau_m \in \{(12), (34), (56), \dots\}$. By indexing the disjoint transpositions by $i = 1, \dots, n$ and for each i, we interpret an entry of -1 as the absence of that transposition in the decomposition of σ and $+1$ as its presence. This is possible due to the commutativity of the disjoint transpositions. Observe that σ has $n - 2m$ fixed points. Thus, σ cannot be expressed in terms of fewer than m *nondisjoint* transpositions, since such an expression would have a strictly greater number of fixed points. Therefore, $d_{\mathrm{Ham}}(\sigma, \mathrm{id}) = d_{S_{2n}}(\sigma, \mathrm{id}) = m$.

Proposition 5.10. *For any $f : (S_n, S) \to (X, \|\cdot\|)$, a bilipschitz embedding with distortion D with $(X, \|\cdot\|)$ of type $p \in [1, 2]$, we have that, up to universal constants depending only on $T_p^{\mathrm{R}}(X)$,*

$$D \gtrsim n^{1-\frac{1}{p}}.$$

Proof. Let us consider $f : (S_{2n}, S) \to (X, \|\cdot\|)$ where f is a bilipschitz embedding with distortion D with respect to the transposition word metric of S_{2n}. We have that

$$D \geq \sup_{x \in S_{2n} \setminus \{\mathrm{id}\}} \frac{\|f(x) - f(\mathrm{id})\|}{d_{S_{2n}}(x, \mathrm{id})} \geq \sup_{x \in H_n \setminus \{\mathrm{id}\}} \frac{\|f(x) - f(\mathrm{id})\|}{d_{S_{2n}}(x, \mathrm{id})}.$$

Therefore,

$$D \geq \sup_{x \in \{-1,1\}^n \setminus \{\mathrm{id}\}} \frac{\|f(x) - f(\mathrm{id})\|}{d_{\mathrm{Ham}}(x, \mathrm{id})}.$$

It follows from (5) that

$$\sup_{x \in \{-1,1\}^n \setminus \{\mathrm{id}\}} \frac{\|f(x) - f(\mathrm{id})\|}{d_{\mathrm{Ham}}(x, \mathrm{id})} \sim n^{1-\frac{1}{p}}.$$

$\square$

Bibliography

1. David Aldous, *Random walks on finite groups and rapidly mixing Markov chains*, Séminaire de probabilités XVII 1981/82 (Jacques Azéma and Marc Yor, eds.), Lecture Notes in Mathematics, vol. 986, Springer, Berlin, Heidelberg, 1983, pp. 243–297.
2. Dominique Bakry and Michel Emery, *Diffusions hypercontractives, Sémin. de probabilités XIX, Univ. Strasbourg*, 1983.
3. George Bennett, *Probability inequalities for the sum of independent random variables*, Journal of the American Statistical Association **57** (1962), no. 297, 33–45.
4. Nathanaël Berestycki and Bati Sengül, *Cutoff for conjugacy-invariant random walks on the permutation group*, Probability Theory and Related Fields **173** (2019), no. 3, 1197–1241 (en).
5. F. R. K. Chung and S.-T. Yau, *Logarithmic harnack inequalities*, Mathematical Research Letters **3** (1996), no. 6, 793–812.
6. Persi Diaconis and Mehrdad Shahshahani, *Generating a random permutation with random transpositions*, Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete **57** (1981), no. 2, 159–179.
7. Lutz Dümbgen, *Symmetrization and decoupling of combinatorial random elements*, Statistics & probability letters **39** (1998), no. 4, 355–361.
8. Per Enflo, *On infinite-dimensional topological groups*, Séminaire Maurey-Schwartz (1978), 1–11.
9. Alexandros Eskenazis, *Some geometric applications of the discrete heat flow*, arXiv preprint arXiv:2310.01868 (2023).
10. Miriam Gordin, *Vector-valued concentration inequalities on the biased discrete cube*, arXiv preprint arXiv:2410.09607 (2024).
11. Leonard Gross, *Logarithmic Sobolev inequalities*, American Journal of Mathematics **97** (1975), no. 4, 1061–1083.
12. Tuomas Hytönen, Jan van Neerven, Mark Veraar, and Lutz Weis, *Analysis in Banach spaces. Vol. I. Martingales and Littlewood-Paley theory*, vol. 63, Springer, Cham, 2016.
13. Paata Ivanisvili, Ramon van Handel, and Alexander Volberg, *Rademacher type and Enflo type coincide*, Annals of Mathematics **192** (2020), no. 2, 665 – 678.

14. Subhash Khot and Assaf Naor, *Nonembeddability theorems via Fourier analysis*, Mathematische Annalen **334** (2006), 821–852.

15. Boaz Klartag, Gady Kozma, Peter Ralli, and Prasad Tetali, *Discrete curvature and abelian groups*, Canadian Journal of Mathematics **68** (2016), no. 3, 655–674.

16. James R. Lee, *Separators in Region Intersection Graphs*, 8th Innovations in Theoretical Computer Science Conference (ITCS 2017) (Dagstuhl, Germany) (Christos H. Papadimitriou, ed.), Leibniz International Proceedings in Informatics (LIPIcs), vol. 67, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2017, pp. 1:1–1:8.

17. David A Levin and Yuval Peres, *Markov chains and mixing times*, vol. 107, American Mathematical Soc., 2017.

18. Peter Matthews, *A strong uniform time for random transpositions*, Journal of Theoretical Probability **1** (1988), 411–423.

19. Manor Mendel and Assaf Naor, *Towards a Calculus for Non-Linear Spectral Gaps: [Extended Abstract]*, Proceedings of the 2010 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), Proceedings, Society for Industrial and Applied Mathematics, January 2010, pp. 236–255.

20. Florentin Münch, *Li-Yau inequality on finite graphs via non-linear curvature dimension conditions*, Journal de Mathématiques Pures et Appliquées **120** (2018), 130–164.

21. Assaf Naor, *An introduction to the Ribe program*, Japanese Journal of Mathematics **7** (2012), 167–233.

22. Assaf Naor and Gideon Schechtman, *Remarks on non linear type and Pisier's inequality*, Journal für die reine und angewandte Mathematik **2002** (2002), no. 552, 213–236.

23. Yann Ollivier, *Ricci curvature of Markov chains on metric spaces*, Journal of Functional Analysis **256** (2009), no. 3, 810–864.

24. Gilles Pisier, *Probabilistic methods in the geometry of Banach spaces*, Probability and Analysis (Berlin, Heidelberg), Springer, 1986, pp. 167–241.

25. Henri Poincaré, *Sur les équations aux dérivées partielles de la physique mathématique*, American Journal of Mathematics (1890), 211–294.

26. Yuri Rabinovich, *On average distortion of embedding metrics into the line*, Discrete & Computational Geometry **39** (2008), no. 4, 720–733.

27. L Saloff-Coste and J Zúniga, *Refined estimates for some basic random walks on the symmetric and alternating groups*, Alea **4** (2008), 359–392.

28. Michel Talagrand, *Isoperimetry, logarithmic Sobolev inequalities on the discrete cube, and Margulis' graph connectivity theorem*, Geometric & Functional Analysis GAFA **3** (1993), no. 3, 295–314.

29. Ramon van Handel, *Probability in high dimension*, Lecture notes, 2016.

30. Roman Vershynin, *High-Dimensional Probability: An Introduction with Applications in Data Science*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, 2018.