Concentration on the Discrete Torus Using Transportation*

M. Sammer† P. Tetali‡

Abstract

The subgaussian constant of a graph arises naturally in bounding the moment generating function of Lipschitz functions on the graph, with a given probability measure on the set of vertices. The closely related spread constant of a graph measures the maximal variance over all Lipschitz functions on the graph. As such they are both useful (as demonstrated in the works of Bobkov-Houdré-Tetali and Alon-Boppana-Spencer) for describing the concentration of measure phenomenon in product graphs. An equivalent formulation of the subgaussian constant using a transportation inequality, introduced by Bobkov-Götze, is investigated here in depth, leading to a new way of bounding the subgaussian constant. A tight concentration result for the discrete torus is given as a concrete application. An infinite family of graphs is also provided here to demonstrate that typically the spread and the subgaussian constants differ by an order of magnitude.

1 Introduction

Let $G = (V, E)$ be a connected finite graph with a distance (or cost) function d between the vertices. Let π be a probability measure on V. The subgaussian constant $\sigma^2_{\pi,d}(G)$ (or just σ^2 when the context is understood) is defined as the smallest constant satisfying:

$$\mathrm{E}_\pi[e^{t(f-\mathrm{E}_\pi f)}] \leq e^{\sigma^2 t^2/2}, \quad \text{for all } t \in \mathbb{R} \tag{1.1}$$

for every function f on V with Lipschitz constant 1 with respect to d. The subgaussian constant was formally introduced in the context of graphs and studied in [5] (while it was perhaps folklore as well as appearing implicitly in the earlier work of [2]) for measure concentration purposes. It satisfies the tensoring property

$$\sigma^2(G_1 \Box G_2) \leq \sigma^2(G_1) + \sigma^2(G_2),$$

with equality when $G_1 = G_2$. (In the above, on the Cartesian product graph $G_1 \Box G_2$, we take the product measure $\pi_1 \times \pi_2$.) This makes it useful for studying concentration

*Research supported in part by the NSF grants DMS-0401239, DMS-0701043

†The Turing Center, University of Washington, Seattle, WA

‡School of Mathematics and School of Computer Science, Georgia Institute of Technology, Atlanta, GA 30332; tetali@math.gatech.edu

in n-fold product graphs, G^n, with a product measure π^n. Indeed, the inequality (1.1) tensorizes to n-dimensions, and yields using the standard Chebyshev argument the following concentration: for every $n \geq 1$, for every Lipschitz $f : V^n \to \mathbb{R}$, one gets

$$\pi^n\{f - \mathrm{E}_{\pi^n} f \geq h\} \leq e^{-h^2/(2n\sigma^2(G))}. \tag{1.2}$$

Here f is Lipschitz with respect to the l_1-type distance d on G^n.

Bobkov and Götze [4] showed that the subgaussian constant has an equivalent dual formulation. In fact, they showed that the subgaussian constant is also the smallest constant $c = c(G, d, \pi)$ for which a so-called transportation inequality

$$W^2(\nu, \pi) \leq 2c\ D(\nu||\pi) \tag{1.3}$$

holds for all probability measures ν that are absolutely continuous with respect to π. Here $W(\nu, \pi)$ is the Wasserstein distance between ν and π and $D(\nu||\pi)$ is the relative entropy of ν with respect to π (see Section 2 for a full description of these terms). Such transportation inequalities have received considerable attention in the continuous settings of Riemannian manifolds etc...

In [2], Alon, Boppana, and Spencer introduced the spread constant $c^2_{\pi,d}(G)$ (or just c^2 when the context is understood), as:

$$c^2_{\pi,d}(G) = \sup_f \mathrm{Var}_\pi f, \tag{1.4}$$

where the supremum is over f that are Lipschtiz with constant 1 with respect to d. A main result of [2] is that $c^2(G)$ is the optimal constant governing an asymptotically tight isoperimetric inequality – more precisely, in bounding the measure of A^t, for any set A (of vertices) in G^n of measure at least $1/2$, for medium-range enlargements: $\sqrt{n} \ll t \ll n$. It is further observed in [2] and [5] that for every G and π, one has $\sigma^2 \geq c^2$.

While the above motivates the study of σ^2 and c^2 for various graphs, specific computations and bounds on these constants are known for very few cases, such as the complete graph and a path graph. The spread constant of a cycle C_l on $l \geq 3$ vertices has been computed (see Proposition 5.6 in [5]), and shown to be :

$$c^2(C_l) = \begin{cases} \frac{(l^2+8)}{48}, & \text{if } l\text{: even} \\ \\ \frac{(l^2-1)(l^2+3)}{48l^2}, & \text{if } l\text{: odd} \end{cases} \tag{1.5}$$

While the above was done using tedious case analysis, the question of computing the subgaussian of a cycle was left open (see Conjecture 5.1 of [5]).

In the present work, we develop the transportation approach further, and prove some general facts about the transportation inequality, and apply these facts towards establishing a tight inequality for the cycle.

Here is a summary of our main results. In Theorem 3.2, we first show that for any G, if $\sigma^2(G) \neq c^2(G)$, then there is a probability measure ν other than π for which $W^2(\nu, \pi) = 2\sigma^2 D(\nu||\pi)$. We then use this fact (in its contrapositive form) to show that for every $k \geq 1$, for the even cycle C_{2k}, indeed $\sigma^2(C_{2k}) = c^2(C_{2k})$, under the uniform probability measure (see Theorem 5.3). To deal with the odd cycle, in Theorem 5.5, we use the (dual) transportation formulation of the subgaussian constant to show that $c^2(C_{2k+1}) < \sigma^2(C_{2k+1}) = c^2(C_{2k+1})(1+o(1))$, for $k \geq 1$, where $o(1)$ goes to zero, k goes to infinity. In the final section we show that the previous results are in some sense unusual, although perhaps not surprising – that for typical families of graphs $\{G_n\}_{n=1}^{\infty}$, we have $c^2(G_n) \ll \sigma^2(G_n)$ – in particular, for bounded degree expander (family of) graphs G_n on n vertices, we show that $\sigma^2 = \Omega(\log n)$ while $c^2 = \Theta(1)$, independent of n. This is proved as our final Theorem 6.2.

The result on cycles bears some similarity to a recent result of Chen and Sheu [8], who found the exact value of the log-Sobolev constant ρ of the even cycle. They prove that for the even cycle, ρ equals the spectral gap λ_1, using the fact that the inequality $\rho \leq \lambda_1$ is actually an equality if there is no minimizer – that is, if there is no function for which the log-Sobolev inequality is satisfied with equality, but is achieved only in the limit. While we see no formal connection between the two results, our result for the odd cycle does raise the intriguing question of whether $\rho(C_{2k+1}) = \lambda_1(C_{2k+1})(1+o(1))$, for $k \geq 1$, where $o(1)$ disappears as k goes to infinity.

Prior to our work, the subgaussian constant of a 3-cycle C_3 with the uniform measure was shown to be $1/(6\log 2)$ in [5], which is the smallest graph with the uniform measure for which σ^2 is distinct from the spread constant c^2. In [5], the exact values of the subgaussian for a few other graphs were computed, including the 2-point space with arbitrary probability measure, and the completely connected graph and the path of arbitrary length with uniform probability measure. They reduce the problem of finding $\sigma^2(C_3)$ with the uniform measure to finding the subgaussian constant on a non-uniformly weighted path of length two. While this approach extends to finding the subgaussian of a completely connected graph on an arbitrary number of vertices, computing the same on cycles of length *larger than four* remained open! In this work, besides establishing various general properties of optimal measures ν arising from (1.3), we solve the problem of estimating $\sigma^2(C_n)$ asymptotically, irrespective of the parity of n.

Following and extending seminal work by Bollobás-Leader [6, 7], Riordan [13] completed solving the isoperimetric problem on the discrete torus consisting of a product of even cycles, by finding an ordering on the torus for which the initial segments are sets of smallest surface area. These authors note that their proof cannot extend to products of cycles of odd length because the extremal sets in powers of an odd cycle (of the cube of the 3-cycle, for example), are not necessarily nested. Our work is pertinent here, since by finding tight bounds on the subgaussian constant of a cycle, and using tensoring, we get a tight concentration result for the discrete torus without going through the isoperimetric problem.

The results for the cycle are proved in Section 5, while much of the work for the last proposition of Section 5 is contained in a collection of lemmas in Section 4. Section 3 contains lemmas concerning the transportation formulation of the subgaussian constant that are not specific to determining the constant on the cycle, although they are used in the proofs in Section 5. To start off the process, Section 2 gives the necessary definitions and facts, and goes through the proof of the equivalence of the two formulations, (1.1) and (1.3), of the subgaussian constant. Section 6 describes the result concerning expander graphs, and we conclude with a brief discussion of some open problems.

2 Preliminaries

For the graph $G = (V, E)$, let $\mathrm{Lip}(G)$ denote the set of functions on V with Lipschitz constant 1 with respect to the distance function d: for $f \in \mathrm{Lip}(G)$, $|f(x) - f(y)| \leq d(x, y)$, for all $x, y \in V$. Let $P(G)$ denote the set of probability measures on V. For $\nu, \pi \in P(G)$, let $P(\nu, \pi)$ denote the set of probability measures on $V \times V$ with first and second marginals ν and π respectively. For fixed $\nu, \pi \in P(G)$ let

$$M(\mu) = \sum_{x,y \in V} d(x, y)\mu(x, y) \tag{2.6}$$

for any $\mu \in P(\nu, \pi)$. The problem of minimizing $M(\mu)$ over all $\mu \in P(\nu, \pi)$ is the mass transportation problem, originally formulated by Monge in the 18th century. We will refer to this problem simply as Monge's problem, and any $\mu \in P(\nu, \pi)$ which minimizes $M(\mu)$ will be referred to as a solution to Monge's problem with respect to ν and π. In our discrete setting, Monge's problem is simply a linear programming problem, so it has a corresponding dual formulation. Again for fixed $\nu, \pi \in P(G)$ let

$$K(f, g) = \sum_{x \in V} f(x)\nu(x) + g(x)\pi(x)$$

for any functions f and g on V. The problem of maximizing $K(f, g)$ over the set of functions f and g for which $f(x) + g(y) \leq d(x, y)$ for all $x, y \in V$ is the linear programming dual of Monge's problem. We refer to the problem as Kantorovich's problem, because in the 1940's Kantorovich formulated the dual of Monge's problem in general (in the continuous setting where it is not simply a linear programming problem). If the distance function d satisfies the triangle inequality, Kantorovich's problem simplifies to minimizing

$$K(f) = \sum_{x \in V} f(x)(\nu(x) - \pi(x)) \tag{2.7}$$

over $f \in \mathrm{Lip}(G)$. Throughout the rest of the paper we will assume that d satisfies the triangle inequality. Hence any single function $f \in \mathrm{Lip}(G)$ that minimizes $K(f)$ will be referred to as a solution to Kantorovich's problem with respect to ν and π. The joint optimal value of Monge's and Kantorovich's problems is called the Wasserstein distance between ν and π and is denoted by $W(\nu, \pi)$.

The Wasserstein distance is found on the left hand side of the transportation inequality (1.3), while the right hand side contains the relative entropy (or informational divergence) of ν with respect to π, defined by:

$$D(\nu||\pi) = \sum_{x \in V} \nu(x) \log\left(\frac{\nu(x)}{\pi(x)}\right) = \sum_{x \in V} [g(x) \log g(x)]\pi(x),$$

where $g = \nu/\pi$ may be considered the (discrete) density of ν with respect to π. To make sure that the relative entropy is well defined, we only consider ν that are absolutely continuous with respect to π. In the present discrete setting this simply means that $\nu(x) = 0$ if $\pi(x) = 0$. This ensures that we do not divide by zero, except for the case of $\frac{0}{0}$. If $\nu(x) = \pi(x) = 0$, then by convention we let $\frac{\nu(x)}{\pi(x)} = 1$. Also by convention, we define $0\log(0) = 0$, so that $D(\nu||\pi)$ is a continuous function of ν on the subset of measures in $P(G)$ that are absolutely continuous with respect to π. When $\nu = \pi$ we have equality in the transportation inequality since both sides of the inequality are zero and so no restriction is placed on σ^2. Hence in finding the optimal constant we may restrict ourselves to considering $\nu \in P(G) \setminus \{\pi\}$.

We use a couple of notational conveniences that merit mention. First, recall a fairly standard notation that, for a positive f, $\mathrm{Ent}_\pi f := \mathrm{E}_\pi(f\log f) - (\mathrm{E}_\pi f)(\log \mathrm{E}_\pi f)$. Observe that when $f = d\nu/d\pi$, we have $\mathrm{Ent}_\pi f = D(\nu\|\pi)$. Second, because E and Ent will always be taken with respect to the measure π, we will usually omit the π in E_π and Ent_π.

We will need the following standard complementary slackness theorem for dual linear programming problems (see for example Theorem 5.2 in [9]).

Lemma 2.1. *Let f be a Lipschitz function on V and let μ be a probability measure on $V \times V$ with marginals ν and π. Then*

$$f(x) - f(y) = d(x, y) \quad \text{for every } x, y \in V \text{ with } \mu(x, y) > 0 \tag{2.8}$$

if and only if f is a solution to Kantorovich's problem and μ is a solution to Monge's problem both with respect to ν and π.

Solutions to Monge's problem are in general not unique. The following lemma shows that we can always consider a solution to Monge's problem in which mass is never moved both into a vertex and out of the same vertex (given without proof). The fact that the following is true, in any metric space setting where we are trying to minimize the average distance transported, can be inferred from the discussion on the dual problem in the original paper of Kantorovich [11]; it also follows from Lemma 2.11 of McCann-Gangbo [10].

Lemma 2.2. *Suppose $\nu \in P(G)$. Then there exists a solution μ to Monge's problem with respect to ν and π with the following properties for every $y \in V$:*

- *If $\nu(y) \geq \pi(y)$ then $\mu(x, y) > 0$ implies that $x = y$.*
- *If $\nu(y) \leq \pi(y)$ then $\mu(y, z) > 0$ implies that $z = y$.*

In the following we rederive the result from [4] in the present discrete context (for completeness), of the equivalence of the two formulations of the subgaussian inequality, paying close attention to the state of optimality in both formulations. Proposition 2.3 below states their result and the following Proposition 2.4 is our refinement of it.

Proposition 2.3 (Bobkov-Götze)**.** *Let σ be a positive real number. Then the following two statements are equivalent.*

1. $\mathrm{E}_\pi\left[e^{t(f-\mathrm{E}_\pi[f])}\right] \leq e^{\sigma^2 t^2/2}$ *for every Lipschitz function* f *and real number* t.
2. $W^2(\nu,\pi) \leq 2\sigma^2 D(\nu||\pi)$ *for every measure* ν *absolutely continuous with respect to* π.

Proposition 2.4. *Suppose that* σ *is a positive real number for which the two statements in Proposition 2.3 are true. Then we have:*

(a) *Suppose that there exists a Lipschitz function* $\tilde{f}$ *and real number* $\tilde{t} > 0$ *with the property that* $\mathrm{E}\left[e^{\tilde{t}(\tilde{f}-\mathrm{E}[\tilde{f}])}\right] = e^{\sigma^2\tilde{t}^2/2}$. *Define* $\tilde{\nu}$ *by* $d\tilde{\nu} = e^{\tilde{t}(\tilde{f}-\mathrm{E}[\tilde{f}])-\sigma^2\tilde{t}^2/2}d\pi$. *Then we have* $\tilde{\nu} \in P(G)$ *with* $\tilde{\nu} \neq \pi$ *and* $W^2(\tilde{\nu},\pi) = 2\sigma^2 D(\tilde{\nu}||\pi)$. *Furthermore,* $\tilde{f}$ *is a solution to Kantorovich's problem with respect to* $\tilde{\nu}$ *and* π *and* $\tilde{t}^2 = \frac{2}{\sigma^2}D(\tilde{\nu}||\pi)$.

(b) *Suppose there exists* $\tilde{\nu} \in P(G)$ *with* $\tilde{\nu} \neq \pi$ *and* $W^2(\tilde{\nu},\pi) = 2\sigma^2 D(\tilde{\nu}||\pi)$. *Let* $\tilde{f}$ *be a solution to Kantorovich's problem with respect to* $\tilde{\nu}$ *and* π. *Then* $\tilde{f}$ *and* $\tilde{\nu}$ *are related by* $d\tilde{\nu} = e^{\tilde{t}(\tilde{f}-\mathrm{E}[\tilde{f}])-\sigma^2\tilde{t}^2/2}d\pi$ *for* $\tilde{t} = \sqrt{\frac{2}{\sigma^2}D(\tilde{\nu}||\pi)}$. *And we have* $\mathrm{E}\left[e^{\tilde{t}(\tilde{f}-\mathrm{E}[\tilde{f}])}\right] = e^{\sigma^2\tilde{t}^2/2}$.

In the following we prove Proposition 2.4, building on the proof of Proposition 2.3 from [4]. First we observe the following easy, but useful, corollary.

Corollary 2.5. *Suppose* $\tilde{\nu} \in P(G)$ *with* $\tilde{\nu} \neq \pi$ *and* $W^2(\tilde{\nu},\pi) = 2\sigma^2 D(\tilde{\nu}||\pi)$. *Then up to translation there exists a unique solution* $\tilde{f}$ *to Kantorovich's problem with respect to* $\tilde{\nu}$ *and* π. *And for each* $x, y \in V$, $\tilde{f}(x) > \tilde{f}(y)$ *if and only if* $\frac{\tilde{\nu}(x)}{\pi(x)} > \frac{\tilde{\nu}(y)}{\pi(y)}$.

Proof of Corollary 2.5. Suppose $\tilde{f}$ and $\tilde{g}$ are solutions to Kantorovich's problem with respect to $\tilde{\nu}$ and π. Then by Proposition 2.4 they are related by $e^{\tilde{t}(\tilde{f}-\mathrm{E}[\tilde{f}])-\sigma^2\tilde{t}^2/2} = e^{\tilde{t}(\tilde{g}-\mathrm{E}[\tilde{g}])-\sigma^2\tilde{t}^2/2}$. Hence $\tilde{f} - \tilde{g} = E[\tilde{g} - \tilde{f}]$ which is a constant, proving the first part. The second part follows directly from the fact that $\frac{\tilde{\nu}(x)}{\pi(x)} = e^{\tilde{t}(\tilde{f}(x)-\mathrm{E}[\tilde{f}])-\sigma^2\tilde{t}^2/2}$ for each $x \in V$. □

We start with a couple of well-known inequalities. The first is Young's inequality and the second is a fairly standard, but extremely convenient (technically) one which we prove for completeness.

$$\textbf{Young's inequality}: \qquad uv \leq u\, log\, u - u + e^v, \quad u \geq 0, \quad v \in \mathbb{R}\,, \tag{2.9}$$

where equality occurs if and only if $u = e^v$.

Lemma 2.6.

$$\mathrm{E}[e^h] \leq 1 \iff \mathrm{E}[gh] \leq \mathrm{Ent}[g] \text{ for every density } g.$$

And if either side is true, then $\mathrm{E}[gh] = \mathrm{Ent}[g]$ *for some density* g *if and only if* $\mathrm{E}[e^h] = 1$ *and* $g = e^h$.

Proof. Suppose g is a probability density on V with respect to π and that $h : V \to \mathbb{R}$. For $x \in V$, apply Young's inequality (2.9) with $u = g(x)$ and $v = h(x)$ to get:

$$g(x)h(x) \leq g(x) \log g(x) - g(x) + e^{h(x)}.$$

Note that equality holds if and only if $g(x) = e^{h(x)}$. Then we can take expectations of both sides to get:

$$\mathrm{E}[gh] \leq \mathrm{Ent}[g] - 1 + \mathrm{E}[e^h]$$

where there is equality if and only if $g = e^h$ (i.e. $g(x) = e^{h(x)}$ for all $x \in V$).

If $E[e^h] \leq 1$, then

$$\mathrm{E}[gh] \leq \mathrm{Ent}[g]$$

with equality if and only if $\mathrm{E}[e^h] = 1$ and $g = e^h$. So $E[e^h] \leq 1$ implies that $\mathrm{E}[gh] \leq \mathrm{Ent}[g]$ for every density g, with equality if and only if $\mathrm{E}[e^h] = 1$ and $g = e^h$.

Now suppose that for some h we have $\mathrm{E}[gh] \leq \mathrm{Ent}[g]$ for every density g. Choose $c > 0$ so that $\mathrm{E}[ce^h] = 1$. Let $g = ce^h$. Then g is a density and $\mathrm{E}[gh] \leq \mathrm{Ent}[g]$ tells us that

$$c\mathrm{E}[he^h] \leq c\mathrm{E}[e^h(\log c + h)]$$

This implies that $(\log c)\mathrm{E}[e^h] \geq 0$, so $c \geq 1$, and hence $\mathrm{E}[e^h] \leq 1$. Then by the previous paragraph, we have $\mathrm{E}[gh] = \mathrm{Ent}[g]$ if and only if $\mathrm{E}[e^h] = 1$ and $g = e^h$. Hence the lemma. □

Proof of Proposition 2.4. Part (a). First suppose there exists a positive real number σ such that for all real t and Lipschitz f we have

$$\mathrm{E}\left[e^{t(f-\mathrm{E}[f])}\right] \leq e^{\sigma^2 t^2/2}$$

or equivalently

$$\mathrm{E}\left[e^{t(f-\mathrm{E}[f])-\sigma^2 t^2/2}\right] \leq 1.$$

Now suppose there exists a Lipschitz function $\tilde{f}$ and a real number $\tilde{t} > 0$ with the property that

$$\mathrm{E}\left[e^{\tilde{t}(\tilde{f}-\mathrm{E}[\tilde{f}])-\sigma^2 \tilde{t}^2/2}\right] = 1.$$

Note that $\tilde{f}$ cannot be a constant function since $\tilde{t} \neq 0$. Now for any real t and Lipschitz f, we can use the preliminary result above to get:

$$\mathrm{E}\left[\left(t(f - \mathrm{E}[f]) - \sigma^2 t^2/2\right) g\right] \leq \mathrm{Ent}[g]$$

for every density g. Let $\tilde{g} = e^{\tilde{t}(\tilde{f}-\mathrm{E}[\tilde{f}])-\sigma^2\tilde{t}^2/2}$. Then there is equality when $f = \tilde{f}$, $t = \tilde{t}$, and $g = \tilde{g}$. Simplifying and rearranging, we get that for all Lipschitz f and $t > 0$:

$$\mathrm{E}\left[fg - f\right] \leq \frac{\sigma^2 t}{2} + \frac{1}{t}\,\mathrm{Ent}[g]$$

for every density g, with equality when $f = \tilde{f}$, $t = \tilde{t}$, and $g = \tilde{g}$. Now for a fixed non-constant density g consider the function

$$\phi_g(t) = \frac{\sigma^2 t}{2} + \frac{1}{t}\,\mathrm{Ent}[g]$$

defined on positive t. Taking derivatives with respect to t shows us that $\phi_g(t)$ has a unique minimum at

$$t = t^*(g) = \frac{\sqrt{2\,\mathrm{Ent}[g]}}{\sigma}.$$

Now

$$\phi_g(t^*(g)) = \sqrt{2\sigma^2\,\mathrm{Ent}[g]}.$$

So for every Lipschitz f and $t > 0$ we have

$$\mathrm{E}\left[fg - f\right] \leq \sqrt{2\sigma^2\,\mathrm{Ent}[g]} \leq \frac{\sigma^2 t}{2} + \frac{1}{t}\,\mathrm{Ent}[g].$$

for every density g, with equality both places when $f = \tilde{f}$, $t = \tilde{t}$, and $g = \tilde{g}$. Since $\tilde{f}$ is not a constant function and $\tilde{t} \neq 0$ we get that $\tilde{g}$ is not a constant density. So $t^*(\tilde{g})$ is the unique minimum of $\phi_{\tilde{g}}(t)$ giving us $\tilde{t} = t^*(\tilde{g}) = \frac{\sqrt{2\,\mathrm{Ent}[\tilde{g}]}}{\sigma}$, since $\tilde{t}$ also minimizes $\phi_{\tilde{g}}(t)$. Let $d\tilde{\nu} = \tilde{g}d\pi$. Then in terms of probability measures ν instead of densities g, we have that for all Lipschitz f:

$$\sum_{x\in V} f(x)(\nu(x) - \pi(x)) \leq \sqrt{2\sigma^2 \mathrm{D}(\nu||\pi)}.$$

for every probability measure ν absolutely continuous with respect to π. There is equality when $f = \tilde{f}$ and $\nu = \tilde{\nu}$. Finally this tells us that

$$\mathrm{W}^2(\nu, \pi) \leq 2\sigma^2 \mathrm{D}(\nu||\pi)$$

for every ν absolutely continuous with respect to π. There is equality when $\nu = \tilde{\nu}$ and in this case $\tilde{f}$ is a solution to Kantorovich's problem with respect to $\tilde{\nu}$ and π. And $\tilde{t}^2 = \frac{2}{\sigma^2} D(\tilde{\nu}||\pi)$.

Part (b). We start by assuming that there exists a positive real number σ with the property that for all probability measures ν absolutely continuous with respect to π we have

$$\mathrm{W}^2(\nu, \pi) \leq 2\sigma^2 \mathrm{D}(\nu||\pi).$$

Next suppose there exists a probability measure $\tilde{\nu} \neq \pi$ with

$$\mathrm{W}^2(\tilde{\nu}, \pi) = 2\sigma^2 \mathrm{D}(\tilde{\nu}||\pi).$$

Let $\tilde{f}$ be a solution to Kantorovich's problem with respect to $\tilde{\nu}$ and π. Then we get

$$\sum_{x\in V} f(x)(\nu(x) - \pi(x)) \leq \sqrt{2\sigma^2 \mathrm{D}(\nu||\pi)}$$

for every Lipschitz f and ν absolutely continuous with respect to π, with equality if $f = \tilde{f}$ and $\nu = \tilde{\nu}$. Let $\tilde{g}$ be the density of $\tilde{\nu}$ with respect to π. Note that $\tilde{g}$ is not a constant function since $\tilde{\nu} \neq \pi$. Then we can rewrite this in terms of densities g with respect to π instead of measures ν getting:

$$\mathrm{E}\left[fg - f\right] \leq \sqrt{2\sigma^2\,\mathrm{Ent}[g]}\,,$$

for every Lipschitz f and density g, with equality if $f = \tilde{f}$ and $g = \tilde{g}$. Equivalently we can write:

$$\mathrm{E}\left[(f - \mathrm{E}[f])g\right] \leq \sqrt{2\sigma^2 \, \mathrm{Ent}[g]}\,,$$

for every Lipschitz f and density g, with equality if $f = \tilde{f}$ and $g = \tilde{g}$. Furthermore,

$$\mathrm{E}\left[(f - \mathrm{E}[f])g\right] \leq \sqrt{2\sigma^2 \, \mathrm{Ent}[g]} \leq \frac{\sigma^2 t}{2} + \frac{1}{t}\,\mathrm{Ent}[g]\,,$$

for every Lipschitz f, density g, and $t > 0$. Let $\tilde{t} = \frac{\sqrt{2\,\mathrm{Ent}[\tilde{g}]}}{\sigma}$, and note that $\tilde{t} > 0$. Then we have equality everywhere if $f = \tilde{f}$, $g = \tilde{g}$, and $t = \tilde{t}$. So we get

$$\mathrm{E}\left[\left(t(f - \mathrm{E}[f]) - \frac{\sigma^2 t^2}{2}\right) g\right] \leq \mathrm{Ent}[g]\,,$$

for every Lipschitz f, density g, and $t > 0$, with equality when $f = \tilde{f}$, $g = \tilde{g}$, and $t = \tilde{t}$. Then by our preliminary result we have:

$$\mathrm{E}\left[e^{t(f - \mathrm{E}[f]) - \frac{\sigma^2 t^2}{2}}\right] \leq 1\,,$$

for every Lipschitz f and $t > 0$, with equality when $f = \tilde{f}$ and $t = \tilde{t}$. And we have $\tilde{g} = e^{\tilde{t}(\tilde{f} - \mathrm{E}[\tilde{f}]) - \frac{\sigma^2 \tilde{t}^2}{2}}$. Finally we have

$$\mathrm{E}\left[e^{t(f - \mathrm{E}[f])}\right] \leq e^{\frac{\sigma^2 t^2}{2}}$$

for every Lipschitz f and real number t, with equality when $f = \tilde{f}$ and $t = \tilde{t}$. $\square$

3 Facts about the Transportation Formulation

The following lemma is useful for proving that the subgaussian constant and the spread constant are different.

Lemma 3.1. *Let f be a Lipschitz function with $\mathrm{E}[f] = 0$ and $\mathrm{Var}[f] = c^2$. If $\mathrm{E}[f^3] \neq 0$ then $\sigma^2 > c^2$.*

Proof. Let f be a Lipschitz function with $\mathrm{E}[f] = 0$ and $\mathrm{Var}[f] = c^2$. On the subset of $P(G)$ for which the denominator is not zero, we define the function F as:

$$F(\nu) = \frac{D(\nu||\pi)}{\left(\sum_{x \in V} f(x)(\nu(x) - \pi(x))\right)^2}$$

Then for each ν for which $F(\nu)$ is defined, because f is not necessarily a solution to Kantorovich's problem with respect to ν and π, we have

$$F(\nu) \geq \frac{D(\nu||\pi)}{W^2(\nu, \pi)} \geq \frac{1}{2\sigma^2}. \tag{3.10}$$

For positive ϵ small enough that $|\epsilon f(x)| < 1$ for every $x \in V$, define the measure ν_ϵ by $d\nu_\epsilon = (1 + \epsilon f)d\pi$. Consider the following limit:

$$\begin{aligned}
\lim_{\epsilon \to 0} F(\nu_\epsilon) &= \lim_{\epsilon \to 0} \frac{\sum_{x \in V} (1 + \epsilon f(x)) \log(1 + \epsilon f(x)) \pi(x)}{\left(\sum_{x \in V} f(x)[(1 + \epsilon f(x))\pi(x) - \pi(x)]\right)^2} \\
&= \frac{1}{2\mathrm{E}[f^2]} = \frac{1}{2c^2}.
\end{aligned}$$

Let I be an open interval around 0 small enough so that $|\epsilon f(x)| < 1$ for every $x \in V$ and $\epsilon \in I$. Define $H : I \to \mathbb{R}$ by

$$H(\epsilon) = \begin{cases} F(\nu_\epsilon), & \epsilon \neq 0 \\ \frac{1}{2c^2}, & \epsilon = 0 \end{cases}$$

As a real-valued function defined over the (real) interval I, note that H is continuous at 0 by the previous limit. Furthermore we have

$$\left.\frac{d}{d\epsilon} H(\epsilon)\right|_{\epsilon=0} = \frac{-\frac{1}{6}\mathrm{E}[f^3]}{\mathrm{E}[f^2]^2}.$$

Now suppose $\mathrm{E}[f^3] \neq 0$. Then $\left.\frac{d}{d\epsilon} H(\epsilon)\right|_{\epsilon=0} \neq 0$, which implies there exists $\epsilon \neq 0$ with $H(\epsilon) < H(0)$. This means there exists $\nu \in P(G)$ with $\nu \neq \pi$ and $F(\nu) < \frac{1}{2c^2}$. With (3.10) this gives us $\sigma^2 > c^2$. ☐

The following theorem provides a sufficient condition for obtaining *equality* in the transportation inequality, and is of independent interest.

Theorem 3.2. *If $\sigma^2 \neq c^2$ then there exists $\nu \in P(G)$ with $\nu \neq \pi$ and $W^2(\nu, \pi) = 2\sigma^2 D(\nu||\pi)$.*

Proof. Define $F : P(G) \setminus \{\pi\} \to \mathbb{R}$ by

$$F(\nu) = \frac{D(\nu||\pi)}{W^2(\nu, \pi)},$$

so that

$$\frac{1}{2\sigma^2} = \inf_{\nu \in P(G) \setminus \{\pi\}} F(\nu).$$

To prove the theorem we must show that the infimum is attained under our assumption that $\sigma^2 \neq c^2$.

First let us note that F is continuous since $D(\cdot||\pi)$ and $W(\cdot, \pi)$ are continuous, and because $W(\nu, \pi) = 0$ only if $\nu = \pi$. At this point, if $P(G) \setminus \{\pi\}$ were compact, we would be done. We will show that if ν is near π then $F(\nu)$ is too large to be relevant to the infimum. Let us use the l^1 norm so that if, for example, we have $\mu_1, \mu_2 \in P(G)$, then

$$\|\mu_1 - \mu_2\| = \sum_{x \in V} |\mu_1(x) - \mu_2(x)|.$$

Since we assume that $\sigma^2 \neq c^2$ and we know in general that $\sigma^2 \geq c^2$, there exists $\epsilon > 0$ such that $\sigma^2 > (1+\epsilon)c^2$. Let $m = \min\{\pi(x) : x \in V \text{ and } \pi(x) \neq 0\}$. Then let K and δ_1 be positive real numbers with

$$\frac{1}{2} - 3\delta_1 \geq K = \frac{1}{2(1+\epsilon)}.$$

Next let δ_2 be a positive real number small enough that $m - \delta_2 > 0$ and

$$\frac{\delta_2}{m - \delta_2} \leq \delta_1.$$

Let $\nu \in P(G) \setminus \{\pi\}$ with $\|\nu - \pi\| \leq \delta_2$. Let $a(x) = 1 - \nu(x)/\pi(x)$ for $x \in V$. Then $\|a\| \leq \frac{1}{m}\|\nu - \pi\| \leq \frac{\delta_2}{m}$. Let f be a solution to Kantorovich's problem with respect to ν and π with $\mathrm{E}[f] = 0$. Then

$$F(\nu) \quad = \quad \frac{\sum_{x\in V}(1-a(x))\log\left(1-a(x)\right)\pi(x)}{\left(\sum_{x\in V} f(x)a(x)\pi(x)\right)^2}.$$

For each $x \in V$ we have $|a(x)| \leq \|a\| \leq \frac{\delta_2}{m} < 1$, so we may use the Taylor expansion of $\log(1-a(x))$ to get $\log(1-a(x)) = -a(x) - \frac{1}{2}a(x)^2 + R_3(-a(x))$, where $R_3(-a(x))$ is the remainder term. From this we obtain $|R_3(-a(x))| \leq a(x)^2 \frac{|a(x)|}{1-|a(x)|} \leq a(x)^2 \frac{\delta_2}{m-\delta_2} \leq a(x)^2\delta_1$. Since $1 - a(x)$ is positive we have:

$$\begin{aligned}(1-a(x))\log(1-a(x)) &= (1-a(x))[-a(x) - (1/2)a(x)^2 + R_3(-a(x))] \\ &\geq -a(x) + Ka(x)^2.\end{aligned}$$

Hence:

$$\begin{aligned}F(\nu) &\geq \frac{\sum_{x\in V}(-a(x) + Ka(x)^2)\pi(x)}{\left(\sum_{x\in V} f(x)a(x)\pi(x)\right)^2} \\ &\geq \frac{K}{\sum_{x\in V} f(x)^2\pi(x)} \geq \frac{K}{c^2} = \frac{1}{2(1+\epsilon)c^2}. \end{aligned} \tag{3.11}$$

where the first inequality in (3.11) is by Cauchy-Schwarz. Let $B(\pi, \delta_2) = \{\mu \in P(G) : \|\mu - \pi\| < \delta_2\}$. Let $\langle \nu_i \rangle_{i=1}^\infty$ be a sequence of measures in $P(G) \setminus \{\pi\}$ with the property that $F(\nu_i) \to 1/(2\sigma^2)$ as $i \to \infty$. Now

$$\frac{1}{2\sigma^2} < \frac{1}{2(1+\epsilon)c^2}$$

so there exists an integer N so that for all integers $i \geq N$ we have

$$F(\nu_i) < \frac{1}{2(1+\epsilon)c^2}$$

So $\nu_i \in P(G) \setminus B(\pi, \delta_2)$ for all integers $i \geq N$. Hence

$$\inf_{\nu \in P(G)\setminus B(\pi,\delta_2)} F(\nu) = \frac{1}{2\sigma^2}, \tag{3.12}$$

Since $P(G) \setminus B(\pi, \delta_2)$ is compact and F is continuous on $P(G) \setminus B(\pi, \delta_2)$, the infimum in (3.12) is attained. Hence there exists $\nu \neq \pi$ with $W^2(\nu, \pi) = 2\sigma^2 D(\nu\|\pi)$. $\square$

The following two lemmas are inspired by Theorem 2.1 of [2], concerning the function achieving the spread constant.

Lemma 3.3. *Suppose that there exists $\nu \in P(G)$ with $\nu \neq \pi$ and $W^2(\nu, \pi) = 2\sigma^2 D(\nu||\pi)$. Let f be a solution to Kantorovich's problem with respect to ν and π. Then for every $Q \subsetneq V$ we have:*

- *If $\sum_{x \in Q} \nu(x) \geq \sum_{x \in Q} \pi(x)$ then there exists a vertex $x \in Q$ and a vertex $y \notin Q$ with $f(x) - f(y) = d(x, y)$.*
- *If $\sum_{x \in Q} \nu(x) \leq \sum_{x \in Q} \pi(x)$ then there exists a vertex $x \in Q$ and a vertex $y \notin Q$ with $f(x) - f(y) = -d(x, y)$.*

We note that if the distance d under consideration is the graph distance, then the vertices x and y may be taken to be neighbors.

Proof. Let $Q \subsetneq V$, and define the function f_ϵ for each real ϵ by:

$$f_\epsilon(x) = \begin{cases} f(x), & x \notin Q \\ f(x) + \epsilon, & x \in Q \end{cases}$$

Then recalling the function K from Section 2, we have

$$\begin{aligned} K(f_\epsilon) &= \sum_{x \in V} f(x)(\nu(x) - \pi(x)) + \epsilon \sum_{x \in Q} (\nu(x) - \pi(x)) \\ &= K(f) + \epsilon \sum_{x \in Q} (\nu(x) - \pi(x)). \end{aligned}$$

First suppose that $\sum_{x \in Q} \nu(x) > \sum_{x \in Q} \pi(x)$. Since the coefficient of ϵ is positive and f is a solution to Kantorovich's problem with respect to ν and π (i.e. it maximizes the function K over Lipschitz functions), we must have that for every positive ϵ, $f_\epsilon \notin \text{Lip}(G)$. Hence there exists $x \in Q$ and $y \notin Q$ with $f(x) - f(y) = d(x, y)$. Now suppose that $\sum_{x \in Q} \nu(x) < \sum_{x \in Q} \pi(x)$. The coefficient of ϵ is now negative, so we must have that for every negative ϵ, $f_\epsilon \notin \text{Lip}(G)$. Hence there exists $x \in Q$ and $y \notin Q$ with $f(x) - f(y) = -d(x, y)$. Finally suppose that $\sum_{x \in Q} \nu(x) = \sum_{x \in Q} \pi(x)$. Then $K(f_\epsilon) = K(f)$ for every ϵ. Hence f_ϵ is a solution to Kantorovich's problem with respect to ν and π whenever $f_\epsilon \in \text{Lip}(G)$. Since Q is a strict subset of V, f_ϵ is not a translation of f for any $\epsilon \neq 0$. By Corollary 2.5, f is the unique solution to Kantorovich's problem up to translation, so we must have that for every non-zero ϵ, $f_\epsilon \notin \text{Lip}(G)$. The conclusion then follows. $\square$

For this lemma we assume that we are using the graph distance.

Lemma 3.4. *Suppose that there exists $\nu \in P(G)$ with $\nu \neq \pi$ and $W^2(\nu, \pi) = 2\sigma^2 D(\nu||\pi)$. Let f be any solution to Kantorovich's problem with respect to ν and π. Then, up to a possible translation, f is integer-valued and has the property that for some $U \subset V$, $f(x) = \pm d(x, U)$ for all $x \in V$.*

Proof. Let f be a solution to the Kantorovich problem with respect to ν and π. We start by showing that a translation of f will be integer valued. Consider the graph G_f with vertex set V and edge set $E_f \subset E$ where $\{x, y\} \in E_f$ if and only if $\{x, y\} \in E$ and $|f(x) - f(y)| = 1$. Assume to the contrary that G_f is not connected and let C be the set of vertices in one of the connected subgraphs. Then Lemma 3.3 applied to the (original connected) graph G, with C playing the role of the set Q in the lemma, gives us neighbors x and y with $x \in C$ and $y \in V \setminus C$ and $f(x) - f(y) = \pm d(x, y) = \pm 1$. Hence $\{x, y\} \in E_f$ contradicting the assertion that E_f contains no edges between C and $V \setminus C$. So G_f is connected. Hence a translation of f will be integer valued.

For the next part we assume f is integer valued and consider the following set:

$$U = \left\{ x \in V : \nu(x) \geq \pi(x) \text{ and } \frac{\nu(x)}{\pi(x)} \leq \frac{\nu(y)}{\pi(y)} \text{ for all } y \in V \text{ with } \nu(y) \geq \pi(y) \right\}$$

If $x, y \in U$, then $\frac{\nu(x)}{\pi(x)} = \frac{\nu(y)}{\pi(y)}$ and hence $f(x) = f(y)$ by Corollary 2.5. By translating f we may assume that $f(x) = 0$ for all $x \in U$. Let $O = \{|f(x)| : x \in V\}$. Then O contains every integer between 0 and some maximum value.

The proof of the lemma can now be done by induction on $|f(x)|$. For the base case let $x \in V$ with $|f(x)| = 0$. Then $f(x) = f(u)$ for some $u \in U$. Hence $\frac{\nu(x)}{\pi(x)} = \frac{\nu(u)}{\pi(u)}$ again by Corollary 2.5. So $x \in U$ and $d(x, U) = 0$, showing that $f(x) = d(x, U)$. Now let $m \in O$ and assume that $f(x) = \pm d(x, U)$ for any $x \in V$ with $|f(x)| \leq m$. Suppose $m + 1 \in O$. Let $z \in V$ with $|f(z)| = m + 1$.

Case (i). Suppose that $f(z) > 0$. Since $f(z) > f(u)$ for some $u \in U$, we have $\frac{\nu(z)}{\pi(z)} > \frac{\nu(u)}{\pi(u)} \geq 1$ by Corollary 2.5. By Lemma 3.3, since $\nu(z) > \pi(z)$ there exists a neighbor x of z with $f(z) - f(x) = 1$. Then $f(x) = m$ and by the induction hypothesis, $f(x) = d(x, U)$. So $f(z) = d(z, x) + d(x, U) \geq d(z, U)$, by the triangle inequality. Let $u \in U$ with $d(z, u) = d(z, U)$. Then since $f \in \mathrm{Lip}(G)$ we have $d(z, U) = d(z, u) \geq |f(z) - f(u)| = f(z) \geq d(z, U)$. Thus $f(z) = d(z, U)$.

Case(ii). Now consider the case that $f(z) < 0$. Since $f(z) < f(u)$ for some $u \in U$, we have by Corollary 2.5 that $\frac{\nu(z)}{\pi(z)} < \frac{\nu(u)}{\pi(u)}$. So $\nu(z) < \pi(z)$ by the definition of U. Hence there exists a neighbor x of z with $f(z) - f(x) = -1$. So $f(x) = -m$ and $f(x) = -d(x, U)$ by the induction hypothesis. This means that $f(z) = -(d(z, x) + d(x, U)) \leq -d(z, U)$. Let $u \in U$ with $d(z, u) = d(z, U)$. Then since f is Lipschitz we have $d(z, U) = d(z, u) \geq |f(z) - f(u)| = -f(z) \geq d(z, U)$. This gives us $f(z) = -d(z, U)$, concluding the induction step. Hence, $f(x) = \pm d(x, U)$ for all $x \in V$, proving the lemma. □

4 Specific Lemmas for the Method

Lemma 4.1. $(x + y - 1) \log(x + y - 1) \geq x \log(x) + y \log(y)$ *for* $(x, y) \in A$ *where* $A = \{(x, y) \in \mathbb{R}^2 : (x + y \geq 1) \text{ AND } (x, y \geq 1 \text{ OR } x, y \leq 1)\}$.

Proof. Let $f(x, y) = x \log(x) + y \log(y) - (x + y - 1) \log(x + y - 1)$. We must show that $f(x, y) \leq 0$ on A. Note that $f(1, y) = f(x, 1) = 0$ for all $(x, y) \in A$. It suffices to show that $\frac{\partial f}{\partial x}(x,y) \geq 0$ for $(x,y) \in A$ with $x,y \leq 1$ and $\frac{\partial f}{\partial x}(x,y) \leq 0$ for $(x,y) \in A$ with $x,y \geq 1$.

$$\frac{\partial f}{\partial x} = \log\left(\frac{x}{x+y-1}\right).$$

So the lemma follows by noting that $\frac{x}{x+y-1} \geq 1$ for $(x,y) \in A$ with $x,y \leq 1$ and $\frac{x}{x+y-1} \leq 1$ for $(x,y) \in A$ with $x,y \geq 1$. $\square$

We will need the following three lemmas for passing from the case of the even cycle to that of the odd cycle. Let $G = (V,E)$ with $z \in V$ and $z_1, z_2 \notin V$. Let π be a probability measure on V and let $\tilde{\pi}$ be a probability measure on $\tilde{V} = (V \setminus \{z\}) \cup \{z_1, z_2\}$. Assume that $\tilde{\pi}(x) = k\pi(x)$ for $x \in V \setminus \{z\}$ and that $\tilde{\pi}(z_1) = \tilde{\pi}(z_2) = k\pi(z)$, where k is the constant necessary to make $\tilde{\pi}$ a probability measure. Let us note that $k = \frac{1}{1+\pi(z)}$, giving us $k = \frac{n}{n+1}$ when π is the uniform measure on G.

Lemma 4.2. *Let g be a probability density on $\tilde{V}$ with respect to $\tilde{\pi}$, with the above definitions. Define f on V as follows: let $f(x) = g(x)$ for $x \in V \setminus \{z\}$ and $f(z) = g(z_1) + g(z_2) - 1$. Then f is a probability density on V with respect to π. More over, if $g(z_1), g(z_2) \leq 1$ or $g(z_1), g(z_2) \geq 1$, then $\operatorname{Ent}_\pi[f] \geq \frac{1}{k} \operatorname{Ent}_{\tilde{\pi}}[g]$.*

Proof. The fact that f is a probability density with respect to π on V follows from the choice of k mentioned above, and then the rest of the proof is a direct application of Lemma 4.1. $\square$

Let $\nu \in P(G)$ and let μ be a solution to Monge's problem with respect to ν and π. Assume z has the following properties:

- If $\nu(z) \geq \pi(z)$ then $x \in V$ with $\mu(x,z) > 0$ implies that $x = z$.
- If $\nu(z) \leq \pi(z)$ then $x \in V$ with $\mu(z,x) > 0$ implies that $x = z$.

Note that Lemma 2.2 guarantees that we can always find a μ so that these properties are satisfied for any z. Suppose $\tilde{G} = (\tilde{V}, \tilde{E})$ is a graph with distance function $\tilde{d}$ satisfying the triangle inequality and the following conditions:

1. $\tilde{d}(x,y) \geq d(x,y)$ for every $x,y \in V \setminus \{z\}$.
2. $\tilde{d}(x,y) = d(x,y)$ for every $x,y \in V \setminus \{z\}$ with $\mu(x,y) > 0$.
3. $\tilde{d}(x,z_1) \geq d(x,z)$ and $\tilde{d}(x,z_2) \geq d(x,z)$ for every $x \in V \setminus \{z\}$.
4. $\tilde{d}(x,z_1) = d(x,z)$ or $\tilde{d}(x,z_2) = d(x,z)$ for every $x \in V \setminus \{z\}$.

Then we get the following result.

Lemma 4.3. *There exists $\tilde{\nu} \in P(\tilde{G})$ satisfying the following properties:*

1. *$W(\tilde{\nu}, \tilde{\pi}) = kW(\nu, \pi)$.*

2. $\dfrac{\tilde{\nu}(x)}{\tilde{\pi}(x)} = \dfrac{\nu(x)}{\pi(x)}$ *for every* $x \in V \setminus \{z\}$.

3. $\dfrac{\tilde{\nu}(z_1)}{\tilde{\pi}(z_1)} + \dfrac{\tilde{\nu}(z_2)}{\tilde{\pi}(z_2)} - 1 = \dfrac{\nu(z)}{\pi(z)}$

4. *If* $\nu(z) \geq \pi(z)$ *then* $\tilde{\nu}(z_1) \geq \tilde{\pi}(z_1)$ *and* $\tilde{\nu}(z_2) \geq \tilde{\pi}(z_2)$. *If* $\nu(z) \leq \pi(z)$ *then* $\tilde{\nu}(z_1) \leq \tilde{\pi}(z_1)$ *and* $\tilde{\nu}(z_2) \leq \tilde{\pi}(z_2)$.

Proof. Let $V_1 = \{x \in \ V \setminus \{z\} : d(x,z) = \tilde{d}(x,z_1)\}$. Let $V_2 = (V \setminus \{z\}) \setminus V_1$. Then $d(x,z) = \tilde{d}(x,z_2)$ for all $x \in V_2$ by Condition 4 above. We now construct a joint distribution $\tilde{\mu}$ on $\tilde{V} \times \tilde{V}$ with marginals $\tilde{\pi}$ and a measure $\tilde{\nu}$, and show that this $\tilde{\nu}$ satisfies the properties in the statement of the lemma. The idea behind the following construction is fairly obvious – for $i = 1, 2$, vertex z_i plays the role of z as far as the vertices in V_i are concerned; it is inefficient to transport mass between z_1 and vertices in V_2, and analogously between z_2 and those in V_1; on the rest, $\tilde{\mu}$ is identical μ, except for the normalizing factor k.

Define $\tilde{\mu} : \tilde{V} \times \tilde{V} \to \mathbb{R}$ by

$$\begin{array}{ll}
\tilde{\mu}(x,y) = k\mu(x,y) & x, y \in V \setminus \{z\} \\
\tilde{\mu}(z_1,x) = k\mu(z,x) & x \in V_1 \\
\tilde{\mu}(x,z_1) = k\mu(x,z) & x \in V_1 \\
\tilde{\mu}(z_1,x) = 0 & x \in V_2 \\
\tilde{\mu}(x,z_1) = 0 & x \in V_2 \\
\tilde{\mu}(z_2,x) = k\mu(z,x) & x \in V_2 \\
\tilde{\mu}(x,z_2) = k\mu(x,z) & x \in V_2 \\
\tilde{\mu}(z_2,x) = 0 & x \in V_1 \\
\tilde{\mu}(x,z_2) = 0 & x \in V_1 \\
\tilde{\mu}(z_1,z_1) = k\mu(z,z) + k\sum_{x \in V_2} \mu(x,z) & \\
\tilde{\mu}(z_2,z_2) = k\mu(z,z) + k\sum_{x \in V_1} \mu(x,z) & \\
\tilde{\mu}(z_2,z_1) = 0 & \\
\tilde{\mu}(z_1,z_2) = 0 &
\end{array}$$

By direct calculation it can be seen that $\tilde{\pi}$ is the second marginal of $\tilde{\mu}$. Now define $\tilde{\nu} : \tilde{V} \to \mathbb{R}$ by $\tilde{\nu}(x) = \sum_{y \in \tilde{V}} \tilde{\mu}(x,y)$. By definition $\tilde{\nu}$ is the first marginial of $\tilde{\mu}$. Properties 2 and 3, once verified, can be used to show that indeed $\tilde{\nu}$ is a probability measure on $\tilde{V}$.

The verification of Properties 2 and 3 is straightforward, but we include the details for proving the latter of the two:

$$\begin{aligned}
\tilde{\nu}(z_1)+\tilde{\nu}(z_2) &= \sum_{y\in\tilde{V}}\tilde{\mu}(z_1,y)+\sum_{y\in\tilde{V}}\tilde{\mu}(z_2,y)\\
&= \sum_{y\in V_1}\tilde{\mu}(z_1,y)+\sum_{y\in V_2}\tilde{\mu}(z_1,y)+\tilde{\mu}(z_1,z_1)+\tilde{\mu}(z_1,z_2)\\
&\quad+\sum_{y\in V_1}\tilde{\mu}(z_2,y)+\sum_{y\in V_2}\tilde{\mu}(z_2,y)+\tilde{\mu}(z_2,z_1)+\tilde{\mu}(z_2,z_2)\\
&= k\left[\sum_{y\in V_1}\mu(z,y)+0+\mu(z,z)+\sum_{x\in V_2}\mu(x,z)+0\right]\\
&\quad+k\left[0+\sum_{y\in V_2}\mu(z,y)+0+\mu(z,z)+\sum_{x\in V_1}\mu(x,z)\right]\\
&= k\sum_{x\in V}\mu(z,x)+k\sum_{x\in V}\mu(x,z)\\
&= k[\nu(z)+\pi(z)].
\end{aligned}$$

This is what we want after dividing both sides by $k\pi(z)$ and recalling that $\tilde{\pi}(z_1)=\tilde{\pi}(z_2)=k\pi(z)$.

Now we will verify Property 4, working only with z_1 since the calculations are similar for z_2:

$$\begin{aligned}
\tilde{\nu}(z_1) &= k\left[\sum_{y\in V_1}\mu(z,y)+\mu(z,z)+\sum_{x\in V_2}\mu(x,z)\right]\\
&= k\left[\sum_{y\in V_1}\mu(z,y)-\sum_{x\in V_1}\mu(x,z)+\pi(z)\right]
\end{aligned}$$

If $\nu(z)\geq\pi(z)$, then $\mu(x,z)>0$ implies that $x=z$ and so

$$\tilde{\nu}(z_1)=k\left[\sum_{y\in V_1}\mu(z,y)+\pi(z)\right]\geq k\pi(z)=\tilde{\pi}(z_1).$$

If $\nu(z)\leq\pi(z)$, then $\mu(z,y)>0$ implies that $y=z$ and so

$$\tilde{\nu}(z_1)=k\left[-\sum_{x\in V_1}\mu(x,z)+\pi(z)\right]\leq k\pi(z)=\tilde{\pi}(z_1).$$

We finally verify Property 1. Let f be a solution to Kantorovich's problem on G with respect to ν and π. Define $\tilde{f}:\tilde{V}\to\mathbb{R}$ by $\tilde{f}(x)=f(x)$ for $x\in\tilde{V}\setminus\{z_1,z_2\}$ and $\tilde{f}(z_1)=\tilde{f}(z_2)=f(z)$. First let us verify that $\tilde{f}$ is Lipschitz with respect to $\tilde{d}$. Suppose $x,y\in\tilde{V}\setminus\{z_1,z_2\}$. Then by Condition 1 we have:

$$|\tilde{f}(x)-\tilde{f}(y)|=|f(x)-f(y)|\leq d(x,y)\leq\tilde{d}(x,y).$$

By Condition 3, for $i \in \{1,2\}$ and for all $x \in \tilde{V} \setminus \{z_1, z_2\}$ we have:

$$|\tilde{f}(x) - \tilde{f}(z_i)| = |f(x) - f(z)| \leq d(x,z) \leq \tilde{d}(x, z_i).$$

Finally $|\tilde{f}(z_1) - \tilde{f}(z_2)| = 0 \leq \tilde{d}(z_1, z_2)$. So $\tilde{f}$ is Lipschitz with respect to $\tilde{d}$. Next we use Lemma 2.1 to show that $\tilde{f}$ is a solution to Kantorovich's problem and $\tilde{\mu}$ is a solution to Monge's problem both with respect to $\tilde{\nu}$ and $\tilde{\pi}$. Suppose $x, y \in \tilde{V} \setminus \{z_1, z_2\}$ with $\tilde{\mu}(x,y) > 0$. Then by the definition of $\tilde{\mu}$ we must also have $\mu(x,y) > 0$. So by the definition of $\tilde{f}$, Lemma 2.1, and Condition 2 we get:

$$\tilde{f}(x) - \tilde{f}(y) = f(x) - f(y) = d(x,y) = \tilde{d}(x,y)$$

If $\tilde{\mu}(x, z_i) > 0$ for some $x \in \tilde{V} \setminus \{z_1, z_2\}$ and $i \in \{1,2\}$, then $\mu(x,z) > 0$ and $x \in V_i$ so that

$$\tilde{f}(x) - \tilde{f}(z_i) = f(x) - f(z) = d(x,z) = \tilde{d}(x, z_i).$$

Similarly, if $\tilde{\mu}(z_i, y) > 0$ for some $y \in \tilde{V} \setminus \{z_1, z_2\}$ and $i \in \{1,2\}$, then $\mu(z,y) > 0$ and $y \in V_i$ so that

$$\tilde{f}(z_i) - \tilde{f}(y) = f(z) - f(y) = d(z,y) = \tilde{d}(z_i, y).$$

Finally we note that $\tilde{\mu}(z_1, z_2) = \tilde{\mu}(z_2, z_1) = 0$, and $\tilde{f}(z_i) - \tilde{f}(z_i) = 0 = \tilde{d}(z_i, z_i)$ for $i \in \{1,2\}$. Hence for $x, y \in \tilde{V}$, $\tilde{\mu}(x,y) > 0$ implies that $\tilde{f}(x) - \tilde{f}(y) = \tilde{d}(x,y)$. So by Lemma 2.1, $\tilde{\mu}$ is a solution to Monge's problem and $\tilde{f}$ is a solution to Kantorovich's problem, both on $\tilde{G}$ with respect to $\tilde{\nu}$ and $\tilde{\pi}$. And now we can finish the verification of property 1 (using properties 2 and 3):

$$\begin{aligned}
W(\tilde{\nu}, \tilde{\pi}) &= \sum_{x \in \tilde{V}} \tilde{f}(x)(\tilde{\nu}(x) - \tilde{\pi}(x)) \\
&= \sum_{x \in V \setminus \{z\}} \tilde{f}(x)(\tilde{\nu}(x) - \tilde{\pi}(x)) + \tilde{f}(z_1)(\tilde{\nu}(z_1) - \tilde{\pi}(z_1)) + \tilde{f}(z_2)(\tilde{\nu}(z_2) - \tilde{\pi}(z_2)) \\
&= \sum_{x \in V \setminus \{z\}} f(x)(k\nu(x) - k\pi(x)) + f(z)(k\nu(z) + k\pi(z) - k\pi(z) - k\pi(z)) \\
&= k \sum_{x \in V} f(x)(\nu(x) - \pi(x)) = kW(\nu, \pi)\,.
\end{aligned}$$

$\square$

For a solution μ to Monge's problem with respect to ν and π, define the equivalence relation $\sim_\mu$ on V to be the smallest equivalence relation for which $x \sim_\mu y$ if $\mu(x,y) > 0$ or $\mu(y,x) > 0$. Let $\{V_i\}_{i=1}^m$ be the equivalence classes generated by $\sim_\mu$. Let $G_i = (V_i, E_i)$ for $i \in [m]$ be the subgraphs of G induced by V_i. Let π_i be a probability measure on V_i defined by $\pi_i(x) = k_i \pi(x)$ for $x \in V_i$, where k_i is the appropriate constant that makes π_i a probability measure. We will note that if π is the uniform measure on V, then π_i is the uniform measure on V_i and $k_i = \frac{n}{|V_i|}$. Let d_i denote the distance function on G_i defined by $d_i(x,y) = d(x,y)$ for $x, y \in V_i$. Define $\nu_i \in P(G_i)$ by $\nu_i(x) = k_i \nu(x)$, for $x \in V_i$. Then we get the following lemma.

Lemma 4.4.

$$W(\nu, \pi) = \sum_{i=1}^{m} \frac{1}{k_i} W(\nu_i, \pi_i) \,.$$

Proof. Let μ be a solution to Monge's problem with respect to ν and π and let f be a solution to Kantorovich's problem with respect to ν and π. Define $\mu_i : V_i \times V_i \to \mathbb{R}$ by

$$\mu_i(x, y) = k_i \mu(x, y), \text{ for } x, y \in V_i$$

and $f : V_i \to \mathbb{R}$ by

$$f_i(x) = f(x), \text{ for } x \in V_i.$$

By direct calculation, μ_i has first and second marginals ν_i and π_i respectively. By Lemma 2.1, μ_i is a solution to Monge's problem and f_i is a solution to Kantorovich's problem both on G_i with respect to ν_i and π_i. We may now verify the statement of the lemma:

$$\begin{aligned} \sum_{i \in [m]} \frac{1}{k_i} W(\nu_i, \pi_i) &= \sum_{i \in [m]} \frac{1}{k_i} \sum_{x,y \in V_i} d_i(x, y) \mu_i(x, y) \\ &= \sum_{i \in [m]} \sum_{x,y \in V_i} d(x, y) \mu(x, y) = \sum_{x,y \in V} d(x, y) \mu(x, y) = W(\nu, \pi) \,. \end{aligned}$$

□

5 Application to the Cycle

For this entire section, we assume that π is the uniform measure on the graph under consideration, and the distance function is always the graph distance.

Lemma 5.1. *Suppose f is an integer valued Lipschitz function on the vertices of a cycle $C = (V, E)$. Then there exist $a, b \in V$ and a permutation p of V that satisfy the following properties:*

- *$f(p(x))$ is Lipschitz.*
- *$f(p(x))$ is non-decreasing along the two internally disjoint paths from a to b.*

Proof. Let $m = \min_{x \in V} f(x)$ and $M = \max_{x \in V} f(x)$. Let $w_1, w_2 \in V$ with $f(w_1) = m$ and $f(w_2) = M$. Let $[m, M]$ denote the integers between and including m and M. Since f is integer valued and Lipschitz, and because C is a connected graph, $f(V) = [m, M]$. Suppose c is an integer with $m < c < M$ and let $x_1 \in V$ with $f(x_1) = c$. Since $C - x_1$ (which is the graph C with vertex x_1 and all edges incident with x_1 deleted) is still a connected graph and f is Lipschitz on $C - x_1$, we also have $f(V \setminus \{x_1\}) = [m, M]$ and so there exists $x_2 \in V \setminus \{x_1\}$ with $f(x_2) = c$. Hence we can find $V_1, V_2 \subset V$ with $V_1 \cup V_2 = V$, $V_1 \cap V_2 = \{w_1, w_2\}$, and $f(V_1) = f(V_2) = [m, M]$. From this we can form paths P_1 and P_2 (not necessarily subgraphs of C) using vertex sets V_1 and V_2 respectively with the property that f is non-decreasing on each path from w_1 to

w_2. It also follows that f is Lipschitz on each path. We can then form the cycle $J = (V, E(P_1) \cup E(P_2))$, which has the property that f is Lipschitz and non-decreasing on the two internally disjoint paths from w_1 to w_2. Let p be an isomorphism between C and J, and let $a = p^{-1}(w_1)$ and $b = p^{-1}(w_2)$. Then a, b and p are the desired vertices and permutation. □

Lemma 5.2. *Let $C = (V, E)$ be a cycle and suppose there exists $\nu \in P(C)$ with $\nu \neq \pi$ and $W^2(\nu, \pi) = 2\sigma^2 D(\nu||\pi)$. Then there exists $z \in V$ with the property that one of the functions $f(x) = d(x, z)$ or $f(x) = -d(x, z)$ is a solution to Kantorovich's problem with respect to ν and π.*

Proof. By Lemma 3.4, let f be an integer valued solution to Kantorovich's problem with respect to ν and π. Then there exist vertices a and b and a permutation p of V satisfying the properties of Lemma 5.1. Let $\tilde{f}(x) = f(p(x))$. Let $\tilde{\nu}(x) = \nu(p(x))$. Since the image of V under $\tilde{\nu}$ is equal as a multiset to the image of V under ν (and since π is the uniform measure), we have $D(\tilde{\nu}||\pi) = D(\nu||\pi)$. Also,

$$\begin{aligned} W(\nu, \pi) &= \sum_{x \in C} f(x)(\nu(x) - \pi(x)) = \sum_{x \in C} f(p(x))(\nu(p(x)) - \pi(p(x))) \\ &= \sum_{x \in C} \tilde{f}(x)(\tilde{\nu}(x) - \pi(x)) \leq W(\tilde{\nu}, \pi). \end{aligned}$$

This leads us to:

$$\frac{1}{2\sigma^2} = \frac{D(\nu||\pi)}{W^2(\nu, \pi)} \geq \frac{D(\tilde{\nu}||\pi)}{W^2(\tilde{\nu}, \pi)} \geq \frac{1}{2\sigma^2}.$$

And so the inequalities must actually be equalities. This means that $\tilde{\nu}$ gives equality in the transportation inequality and $\tilde{f}$ is a solution to Kantorovich's problem with respect to $\tilde{\nu}$ and π.

Let P_1 and P_2 be the two internally disjoint paths from a to b. Since $\tilde{f}$ is non-decreasing along P_1 and P_2 from a to b, we have $\tilde{f}(a) \leq \tilde{f}(x) \leq \tilde{f}(b)$ for every $x \in V$. Since $\tilde{f}$ is Lipschitz, for any integer c with $\tilde{f}(a) < c < \tilde{f}(b)$ there must exist a (unique) $x_1 \in P_1$ and a (unique) $x_2 \in P_2$ with $\tilde{f}(x_1) = \tilde{f}(x_2) = c$.

Suppose to the contrary that there exist vertices x' and x'' both in the same path P_i (for $i = 1$ or 2) with $\tilde{f}(x') = \tilde{f}(x'') = c$. Since $\tilde{f}$ is non-decreasing along P_i, there exist adjacent x' and x'' with this property. By Corollary 2.5, $\tilde{\nu}(x') = \tilde{\nu}(x'')$ since $\tilde{f}(x') = \tilde{f}(x'')$. Let v' be the neighbor of x' other than x'' and let v'' be the neighbor of x'' other than x', so that the vertices between a and b in P_i appear in the order, $a \cdots v'x'x''v'' \cdots b$, as shown in Figure 1. Then $\tilde{f}(v') \leq \tilde{f}(x') = \tilde{f}(x'') \leq \tilde{f}(v'')$ since $\tilde{f}$ is non-decreasing along P_i. If $\tilde{\nu}(x') = \tilde{\nu}(x'') \geq \pi(x'')$ then by Lemma 3.3 there must exist a neighbor y'' of x'' with $f(y'') < f(x'')$. But x'' has only two neighbors, so this is a contradiction. Similarly, if $\pi(x') \geq \tilde{\nu}(x') = \tilde{\nu}(x'')$ then there exists a neighbor y' of x' with $f(y') > f(x')$ which is again a contradiction. Hence for every integer c with $\tilde{f}(a) < c < \tilde{f}(b)$, for each $i \in \{1, 2\}$, there exists exactly one vertex $x \in P_i$ with $\tilde{f}(x) = c$.

Next assume that there exist three adjacent vertices $r \sim s \sim t \in V$ with $\tilde{f}(r) = \tilde{f}(s) = \tilde{f}(t)$, noting that this is only possible if this joint value is $\tilde{f}(a)$ or $\tilde{f}(b)$. By

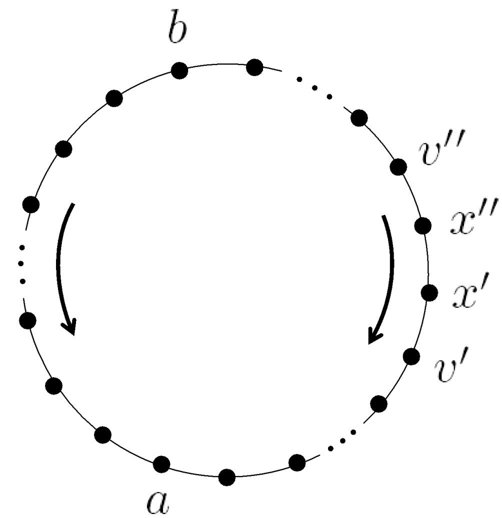

Figure 1: Distance from a Point

Lemma 3.3 (applied with $Q = \{s\}$), we derive that s must have a neighbor x with $\tilde{f}(x) > \tilde{f}(s)$ or $\tilde{f}(x) < \tilde{f}(s)$, which is a contradiction.

Now there are an even number of vertices for which $\tilde{f}$ attains values strictly between $\tilde{f}(a)$ and $\tilde{f}(b)$, and there are at most two vertices which attain the maximum value, $\tilde{f}(b)$, and at most two which attain the minimum value, $\tilde{f}(a)$. So if $|V|$ is odd, there exists only one vertex on which the maximum value of $\tilde{f}$ is attained or one vertex on which the minimum value of $\tilde{f}$ is attained. By translating $\tilde{f}$ so that respectively either the maximum or minimum value is zero, we get that $\tilde{f}(x) = -d(x, b)$ or $\tilde{f}(x) = d(x, a)$.

If $|V|$ is even then we must show there cannot be two vertices on which $\tilde{f}$ attains the maximum value and two vertices on which $\tilde{f}$ attains the minimum value. Let P be a path in C along which $\tilde{f}$ is strictly increasing, starting from one of the vertices on which $\tilde{f}$ attains the minimum value and ending on one of the vertices on which $\tilde{f}$ attains the maximum value. Then by Lemma 3.3 (applied with the vertices of P playing the role of Q in the lemma) one of the following must be true: either the vertex of P which attains the maximum value of $\tilde{f}$ must have a neighbor outside P with a different value of $\tilde{f}$ or the vertex of P which attains the minimum value of $\tilde{f}$ must have a neighbor outside P with a different value of $\tilde{f}$. Hence there can only be one vertex on which $\tilde{f}$ attains the maximum value or one vertex on which $\tilde{f}$ attains the minimum value. But since $|V|$ is even, $\tilde{f}$ attains both the maximum and minimum values at only one vertex. By translating $\tilde{f}$ we can choose to get either $\tilde{f}(x) = -d(x, b)$ or $\tilde{f}(x) = d(x, a)$.

Now $f(V)$ is equal to $\tilde{f}(V)$ as a multiset. But up to rotations of the cycle and translations of the function, there is only one integer valued Lipschitz function on C with this image as a multiset. So f is just a translation and a rotation of $\tilde{f}$. Hence after a possible translation, our original function f must satisfy either $f(x) = d(x, z)$ for some vertex $z \in V$ or $f(x) = -d(x, z)$ for some vertex $z \in V$. $\square$

We now prove the main results on the subgaussian constant of even and odd cycles. The first proof benefits from some observations made in [5] in the course of showing

$\sigma^2(C_4) = c^2(C_4)$.

Theorem 5.3. *If C is a cycle with an even number of vertices, then $\sigma^2(C) = c^2(C)$.*

Proof. Let $C = (V, E)$ be a cycle on $2n$ vertices. Let π be the uniform measure on V so that $\pi(x) = \frac{1}{2n}$ for every $x \in V$. Assume to the contrary that $\sigma^2 \neq c^2$. Let x_0 be an arbitrary vertex in V. Let $f(x) = d(x, x_0)$. From Theorem 3.2, Lemma 5.2 and Proposition 2.4 we know that there exists a $t \neq 0$ such that $\mathrm{E}\left[e^{t(f-\mathrm{E}f)}\right] = e^{\sigma^2 t^2/2}$. So $\sigma^2(C)$ is actually the smallest constant s so that for this particular f, $\mathrm{E}\left[e^{t(f-\mathrm{E}f)}\right] \leq e^{st^2/2}$ for every real number t. Let $L_f(t) = \log \mathrm{E}\left[e^{t(f-\mathrm{E}f)}\right]$. Now $\mathrm{E}[f] = n/2$. So

$$L_f(t) = \log\left[\frac{1}{2n}\left(e^{t(0-n/2)} + 2\sum_{i=1}^{n-1} e^{t(i-n/2)} + e^{t(n-n/2)}\right)\right] \tag{5.13}$$

$$= \begin{cases} \log\left[\frac{1}{2n}\left(e^{\frac{nt}{2}} + e^{-\frac{nt}{2}} - \frac{2e^{-\frac{nt}{2}}(e^t - e^{nt})}{e^t - 1}\right)\right] & t \neq 0 \\ 0 & t = 0\,. \end{cases} \tag{5.14}$$

Consider the function

$$\phi(t) = L_f(t) - \frac{st^2}{2}.$$

Then $\sigma^2(C)$ is the smallest constant s for which $\phi(t) \leq 0$ for every real number t. We will consider the following derivatives of ϕ:

$$\phi'(t) = L_f'(t) - s^2 t, \quad \phi''(t) = L_f''(t) - s^2, \quad \text{and} \quad \phi'''(t) = L_f'''(t)$$

Since $L_f(t)$ is an even function, $\phi(t)$ is also an even function, so $\phi'(t)$ is an odd function and $\phi'(0) = 0$. We also have that $\phi(0) = 0$. Then in order to have $\phi(t) \leq 0$ for all real t, we must have $\phi''(0) \leq 0$, implying $s \geq L_f''(0)$. But, in fact, we will show that if we set $s = L_f''(0)$ then $\phi(t) \leq 0$ for all real t. So the smallest constant s for which $\phi(t) \leq 0$ for all real t is $L_f''(0)$, meaning that $\sigma^2(C) = L_f''(0)$.

To show that $\phi(t) \leq 0$ for all real t when $s = L_f''(0)$ we first restrict ourselves to $t \geq 0$ since $\phi(t)$ is an even function. Then we show that $L_f''(t) < L_f''(0)$ for every $t > 0$. Hence $\phi''(0) = 0$ and $\phi''(t) < 0$ for all $t > 0$, giving us $\phi'(0) = 0$ and $\phi'(t) < 0$ for all $t > 0$, finally giving us $\phi(0) = 0$ and $\phi(t) < 0$ for all $t > 0$.

Now to show that $\phi''(t) < \phi''(0)$ for all $t > 0$, we note that $\phi'''(0) = 0$ again because $\phi(t)$ is an even function. Then we show that $\phi'''(t) < 0$ for all $t > 0$. Now for $t \neq 0$ we have:

$$\phi'''(t) = L_f'''(t) = \frac{1}{(e^t - 1)^3 (e^t + 1)^3 (e^{nt} - 1)^3} \cdot A(t, n),$$

where

$$\begin{aligned} A(t,n) = \Big(& 2e^t + 12e^{3t} + 2e^{5t} + 36e^{(3+2n)t} + 6e^{(5+2n)t} + 6e^{(1+2n)t} \\ & + 3n^3 e^{(2+2n)t} + 3n^3 e^{(4+2n)t} + n^3 e^{(6+2n)t} + n^3 e^{(6+n)t} \\ & - 12e^{(3+3n)t} - 36e^{(3+n)t} - 6e^{(5+n)t} - 2e^{(5+3n)t} - 6e^{(1+n)t} \\ & - 2e^{(1+3n)t} - n^3 e^{nt} - n^3 e^{2nt} - 3n^3 e^{(4+2n)t} - 3n^3 e^{(4+n)t} \Big). \end{aligned}$$

Although we are only interested in positive integers $n \geq 2$, for a fixed t, if we allow n to take on any positive real value then $\phi(t)$ is a differentiable function of n. Now for $n = 2$ we have:

$$\phi'''(t) = -\frac{2e^t(e^t - 1)}{(e^t + 1)^3},$$

and so $\phi'''(t) < 0$ for any $t > 0$ in this case. Finally we show that for every $t > 0$

$$\frac{\partial}{\partial n}\phi'''(t) < 0,$$

for all $n \geq 2$. Hence $\phi(t) < 0$ for all integers $n \geq 2$ and real $t > 0$.

So for $t > 0$ we calculate:

$$\frac{\partial}{\partial n}\phi'''(t) = \frac{-n^2}{8\left(\sinh(nt/2)\right)^4}(nt(2 + \cosh(nt)) - 3\sinh(nt))$$

To show that $\frac{\partial}{\partial n}\phi'''(t) < 0$ for every real $n \geq 2$ and $t > 0$ it suffices to show that

$$\psi(t) = nt(2 + \cosh(nt)) - 3\sinh(nt) > 0,$$

for every real $n \geq 2$ and $t > 0$. We'll start by taking some derivatives:

$$\tfrac{d}{dt}\psi(t) = n(2 - 2\cosh(nt) + nt\sinh(nt)),$$

$$\tfrac{d^2}{dt^2}\psi(t) = n^2(nt\cosh(nt) - \sinh(nt)),$$

$$\tfrac{d^3}{dt^3}\psi(t) = n^4 t\sinh(nt)).$$

Now $\psi(t)$, $\frac{d}{dt}\psi(t)$, and $\frac{d^2}{dt^2}\psi(t)$ are zero when $t = 0$, and $\frac{d^3}{dt^3}\psi(t)$ is strictly positive for $t > 0$. Hence $\frac{d^2}{dt^2}\psi(t)$, $\frac{d}{dt}\psi(t)$ and $\psi(t)$ are all strictly positive for $t > 0$.

Now that we have shown that $\sigma^2(C) = L_f''(0)$ we may take our pick of contradictions. First, we have shown that $\phi(t) < 0$ for $t \neq 0$ when $s = \sigma^2(C) = L_f''(0)$. This contradicts the fact that there exists a $t \neq 0$ for which $\mathrm{E}\left[e^{t(f - \mathrm{E}f)}\right] = e^{\sigma^2 t^2/2}$. Or we could note that $L_g''(0) = \mathrm{Var}[g]$ for any g. From [5], it is known that $\mathrm{Var}[f] = c^2(C)$ for our particular function f. Hence we have shown that $\sigma^2(C) = c^2(C)$, yielding another contradiction.

Thus we may safely conclude that, in fact, $\sigma^2(C) = c^2(C)$. □

Proposition 5.4. *If C is a cycle with an odd number of vertices, then $\sigma^2(C) > c^2(C)$.*

Proof. Let C be a cycle with $2n + 1$ vertices. It is known from [5] that for any vertex x_0, the function $f(x) = d(x, x_0)$ is optimal for the spread constant of this graph, meaning that $\mathrm{Var}[f] = c^2(C)$. Now $\mathrm{E}[f] = \frac{n(1+n)}{1+2n}$. If we set $g(x) = f(x) - \mathrm{E}[f]$, then $\mathrm{Var}[g] = c^2(C)$ and $\mathrm{E}[g] = 0$, but $\mathrm{E}[g^3] = -\frac{n^2(1+n)^2}{2(1+2n)^2} \neq 0$. So by Lemma 3.1 we have $\sigma^2(C) > c^2(C)$. □

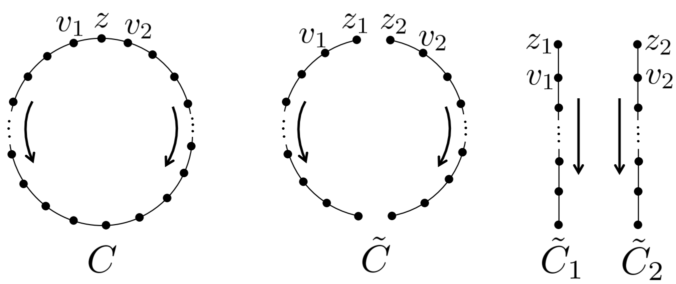

Figure 2: Bounding the Subgaussian of the Odd Cycle

Theorem 5.5. *Suppose that C is a cycle with an odd number n of vertices. Then $\sigma^2(C) = c^2(C)(1 + O(1/n))$.*

Proof. Let $C = (V, E)$ be a cycle on n vertices, where $n \geq 3$ is an odd integer. From Proposition 5.4 we know that $\sigma^2(C) \neq c^2(C)$. Hence, by Theorem 3.2, there exists $\nu \in P(G)$ with $\nu \neq \pi$ and $W^2(\nu, \pi) = 2\sigma^2 D(\nu||\pi)$. Let μ be a solution to Monge's problem with respect to ν and π given to us by Lemma 2.2. By Lemma 5.2, there exists $z \in V$ so that either the function $f(x) = d(x, z)$ or the function $f(x) = -d(x, z)$ is a solution to Kantorovich's problem with respect to ν and π. Let v_1 and v_2 be the two neighbors of z. For $z_1, z_2 \notin V$, let $\tilde{C} = (\tilde{V}, \tilde{E})$ be the graph obtained from C by

- $\tilde{V} = (V \setminus \{z\}) \cup \{z_1, z_2\}$.
- $\tilde{E}$ retains all edges of E, except for the edges $\{v_1, z\}$ and $\{z, v_2\}$. Additionally, we have new edges $\{z_1, v_1\}, \{z_2, v_2\} \in \tilde{E}$.

We will verify that $\tilde{C}$ satisfies the conditions before Lemma 4.3.

For Condition 1, suppose $x, y \in V \setminus \{z\}$. If the distance between x and y in $\tilde{C}$ is infinite, then we are done. Otherwise suppose P is a shortest path between x and y in $\tilde{C}$. Since z_1 and z_2 each have only one neighbor and they are not the endpoints of P, they cannot appear in P. Hence P only contains edges and vertices that appear in C, meaning P is a path in C between x and y. So $\tilde{d}(x, y) \geq d(x, y)$.

For Condition 2 suppose that $x, y \in V \setminus \{z\}$ with $\mu(x, y) > 0$. Let P be a shortest path in C between x and y. We will show that P is also a path in $\tilde{C}$. Since $\mu(x, y) > 0$

we have $f(x) - f(y) = d(x, y)$. Suppose to the contrary that z is a vertex in P. We need to consider two possible cases.

First suppose $f(\cdot) = d(\cdot, z)$. Then we would have $f(x) - f(y) < f(x) - f(z) \leq d(x, z)$. But then we have $d(x, y) < d(x, z)$ contradicting the fact that z is a vertex in the shortest path from x to y. Next assume $f(\cdot) = -d(\cdot, z)$, then we would have $f(x) - f(y) < f(z) - f(y) \leq d(z, y)$. But then we have $d(x, y) < d(z, y)$ again contradicting the fact that $z \in P$. So in fact $z \notin P$, and P only contains vertices that are also vertices of $\tilde{C}$. Because f is Lipschitz and $d(x, y) = f(x) - f(y)$ we get that for every edge $\{s, t\}$ in P, $|f(s) - f(t)| = 1$. This means that $\{s, t\}$ is also an edge of $\tilde{C}$. So P is also a path in $\tilde{C}$.

Finally we verify Conditions 3 and 4. Let $x \in V \setminus \{z\}$. $\tilde{C}$ has two connected components, one containing z_1 and the other containing z_2. By construction, if x is in the component containing z_1 then $\tilde{d}(x, z_1) = d(x, z)$ and $\tilde{d}(x, z_2) = \infty$. Likewise, if x is in the component containing z_2 then $\tilde{d}(x, z_1) = \infty$ and $\tilde{d}(x, z_2) = d(x, z)$.

So Lemma 4.3 gives us a probability measure $\tilde{\nu}$ satisfying the four properties of the lemma. If we look at the equivalence relation $\sim_{\tilde{\mu}}$ defined before Lemma 4.4, we obtain the two graphs $\tilde{C}_1 = (\tilde{V}_1, \tilde{E}_1)$ and $\tilde{C}_2 = (\tilde{V}_2, \tilde{E}_2)$, which are both paths on $(n+1)/2$ vertices. Next, we may apply Lemmas 4.2, 4.3, and 4.4 to help complete the proof of the theorem. Note that in the computation below, while making use of Lemmas 4.2 and 4.3, we recall that $k = \frac{n}{n+1}$, while in making use of Lemma 4.4, we use $k_1 = k_2 = 2$:

$$
\begin{aligned}
W^2(\nu, \pi) &= \left[\frac{n+1}{n} W(\tilde{\nu}, \tilde{\pi})\right]^2 = \left(\frac{n+1}{n}\right)^2 \left(\frac{1}{2} W(\tilde{\nu}_1, \tilde{\pi}_1) + \frac{1}{2} W(\tilde{\nu}_2, \tilde{\pi}_2)\right)^2 \\
&\leq \left(\frac{n+1}{n}\right)^2 \left(\frac{1}{2} W^2(\tilde{\nu}_1, \tilde{\pi}_1) + \frac{1}{2} W^2(\tilde{\nu}_2, \tilde{\pi}_2)\right) \\
&\leq \left(\frac{n+1}{n}\right)^2 \left(\sigma^2(\tilde{C}_1) D(\tilde{\nu}_1 || \tilde{\pi}_1) + \sigma^2(\tilde{C}_2) D(\tilde{\nu}_2 || \tilde{\pi}_2)\right) \\
&= \left(\frac{n+1}{n}\right)^2 \sigma^2\left(P_{\frac{n+1}{2}}\right) \left(D(\tilde{\nu}_1 || \tilde{\pi}_1) + D(\tilde{\nu}_2 || \tilde{\pi}_2)\right) \\
&= \left(\frac{n+1}{n}\right)^2 \sigma^2\left(P_{\frac{n+1}{2}}\right) \\
&\quad \left(\sum_{x \in \tilde{V}_1} \frac{\tilde{\nu}_1(x)}{\tilde{\pi}_1(x)} \log\left(\frac{\tilde{\nu}_1(x)}{\tilde{\pi}_1(x)}\right) \tilde{\pi}_1(x) + \sum_{x \in \tilde{V}_2} \frac{\tilde{\nu}_2(x)}{\tilde{\pi}_2(x)} \log\left(\frac{\tilde{\nu}_2(x)}{\tilde{\pi}_2(x)}\right) \tilde{\pi}_2(x)\right) \\
&= 2\left(\frac{n+1}{n}\right)^2 \sigma^2\left(P_{\frac{n+1}{2}}\right) D(\tilde{\nu} || \tilde{\pi}) \;\leq\; 2\left(\frac{n+1}{n}\right) \sigma^2(P_{\frac{n+1}{2}}) D(\nu || \pi).
\end{aligned}
$$

Hence

$$
\begin{aligned}
\sigma^2(C_n) &\leq \sigma^2(P_{\frac{n+1}{2}}) \left(\frac{n+1}{n}\right) = \frac{n^2 + 3n - 1 - \frac{3}{n}}{48} \\
&= c^2(C_n) \left(1 + \frac{3n^3 - 3n^2 - 3n + 3}{n^4 + 2n^2 - 3}\right) = c^2(C_n)\Big(1 + O\big(\frac{1}{n}\big)\Big).
\end{aligned}
$$

$\square$

Remark. We could have gotten a slightly easier, but weaker, bound :

$$W^2(\nu,\pi) = \left[\frac{n+1}{n}W(\tilde{\nu},\tilde{\pi})\right]^2 \leq \left(\frac{n+1}{n}\right)^2 2\sigma^2(C_{n+1})D(\tilde{\nu}||\tilde{\pi}) \leq 2\frac{n+1}{n}\sigma^2(C_{n+1})D(\nu||\pi), \tag{5.15}$$

resulting in $\sigma^2(C_n) \leq \frac{n+1}{n}\sigma^2(C_{n+1})$. The bound is weaker since $\sigma^2\left(P_{\frac{n+1}{2}}\right) < \sigma^2(C_{n+1})$.

6 Subgaussian and Spread Constant of Expanders

Here we show that the above result, $\sigma^2(C_v) = c^2(C_v)(1+o(1))$ for the cycles C_v on v vertices, is in some sense unusual. We begin with a family of graphs $\{G_n = (V_n, E_n)\}_{n=1}^{\infty}$. For each graph G_n we associate a probability measure π_n on V_n, and we let v_n denote the number of vertices in G_n. We now discover circumstances under which $c^2(G_n) \ll \sigma^2(G_n)$, starting with a lemma bounding σ^2 from below in a manner similar to the lower bound on the spectral gap λ_1 by Alon and Milman in [3].

Lemma 6.1. *Suppose $G = (V, E)$ is a graph with an associated probability measure π. Suppose further that $\pi^* = \min_{x\in V}\pi(x)$ is strictly positive. Then $\sigma^2(G) \geq \frac{D^2}{32\log\frac{1}{\pi^*}}$, where D is the diameter of G.*

Proof. Let $x, y \in V$ with $d(x,y) = D$. Let

$$A_x = \{v \in V : d(x,v) \leq d(v,y)\} \quad \text{and} \quad A_y = \{v \in V : d(x,v) \geq d(v,y)\}$$

Then $\pi(A_x) \geq \frac{1}{2}$ or $\pi(A_y) \geq \frac{1}{2}$. Without loss of generality suppose $\pi(A_x) \geq \frac{1}{2}$. Let $v^* \in A_x$ with the property that $d(y, A_x) = d(y, v^*)$. Then

$$D = d(x,y) \leq d(x,v^*) + d(v^*,y) \leq 2d(v^*,y) = 2d(y,A_x)$$

giving $d(y, A_x) \geq D/2$. So $\{v \in V : d(v, A_x) \geq D/2\}$ is not empty. Then by a standard application (see, for example, Proposition 1.7 of [12]) of (1.2), with $n = 1$, and f being the distance to the set A_x, we get:

$$\pi^* \leq \pi(\{v \in V : d(v,A_x) \geq D/2\}) \leq e^{-\frac{(D/2)^2}{8\sigma^2}}.$$

Solving for σ^2 gives the result. $\square$

Theorem 6.2. *There exist graphs G_n on n vertices, for which $\sigma^2(G_n) = \Omega(\log n)$, while $c^2(G_n) = \Theta(1)$, independent of n.*

Proof. Applying the above lemma to our family of graphs, letting D_n be the diameter of G_n and assuming that $\pi(\cdot)$ is the uniform measure, we get:

$$\sigma^2(G_n) \geq \frac{D_n^2}{32\log v_n}.$$

If $D_n \gg \sqrt{\log v_n}$, then $\sigma^2(G_n)$ goes to infinity with n.

Following [1], recall that there exist (explicit constructions of) bounded degree *expander* graphs on n vertices, for n arbitrarily large, with diameter $\Theta(\log n)$. Thus the subgaussian constant grows at least logarithmically with n for these graphs.

On the other hand, it is easy to see that the spread constant is bounded from above by a constant, independent of n. One way to see this is by observing that the spectral gap $\lambda(G_n)$ of any random walk with uniform stationary distribution $\pi(\cdot)$ on the vertices of such a graph is at most inverse of the spread constant: Indeed, letting $P(x,y)$ denote the probability on edge $\{x,y\} \in E$ of such a walk, we recall that the spectral gap may be written as:

$$\begin{aligned}
\lambda(G_n) &= \inf_{f:E_\pi f=0} \frac{\sum_{x,y\in E}\Big((f(x)-f(y)\Big)^2 \pi(x)P(x,y)}{\operatorname{Var}_\pi f} \\
&\leq \inf_{\substack{f:E_\pi f=0 \\ f:\text{Lipschitz}}} \frac{\sum_{x,y\in E}\Big((f(x)-f(y)\Big)^2 \pi(x)P(x,y)}{\operatorname{Var}_\pi f} \\
&\leq \inf_{\substack{f:E_\pi f=0 \\ f:\text{Lipschitz}}} \frac{1}{\operatorname{Var}_\pi f} = \frac{1}{c^2(G_n)} .
\end{aligned}$$

Since the expander graphs have spectral gap bounded from below by a constant independent of n, this concludes the proof. □

7 Conclusion

We hope the transportation approach developed in the discrete context here finds several more applications in future. Some concrete questions remain unsolved; the exact value of the subgaussian constant of odd cycles is a natural open problem. In [5] tight bounds (up to a factor of four) have been provided for the subgaussian constant of the graph (S_n, ρ) on the set S_n of permutations of n elements, under transposition metric ρ – with adjacency defined by transpositions; it seems a challenging problem to determine the precise value here. A tight transportation inequality for this example or for the closely related one, with ρ replaced by Hamming distance, as studied by Maurey and others (see [5] for comments and references therein), will be quite useful.

Acknowledgments. We thank Wilfrid Gangbo for many helpful discussions. We are also grateful to Robert McCann for providing key references to Lemma 2.2, and to an anonymous referee for a careful reading and valuable feedback.

References

[1] Noga Alon. Eigenvalues and expanders. *Combinatorica*, 6(2):83–96, 1986. Theory of computing (Singer Island, Fla., 1984).

[2] Noga Alon, Ravi Boppana, and Joel Spencer. An asymptotic isoperimetric inequality. *Geom. Funct. Anal.*, 8(3):411–436, 1998.

[3] Noga Alon and Vitali D. Milman. λ_1, isoperimetric inequalities for graphs, and superconcentrators. *J. Combin. Theory Ser. B*, 38(1):73–88, 1985.

[4] Sergey G. Bobkov and Friedrich Götze. Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. *J. Funct. Anal.*, 163(1):1–28, 1999.

[5] Sergey G. Bobkov, Christian Houdré, and Prasad Tetali. The subgaussian constant and concentration inequalities. *Israel J. Math.*, 156:255–283, 2006.

[6] Béla Bollobás and Imre Leader. An isoperimetric inequality on the discrete torus. *SIAM J. Discrete Math.*, 3(1):32–37, 1990.

[7] Béla Bollobás and Imre Leader. Compressions and isoperimetric inequalities. *J. Combin. Theory Ser. A*, 56(1):47–62, 1991.

[8] Guan-Yu Chen and Yuan-Chung Sheu. On the log-Sobolev constant for the simple random walk on the n-cycle: the even cases. *J. Funct. Anal.*, 202(2):473–485, 2003.

[9] Vasek Chvatal. *Linear Programming (Series of Books in the Mathematical Sciences)*. W. H. Freeman, September 1983.

[10] Wilfrid Gangbo and Robert J. McCann. The geometry of optimal transportation. *Acta Math.*, 177(2):113–161, 1996.

[11] Leonid V. Kantorovich. On the translocation of masses. *Dokl. Akad. Nauk SSSR*, 37(7–8):227–229, 1942.

[12] Michel Ledoux. *The Concentration of Measure Phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2001.

[13] Oliver Riordan. An ordering on the even discrete torus. *SIAM J. Discrete Math.*, 11(1):110–127 (electronic), 1998.