APPENDIX

A. Supplementary Proofs

1) Refresher on Euclidean Optimization: Suppose that we wish to solve

$$\min_{x \in \mathcal{C}} f(x), \tag{A.1}$$

where $\mathcal{C}$ is a subset of $\mathbb{R}^d$. Two popular methods to solve this problem in practice are the Frank-Wolfe method and the mirror descent method. In both cases, one relies on an efficiently implementable update.

2) Frank-Wolfe Method: The Frank-Wolfe, or conditional gradient descent method, over a Euclidean space, maintains feasibility by solving a linear program at each iteration. In particular, Frank-Wolfe iterates by solving

$$x_{k+1} = (1-\eta_k)x_k + \eta_k g_k, \tag{A.2}$$

$$g_k = \operatorname*{argmin}_{g \in \mathcal{C}} \langle g, \nabla f(x_k) \rangle. \tag{A.3}$$

Here, if $g_k = \nabla f(x_k)$. then we refer to the method as the Frank-Wolfe (FW) algorithm. On the other hand, if one uses a stochastic approximation to the gradient for g_k, as is common in modern optimization, we refer to the method as the Stochastic Frank-Wolfe (SFW) algorithm.

We have the following standard rate of convergence. For a proof, see Guélat and Marcotte [1986].

Theorem 4. *Suppose that f is convex and has β-Lipschitz gradient and $\eta_k = \frac{2}{2+k}$. Then,*

$$f(x_k) - f(x_*) \leq \frac{L}{k+\xi}, \tag{A.4}$$

for some constant ξ. If it is further assumed that f is α-strongly convex and x_* is the in the relative interior of $\mathcal{C}$, then x_k converges linearly to x_*.

Notice that even in the strongly convex case, if the optimal point lies on the boundary, then convergence is slow. To address this, Wolfe [1970] introduced the away step variant of Frank-Wolfe, which uses an active set of vertices to help the method move in directions that are less parallel to the boundary (i.e., it deals with the zig-zagging phenomenon). In Lacoste-Julien and Jaggi [2015], the authors study this and some other variants of Franke-Wolfe and prove linear convergence in the strongly convex setting.

In the stochastic setting, one must use other strategies to ensure convergence. Another popular idea involves average the gradient at each iteration before solving the linear program Zhang et al. [2020]:

$$g_t = (1-\eta_t)(g_{t-1} + \tilde{\Delta}_t) + \eta_t \nabla \tilde{f}(x_t), \tag{A.5}$$

where $\nabla \tilde{f}(x_t)$ is the stochastic approximation to the gradient of f and $\tilde{\Delta}_t$ is an unbiased estimator of $\nabla f(x_t) - \nabla f(x_{t-1})$. These methods typically converge at a $1/\sqrt{t}$ sublinear rate in the convex setting and a $1/t$ sublinear rate in the strongly convex setting.

3) Mirror Descent: Over Euclidean space, given a strictly convex function Φ, the Bregman divergence is given by

$$D_\Phi(x,y) = \Phi(x) - \Phi(y) - \langle \nabla\Phi(y), x-y \rangle. \tag{A.6}$$

By convexity, $D_\Phi \geq 0$.

The mirror descent (MD) iteration [Beck and Teboulle, 2003] for a function f is

$$x_{k+1} = \operatorname*{argmin}_{x \in \mathcal{C}} \langle \eta \nabla f(x_k), x - x_k \rangle + D_\Phi(x, x_k). \tag{A.7}$$

Solving the minimization yields the other familiar MD iteration

$$\nabla\Phi(x_{k+1}) = \nabla\Phi(x_k) - \eta\nabla f(x_k). \tag{A.8}$$

One popular choice of mirror map is the entropy $\Phi(x) = \sum_i x_i \log x_i$. In this case, the optimization algorithm over the simplex is the exponential weights algorithm [Cesa-Bianchi and Lugosi, 2006]. Indeed, it is not hard to see that

$$\nabla\Phi(x) = (\log x_i + 1)_{i=1}^n, \ \nabla\Phi^*(y) = (e^{y_i - 1})_{i=1}^n \tag{A.9}$$

yields the update

$$x_{k+1} = \exp(\log x_k - \eta\nabla f(x_k)) = x_k \exp(-\eta \nabla f(x_k)) \tag{A.10}$$

If the optimization is constrained over the simplex, where the elements of x_k must sum to one, an additional normalization is considered as $x_{k+1} = x_{k+1}/\|x_{k+1}\|_1$.

B. Laplacian Linear Programs

We begin by consider linear programs over graph Laplacians. This is used as a subroutine in the Frank-Wolfe algorithm. In its most general form, a linear program over graph Laplacians would take the form

$$\min_{\boldsymbol{L} \in \mathcal{L}_n} \langle \boldsymbol{L}, \boldsymbol{C} \rangle. \tag{A.11}$$

Notice that this can be written as a linear program in the variable $\boldsymbol{A}$ by

$$\min_{\operatorname{diag}(\boldsymbol{A}\mathbf{1}) - \boldsymbol{A} \in \mathcal{L}_n} \langle \operatorname{diag}(\boldsymbol{A}\mathbf{1}) - \boldsymbol{A}, \boldsymbol{C} \rangle \tag{A.12}$$

Formulation	Solution
Fixed Trace Ent. Reg.	$\boldsymbol{A}_{ij} = \sum_{ij} \exp\left(\frac{1}{2\epsilon}(\boldsymbol{C}_{ii} + \boldsymbol{C}_{jj} - \boldsymbol{C}_{ij})\right)$ $\sum_{ij} \boldsymbol{A}_{ij} = T$
Fixed Trace	$\mathcal{I} = \text{argmax}_{ij}\, \boldsymbol{C}_{ii} + \boldsymbol{C}_{jj} - \boldsymbol{C}_{ij} - \boldsymbol{C}_{ji}$ $\boldsymbol{A}_{ij} = a_{ij},\ ij \in \mathcal{I},\ a_{ij} \geq 0$ $\sum_{ij} \boldsymbol{A}_{ij} = T$
Fixed Degree Ent. Reg	Sinkhorn [Cuturi, 2013]
Fixed Degree	Interior Point Method
No Constraints	Ill-posed

TABLE I
SOLUTIONS TO REGULARIZED AND UNREGULARIZED LAPLACIAN LINEAR PROGRAMS.

$$\equiv \min_{\substack{\boldsymbol{A}\in\mathcal{H}_n \\ \boldsymbol{A}\geq 0,\ \text{diag}(\boldsymbol{A})=\boldsymbol{0}}} \langle \text{diag}(\boldsymbol{A}\boldsymbol{1}) - \boldsymbol{A}, \boldsymbol{C}\rangle$$

However, this problem may be ill-posed in general without additional constraints. Therefore, we consider optimization over fixed trace and fixed degree Laplacian matrices

$$\min_{\boldsymbol{L}\in\mathcal{L}_n(T)} \langle \boldsymbol{L}, \boldsymbol{C}\rangle \equiv \min_{\substack{\boldsymbol{A}\in\mathcal{H}_n \\ \boldsymbol{A}\geq 0,\ \text{diag}(\boldsymbol{A})=\boldsymbol{0} \\ \sum_{ij}\boldsymbol{A}_{ij}=T}} \langle \text{diag}(\boldsymbol{A}\boldsymbol{1}) - \boldsymbol{A}, \boldsymbol{C}\rangle. \tag{A.13}$$

$$\min_{\boldsymbol{L}\in\mathcal{L}_n(\boldsymbol{d})} \langle \boldsymbol{L}, \boldsymbol{C}\rangle \equiv \min_{\substack{\boldsymbol{A}\in\mathcal{H}_n \\ \boldsymbol{A}\geq 0,\ \text{diag}(\boldsymbol{A})=\boldsymbol{0} \\ \boldsymbol{A}\boldsymbol{1}=\boldsymbol{d}}} \langle \text{diag}(\boldsymbol{A}\boldsymbol{1}) - \boldsymbol{A}, \boldsymbol{C}\rangle. \tag{A.14}$$

We also define some regularized surrogate convex programs using an entropic regularizer. The use of an entropic regularizer has become popular in linear programing to find solutions to problems such as optimal transport Cuturi [2013]. We define the off-diagonal entropy as

$$H_{od}(\boldsymbol{P}) = \sum_{i\neq j} \boldsymbol{P}_{ij}(\log \boldsymbol{P}_{ij} - 1). \tag{A.15}$$

and consider the surrogate convex programs which add $-\lambda H$ to the objectives (A.13) and (A.14).

We collect all of the solutions for the Laplacian linear programs in Table I. We see that the fixed trace linear programs have closed form solutions while the fixed degree linear programs must be solved using other algorithms. These solutions are derived in the appendix.

1) Results in Table I: We first discuss solutions to (A.13) and its regularized variant.

Lemma 5. *The entropically regularized variant of (A.13) has solution $T\boldsymbol{A}/\|\boldsymbol{A}\|_1$, where*

$$\boldsymbol{A}_{ij} = \sum_{ij} \exp\left(\frac{1}{2\epsilon}(\boldsymbol{C}_{ii} + \boldsymbol{C}_{jj} - \boldsymbol{C}_{ij})\right) \tag{A.16}$$

Furthermore, define the set

$$\mathcal{I} = \underset{ij}{\text{argmax}}\, \boldsymbol{C}_{ii} + \boldsymbol{C}_{jj} - \boldsymbol{C}_{ij} - \boldsymbol{C}_{ji}. \tag{A.17}$$

The solution to the linear program (A.11) is any matrix $\boldsymbol{B} = T\boldsymbol{A}/\|\boldsymbol{A}\|_1$, where

$$\boldsymbol{A}_{ij} = \begin{cases} a_{ij}, & ij \in \mathcal{I} \\ 0, & \textit{else}. \end{cases} \tag{A.18}$$

where $a_{ij} \geq 0$, with at least one $ij \in \mathcal{I}$ such that $a_{ij} > 0$.

For the degree constrained case as well as its entropically regularized variants, we can follow the literature on optimal transport and derive a Sinkhorn style algorithm to solve an entropically regularized version of this problem.

In other words, we instead propose to solve the surrogate problem

$$\min_{\substack{\boldsymbol{A}\boldsymbol{1}=\boldsymbol{d} \\ \boldsymbol{A}=\boldsymbol{A}^T}} \langle \boldsymbol{A}, -\boldsymbol{C}\rangle - \epsilon H_{od}(\boldsymbol{A}) + \mathbb{1}(\text{diag}(\boldsymbol{A}) = 0). \tag{A.19}$$

Examining the KKT conditions are

$$\begin{aligned} \boldsymbol{A}_{ij} &= \exp\left(\frac{1}{\epsilon}(\boldsymbol{C}_{ij}) + \boldsymbol{f}_i + \boldsymbol{g}_j\right),\ i \neq j, \\ \boldsymbol{A}\boldsymbol{1} &= \boldsymbol{d}. \end{aligned} \tag{A.20}$$

which is what one would get if one considered an entropic regularization of the second version of the linear program. In any case, we can solve the entropically regularized Laplacian linear program with fixed degree by appealing to the Sinkhorn algorithm [Cuturi, 2013]. We refer to the resulting method as Fixed Degree Laplacian Sinkhorn.

2) Connection Between Laplacian Linear Programs and Optimal Transport: In this case, the linear program is equivalent to

$$\begin{aligned} &\min_{\substack{\boldsymbol{A}\in\mathcal{H}_n \\ \boldsymbol{A}\geq 0,\ \text{diag}(\boldsymbol{A})=\boldsymbol{0} \\ \boldsymbol{A}\boldsymbol{1}=\boldsymbol{d}}} \langle \text{diag}(\boldsymbol{A}\boldsymbol{1}) - \boldsymbol{A}, \boldsymbol{C}\rangle \\ &= \langle \text{diag}(\boldsymbol{d}), \boldsymbol{C}\rangle + \min_{\substack{\boldsymbol{A}\in\mathcal{H}_n \\ \boldsymbol{A}\geq 0,\ \text{diag}(\boldsymbol{A})=\boldsymbol{0} \\ \boldsymbol{A}\boldsymbol{1}=\boldsymbol{d}}} \langle \boldsymbol{A}, -\boldsymbol{C}\rangle. \end{aligned} \tag{A.21}$$

While we could employ various linear programming techniques to solve this problem.

$$\min_{\substack{\boldsymbol{A}\in\mathcal{H}_n \\ \boldsymbol{A}\geq 0,\ \mathrm{diag}(\boldsymbol{A})=\boldsymbol{0} \\ \boldsymbol{A}\boldsymbol{1}=\boldsymbol{d}}} \langle \boldsymbol{A}, -\boldsymbol{C}\rangle. \tag{A.22}$$

We recognize this as an optimal transportation problem with cost matrix $-\boldsymbol{C}$, symmetric marginals $\boldsymbol{d}$, with the additional constraint that the diagonal of the coupling must be zero.

3) Proof of Lemmas 5:

Proof. The constrained convex optimization program (**??**) is equivalent to

$$\min_{\substack{\mathrm{diag}(\boldsymbol{A})=0 \\ \boldsymbol{1}^T\boldsymbol{A}\boldsymbol{1}=T \\ \boldsymbol{A}=\boldsymbol{A}^T}} \langle \mathrm{diag}(\boldsymbol{A}\boldsymbol{1}), \boldsymbol{C}\rangle + \langle \boldsymbol{A}, -\boldsymbol{C}\rangle - \epsilon H_o(\boldsymbol{A}). \tag{A.23}$$

To find the solution subject to the trace T constraint, we will find the KKT conditions. The Lagrangian is

$$\begin{aligned}\mathcal{L}(\boldsymbol{A},\lambda) =& \langle \mathrm{diag}(\boldsymbol{A}\boldsymbol{1}), \boldsymbol{C}\rangle + \langle \boldsymbol{A}, -\boldsymbol{C}\rangle - \epsilon H_o(\boldsymbol{A}) \\ & + \lambda(\sum_{ij} \boldsymbol{A}_{ij} - 1).\end{aligned} \tag{A.24}$$

Using the symmetry of $\boldsymbol{A}$, the first-order KKT condition is

$$\begin{aligned}\frac{\partial}{\partial \boldsymbol{A}_{ij}}\mathcal{L}(\boldsymbol{A}) &= \boldsymbol{C}_{ii} + \boldsymbol{C}_{jj} - \boldsymbol{C}_{ij} - \boldsymbol{C}_{ji} \\ &\quad + \epsilon \log(\boldsymbol{A}_{ij}) + \epsilon\log(\boldsymbol{A}_{ji}) - 2\lambda \\ &= 0.\end{aligned} \tag{A.25}$$

The KKT conditions are therefore

$$\boldsymbol{A}_{ij} = \exp\Big(-\frac{1}{2\epsilon}(\boldsymbol{C}_{ii} + \boldsymbol{C}_{jj} - 2\boldsymbol{C}_{ij}) + \lambda\Big), \tag{A.26}$$

$$\sum_{ij} \boldsymbol{A}_{ij} = T. \tag{A.27}$$

Therefore, λ is the unique real number number such that $\sum_{ij} \boldsymbol{A}_{ij} = T$, or

$$\sum_{ij} \exp\Big(-\frac{1}{2\epsilon}(\boldsymbol{C}_{ii} + \boldsymbol{C}_{jj} - 2\boldsymbol{C}_{ij}) + \lambda\Big) = T. \tag{A.28}$$

Equivalently, the solution is $T\boldsymbol{A}/\|\boldsymbol{A}\|_1$.

We now proceed with the solution to the linear program (A.13). Examining the linear program, we see that we can rewrite the problem as

$$\min_{\substack{\sum_{i\neq j}\boldsymbol{A}_{ij}=T \\ \boldsymbol{A}_{ij}\geq 0}} \sum_{ij} \boldsymbol{A}_{ij}(\boldsymbol{C}_{ii} + \boldsymbol{C}_{jj} - 2\boldsymbol{C}_{ij}). \tag{A.29}$$

This linear program has a well-known solution (see, for example, Exercise 4.8 in Boyd et al. [2004a]).

We note that taking $\epsilon \to 0$ in the solution to the entropically regularized LP, we see that $\boldsymbol{A}$ becomes a matrix supported on the entries $ij \in \mathcal{I}$, where $\mathcal{I}$ is defined by (A.17) and all of the a_{ij} are equal. □

4) Proof of Lemma 1: Therefore, the path between these parametrizations is really just the Euclidean path in $\mathbb{S}^n_+$. In particular, this means that if F is convex as a function of $\boldsymbol{L}$, F is convex as a function of $\mathrm{diag}(\boldsymbol{A}\boldsymbol{1}) - \boldsymbol{A}$, or $F \circ \pi$ is convex. Indeed, letting $\boldsymbol{L}(t) = (1-t)\boldsymbol{L} + t\boldsymbol{L}'$ be a path over Laplacians, we see that $\boldsymbol{L}(t) = \pi(\boldsymbol{A}(t))$, where $\boldsymbol{A}(t) = (1-t)\boldsymbol{A} + t\boldsymbol{A}'$. The derivatives also match

$$\begin{aligned}&\partial_t F(\boldsymbol{L}(t)) = \langle \nabla F(\boldsymbol{L}(t)), \boldsymbol{L}'(t)\rangle \\ &= \langle \nabla F(\pi(\boldsymbol{A}(t))), \boldsymbol{L}' - \boldsymbol{L}\rangle \\ &= \langle \nabla F(\pi(\boldsymbol{A}(t))), \mathrm{diag}((\boldsymbol{A}' - \boldsymbol{A})\boldsymbol{1}) - (\boldsymbol{A}' - \boldsymbol{A})\rangle \\ &= \langle \nabla F(\pi(\boldsymbol{A}(t))), \partial_t \pi(\boldsymbol{A}(t))\rangle \\ &= \partial_t F(\pi(\boldsymbol{A}(t))).\end{aligned} \tag{A.30}$$

C. Other Properties of Graph Laplacians

In practice, nodes with large degree may have undue influence on the spectral properties of the graph Laplacian. Therefore, it is useful to also consider normalized versions of the graph Laplacian. In particular, the symmetric normalized graph Laplacian is given by

$$\overline{\boldsymbol{L}} = \boldsymbol{D}^{-1/2}(\boldsymbol{D} - \boldsymbol{A})\boldsymbol{D}^{-1/2} = \boldsymbol{I} - \boldsymbol{D}^{-1/2}\boldsymbol{A}\boldsymbol{D}^{-1/2}. \tag{A.31}$$

Alternative normalizations include the left and right normalized graph Laplacians, which are $\boldsymbol{D}^{-1}\boldsymbol{L}$ and $\boldsymbol{L}\boldsymbol{D}^{-1}$, respectively.

D. Less Constraints: Variable Trace Linear Programs

In this section, we discuss in more detail what happens to linear programs when we do not constrain the trace or the degree. Suppose we now want to solve the generalization of (A.11) where the trace T is not fixed. In this case, the linear program becomes

$$\min_{\mathrm{diag}(\boldsymbol{A})=0,\ \boldsymbol{A}_{ij}\geq 0} \langle \mathrm{diag}(\boldsymbol{A}\boldsymbol{1}), \boldsymbol{C}\rangle + \langle \boldsymbol{A}, -\boldsymbol{C}\rangle. \tag{A.32}$$

In general, this problem is not well posed. Indeed, writing this linear program in the equivalent form

$$\min_{\boldsymbol{A}_{ij}\geq 0,\ i\neq j} \sum_{ij} \boldsymbol{A}_{ij}(\boldsymbol{C}_{ii} + \boldsymbol{C}_{jj} - 2\boldsymbol{C}_{ij}), \tag{A.33}$$

the solution is either 0 or $-\infty$ depending on the signs of the adjoint operator $\mathcal{A}(\boldsymbol{C})$. Furthermore, the entropically regularized program becomes

$$\min_{\mathrm{diag}(\boldsymbol{A})=0} \langle \mathrm{diag}(\boldsymbol{A}\mathbf{1}), \boldsymbol{C}\rangle + \langle \boldsymbol{A}, -\boldsymbol{C}\rangle - \epsilon H_o(\boldsymbol{A}) \quad \text{(A.34)}$$

Notice that this is now an unconstrained minimization problem. The solution is given by

$$\boldsymbol{A}_{ij} = \exp\Big(- \frac{1}{2\epsilon}(\boldsymbol{G}_{ii} + \boldsymbol{G}_{jj} - 2\boldsymbol{G}_{ij})\Big), \quad \text{(A.35)}$$

Again taking $\epsilon \to 0$, we see that all elements ij such that $\boldsymbol{G}_{ii} + \boldsymbol{G}_{jj} - 2\boldsymbol{G}_{ij} < 0$ blow up, and so the solution to the original LP is again 0 or $-\infty$. We note that the original LP is equivalent to

$$\min_{T, \boldsymbol{L} \in \mathcal{L}_n(T)} \langle \boldsymbol{L}, \boldsymbol{C}\rangle. \quad \text{(A.36)}$$

We could consider adding a Laplacian trace regularization term, which would yield the augmented linear program

$$\min_{\mathrm{diag}(\boldsymbol{A})=0,\ \boldsymbol{A}_{ij}\geq 0} \langle \mathrm{diag}(\boldsymbol{A}\mathbf{1}), \boldsymbol{C}\rangle + \langle \boldsymbol{A}, -\boldsymbol{C}\rangle - \lambda \sum_{ij} \boldsymbol{A}_{ij}. \quad \text{(A.37)}$$

and the entropically regularized program

$$\min_{\mathrm{diag}(\boldsymbol{A})=0} \langle \mathrm{diag}(\boldsymbol{A}\mathbf{1}), \boldsymbol{C}\rangle + \langle \boldsymbol{A}, -\boldsymbol{C}\rangle - \epsilon H_o(\boldsymbol{A}) - \lambda \sum_{ij} \boldsymbol{A}_{ij}. \quad \text{(A.38)}$$

The entropically regularized program has solution

$$\boldsymbol{A}_{ij} = \exp\Big(- \frac{1}{2\epsilon}(\boldsymbol{C}_{ii} + \boldsymbol{C}_{jj} - 2\boldsymbol{C}_{ij}) + \lambda\Big), \quad \text{(A.39)}$$