# MIT Open Access Articles

# High-precision Estimation of Random Walks in Small Space

**Massachusetts Institute of Technology**

# High-precision Estimation of Random Walks in Small Space

AmirMahdi Ahmadinejad
Stanford University
ahmadi@stanford.edu

Jonathan Kelner
Massachusetts Institute of Technology
kelner@mit.edu

Jack Murtagh[*]
Harvard University
jmurtagh@g.harvard.edu

John Peebles[†]
Yale University
john.peebles@yale.edu

Aaron Sidford[‡]
Stanford University
sidford@stanford.edu

Salil Vadhan[§]
Harvard University
salil_vadhan@harvard.edu

April 21, 2020

## Abstract

In this paper, we provide a deterministic $\tilde{O}(\log N)$-space algorithm for estimating random walk probabilities on undirected graphs, and more generally Eulerian directed graphs, to within inverse polynomial additive error ($\epsilon = 1/\mathrm{poly}(N)$) where $N$ is the length of the input. Previously, this problem was known to be solvable by a randomized algorithm using space $O(\log N)$ (following Aleliunas et al., FOCS '79) and by a deterministic algorithm using space $O(\log^{3/2} N)$ (Saks and Zhou, FOCS '95 and JCSS '99), both of which held for arbitrary directed graphs but had not been improved even for undirected graphs. We also give improvements on the space complexity of both of these previous algorithms for non-Eulerian directed graphs when the error is negligible ($\epsilon = 1/N^{\omega(1)}$), generalizing what Hoza and Zuckerman (FOCS '18) recently showed for the special case of distinguishing whether a random walk probability is 0 or greater than $\epsilon$.

We achieve these results by giving new reductions between powering Eulerian random-walk matrices and inverting Eulerian Laplacian matrices, providing a new notion of spectral approximation for Eulerian graphs that is preserved under powering, and giving the first deterministic $\tilde{O}(\log N)$-space algorithm for inverting Eulerian Laplacian matrices. The latter algorithm builds on the work of Murtagh et al. (FOCS '17) that gave a deterministic $\tilde{O}(\log N)$-space algorithm for inverting undirected Laplacian matrices, and the work of Cohen et al. (FOCS '19) that gave a randomized $\tilde{O}(N)$-time algorithm for inverting Eulerian Laplacian matrices. A running theme throughout these contributions is an analysis of "cycle-lifted graphs," where we take a graph and "lift" it to a new graph whose adjacency matrix is the tensor product of the original adjacency matrix and a directed cycle (or variants of one).

**Keywords:** derandomization, space complexity, random walks, Markov chains, Laplacian systems, spectral sparsification, Eulerian graphs

# 1 Introduction

In this paper, we give the first deterministic, nearly logarithmic-space algorithm for accurately estimating random walk probabilities on undirected graphs. Our algorithm extends to Eulerian digraphs, which are directed graphs where the indegree of a vertex $v$ is equal to its outdegree for every vertex $v$. (Note that a random walk on an undirected graph is equivalent to a random walk on the associated Eulerian digraph obtained by replacing each undirected edge $\{u, v\}$ with two directed edges $(u, v)$ and $(v, u)$.) In more detail, our main result is as follows:

**Theorem 1.1** (informally stated (see also Theorem 8.1)). *There is a deterministic, $\tilde{O}(\log(k \cdot N))$-space algorithm that given an Eulerian digraph $G$ (or an undirected graph $G$), two vertices $s, t$, and a positive integer $k$, outputs the probability that a $k$-step random walk in $G$ started at $s$ ends at $t$, to within additive error of $\epsilon$, where $N$ is the length of the input and $\epsilon = 1/\text{poly}(N)$ is any desired polynomial accuracy parameter.*

Estimating random walk probabilities to inverse polynomial accuracy, even in general digraphs, can easily be done by randomized algorithms running in space $O(\log N)$. Since that much space is sufficient to simulate random walks [AKL+]. In fact, estimating random walk probabilities in general digraphs is complete for randomized logspace.[1] The best known deterministic algorithm prior to our work was that of Saks and Zhou [SZ1], which runs in space $O(\log^{3/2} N)$, and works for general digraphs. (See the excellent survey of Saks [Sak] for more discussion of the close connection between randomized space-bounded computation and random walks, as well as the state-of-art in derandomizing such computations up to the mid-1990's.)

For undirected graphs, Murtagh et al. [MRSV2] recently gave a deterministic $\tilde{O}(\log(k \cdot N))$-space algorithm that computes a much weaker form of approximation for random walks: given any subset $S$ of vertices, the algorithm estimates, to within a multiplicative factor of $(1 \pm 1/\text{polylog}(N))$, the *conductance* of the set $S$, namely the probability that a $k$-step random walk started at a random vertex of $S$ (with probability proportional to vertex degrees) ends outside of $S$. Our result is stronger because in undirected graphs, all nonzero conductances can be shown to be of magnitude at least $1/\text{poly}(N)$ and can be expressed as a sum of at most $N^2/4$ random walk probabilities. Consequently, with the same space bound our algorithm can estimate the conductance of any set $S$ to within a multiplicative factor of $(1 \pm 1/\text{poly}(N))$.

Like [MRSV2], our work is part of a larger project, initiated in [MRSV1], that seeks to make progress on the derandomization of space-bounded computation by importing ideas from the literature on time-efficient randomized algorithms for solving graph Laplacian systems [ST1, KMP1, KMP2, KOSZ, LS, PS, CKM+, KLP+, KS, CKP+2, CKP+1, CKK+, AJSS] . While we consider Theorem 1.1 to be a natural derandomization result in its own right, it and our analogous result for solving Eulerian Laplacian systems (Theorem 1.2 below) can also be viewed as a step toward handling general directed graphs and thereby having an almost-complete derandomization of randomized logspace. Indeed, in recent years, nearly linear-time randomized algorithms for estimating properties of random walks on general directed graphs (with polynomial mixing time, which also yield complete problems for randomized logspace [RTV]) were obtained by reduction to solving Eulerian Laplacian systems [CKP+2, CKP+1, CKK+]. A deterministic and sufficiently space-efficient analogue of such a reduction, combined with our results, would put randomized logspace in deterministic space $\tilde{O}(\log N)$.

---

[1] Formally, given $G, s, t, k$, a threshold $\tau$, and a unary accuracy parameter $1^a$, the problem of deciding whether the $k$-step random walk probability from $s$ to $t$ is larger than $\tau + 1/a$ or smaller than $\tau$ is complete for the class BPL of promise problems having randomized logspace algorithms with two-sided error. By binary search over the threshold $\tau$, this problem is log-space equivalent to estimating the same probability to within error $1/a$.

To achieve our main result and prove Theorem 1.1, we provide several results that may be of interest even outside of the space-bounded derandomization context, such as a new notion of spectral approximation and new reductions between estimating random walk probabilities and inverting Laplacian systems. Below we provide more details on how our work relates to both the space-bounded derandomization and the Laplacian solving literature, and describe some of our other contributions.

## 1.1 Derandomization of Space-Bounded Computation

It is widely conjectured that every problem solvable in randomized logarithmic space can also be solved in deterministic logarithmic space (i.e. RL = L, BPL=L for the one-sided and two-sided error versions, respectively) [Sak]. Though this is known to follow from mild assumptions in complexity theory (e.g. that there is a Boolean function in **DSPACE**$(n)$ that requires branching programs of size $2^{\Omega(n)}$ [KvM]), the best known unconditional derandomization is the aforementioned, quarter-century-old result of Saks and Zhou [SZ1], which places randomized logspace in deterministic space $O(\log^{3/2} N)$.

Most of the effort on derandomizing logspace computations over the past three decades has gone towards the design of *pseudorandom generators* that fool *ordered branching programs*. An ordered branching program of *width $w$* and *length $k$* is given by a directed graph on vertex set $[k] \times [w]$. which we view as consisting of $k$ layers of $w$ vertices. All of the edges from the $i$th layer go to the $(i + 1)$'st layer (so there are no edges entering the first layer or exiting the last layer). We call the first vertex of the first layer (i.e. vertex $(1, 1)$) the *start vertex*, and the first vertex of the last layer (i.e. vertex $(k, 1)$) the *accept vertex $t$*. Typically, every vertex in layers 1 to $k - 1$ has outdegree 2, with the two edges labelled by 0 and 1. Intuitively, the vertices in the $i$th layer correspond to possible states of a space-bounded algorithm before it makes its $i$th coin toss, and the two edges lead to its two possible states after that coin toss. The acceptance probability of an ordered branching program is exactly the probability that a random walk from the start vertex $s$ of length $k - 1$ ends at accept vertex $t$. Generating such a truly random walk takes $k - 1$ random bits, so the goal of a pseudorandom generator for ordered branching programs is to generate a walk of length $k - 1$ using a much shorter random seed such that the acceptance probability is preserved up to an additive $\epsilon$. Given such a pseudorandom generator, we can obtain a deterministic algorithm for estimating the acceptance probability by enumerating all seeds of the pseudorandom generator.

A general $O(\log N)$-space computation can have $w = 2^{O(\log N)} = \text{poly}(N)$ states and toss $k = \text{poly}(N)$ coins. The best known pseudorandom generator for such ordered branching programs is the classic generator of Nisan [Nis], which has a seed length of $O(\log^2 N)$ (for any error $\epsilon \geq 1/\text{poly}(N)$) and thus does not directly yield a derandomization of space complexity better than $O(\log^2 N)$ (due to enumerating the seeds), which can be achieved more easily by recursively multiplying the transition matrices between layers. (Multiplying $k$ boolean $w \times w$ matrices to within a final accuracy of $\epsilon$ can be done recursively in space $O((\log k) \cdot (\log w + \log \log(k/\epsilon)))$.) Nisan's generator is also a crucial tool in the algorithm of Saks and Zhou [SZ1].

Due to the long lack of progress in improving Nisan's generator, effort has turned to restricted classes of branching programs, such as those of constant width ($w = O(1)$), with there being significant progress in recent years for the case of width $w = 3$. [SZ2, GMR$^+$, MRT]. Another restriction that has been studied is that of *regular* branching programs, where every vertex in layers $2, \ldots, k$ in the branching program has indegree 2. For this case, Braverman, Rao, Raz, and Yehudayoff [BRRY] give a pseudorandom generator with seed length $\tilde{O}(\log N)$ when $w \leq \text{polylog}(N)$ and $\epsilon \geq 1/\text{polylog}(N)$, which again does not yield a deterministic algorithm that improves upon recursive matrix multiplications.

In contrast, our algorithm for Eulerian graphs can be used to estimate the acceptance probability of a regular branching program in space $\tilde{O}(\log N)$ even when $w = \text{poly}(N)$ and $\epsilon = 1/\text{poly}(N)$. Indeed, by adding edges from the $k$th layer back to the first layer, a regular branching program can be made into an Eulerian graph, without changing the probability that a random walk of length $k - 1$ from $s$ ends at $t$. In addition, our techniques also yield an improved pseudorandom generator for *permutation branching programs* (regular branching programs where the labelling is constrained so that for each $b \in \{0, 1\}$, the edges labelled $b$ form perfect matchings between the vertices in consecutive layers). Specifically, [MPV] use our results to derive a pseudorandom generator with seed length $\tilde{O}(\log N)$ for permutation branching programs (with a single accept vertex in layer $k$) of width $w = \text{poly}(N)$, albeit with error only $\epsilon = 1/\text{polylog}(N)$. Even for the special case of permutation branching programs, seed length $\tilde{O}(\log N)$ was previously only achieved for width $w = \text{polylog}(N)$ [KNP, De, Ste].

It is also worth comparing our result with Reingold's Theorem, which gives a deterministic logspace algorithm for deciding *s-t* connectivity in undirected graphs. Reingold, Trevisan, and Vadhan [RTV] interpreted and generalized Reingold's methods to obtain "pseudoconverging walk generators" for regular digraphs where each edge label forms a permutation of the vertices, as in the permutation branching programs described above. These generators provide a way to use a seed of $O(\log N)$ random bits to generate walks of length $\text{poly}(N)$ that converge to the uniform distribution on the connected component of the start vertex (just like a truly random walk would). Such generators turn out to suffice for deciding *s-t* connectivity on arbitrary Eulerian digraphs. However, these generators are not guaranteed to closely approximate the behavior of a truly random walk at shorter walk lengths. Indeed, even the length of the walks needed for mixing is polynomially larger than with a truly random walk. Nevertheless, one of the techniques we use, the *derandomized square*, originated from an effort to simplify Reingold's algorithm and these pseudoconverging walk generators [RV].

Our work builds on recent papers of Murtagh et al. [MRSV1, MRSV2], which gave deterministic, nearly logarithmic-space algorithms for estimating certain quantities associated with random walks on undirected graphs. Specifically, the first of these papers [MRSV1] gave a "Laplacian solver" (defined below) that implied accurate (to within $1/\text{poly}(N)$ error) estimates of *escape probabilities* (the probability that a random walk from $s$ visits $t$ before visiting another vertex $v$); these again refer to the long-term behavior of random walks, rather than the behavior at a given time below mixing. The second paper [MRSV2] dealt with random walks of a fixed length $k$, but as discussed earlier, only gave a weak approximation to the conductance of subsets of vertices.

## 1.2 Inverting Laplacian Systems

We prove Theorem 1.1 by a novel reduction from estimating $k$-step random walk probabilities to solving linear systems given by graph Laplacians, and giving a small-space algorithm for the latter in the case of Eulerian graphs.

Let $G$ be a digraph on $n$ vertices, and let $\mathbf{W}$ be the $n \times n$ transition matrix for the random walk on $G$. Then we will call $\mathbf{L} = \mathbf{I} - \mathbf{W}$ the *random-walk Laplacian of $G$*.[2] Solving Laplacian systems refers to the problem of given a vector $b \in \mathbb{R}^n$, finding a vector $x \in \mathbb{R}^n$ such that $\mathbf{L}x = b$ (if one exists). Since the matrix $\mathbf{L}$ is not invertible (a stationary distribution of the random walk on $G$ is

---

[2]The standard Laplacian of $G$, which we simply refer to as the *Laplacian of $G$* is $\mathbf{D} - \mathbf{A}$, where $\mathbf{D}$ is the diagonal matrix of outdegrees and $\mathbf{A}$ is the adjacency matrix (where we define $(\mathbf{A})_{ij}$ to be the number of edges from $j$ to $i$ in $G$). Notice that $\mathbf{I} - \mathbf{W} = (\mathbf{D} - \mathbf{A})\mathbf{D}^{-1}$. For undirected graphs, it is common to use a symmetric normalization of the Laplacian given by $\mathbf{D}^{-1/2}(\mathbf{D} - \mathbf{A})\mathbf{D}^{-1/2} = \mathbf{D}^{-1/2}(\mathbf{I} - \mathbf{W})\mathbf{D}^{1/2}$, but the $\mathbf{I} - \mathbf{W}$ formulation will be more convenient for us.

in the kernel), a Laplacian system can be solved by instead computing its *pseudoinverse* $\mathbf{L}^+$, which acts as an inverse on $\text{Image}(\mathbf{L})$, and is zero on the orthogonal complement of $\text{Image}(\mathbf{L})$.

We show that we can compute the pseudoinverse of an Eulerian Laplacian in nearly logarithmic space.

**Theorem 1.2** (informally stated (see also Theorem 7.1))**.** *There is a deterministic, $\tilde{O}(\log N)$-space algorithm that given an Eulerian digraph $G$ with random-walk transition matrix $\mathbf{W}$, outputs a matrix $\widetilde{\mathbf{L}^+}$ whose entries differ from $\mathbf{L}^+$ by at most an additive $\epsilon$, where $N$ is the length of the input and $\epsilon = 1/\text{poly}(N)$ is any desired polynomial accuracy parameter.*

Previously, Cohen et al. [CKP$^+$1, CKK$^+$] showed how to solve Eulerian Laplacian systems by randomized, nearly linear-time algorithms, and Murtagh et al. [MRSV1] showed how to solve undirected Laplacian systems by deterministic, nearly logarithmic-space algorithms. Our proof of Theorem 1.2 draws on all of these works.

As explained below, the extension from undirected graphs (handled by [MRSV1]) to Eulerian graphs (as in Theorem 1.2) is crucial for obtaining our high-precision estimation of random walks (Theorem 1.1) even for the case of undirected graphs. In addition, as discussed earlier, this extension can also be viewed as a step toward handling general directed graphs and thereby having an almost-complete derandomization of randomized logspace. Recall that nearly linear-time randomized algorithms for estimating properties of random walks on general directed graphs were obtained by reduction to solving Eulerian Laplacian systems [CKP$^+$2, CKP$^+$1, CKK$^+$]. A deterministic and sufficiently space-efficient analogue of such a reduction, combined with Theorem 1.2, would put randomized logspace in deterministic space $\tilde{O}(\log N)$, i.e. $\text{BPL} \subseteq \widetilde{\text{L}}$.

We will describe the ideas in the proof of Theorem 1.2 below in Section 1.3. Here we describe our reduction from powering (Theorem 1.1) to computing the pseudoinverse of a Laplacian (Theorem 1.2).

Let $G$ be an $n$-vertex digraph with random-walk transition matrix $\mathbf{W}$. Let $\mathbf{P}_k$ be the adjacency matrix of a $k$-vertex directed path. Note that $\mathbf{P}_k$ is not stochastic, but rather *substochastic* (nonnegative with column sums at most 1), since there are no edges leaving the last vertex of the path (i.e. random walks "die off" when leaving that vertex). Then the $kn \times kn$ matrix $\mathbf{W}' = \mathbf{P}_k \otimes \mathbf{W}$, i.e. the Kronecker product of $\mathbf{P}_k$ and $\mathbf{W}$, is a $k \times k$ block matrix consisting of $n \times n$ blocks that equal $\mathbf{W}$ just below the diagonal and are zero elsewhere. For example, when $k = 4$, we have:

$$\mathbf{W}' = \mathbf{P}_k \otimes \mathbf{W} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ \mathbf{W} & 0 & 0 & 0 \\ 0 & \mathbf{W} & 0 & 0 \\ 0 & 0 & \mathbf{W} & 0 \end{bmatrix}.$$

$\mathbf{W}'$ is also a substochastic matrix describing random walks on a graph with $k$ layers of $n$ vertices each, where there is a bipartite version of $G$ going from the $i$th layer to the $(i+1)$'st layer for each $i = 1, \ldots, k-1$, and again there are no edges leaving the $k$th layer. We call this the *path-lift of G of length k*, or the *path-lifted graph* when $G$ and $k$ are clear from context. (This construction is inspired by the ordered branching programs that arise in space-bounded computation, as described above.)

The "Laplacian" of this layered graph, $\mathbf{L} = \mathbf{I}_{kn} - \mathbf{W}'$, is invertible, and noting that $(\mathbf{W}')^k = 0$, we can calculate its inverse as:

$$\mathbf{L}^{-1} = \mathbf{I}_{nk} + \mathbf{W}' + (\mathbf{W}')^2 + \cdots + (\mathbf{W}')^{k-1} = \mathbf{I}_k \otimes \mathbf{I}_n + \mathbf{P}_k \otimes \mathbf{W} + \mathbf{P}_k^2 \otimes \mathbf{W}^2 + \cdots \mathbf{P}_k^{k-1} \otimes \mathbf{W}^{k-1}. \quad (1)$$

Thinking of $\mathbf{L}^{-1}$ as a block matrix, the term $\mathbf{P}_k^j \otimes \mathbf{W}^j$ places a copy of $\mathbf{W}^j$ in each of the blocks that are in the $j$th diagonal below the main diagonal. (So on the main diagonal are blocks of $\mathbf{I}_n$,

just below the main diagonal are blocks of $\mathbf{W}$, below that $\mathbf{W}^2$, and so on.) For example, for $k = 4$, we can write $\mathbf{L}^{-1}$ in block form as

$$\mathbf{L}^{-1} = \begin{bmatrix} \mathbf{I} & 0 & 0 & 0 \\ \mathbf{W} & \mathbf{I} & 0 & 0 \\ \mathbf{W}^2 & \mathbf{W} & \mathbf{I} & 0 \\ \mathbf{W}^3 & \mathbf{W}^2 & \mathbf{W} & \mathbf{I} \end{bmatrix}.$$

Thus from an accurate estimate of $\mathbf{L}^{-1}$, we can read off accurate estimates of the powers of $\mathbf{W}$. For example, entry $((\ell, t), (1, s))$ of $\mathbf{L}^{-1}$ is exactly the probability that a length $\ell - 1$ random walk in $G$ started at $s$ ends at $t$.

However, we can only apply Theorem 1.2 directly if $\mathbf{L}$ is the Laplacian of an Eulerian graph, and the above $\mathbf{W}'$ is not even stochastic. We can fix this by (a) starting with an Eulerian graph $G$ and (b) considering a *cycle-lifted graph* instead of a path-lifted graph, i.e. considering transition matrix $\mathbf{C}_k \otimes \mathbf{W}$. Additionally, it is convenient to collapse all of the vertices in layer $k$ to a single vertex $v$. Then it turns out from the Laplacian pseudoinverse $\mathbf{L}^+$, it is possible to read off *escape probabilities* — the probability that a random walk from one vertex, say $(1, s)$, visits another vertex, say $(\ell, t)$, before visiting a third vertex, say $v$. The condition "before visiting $v$" allows us to not worry about walks that traverse all the way around the cycle, and thus we get exactly the probability that a length $\ell - 1$ random walk in $G$ started at $s$ ends at $t$.[3]

Note that even if $G$ is undirected, this reduction requires inverting a Laplacian of a directed layered graph. Thus, our extension of the small-space undirected Laplacian solver of [MRSV1] to Eulerian graphs (Theorem 1.2) seems essential for obtaining high-precision estimates of powers even for undirected graphs (Theorem 1.1).

The reduction above from computing powers to inverting also allows us to obtain new algorithms for general digraphs and Markov chains:

**Theorem 1.3** (informally stated (see also Theorem 8.6)). *Given a Markov chain $G$ specified by a stochastic matrix $\mathbf{W}$, two states $s, t$, and a positive integer $k$, we can compute the probability that a $k$-step random walk in $G$ started at $s$ ends at $t$, to within an additive error of $\epsilon$:*

1. *In randomized space $O((\log Nk) \cdot \log(\log_{Nk}(1/\epsilon)))$, or*

2. *In deterministic space $O(\log^{3/2}(Nk) + (\log(Nk)) \cdot \log(\log_{Nk}(1/\epsilon)))$.*

*where $N$ is the length of the input.*

This theorem generalizes one of the results from recent work of Hoza and Zuckerman [HZ] that gave the same bounds for the 1-sided version of the problem, namely distinguishing probability 0 from probability greater than $\epsilon$. For the two-sided version of the problem that we consider, a randomized algorithm using space $O(\log(Nk/\epsilon))$ follows from performing poly$(1/\epsilon)$ random walks and counting how many end at $t$. For deterministic algorithms, the best previous algorithm is from Saks and Zhou [SZ1], which uses space $O(\log(Nk/\epsilon) \cdot \log^{1/2} k)$. Note that Theorem 1.3 has a doubly-logarithmic dependence on $\epsilon$ rather than a singly-logarithmic one. In particular, Saks and Zhou [SZ1] only achieves space $O(\log^{3/2} Nk)$ for $\epsilon = 1/\text{poly}(Nk)$ whereas we achieve it for a much smaller $\epsilon = 1/\exp(\exp(\sqrt{\log Nk}))$.

The proof of Theorem 1.3 begins with the observation that we can approximate $\mathbf{L}^{-1} = (\mathbf{I}_{nk} - \mathbf{P}_k \otimes \mathbf{W})^{-1}$ to within accuracy $1/\text{poly}(N, k)$ in randomized space $O(\log Nk)$ or deterministic space

---

[3]Alternatively (and essentially equivalently), we could note that if $D - A$ is the Laplacian of $G$, then $I_k \otimes D - P_k \otimes A = I_k \otimes D \cdot (I_{nk} - P_k \otimes W)$ is a "row-column diagonally dominant matrix" and apply the reduction from inverting such matrices to pseudo-inverting Eulerian Laplacian systems [CKP$^+$2].

$O(\log^{3/2} Nk)$. Indeed, by Equation (1), it suffices to estimate $\mathbf{I}, \mathbf{W}, \mathbf{W}^2, \ldots, \mathbf{W}^{k-1}$ up to accuracy $\pm 1/\text{poly}(N, k)$.

We then use the fact that matrix inversion has a very efficient error reduction procedure, equivalent to what is commonly known as "preconditioned Richardson iterations". Let $\widetilde{\mathbf{L}^{-1}}$ denote our approximation to $\mathbf{L}^{-1}$ with error $1/\text{poly}(N, k)$. For an appropriate choice of the polynomial error bound, it follows that the "error matrix" $\mathbf{E} = \mathbf{I}_{nk} - \widetilde{\mathbf{L}^{-1}}\mathbf{L}$ has norm at most $1/Nk$ (in, say, spectral norm). Then we can obtain a more accurate estimate of $\mathbf{L}^{-1}$ by using the identity:

$$\mathbf{L}^{-1} = (\mathbf{I}_{nk} - \mathbf{E})^{-1}\widetilde{\mathbf{L}^{-1}} = (\mathbf{I}_{nk} + \mathbf{E} + \mathbf{E}^2 + \mathbf{E}^3 + \cdots)\widetilde{\mathbf{L}^{-1}}.$$

Since $\mathbf{E}$ has norm at most $1/Nk$, the series converges very quickly, and can be truncated at $O(\log_{Nk}(1/\epsilon))$ terms to achieve an approximation to within $\pm\epsilon$. As noted earlier, from such an accurate estimation of $\mathbf{L}^{-1} = (\mathbf{I}_{nk} - \mathbf{P}_k \otimes \mathbf{W})^{-1}$, we can accurately estimate random walks of length $k - 1$.

This same error reduction procedure is also used in our proof of Theorem 1.2 (and also throughout the literature on time-efficient Laplacian solvers), and thus is also key to the high precision estimates we obtain in Theorem 1.1. Although we fixed error $1/\text{poly}(N)$ in the statement of the theorem, we can also achieve smaller error $\epsilon$ at a price of $O((\log N) \cdot \log(\log_N(1/\epsilon)))$ in the space complexity.

Interestingly, early work on randomized space-bounded computation [Gil, Sim, BCP, Jun] reduced the problem of *exactly* computing *hitting probabilities* of Markov chains (the probability that an *infinite* random walk from $s$ ever hits $t$) to computing $(\mathbf{I} - \mathbf{W})^{-1}$ for a substochastic matrix $\mathbf{W}$, and used this to show that *unbounded-error* and *non-halting* randomized logspace is contained in deterministic space $O(\log^2 N)$. As far as we know, ours is the first application of inverting Laplacian systems to estimating finite-time random-walk probabilities to within polynomially small error, and consequently it is also the first application of inverting Laplacian systems to the commonly accepted formulation of randomized logspace (i.e. bounded error and halting).

## 1.3   Complex spectral approximation, cycle-lifted graphs, and powering

We now describe the techniques underlying our space-efficient Eulerian Laplacian inverter (Theorem 1.2). Let $\mathbf{W}$ be the transition matrix for the random walk on an $n$-vertex Eulerian graph $G$, for which we want to estimate $(\mathbf{I} - \mathbf{W})^{-1}$. Because of (a generalization of) the error-reduction procedure described above, it suffices to compute a rough approximation to $(\mathbf{I} - \mathbf{W})^{-1}$. For symmetric matrices (as arising from undirected graphs), a sufficient notion of approximation is *spectral approximation* as introduced by Spielman and Teng [ST2]. Applying the Spielman–Teng notion to symmetric Laplacians $\mathbf{I} - \mathbf{W}$ and $\mathbf{I} - \widetilde{\mathbf{W}}$, we say that $\widetilde{\mathbf{W}}$ is an $\epsilon$-*approximation* of $\mathbf{W}$ (written $\widetilde{\mathbf{W}} \approx_\epsilon \mathbf{W}$) if

$$\forall x \in \mathbb{R}^n \qquad \left| x^\top(\mathbf{W} - \widetilde{\mathbf{W}})x \right| \leq \epsilon \cdot x^\top(\mathbf{I} - \mathbf{W})x = \epsilon \cdot \left( \|x\|^2 - x^\top\mathbf{W}x \right). \qquad (2)$$

Cohen et al. [CKP+1] introduced a generalization of spectral approximation for asymmetric matrices and directed graphs:

$$\forall x, y \in \mathbb{R}^n \qquad \left| x^\top(\mathbf{W} - \widetilde{\mathbf{W}})y \right| \leq \frac{\epsilon}{2} \cdot \left( \|x\|^2 + \|y\|^2 - x^\top\mathbf{W}x - y^\top\mathbf{W}y \right). \qquad (3)$$

In the case of symmetric matrices, their definition is equivalent to the Spielman–Teng notion, and thus we use the same terminology $\epsilon$-*approximation* and notation $\widetilde{\mathbf{W}} \approx_\epsilon \mathbf{W}$ for their notion.

In this paper, we introduce a stronger notion of spectral approximation. Specifically, we say $\widetilde{\mathbf{W}}$ is a *unit-circle $\epsilon$-approximation* of $\mathbf{W}$ (written $\widetilde{\mathbf{W}} \overset{\circ}{\approx}_\epsilon \mathbf{W}$) if

$$\forall x, y \in \mathbb{C}^n \qquad \left| x^*(\mathbf{W} - \widetilde{\mathbf{W}})y \right| \leq \frac{\epsilon}{2} \cdot \left( \|x\|^2 + \|y\|^2 - |x^*\mathbf{W}x + y^*\mathbf{W}y| \right), \tag{4}$$

where $v^*$ refers to the conjugate transpose of $v$. Note that we now allow the vectors to range over $\mathbb{C}^n$ rather than $\mathbb{R}^n$. As we show (see Section 3) this in itself does not make the definition stronger as the earlier notions of [ST1, CKP$^+$1] have equivalent formulations using complex vectors. The more important change is the introduction of the complex magnitude $|\cdot|$ in the term $|x^*\mathbf{W}x + y^*\mathbf{W}y|$.

The significance of this change can be seen by considering an eigenvector $v$ of $\mathbf{W}$ whose eigenvalue $\lambda \in \mathbb{C}$ has magnitude 1. Consider what happens if we set $x = y = v$ in both the Spielman–Teng definition (2) and our definition (4). If $\lambda = 1$ (e.g. $v$ is a stationary distribution of the random walk specified by $\mathbf{W}$), then the right-hand side of the inequality in both cases is zero, so we must have exact equality on the left-hand side, i.e. $v^*\widetilde{\mathbf{W}}v = v^*\mathbf{W}v$. On the other hand, if $\lambda$ is some other root of unity (e.g. an eigenvalue of the $k$-cycle $\mathbf{C}_k$, or in any random walk with periodicity), then only our definition requires exact equality. This also explains our terminology *unit-circle* approximation. It also can be shown that $\widetilde{\mathbf{W}} \overset{\circ}{\approx}_\epsilon \mathbf{W}$ if and only if $z\widetilde{\mathbf{W}} \approx_\epsilon z\mathbf{W}$ for all complex $z$ of magnitude 1. That is, our definition amounts to demanding that all unit-circle multiples of the matrices approximate each other in the previous sense. In the case of symmetric matrices (undirected graphs), it suffices to consider $z = \pm 1$, corresponding to the fact that the eigenvalues are all real and the only periodicity that can occur is 2.

The benefit of unit-circle approximation is that, unlike the previous notions of spectral approximation, it is preserved under cycle-lifts and powers.

**Lemma 1.4.** *Suppose $\widetilde{\mathbf{W}} \overset{\circ}{\approx}_\epsilon \mathbf{W}$. Then for all $k \in \mathbb{N}$, we have:*

*1. $\mathbf{C}_k \otimes \widetilde{\mathbf{W}} \overset{\circ}{\approx}_\epsilon \mathbf{C}_k \otimes \mathbf{W}$, and*

*2. $\widetilde{\mathbf{W}}^k \overset{\circ}{\approx}_{\epsilon/(1-\epsilon)} \mathbf{W}^k$.*

We note that in previous work ([CCL$^+$, MRSV1]), it was observed that for symmetric matrices, if $\widetilde{\mathbf{W}} \approx_\epsilon \mathbf{W}$ and $-\widetilde{\mathbf{W}} \approx_\epsilon -\mathbf{W}$ then we do get $\widetilde{\mathbf{W}}^2 \approx_\epsilon \mathbf{W}^2$. Lemma 1.4 holds even for asymmetric matrices and handles all powers $k$.

Item 1 is proven by observing that the diagonalization of $\mathbf{C}_k$ (using the discrete Fourier basis, which are its eigenvectors) has all $k$th roots of unity along the diagonal, so approximation of the cycle-lifted graphs $\mathbf{C}_k \otimes \widetilde{\mathbf{W}}$ and $\mathbf{C}_k \otimes \mathbf{W}$ amounts to requiring that the approximation of $\widetilde{\mathbf{W}}$ and $\mathbf{W}$ is preserved under multiplication by $k$'th roots of unity. Item 2 is derived from Item 1 by observing that the $k$th powers can be obtained by "shortcutting" random walks through all but one layer of the cycle-lifted graphs. This shortcutting amounts to taking the Schur complements of the corresponding Laplacians, and we show that taking Schur complements of Eulerian Laplacians preserves spectral approximation (generalizing analogous results for undirected and symmetrized Schur complements [MP, CKK$^+$]).

Now we can sketch our algorithm that we use to prove Theorem 1.2. Given an Eulerian digraph $G$, we want to approximate the pseudoinverse of the Laplacian $\mathbf{I}_n - \mathbf{W}$. By standard reductions, we may assume that $G$ is regular, connected, and aperiodic, and therefore, it has polynomial mixing time. Rather than directly approximating the inverse of the Laplacian $\mathbf{I}_n - \mathbf{W}$ of the original graph, we instead approximate the inverse of the Laplacian of the cycle-lifted graph, i.e. $\mathbf{I}_{2^k \cdot n} - \mathbf{C}_{2^k} \otimes \mathbf{W}$, for $2^k$ larger than the mixing time of $\mathbf{W}$. Then the pseudoinverse of $\mathbf{I}_n - \mathbf{W}$ can be well-approximated by an appropriate $n \times n$ projection of the pseudoinverse of $\mathbf{I}_{2^k \cdot n} - \mathbf{C}_{2^k} \otimes \mathbf{W}$.

To approximate the pseudoinverse of $\mathbf{I}_{2^k \cdot n} - \mathbf{C}_{2^k} \otimes \mathbf{W}$, we follow the recent approach of [CKK$^+$] and recursively compute an LU factorization (i.e. a product of a lower-triangular and upper-triangular matrix) that approximates $\mathbf{I}_{2^k \cdot n} - \mathbf{C}_{2^k} \otimes \mathbf{W}$, as LU factorizations can be easily inverted. Each recursive step reduces the task to computing an LU factorization of the Laplacian of the random-walk on a chosen subset $S$ of the vertices, where we short-cut steps of the walk through $S^c$. For our algorithm, we choose the set $S$ to consist of every other layer of the cycle-lifted graph, as opposed to using a randomly chosen and pruned set of vertices as in [CKK$^+$]. Then shortcutting walks through $S^c$ yields a graph on $S$ whose transition matrix is equal to $\mathbf{C}_{2^{k-1}} \otimes \mathbf{W}^2$ — a cycle-lifted version of the two-step random walk, with a cycle of half the length! Unfortunately, we can't just directly recurse, because repeatedly squaring $\mathbf{W}$ $k$ times takes space $O((k \cdot \log N))$. Thus, following [MRSV1], we utilize the "derandomized square" of [RV], which produces an explicit sparse $\epsilon$-approximation to $\mathbf{W}^2$ such that $k$ iterated derandomized squares can be computed in space $O(\log N + k \cdot \log(1/\epsilon)) = \tilde{O}(\log N)$. (We take $\epsilon = 1/O(k)$ so that we can tolerate incurring an $\epsilon$ error in approximation for each of the $k$ levels of recursion.) To make the analysis work, we prove that for regular digraphs, the derandomized square produces a graph $\widetilde{\mathbf{W}^2}$ that is a unit-circle approximation to $\mathbf{W}^2$, so that we can deduce that $\mathbf{C}_{2^{k-1}} \otimes \widetilde{\mathbf{W}^2}$ approximates $\mathbf{C}_{2^{k-1}} \otimes \mathbf{W}^2$ via Lemma 1.4. Previous work [MRSV1] only showed approximation for undirected graphs, and with respect to the original Spielman-Teng notion of spectral approximation. (Rozenman and Vadhan [RV] showed that for regular digraphs, the derandomized square improves spectral expansion nearly as much as the true square, but that is weaker than spectral approximation, as it only refers to the 2nd singular value rather than the entire spectrum.)

We remark that the $n \times n$ projection of the pseudoinverse of the approximate LU factorization we obtain is exactly the matrix we would get if we applied the repeated-squaring-based Laplacian inversion algorithm of Peng and Spielman [PS] (or, more accurately, its space-efficient implementation via derandomized squaring [MRSV1]) to the original Laplacian $\mathbf{I} - \mathbf{W}$. Thus, another conceptual contribution of our paper is connecting the LU factorization approach of [CKK$^+$] to the squaring-based approach of [PS] via the cycle-lifted graph. (However, for technical reasons in our analysis, we don't do the $n \times n$ projection until after applying the error-reduction procedure to obtain a highly accurate pseudoinverse of the cycle-lifted Laplacian.)

# 2 Preliminaries

In this section we introduce notation and facts we use through out the paper.

## 2.1 Notation

We denote by $\mathbb{C}$ the set of complex numbers. For $w = x + yi \in \mathbb{C}$, we use $w^* = x - yi$ to denote the conjugate of $w$. We use $|w| = \sqrt{x^2 + y^2}$ to denote the magnitude of $w$.

**Matrices and vectors.** We use bold capital letters to denote matrices. We use $\mathbf{I}_n \in \mathbb{R}^{n \times n}$ to denote the identity matrix. For a matrix $\mathbf{A} \in \mathbb{C}^{n \times n}$ we use $\mathbf{A}^*$ to denote the conjugate transpose of $\mathbf{A}$ and we write $\mathbf{U_A} = \frac{\mathbf{A} + \mathbf{A}^*}{2}$ to denote its *symmetrization*. We use $\vec{\mathbf{1}}_k \in \mathbb{R}^k$ to denote the all 1's vector or just $\vec{\mathbf{1}}$ when $k$ is clear from context. We denote the conjugate transpose of a vector similarly. We use $\mathbf{A}^\top$ to denote the transpose of a real matrix.

**Positive Semidefinite (PSD) matrices.** For Hermitian matrices $\mathbf{A}, \mathbf{B} \in \mathbb{C}^{n \times n}$ we say $\mathbf{A}$ is PSD or write $\mathbf{A} \succeq 0$ if $x^* \mathbf{A} x \geq 0$ for all $x \in \mathbb{C}^n$. If $\mathbf{A}$ is real the condition is equivalent to

$x^\top \mathbf{A} x \geq 0$ for all $x \in \mathbb{R}^n$. Further we use $\mathbf{A} \succeq \mathbf{B}$ to denote $\mathbf{A} - \mathbf{B} \succeq 0$. We define $\preceq$, $\prec$, and $\succ$ analogously.

**Proposition 2.1.** *Given a PSD matrix $\mathbf{A} \in \mathbb{C}^{n \times n}$ and matrix $\mathbf{B} \in \mathbb{C}^{n \times m}$*

$$\mathbf{B}^* \mathbf{A} \mathbf{B} \succeq 0.$$

**Pseudo-inverse and square root of matrices.** For a matrix $\mathbf{A}$, we use $\mathbf{A}^+$ to denote the (Moore-Penrose) pseudo-inverse of $\mathbf{A}$. For a PSD matrix $\mathbf{B}$, we let $\mathbf{B}^{1/2}$ to denote the square root of $\mathbf{B}$, which is the unique PSD matrix such that $\mathbf{B}^{1/2} \mathbf{B}^{1/2} = \mathbf{B}$. Furthermore, we let $\mathbf{B}^{+/2}$ denote the pseudo-inverse of the square root of $\mathbf{B}$.

**Operator norms.** For any vector norm $\|\cdot\|$ defined on $\mathbb{C}^n$ we define the operator semi-norm it induces on $\mathbb{C}^{n \times n}$ by $\|\mathbf{A}\| = \max_{x \neq 0} \frac{\|\mathbf{A}x\|}{\|x\|}$. For a PSD matrix $\mathbf{H}$ and vector $x$ we let $\|x\|_{\mathbf{H}} = \sqrt{x^* \mathbf{H} x}$, and define the operator semi-norm $\|\mathbf{A}\|_{\mathbf{H}}$ accordingly. We can relate the $\|\cdot\|_{\mathbf{H}}$ and $\|\cdot\|_2$ operator norms as follows. For a matrix $\mathbf{A}$, we have $\|\mathbf{A}\|_{\mathbf{H}} = \|\mathbf{H}^{1/2} \mathbf{A} \mathbf{H}^{+/2}\|_2$. We use the term *spectral norm* to refer to the operator norm induced by $\|\cdot\|_2$. We write $\|\mathbf{A}|_V\|_2$ to denote the spectral norm restricted to a subspace $V$. That is, $\|\mathbf{A}|_V\|_2 = \max_{x \in V, x \neq 0} \frac{\|\mathbf{A}x\|}{\|x\|}$.

**Lemma 2.2.** *Let $\mathbf{M} \colon \mathbb{R}^n \to \mathbb{R}^n$ be a linear operator and extend it to $\mathbf{M} \colon \mathbb{C}^n \to \mathbb{C}^n$ by defining $\mathbf{M}(u + iv) = \mathbf{M}u + i\mathbf{M}v$ for all $z = u + iv \in \mathbb{C}^n$. Then $\|\mathbf{M}\|_{\mathbb{C}^n \to \mathbb{C}^n} = \|\mathbf{M}\|_{\mathbb{R}^n \to \mathbb{R}^n}$.*

**Graphs** Throughout this paper we work with unweighted directed multigraphs (digraphs). These graphs can have parallel edges and self loops and can be viewed as digraphs with integer edge weighs. We specify graphs by $G = (V, E)$ where $V$ is the set of vertices and $E$ is the multiset of edges.

**Adjacency and Random Walk Matrices.** The adjacency matrix of a digraph $G$ on $n$ vertices is the matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ where $\mathbf{A}_{ij}$ is the number of edges from vertex $j$ to vertex $i$ in $G$.[4] The degree matrix $\mathbf{D}$ of a digraph $G$ is the diagonal matrix containing the out-degrees of the vertices in $G$. The random walk matrix or transition matrix of a digraph $G$ is $\mathbf{W} = \mathbf{A}\mathbf{D}^{-1}$. $\mathbf{W}_{ij}$ is the probability that a random step from vertex $j$ leads to $i$ in $G$. Note that $\vec{1}^\top \mathbf{W} = \vec{1}^\top$. A matrix $\mathbf{W} \in \mathbb{R}_{\geq 0}^{n \times n}$ is called substochastic if $\vec{1}^\top \mathbf{W} \leq \vec{1}^\top$ (the inequality is entry-wise).

**Directed Laplacians.** We follow the approach in [CKK$^+$] to define graph Laplacians. A matrix $\mathbf{L} \in \mathbb{R}^{n \times n}$ is a directed Laplacian, if its off-diagonal entries are non-positive, i.e. $\mathbf{L}_{ij} \leq 0$ for $i \neq j$, and $\vec{1}^\top \mathbf{L} = 0$. Every digraph is associated with a directed Laplacian. Occasionally we write $\mathbf{L} = \mathbf{D} - \mathbf{A}$ to express the decomposition of $\mathbf{L}$ into the degree matrix and adjacency matrix of the corresponding digraph. The random-walk Laplacian of a digraph with Laplacian $\mathbf{D} - \mathbf{A}$ is the matrix $(\mathbf{D} - \mathbf{A})\mathbf{D}^{-1} = \mathbf{I} - \mathbf{W}$, where $\mathbf{W}$ is the transition matrix of $G$. We will often write $\mathbf{C}_k$ to denote the adjacency matrix of the $k$-vertex uni-directional directed cycle.

**Eulerian graphs and Eulerian Laplacians.** A directed graph is Eulerian if the in-degree of every node is equal to its out-degree. A directed Laplacian $\mathbf{L}$ is Eulerian if $\mathbf{L}\vec{1} = 0$. A graph is Eulerian if and only if its Laplacian is Eulerian.

---

[4]Often the adjacency matrix is defined to be $\mathbf{A}^\top$ but we find the current formulation more convenient for our purposes.

## 2.2 Kronecker Product

Given matrices $\mathbf{A} \in \mathbb{C}^{n \times m}, \mathbf{B} \in \mathbb{C}^{p \times q}$, the Kronecker product or tensor product of $\mathbf{A}$ and $\mathbf{B}$ denoted by $\mathbf{A} \otimes \mathbf{B} \in \mathbb{C}^{pn \times qm}$ is

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} \mathbf{A}_{11}\mathbf{B} & \mathbf{A}_{12}\mathbf{B} & \cdots & \mathbf{A}_{1m}\mathbf{B} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{A}_{n1}\mathbf{B} & \mathbf{A}_{n2}\mathbf{B} & \cdots & \mathbf{A}_{nm}\mathbf{B} \end{bmatrix}.$$

**Proposition 2.3.** *Given four matrices* $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, *and* $\mathbf{D}$, *if the matrix dimensions make* $\mathbf{AC}$ *and* $\mathbf{BD}$ *well-defined, then*

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC}) \otimes (\mathbf{BD}).$$

## 2.3 Schur Complement

For a matrix $\mathbf{A} \in \mathbb{C}^{n \times n}$ and sets $F, C \subseteq [n]$, let $\mathbf{A}_{FC}$ denote the submatrix corresponding to the rows in $F$ and columns in $C$. Similarly, for a vector $v \in \mathbb{C}^n$ let $v_F \in \mathbb{C}^{|F|}$ be the restriction of $v$ onto coordinates in $F$. If $F, C$ partition $[n]$ and $\mathbf{A}_{FF}$ is invertible, then we denote the Schur complement of $A$ onto the set $C$ by

$$\mathrm{Sc}(\mathbf{A}, C) \stackrel{\mathrm{def}}{=} \mathbf{A}_{CC} - \mathbf{A}_{CF}\mathbf{A}_{FF}^{-1}\mathbf{A}_{FC}.$$

When it is clear from context we may reload this notation as follows to make the Schur complement dimension consistent with $\mathbf{A}$.

$$\mathrm{Sc}(\mathbf{A}, C) \stackrel{\mathrm{def}}{=} \begin{bmatrix} 0_{FF} & 0_{FC} \\ 0_{CF} & \mathbf{A}_{CC} - \mathbf{A}_{CF}\mathbf{A}_{FF}^{-1}\mathbf{A}_{FC} \end{bmatrix}.$$

# 3 Spectral Approximation

Since its introduction by Spielman and Teng [ST2], spectral approximation of graphs and their associated matrices [ST2] has served as a powerful tool for graph-theoretic algorithm development. Below we review the original definition and later generalizations to directed graphs and asymmetric matrices [CKP+1], and then present our new, stronger definition of unit-circle approximation in several equivalent formulations.

## 3.1 Definitions

**Definition 3.1** (Undirected Spectral Approximation [ST2])**.** Let $\mathbf{W}, \widetilde{\mathbf{W}} \in \mathbb{R}^{n \times n}$ be symmetric matrices. We say that $\widetilde{\mathbf{W}}$ is an *undirected $\epsilon$-approximation* of $\mathbf{W}$ (written $\widetilde{\mathbf{W}} \approx_\epsilon \mathbf{W}$) if

$$\forall x \in \mathbb{R}^n, \ (1 - \epsilon) \cdot x^\top(\mathbf{I} - \mathbf{W})x \leq x^\top(\mathbf{I} - \widetilde{\mathbf{W}})x \leq (1 + \epsilon) \cdot x^\top(\mathbf{I} - \mathbf{W})x$$

or equivalently,

$$\forall x \in \mathbb{R}^n, \ \left| x^\top(\mathbf{W} - \widetilde{\mathbf{W}})x \right| \leq \epsilon \cdot x^\top(\mathbf{I} - \mathbf{W})x = \epsilon \cdot \left( \|x\|^2 - x^\top\mathbf{W}x \right).$$

Typically Definition 3.1 is phrased in terms of Laplacian matrices of the form $\mathbf{I} - \mathbf{W}$ and approximation is denoted by $\mathbf{I} - \widetilde{\mathbf{W}} \approx_\epsilon \mathbf{I} - \mathbf{W}$ to indicate the multiplicative approximation between the quadratic forms defined by $\mathbf{I} - \widetilde{\mathbf{W}}$ and $\mathbf{I} - \mathbf{W}$. However, in the more general definitions of this paper it will be more convenient to think of spectral approximation as a measure of approximation between $\widetilde{\mathbf{W}}$ and $\mathbf{W}$ rather than between $\mathbf{I} - \widetilde{\mathbf{W}}$ and $\mathbf{I} - \mathbf{W}$. Note that the definition is asymmetric in $\widetilde{\mathbf{W}}$ and $\mathbf{W}$ but $\widetilde{\mathbf{W}} \approx_\epsilon \mathbf{W}$ for $\epsilon < 1$ implies $\mathbf{W} \approx_{\epsilon/(1-\epsilon)} \widetilde{\mathbf{W}}$.

Spectral approximation is a strong definition that guarantees the two matrices have similar eigenvalues, and their corresponding graphs have similar cuts and random walk behavior [ST2, BSST]. Below we show the generalization to directed graphs from [CKP$^+$1].

**Definition 3.2** (Directed Spectral Approximation [CKP$^+$1]). Let $\mathbf{W}, \widetilde{\mathbf{W}} \in \mathbb{R}^{n \times n}$ be (possibly asymmetric) matrices. We say that $\widetilde{\mathbf{W}}$ is a *directed $\epsilon$-approximation* of $\mathbf{W}$ (written $\widetilde{\mathbf{W}} \approx_\epsilon \mathbf{W}$) if

$$
\begin{aligned}
\forall x, y \in \mathbb{R}^n, \ \left| x^\top (\mathbf{W} - \widetilde{\mathbf{W}}) y \right| &\leq \frac{\epsilon}{2} \cdot \left( x^\top (\mathbf{I} - \mathbf{W}) x + y^\top (\mathbf{I} - \mathbf{W}) y \right) \\
&= \frac{\epsilon}{2} \cdot \left( \|x\|^2 + \|y\|^2 - x^\top \mathbf{W} x - y^\top \mathbf{W} y \right) \\
&= \frac{\epsilon}{2} \cdot \left( \|x\|^2 + \|y\|^2 - x^\top \mathbf{U_W} x - y^\top \mathbf{U_W} y \right).
\end{aligned}
$$

The main difference between the above and Definition 3.1 is the introduction of the $y$ vector instead of having $y = x$. Indeed, using the same vector on both sides would lose the asymmetric information in the matrices $\mathbf{W}$ and $\widetilde{\mathbf{W}}$). However, note that the last inequality shows that the right-hand side depends only on the symmetrization $\mathbf{U_W}$.

We are justified using the same notation for undirected and directed spectral approximation because of the following lemma

**Lemma 3.3** ([MRSV2] Lemma 2.9). *Let $\mathbf{W}, \widetilde{\mathbf{W}} \in \mathbb{R}^{n \times n}$ be symmetric matrices. Then $\widetilde{\mathbf{W}}$ is a directed $\epsilon$-approximation of $\mathbf{W}$ if and only if it is an undirected $\epsilon$-approximation of $\mathbf{W}$.*

It will be convenient for us to generalize Definition 3.2 to complex matrices. In that case, we will quantify over $x, y \in \mathbb{C}^n$ and replace the transposes with Hermitian transposes.

**Definition 3.4** (Complex Spectral Approximation). Let $\mathbf{W}, \widetilde{\mathbf{W}} \in \mathbb{C}^{n \times n}$ be (possibly asymmetric) matrices. We say that $\widetilde{\mathbf{W}}$ is a *complex $\epsilon$-approximation* of $\mathbf{W}$ (written $\widetilde{\mathbf{W}} \approx_\epsilon \mathbf{W}$) if

$$
\begin{aligned}
\forall x, y \in \mathbb{C}^n, \ \left| x^* (\mathbf{W} - \widetilde{\mathbf{W}}) y \right| &\leq \frac{\epsilon}{2} \cdot \left( \|x\|^2 + \|y\|^2 - x^* \mathbf{U_W} x - y^* \mathbf{U_W} y \right) \\
&= \frac{\epsilon}{2} \cdot \left( \|x\|^2 + \|y\|^2 - \mathrm{Re}\left( x^* \mathbf{W} x + y^* \mathbf{W} y \right) \right)
\end{aligned}
$$

The equality in Definition 3.4 comes from the observation that for all $v \in \mathbb{C}^n$ and all matrices $\mathbf{A} \in \mathbb{C}^{n \times n}$ we have

$$
v^* \mathbf{U_A} v = \frac{1}{2} \left( v^* \mathbf{A} v + v^* \mathbf{A}^* v \right) = \frac{1}{2} \left( v^* \mathbf{A} v + (v^* \mathbf{A} v)^* \right) = \mathrm{Re}(v^* \mathbf{A} v)
$$

because the average of a complex number and its conjugate is simply its real part. Notice that the definitions of undirected, directed, and complex spectral approximation are only achievable when the matrix $\mathbf{W}$ has the property that $\mathrm{Re}(x^* \mathbf{W} x) \leq \|x\|^2$ for all vectors $x \in \mathbb{C}^n$ (when $\mathbf{W}$ is real and symmetric as in the case of undirected spectral approximation, this requirement is equivalent to $x^\top \mathbf{W} x \leq \|x\|^2$ for all $x \in \mathbb{R}^n$). When working with these types of approximation, we will often implicitly restrain the matrices to have this property.

Again, we are justified in using the same notation for complex approximation that we use for directed and undirected approximations because of the following lemma.

**Lemma 3.5.** *Let $\mathbf{W}, \widetilde{\mathbf{W}} \in \mathbb{R}^{n \times n}$ be (possibly asymmetric) matrices. Then $\widetilde{\mathbf{W}}$ is a directed $\epsilon$-approximation of $\mathbf{W}$ if and only if $\widetilde{\mathbf{W}}$ is a complex $\epsilon$-approximation of $\mathbf{W}$.*

A proof of Lemma 3.5 can be found in Appendix A. Now we introduce our new stronger definition, which we call *unit-circle spectral approximation.*

**Definition 3.6** (Unit-circle Spectral Approximation)**.** Let $\mathbf{W}, \widetilde{\mathbf{W}} \in \mathbb{C}^{n \times n}$ be (possibly asymmetric) matrices. We say that $\widetilde{\mathbf{W}}$ is a *unit-circle $\epsilon$-approximation* of $\mathbf{W}$ (written $\widetilde{\mathbf{W}} \overset{\circ}{\approx}_\epsilon \mathbf{W}$) if

$$\forall x, y \in \mathbb{C}^n, \; \left| x^*(\mathbf{W} - \widetilde{\mathbf{W}})y \right| \leq \frac{\epsilon}{2} \cdot \left( \|x\|^2 + \|y\|^2 - |x^*\mathbf{W}x + y^*\mathbf{W}y| \right).$$

The change from Definition 3.4 is that we have replaced the real part with the complex magnitude $|\cdot|$ on the quadratic forms $x^*\mathbf{W}x + y^*\mathbf{W}y$ on the right-hand side. To understand what we gain from this, suppose $x = y$ is an eigenvector of $\mathbf{W}$ with eigenvalue $\lambda$ such that $|\lambda| = 1$. Then the right-hand side of the inequality equals zero and so we must have $x^*\widetilde{\mathbf{W}}y = x^*\mathbf{W}y$. In other words, $\widetilde{\mathbf{W}}$ and $\mathbf{W}$ must behave identically on the entire unit circle of eigenvalues with magnitude 1. This is in contrast to the previous definitions, which only required exact preservation in the case where $\lambda = 1$. For example, can an undirected bipartite graph (which has a periodicity of 2 and an eigenvalue of $-1$) have a non-bipartite spectral approximation? Under previous definitions, the answer is yes but under unit-circle approximation, the answer is no because we require exact preservation on all eigenvalues of magnitude 1, not just $\lambda = 1$.

Unit circle approximation applies to a smaller class of matrices than the previous definitions of spectral approximation. While the previous definitions only required that $\text{Re}(x^*\mathbf{W}x) \leq \|x\|^2$ for all $x \in \mathbb{C}^n$, unit circle approximation requires that $|x^*\mathbf{W}x| \leq \|x\|^2$ for all $x \in \mathbb{C}^n$. Again, we will often implicitly restrict our matrices to have this property. Note that all complex matrices $\mathbf{W}$ such that $\|\mathbf{W}\|_1 \leq 1$ and $\|\mathbf{W}\|_\infty \leq 1$ satisfy this property. In particular, if $\mathbf{W}$ is the transition matrix of an Eulerian graph, then $\mathbf{W}$ satsifies the property as does $z \cdot \mathbf{W}$ for all $z \in \mathbb{C}$ such that $|z| \leq 1$.

We will see in the coming sections that unit-circle approximation is preserved under powering of $\widetilde{\mathbf{W}}$ and $\mathbf{W}$ and is useful for achieving spectral approximation of a class of graphs we call *cycle-lifted graphs*, which are essential for the anlaysis of our Eulerian Laplacian solver.

## 3.2  Equivalent Formulations

There are many useful equivalent formulations of Definition 3.6. First we look at what our definition gives in the case of real and symmetric matrices.

**Lemma 3.7** (Real, Symmetric Equivalence)**.** *Let $\mathbf{W}, \widetilde{\mathbf{W}} \in \mathbb{R}^{n \times n}$ be symmetric matrices. Then the following are equivalent:*

*1.* $\widetilde{\mathbf{W}} \overset{\circ}{\approx}_\epsilon \mathbf{W}$.

*2.* $\widetilde{\mathbf{W}} \approx_\epsilon \mathbf{W}$ *and* $-\widetilde{\mathbf{W}} \approx_\epsilon -\mathbf{W}$.

*3. For all $x \in \mathbb{R}^n$ we have*

$$\left| x^\top (\mathbf{W} - \widetilde{\mathbf{W}})x \right| \leq \epsilon \cdot \left( \|x\|^2 - |x^\top \mathbf{W}x| \right).$$

A proof of Lemma 3.7 can be found in Appendix A. In the original [ST2] formulation of spectral approximation as multiplicative approximation between quadratic forms, Lemma 3.7 says that in the real, symmetric setting, unit-circle spectral approximation is equivalent to $\mathbf{I} - \widetilde{\mathbf{W}}$ approximating $\mathbf{I} - \mathbf{W}$ *and* $\mathbf{I} + \widetilde{\mathbf{W}}$ approximating $\mathbf{I} + \mathbf{W}$. This makes intuitive sense because symmetric matrices have real eigenvalues so the only eigenvalues that can lie on the unit circle are $+1$ and $-1$.

This "plus and minus" approximation has been studied before in [CCL$^+$, MRSV2], where it was found to be useful because spectral approximation is preserved under squaring when both the "plus" and "minus" approximations hold. We will see in Section 4 that even in the general directed, complex case, unit-circle approximation is preserved under all powering.

We now show some convenient equivalent formulations of unit-circle spectral approximation.

**Lemma 3.8.** *Let* $\mathbf{W}, \widetilde{\mathbf{W}} \in \mathbb{C}^{n \times n}$ *be (possibly asymmetric) matrices. Then the following are equivalent*

1. $\widetilde{\mathbf{W}} \overset{\circ}{\approx}_\epsilon \mathbf{W}$

2. *For all* $z \in \mathbb{C}$ *such that* $|z| = 1$, $z \cdot \widetilde{\mathbf{W}} \approx_\epsilon z \cdot \mathbf{W}$

3. *For all* $z \in \mathbb{C}$ *such that* $|z| = 1$,

   - $\ker(\mathbf{U}_{\mathbf{I} - z \cdot \mathbf{W}}) \subseteq \ker(\widetilde{\mathbf{W}} - \mathbf{W}) \cap \ker((\widetilde{\mathbf{W}} - \mathbf{W})^\top)$ *and*
   - $\left\| \mathbf{U}_{\mathbf{I} - z \cdot \mathbf{W}}^{+/2} (\widetilde{\mathbf{W}} - \mathbf{W}) \mathbf{U}_{\mathbf{I} - z \cdot \mathbf{W}}^{+/2} \right\| \leq \epsilon$

A proof of Lemma 3.8 can be found in Appendix A.

# 4 Approximating Cycle-Lifted Graphs and Powers

In this section we discuss how unit-circle spectral approximation allows us to approximate powers of random walk matrices of digraphs and a class of graphs we call *cycle-lifted graphs*, which play an essential role in our Eulerian Laplacian solver. Preservation under powering is a useful property for a definition of matrix approximation but even in the case of symmetric transition matrices, the original definition of spectral approximation does not guarantee this, as is seen in the following proposition.

**Proposition 4.1.** *For all rational* $\epsilon \in (0, 1)$, *there exist undirected graphs with transition matrices* $\widetilde{\mathbf{W}}, \mathbf{W}$ *such that* $\widetilde{\mathbf{W}} \approx_\epsilon \mathbf{W}$ *but* $\widetilde{\mathbf{W}}^2 \not\approx_c \mathbf{W}^2$ *for any finite* $c > 0$.

*Proof.* Let $\mathbf{W}$ be the transition matrix of a connected undirected bipartite graph and define $\widetilde{\mathbf{W}} = (1 - \epsilon) \cdot \mathbf{W} + \epsilon \cdot \mathbf{I}$. Fix a vector $x$ and observe

$$
\begin{aligned}
\left| x^\top (\mathbf{W} - \widetilde{\mathbf{W}}) x \right| &= \epsilon \cdot \left| x^\top (\mathbf{I} - \mathbf{W}) x \right| \\
&= \epsilon \cdot \left( \|x\|^2 - x^\top \mathbf{W} x \right)
\end{aligned}
$$

where we can drop the absolute value in the second line because $\mathbf{I} - \mathbf{W}$ is PSD and hence has a non-negative quadratic form. So $\widetilde{\mathbf{W}} \approx_\epsilon \mathbf{W}$. However $\mathbf{W}^2$ has 2 connected components (because walks of length 2 must start and end at the same side of the bipartition) while $\widetilde{\mathbf{W}}^2$ is strongly connected. This means that $\mathbf{W}^2$ has eigenvalue $\lambda = 1$ with multiplicity 2 while $\widetilde{\mathbf{W}}^2$ has eigenvalue $\lambda = 1$ with multiplicity 1. Hence, they cannot spectrally approximate one another (see Lemma 4.2 below). $\square$

**Lemma 4.2.** *Fix* $\widetilde{\mathbf{W}}, \mathbf{W} \in \mathbb{C}^{n \times n}$. *For any matrix* $M \in \mathbb{C}^{n \times n}$, *let* $V_\lambda(M)$ *denote the eigenspace of* $M$ *of eigenvalue* $\lambda$.

1. *If* $\widetilde{\mathbf{W}} \approx_c \mathbf{W}$ *for a finite* $c > 0$, *then* $V_1(\mathbf{W}) \subseteq V_1(\widetilde{\mathbf{W}})$. *If* $c < 1$, *then* $V_1(\mathbf{W}) = V_1(\widetilde{\mathbf{W}})$.

13

2. If $\widetilde{\mathbf{W}} \overset{\circ}{\approx}_c \mathbf{W}$ for a finite $c > 0$, then for all $\lambda \in \mathbb{C}$ such that $|\lambda| = 1$, $V_\lambda(\mathbf{W}) \subseteq V_\lambda(\widetilde{\mathbf{W}})$ and if $c < 1$ then $V_\lambda(\mathbf{W}) = V_\lambda(\widetilde{\mathbf{W}})$.

A proof of Lemma 4.2 can be found in Appendix B. In previous work ([CCL$^+$, MRSV1]), it was observed that for symmetric matrices, if $\widetilde{\mathbf{W}} \approx_\epsilon \mathbf{W}$ and $-\widetilde{\mathbf{W}} \approx_\epsilon -\mathbf{W}$ then we do get $\widetilde{\mathbf{W}}^2 \approx_\epsilon \mathbf{W}^2$. Furthermore, when $\mathbf{W}$ is PSD we have that $\widetilde{\mathbf{W}} \approx_\epsilon \mathbf{W}$ implies $-\widetilde{\mathbf{W}} \approx_\epsilon -\mathbf{W}$. Since $\mathbf{W}^2$ is trivially PSD, the above can be applied recursively to conclude that $\widetilde{\mathbf{W}}^{2^k} \approx_\epsilon \mathbf{W}^{2^k}$ for all positive integers $k$. However, we observe that analogous approximation guarantees do not hold for Eulerian graphs (or even regular digraphs).

**Proposition 4.3.** *For all rational $\epsilon \in (0,1)$, there exist regular digraphs with transition matrices $\widetilde{\mathbf{W}}, \mathbf{W}$ such that $\widetilde{\mathbf{W}} \approx_\epsilon \mathbf{W}$ and $-\widetilde{\mathbf{W}} \approx_\epsilon -\mathbf{W}$ but $\widetilde{\mathbf{W}}^4 \not\approx_c \mathbf{W}^4$ for any finite $c$.*

*Proof.* Let $\mathbf{C}_4$ be the transition matrix of the directed 4-cycle. Fix $\epsilon \in (0,1)$ and define $\widetilde{\mathbf{C}} = (1 - \epsilon/2) \cdot \mathbf{C}_4 + (\epsilon/2) \cdot \mathbf{C}_4^\top$. In other words, $\widetilde{\mathbf{C}}$ is the directed 4-cycle with an $\epsilon/2$ probability of traversing backwards. We claim that $\widetilde{\mathbf{C}} \approx_\epsilon \mathbf{C}_4$ and $-\widetilde{\mathbf{C}} \approx_\epsilon -\mathbf{C}_4$ but $\widetilde{\mathbf{C}}^4$ does not approximate $\mathbf{C}_4^4$.

First, note that each of $\mathbf{U}_{\mathbf{I} \pm \mathbf{C}_4}$ has a one-dimensional kernel, namely

$$\ker(\mathbf{U}_{\mathbf{I}-\mathbf{C}_4}) = \mathrm{span}(\vec{\mathbf{1}})$$
$$\ker(\mathbf{U}_{\mathbf{I}+\mathbf{C}_4}) = \mathrm{span}\left([1, -1, 1, -1]^\top\right).$$

Furthermore we have,

$$(\widetilde{\mathbf{C}} - \mathbf{C}_4) = (\epsilon/2) \cdot (\mathbf{C}_4 - \mathbf{C}_4^\top),$$

which has a two-dimensional kernel, $\mathrm{span}(\vec{\mathbf{1}}, [1, -1, 1, -1])$. Observe that $\mathbf{U}_{\mathbf{I} \pm \mathbf{C}_4}$ each has eigenvalues of 2, 1, 1, and 0. Hence $\|\mathbf{U}_{\mathbf{I} \pm \mathbf{C}}^{+/2}\| = 1$. Finally, we have that

$$\|\mathbf{C}_4 - \widetilde{\mathbf{C}}\| = (\epsilon/2) \cdot \|\mathbf{C}_4 - \mathbf{C}_4^\top\| = \epsilon$$

Putting this together gives

$$\|\mathbf{U}_{\mathbf{I}-\mathbf{C}_4}^{+/2}(\mathbf{C}_4 - \widetilde{\mathbf{C}})\mathbf{U}_{\mathbf{I}-\mathbf{C}_4}^{+/2}\| \leq \|\mathbf{U}_{\mathbf{I}-\mathbf{C}_4}^{+/2}\| \cdot \|\mathbf{C}_4 - \widetilde{\mathbf{C}}\| \cdot \|\mathbf{U}_{\mathbf{I}-\mathbf{C}_4}^{+/2}\|$$
$$= \epsilon.$$

It follows that $\widetilde{\mathbf{C}} \approx_\epsilon \mathbf{C}_4$. A similar calculation lets us conclude that $-\widetilde{\mathbf{C}} \approx_\epsilon -\mathbf{C}_4$, as desired.

Notice that $\mathbf{C}_4^4$ has 4 connected components (all of the vertices become isolated with self loops) while $\widetilde{\mathbf{C}}^4$ has 2 connected components. So $\mathbf{C}_4^4$ and $\widetilde{\mathbf{C}}^4$ have eigenvalues of 1 with different multiplicities and hence they fail to spectrally approximate of one another by Lemma 4.2. $\square$

Now we show that if a matrix is a unit-circle approximation of another, then all of their powers are as well (with small loss in approximation quality). In fact, we show something stronger, namely that their *cycle-lifted graphs* approximate each other.

**Definition 4.4** (Cycle-Lifted Graph). Let $\mathbf{C}_k$ denote the transition matrix of the $k$-vertex directed cycle. Given a graph $G = (V, E)$ on $n$ vertices with transition matrix $\mathbf{W}$ the *cycle-lifted graph of length $k$, $C_k(G)$,* is a layered graph with $k$ layers (numbered 1 to $k$) of $n$ vertices each, where for every $i \in [k]$, there is an edge from vertex $u$ in layer $i$ to vertex $v$ in layer $(i+1) \mod k$ with multiplicity $\ell$ if and only if $(u, v)$ exists with multiplicity $\ell$ in $G$. That is, $C_k(G) = (V', E')$ with $V' = [k] \times V$ and $E' = \{((i, u), (i+1 \mod k, v)): (u, v) \in E\}$. The transition matrix of $C_k(G)$ is $\mathbf{C}_k \otimes \mathbf{W}$.

**Theorem 4.5.** *Fix* $\mathbf{W}, \widetilde{\mathbf{W}} \in \mathbb{C}^{n \times n}$ *and let* $\mathbf{C}_k$ *be the transition matrix for the directed cycle on* $k$ *vertices. Then* $\mathbf{C}_k \otimes \widetilde{\mathbf{W}} \approx_\epsilon \mathbf{C}_k \otimes \mathbf{W}$ *if and only if for all* $z$ *such that* $z^k = 1$, *we have* $z \cdot \widetilde{\mathbf{W}} \approx_\epsilon z \cdot \mathbf{W}$.

Recall that unit-circle spectral approximation requires that for all $z \in \mathbb{C}$ with $|z| = 1$ we have $z \cdot \widetilde{\mathbf{W}} \approx_\epsilon z \cdot \mathbf{W}$. Theorem 4.5 then tells us that unit-circle spectral approximation implies approximations of the corresponding cycle-lifted graphs of *every length*.

**Corollary 4.6.** *Fix* $\mathbf{W}, \widetilde{\mathbf{W}} \in \mathbb{C}^{n \times n}$. *If* $\widetilde{\mathbf{W}} \overset{\circ}{\approx}_\epsilon \mathbf{W}$ *then for all positive integers* $k$, $\mathbf{C}_k \otimes \widetilde{\mathbf{W}} \overset{\circ}{\approx}_\epsilon \mathbf{C}_k \otimes \mathbf{W}$.

*Proof.* Since $\widetilde{\mathbf{W}} \overset{\circ}{\approx}_\epsilon \mathbf{W}$, we have that for all $\omega \in \mathbb{C}$ such that $|\omega| = 1$ and all $z$ such that $z^k = 1$, we have $z \cdot w \cdot \widetilde{\mathbf{W}} \approx_\epsilon z \cdot w \cdot \mathbf{W}$. By Theorem 4.5, for all $\omega \in \mathbb{C}$ such that $|\omega| = 1$, we have

$$\omega \cdot (\mathbf{C}_k \otimes \widetilde{\mathbf{W}}) = \mathbf{C}_k \otimes (\omega \cdot \widetilde{\mathbf{W}}) \approx_\epsilon \mathbf{C}_k \otimes (\omega \cdot \mathbf{W}) = \omega \cdot (\mathbf{C}_k \otimes \mathbf{W}).$$

Thus, $\mathbf{C}_k \otimes \widetilde{\mathbf{W}} \overset{\circ}{\approx}_\epsilon \mathbf{C}_k \otimes \mathbf{W}$. $\qquad\square$

To prove Theorem 4.5, we use the following lemma, which simplifies calculating spectral norm and hence spectral approximation using restrictions onto invariant subspaces.

**Lemma 4.7.** *Let* $\mathbf{M} \colon \mathbb{C}^n \to \mathbb{C}^n$ *be a linear operator and* $V_1, \ldots, V_\ell \subseteq \mathbb{C}^n$ *subspaces such that*

1. $V_j \perp V_k$ *for all* $j \neq k$

2. $V_1 \oplus \ldots \oplus V_\ell = \mathbb{C}^n$

3. $\mathbf{M} V_j \subseteq V_j$ *for all* $j \in [\ell]$.

*i.e.,* $M$ *is block diagonal with respect to the subspaces* $V_1, \ldots, V_\ell$. *Then,*

$$\|\mathbf{M}\| = \max_{j \in [\ell]} \|\mathbf{M}|_{V_j}\|$$

A proof of Lemma 4.7 can be found in Appendix B. Now we can prove Theorem 4.5.

*Proof of Theorem 4.5.* The intuition behind the proof is to observe that the diagonalization of the directed $k$-cycle $\mathbf{C}_k$ using the discrete Fourier basis (which are its eigenvectors), has all $k$th roots of unity along the diagonal. This means that approximation of cycle-lifted graphs $\mathbf{C}_k \otimes \widetilde{\mathbf{W}}$ and $\mathbf{C}_k \otimes \mathbf{W}$ amounts to requiring that the approximation of $\widetilde{\mathbf{W}}$ and $\mathbf{W}$ is preserved under multiplication by $k$'th roots of unity.

For each $z \in \mathbb{C}$ with $z^k = 1$, define the Fourier basis vector $\chi_z = [1, z, z^2, \ldots, z^{k-1}]^\top$. Consider the subspaces $V_z = \mathrm{span}(\chi_z \otimes \mathbb{C}^n)$. These subspaces are orthogonal, span $\mathbb{C}^k \otimes \mathbb{C}^n$ and are invariant under $\mathbf{C}_k \otimes \mathbf{W}$, $\mathbf{C}_k \otimes \widetilde{\mathbf{W}}$, $(\mathbf{C}_k \otimes \mathbf{W})^* = \mathbf{C}_k^* \otimes \mathbf{W}^*$, $(\mathbf{C}_k \otimes \widetilde{\mathbf{W}})^* = \mathbf{C}_k^* \otimes \widetilde{\mathbf{W}}^*$, and $\mathbf{U}_{\mathbf{I}_{k \cdot n} - \mathbf{C}_k \otimes \mathbf{W}}^{+/2}$. So, by Lemma 4.7 we have

$$\left\| \mathbf{U}_{\mathbf{I}_{k \cdot n} - \mathbf{C}_k \otimes \mathbf{W}}^{+/2} \left( \mathbf{C}_k \otimes \widetilde{\mathbf{W}} - \mathbf{C}_k \otimes \mathbf{W} \right) \mathbf{U}_{\mathbf{I}_{k \cdot n} - \mathbf{C}_k \otimes \mathbf{W}}^{+/2} \right\| =$$
$$\max_{z \,:\, z^k = 1} \left\| \mathbf{U}_{\mathbf{I}_{k \cdot n} - \mathbf{C}_k \otimes \mathbf{W}}^{+/2} \left( \mathbf{C}_k \otimes \widetilde{\mathbf{W}} - \mathbf{C}_k \otimes \mathbf{W} \right) \mathbf{U}_{\mathbf{I}_{k \cdot n} - \mathbf{C}_k \otimes \mathbf{W}}^{+/2}|_{V_z} \right\|.$$

Now observe that since $\mathbf{C}_k \chi_z = z^{-1} \chi_z$, we have that for any $\vec{v} \in \mathbb{C}^n$

$$(\mathbf{C}_k \otimes \mathbf{W})(\chi_z \otimes \vec{v}) = \mathbf{C}_k \chi_z \otimes \mathbf{W} \vec{v}$$
$$= \chi_z \otimes (z^{-1} \cdot \mathbf{W} \vec{v}).$$

15

Similarly we have,

$$(\mathbf{C}_k \otimes \widetilde{\mathbf{W}})(\chi_z \otimes \vec{v}) = \chi_z \otimes (z^{-1} \cdot \widetilde{\mathbf{W}}\vec{v})$$

$$\mathbf{U}_{\mathbf{I}_{k \cdot n} - \mathbf{C}_k \otimes \mathbf{W}}(\chi_z \otimes \vec{v}) = \chi_z \otimes \mathbf{U}_{\mathbf{I}_n - z^{-1} \cdot \mathbf{W}}\vec{v}.$$

It follows that

$$\left\| \mathbf{U}^{+/2}_{\mathbf{I}_{k \cdot n} - \mathbf{C}_k \otimes \mathbf{W}} \left( \mathbf{C}_k \otimes \widetilde{\mathbf{W}} - \mathbf{C}_k \otimes \mathbf{W} \right) \mathbf{U}^{+/2}_{\mathbf{I}_{k \cdot n} - \mathbf{C}_k \otimes \mathbf{W}} \right\|$$

$$= \max_{z \,:\, z^k = 1} \left\| \mathbf{U}^{+/2}_{\mathbf{I}_n - z^{-1} \cdot \mathbf{W}} \left( z^{-1} \cdot \mathbf{W} - z^{-1} \cdot \widetilde{\mathbf{W}} \right) \mathbf{U}^{+/2}_{\mathbf{I}_n - z^{-1} \cdot \mathbf{W}} \right\|.$$

So we get that $\mathbf{C}_k \otimes \widetilde{\mathbf{W}} \approx_\epsilon \mathbf{C}_k \otimes \mathbf{W}$ if and only if for all $z$ such that $z^k = 1$, $z^{-1}\widetilde{\mathbf{W}} \approx_\epsilon z^{-1}\mathbf{W}$. Since for all $k$th roots of unity $z$, $z^{-1}$ is also a $k$th root of unity, the result follows. $\qquad\square$

Theorem 4.5 allows us to reason about approximation under powering by observing that the $k$th power of a matrix can be expressed in terms of the Schur complement of its cycle-lifted graph of length $k$. In [MP], they showed that undirected spectral approximation is preserved under Schur complements. Here we show that the same is true of directed spectral approximation (with a small loss in approximation quality).

**Theorem 4.8.** *Fix $\mathbf{W}, \widetilde{\mathbf{W}} \in \mathbb{C}^n$ and suppose that $\widetilde{\mathbf{W}} \approx_\epsilon \mathbf{W}$ for $\epsilon \in (0, 2/3)$. Let $F \subseteq [n]$ such that $(\mathbf{I}_{|F|} - \mathbf{W}_{FF})$ is invertible and let $C = [n] \setminus F$. Then*

$$\mathbf{I}_{|C|} - \mathrm{Sc}(\mathbf{I}_n - \widetilde{\mathbf{W}}, C) \approx_{\epsilon/(1-3\epsilon/2)} \mathbf{I}_{|C|} - \mathrm{Sc}(\mathbf{I}_n - \mathbf{W}, C)$$

A proof of Theorem 4.8 can be found in Appendix B. The expression in Theorem 4.8 has a natural interpretation in terms of random walks. Indeed, notice that

$$\mathbf{I}_{|C|} - \mathrm{Sc}(\mathbf{I}_n - \mathbf{W}, C) = \mathbf{W}_{CC} + \mathbf{W}_{CF}(\mathbf{I}_{|F|} - \mathbf{W}_{FF})^{-1}\mathbf{W}_{FC}.$$

When $\mathbf{W}$ is the transition matrix for a random walk, the right-hand side above can be interpreted as the transition matrix for the random walk induced by "short-cutting" walks that traverse through the set of vertices in $F$. In other words, walk behavior on $C$ remains the same (the $\mathbf{W}_{CC}$ term) and walks that go from $C$ to $F$ (via $\mathbf{W}_{FC}$) can instantly take arbitrary length walks in $F$ (the $(\mathbf{I}_{|F|} - \mathbf{W}_{FF})^{-1}$ term) before returning to $C$ (via $\mathbf{W}_{CF}$). The theorem above says that spectral approximation is preserved under such "short-cutting".

Now we get the following corollary, which says that unit-circle approximation is preserved under powering.

**Corollary 4.9.** *Let $\mathbf{W}$, $\widetilde{\mathbf{W}}$ be the transition matrices of digraphs $G, \widetilde{G}$. If $\widetilde{\mathbf{W}} \overset{\circ}{\approx}_\epsilon \mathbf{W}$ then for all $k \in \mathbb{N}$ we have $\widetilde{\mathbf{W}}^k \overset{\circ}{\approx}_{\epsilon/(1-3\epsilon/2)} \mathbf{W}^k$.*

*Proof.* Fix $k \in \mathbb{N}$ and $z \in \mathbb{C}$ such that $|z| = 1$. Let $\omega$ be such that $\omega^k = z$. Let $\mathbf{M} = \mathbf{C}_k \otimes \omega \cdot \mathbf{W}$ and $\widetilde{\mathbf{M}} = \mathbf{C}_k \otimes \omega \cdot \widetilde{\mathbf{W}}$. From Corollary 4.6 we have that $\widetilde{\mathbf{M}} \overset{\circ}{\approx}_\epsilon \mathbf{M}$. We can think of $\mathbf{M}$ and $\widetilde{\mathbf{M}}$ as the transition matrices of the cycle-lifted graph of graphs with transition matrices $\mathbf{W}$ and $\widetilde{\mathbf{W}}$. Define $F$ to be the set of vertices in the first layer of these cycle-lifted graphs. Notice that

$$\mathbf{I}_{|F|} - \mathrm{Sc}(\mathbf{I}_n - \mathbf{M}, F) = (\omega \cdot \mathbf{W})^k = z \cdot \mathbf{W}^k$$

$$\mathbf{I}_{|F|} - \mathrm{Sc}(\mathbf{I}_n - \widetilde{\mathbf{M}}, F) = (\omega \cdot \widetilde{\mathbf{W}})^k = z \cdot \widetilde{\mathbf{W}}^k.$$

It follows from Theorem 4.8 that $\widetilde{\mathbf{W}}^k \overset{\circ}{\approx}_{\epsilon/(1-3\epsilon/2)} \mathbf{W}^k$. $\qquad\square$

16

Interestingly, in the case of undirected graphs, Corollary 4.9 says that if $\widetilde{\mathbf{W}} \approx_\epsilon \mathbf{W}$ and $-\widetilde{\mathbf{W}} \approx_\epsilon -\mathbf{W}$ then all $k$th powers approximate one another (with small loss in approximation quality). This was not known (to the best of our knowledge) for any $k$ other than powers of 2.

# 5    Derandomized Square of Regular Digraphs

In order to achieve a space-efficient and deterministic implementation of our algorithm, we need a way to efficiently approximate high powers of regular digraphs. To do this, we use the *derandomized square* graph operation of Rozenman and Vadhan [RV], which uses expander graphs to give sparse approximations to the graph square. We define expander graphs in terms of the measure below.

**Definition 5.1** ([Mih])**.** Let $G$ be a regular directed multigraph with transition matrix $\mathbf{W}$. We define

$$\lambda(G) = \max_{v \perp \vec{\mathbf{1}}} \frac{\|\mathbf{W}v\|}{\|v\|} \in [0, 1],$$

where the maximum can be taken over either real or complex vectors $v$ (equivalent by applying Lemma 2.2 to the restriction of $\mathbf{W}$ to the subspace orthogonal to $\vec{\mathbf{1}}$). The *spectral gap* of $G$ is defined to be $\gamma(G) = 1 - \lambda(G)$ and when $\gamma(G) \geq \gamma$, we say that $G$ has *spectral expansion* $\gamma$. When $G$ is undirected, $\lambda(G)$ equals the second largest eigenvalue of $W$ in absolute value.

It is well known that the larger $\gamma(G)$ is, the faster a random walk on $G$ converges to the stationary distribution. Families of graphs with $\lambda(G) \leq 1 - \Omega(1)$ are called *expanders*.

$\lambda(G)$ also relates naturally to unit-circle approximation as shown in the following lemma, which says that the smaller $\lambda(G)$, the better $G$ approximates the complete graph.

**Lemma 5.2.** *Let $G$ be a strongly connected, regular directed multigraph on $n$ vertices with transition matrix $\mathbf{W}$ and let $\mathbf{J} \in \mathbb{R}^{n \times n}$ be a matrix with $1/n$ in every entry (i.e. $\mathbf{J}$ is the transition matrix of the complete graph with a self loop on every vertex). Then $\lambda(G) \leq \lambda$ if and only if $\mathbf{W} \overset{\circ}{\approx}_\lambda \mathbf{J}$.*

A proof of Lemma 5.2 can be found in Appendix C. Before defining the derandomized square operation, we introduce two-way labelings and rotation maps.

**Definition 5.3** ([RVW, RV])**.** A *two-way labeling* of a $d$-regular directed multigraph $G$ is a labeling of the edges in $G$ such that

1. Every edge $(u, v)$ has two labels in $[d]$, one as an edge incident to $u$ and one as an edge incident to $v$,

2. For every vertex $v$ the labels of the edges incident to $v$ are distinct.

In a two-way labeling, each vertex $v$ has its own labeling from 1 to $d$ for the $d$ edges leaving it and its own labeling from 1 to $d$ for the $d$ edges entering it. Since every edge is incident to two vertices, each edge receives two labels, which may or may not be the same. It is convenient to specify a multigraph with a two-way labeling by a rotation map:

**Definition 5.4** ([RVW, RTV])**.** Let $G$ be a $d$-regular directed multigraph on $n$ vertices with a two-way labeling. The *rotation map* $\mathrm{Rot}_G \colon [n] \times [d] \to [n] \times [d]$ is defined as follows: $\mathrm{Rot}_G(v, i) = (w, j)$ if the $i$th edge leaving vertex $v$ leads to vertex $w$ and this edge is the $j$th edge entering $w$.

Now we can define the derandomized square. Recall that the square of a graph $G^2$ is a graph on the same vertex set whose edges correspond to all walks of length 2 in $G$. The derandomized square picks out a pseudorandom subset of the walks of length 2 by correlating the 2 steps via edges on an expander graph.

**Definition 5.5** ([RV]). Let $G$ be a $d$-regular multigraph on $n$ vertices with a two-way labeling. Let $H$ be a $c$-regular undirected graph on $d$ vertices. The *derandomized square* $G\text{\textcircled{s}}H$ is a $c \cdot d$-regular graph on $n$ vertices with rotation map $\text{Rot}_{G\text{\textcircled{s}}H}$ defined as follows: For $v_0 \in [n], i_0 \in [d]$, and $j_0 \in [c]$, we compute $\text{Rot}_{G\text{\textcircled{s}}H}(v_0, (i_0, j_0))$ as

1. Let $(v_1, i_1) = \text{Rot}_G(v_0, i_0)$

2. Let $(i_2, j_1) = \text{Rot}_H(i_1, j_0)$

3. Let $(v_2, i_3) = \text{Rot}_G(v_1, i_2)$

4. Output $(v_2, (i_3, j_1))$

In the square of a directed graph, for each vertex $v$, there exists a complete, uni-directional bipartite graph from the in-neighbors of $v$, to it's out-neighbors. This corresponds to a directed edge for every two-step walk that has $v$ in the middle of it. A useful way to view the derandomized square is that it replaces each of these complete bipartite graphs with a uni-directional bipartite *expander*.

**Definition 5.6.** Let $H = (V, E)$ be an undirected graph on $d$ vertices. We define $\text{Bip}(H)$ to be a bipartite graph with $d$ vertices on each side of the bipartition and an edge $(u, v)$ from vertex $u$ on the left to vertex $v$ on the right if and only if $(u, v) \in E$.

Note that since we're working with multigraphs, the incoming or outgoing neighbors to/from a vertex may form a multi-set rather than a set due to parallel edges. So when we say that a "copy" of $\text{Bip}(H)$ exists from the in-neighbors of $v$ to its out-neighbors, we mean that if we were to split all of the in-neighbors of $v$ and out-neighbors of $v$ into two sets of $d$ distinct vertices, place the edges from $\text{Bip}(H)$ across the sets, and then re-merge vertices that correspond to a repeat neighbor of $v$, then a copy of that resulting graph can be found across vertex $v$ in $G\text{\textcircled{s}}H$. We formalize this view of the derandomized square in the following lemma.

**Lemma 5.7.** *Let $G$ be a $d$-regular directed multigraph on $n$ vertices with a two-way labeling and transition matrix $\mathbf{W}$. Let $H$ be a $c$-regular undirected graph on $d$ vertices with a two-way labeling and transition matrix $\mathbf{B}$. Let $\mathbf{J}$ be the $d \times d$ matrix with $1/d$ in every entry and let $\widetilde{\mathbf{W}}$ be the transition matrix of $G\text{\textcircled{s}}H$. Define the $2d \times 2d$ matrices*

$$\mathbf{M} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{J} & \mathbf{0} \end{bmatrix}$$

*and*

$$\widetilde{\mathbf{M}} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{B} & \mathbf{0} \end{bmatrix}.$$

*Furthermore, for each $v \in [n]$ define $\mathbf{P}^{(v)}$, $\mathbf{Q}^{(v)}$, and $\mathbf{T}^{(v)}$ as follows*

$$\mathbf{P}^{(v)}_{j,w} = \begin{cases} 1 & \text{if } j\text{th edge entering } v \text{ in } G \text{ comes from } w \\ 0 & \text{otherwise} \end{cases}$$

$$\mathbf{Q}^{(v)}_{j,w} = \begin{cases} 1 & \text{if } j\text{th edge leaving } v \text{ in } G \text{ goes to } w \\ 0 & \text{otherwise} \end{cases}$$

$$\mathbf{T}^{(v)} = \begin{bmatrix} \mathbf{P}^{(v)} \\ \hline \mathbf{Q}^{(v)} \end{bmatrix}.$$

18

*Then we have*

$$\mathbf{W}^2 = \frac{1}{d} \cdot \sum_{v \in [n]} (\mathbf{T}^{(v)})^\top \mathbf{M} \mathbf{T}^{(v)}$$

*and*

$$\widetilde{\mathbf{W}} = \frac{1}{d} \cdot \sum_{v \in [n]} (\mathbf{T}^{(v)})^\top \widetilde{\mathbf{M}} \mathbf{T}^{(v)}$$

*Proof of Lemma 5.7.* First we show that in $G \circledS H$ there exists a "copy" of $\mathrm{Bip}(H)$ from the in-neighbors of each vertex to the out-neighbors. Fix a vertex $v \in G$. Let $u_1, \ldots, u_d$ be the multi-set of incoming neighbors of $v$ and let $w_1, \ldots, w_d$ be the multiset of outgoing neighbors. Without loss of generality suppose that for each $i \in [d]$, the edge from $u_i$ to $v$ has label $i^{\mathrm{in}}$ as an edge incident to $v$ and the edge from $v$ to $w_i$ has label $i^{\mathrm{out}}$ as an edge incident to $v$ in $G$. We claim that in the graph $G \circledS H$ there exists a copy of $\mathrm{Bip}(H)$ from the $u_i$'s to the $w_i$'s such that for each $i \in [d]$, $u_i$ is the $i$th vertex on the left side of $\mathrm{Bip}(H)$ and $w_i$ is the $i$th vertex on the right side of $\mathrm{Bip}(H)$ (where edge multiplicities correspond to merging vertices in $\mathrm{Bip}(H)$).

To see this, suppose $(a, b)$ is an edge in $H$ with labels $\ell_a$ and $\ell_b$, respectively. Also, let $j_a, j_b$ be the labels of edges $(u_a, v)$ and $(v, w_b)$ as edges incident to $u_a$ and $w_b$, respectively (recall that these are labelled $a$ and $b$ from $v$'s perspective). We will show that there is an edge corresponding to $(a, b)$ in $H$ from $u_a$ to $w_b$ in $G \circledS H$, namely the one labeled $(j_a, \ell_a)$, which will complete the proof. We compute the rotation map $\mathrm{Rot}_{G \circledS \mathcal{H}}(u_a, (j_a, \ell_a))$:

1. $\mathrm{Rot}_G(u_a, j_a) = (v, a)$

2. $\mathrm{Rot}_H(a, \ell_a) = (b, \ell_b)$

3. $\mathrm{Rot}_G(v, b) = (w_b, j_b)$

4. Output $(w_b, (j_b, \ell_b))$

So edge $(j_a, \ell_a)$ leaving vertex $u_a$ indeed leads to vertex $w_b$ in $G \circledS H$.

Let

$$\mathbf{W}^{(v)} = (\mathbf{T}^{(v)})^\top \mathbf{M} \mathbf{T}^{(v)}$$
$$\widetilde{\mathbf{W}}^{(v)} = (\mathbf{T}^{(v)})^\top \widetilde{\mathbf{M}} \mathbf{T}^{(v)}$$

From the way we have defined $\mathbf{T}^{(v)}$, we have that $\mathbf{W}^{(v)}$ is exactly the transition matrix of the uni-directional bipartite complete graph from the in-neighbors of vertex $v$ to its out-neighbors in $G^2$ (with all vertices that are not neighbors of $v$ isolated). Likewise $\widetilde{\mathbf{W}}^{(v)}$ is the transition matrix of the uni-directional bipartite expander from the in-neighbors of vertex $v$ to its out-neighbors in $G \circledS H$ (with non-neighbors of $v$ isolated). It follows from the reasoning above that

$$\frac{1}{d} \cdot \sum_{v \in [n]} \mathbf{W}^{(v)} = \mathbf{W}^2$$

and likewise

$$\frac{1}{d} \cdot \sum_{v \in [n]} \widetilde{\mathbf{W}}^{(v)} = \widetilde{\mathbf{W}}.$$

$\square$

Before showing that the derandomized square produces a unit-circle approximation of the true square, we need a lemma about these uni-directional bipartite expanders. The following lemma says that if we convert an expander $H$ into a uni-directional bipartite graph, then it unit-circle approximates the complete uni-directional bipartite expander, where the quality of approximation depends on how good of an expander $H$ is.

**Lemma 5.8.** *Let $\mathbf{W} \in \mathbb{R}^{n \times n}$ be the transition matrix of a regular multigraph $H$ with $\lambda(H) \leq \epsilon$. Let $\mathbf{J}$ be the $n \times n$ matrix with $1/n$ in every entry. Then we have*

$$\begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{W} & \mathbf{0} \end{bmatrix} \overset{\circ}{\approx}_\epsilon \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{J} & \mathbf{0} \end{bmatrix}.$$

*Proof.* We want to show that for all $x, y \in \mathbb{C}^{2n}$ we have

$$\left| x^* \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{W} - \mathbf{J} & \mathbf{0} \end{bmatrix} y \right| \leq \frac{\epsilon}{2} \cdot \left( \|x\|^2 + \|y\|^2 - \left| x^* \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{J} & \mathbf{0} \end{bmatrix} x + y^* \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{J} & \mathbf{0} \end{bmatrix} y \right| \right) \tag{5}$$

For a vector $v \in \mathbb{C}^{2n}$, we will write $v_1$ for the first $n$ components of $v$ and $v_2$ for the last $n$ components. Observe that if

$$v \in V = \left\{ v \in \mathbb{C}^{2n} \colon v_1 \in \mathrm{Span}(\vec{\mathbf{1}}) \text{ and } v_2 \in \mathrm{Span}(\vec{\mathbf{1}}) \right\}$$

then

$$\begin{bmatrix} \mathbf{0} & \mathbf{0} \\ (\mathbf{W} - \mathbf{J}) & \mathbf{0} \end{bmatrix} v = \left( \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ (\mathbf{W} - \mathbf{J}) & \mathbf{0} \end{bmatrix} \right)^\top v = \vec{\mathbf{0}}.$$

Notice that the right-hand side of Equation 5 is always non-negative. Thus, by decomposing $x$ and $y$ into vectors in $V$ and $V^\perp$, we see that it suffices to consider vectors in $V^\perp = \{ v \in \mathbb{C}^{2n} \colon v_1 \perp \vec{\mathbf{1}}, v_2 \perp \vec{\mathbf{1}} \}$. Fix $x, y \in V^\perp$. Let $x_1, x_2$ be the first (respectively last) $n$-coordinates of $x$ and define $y_1, y_2$ analogously. Since $\mathbf{J}v = \vec{\mathbf{0}}$ for all $v \perp \vec{\mathbf{1}}$ it follows that

$$\left| x^* \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ (\mathbf{W} - \mathbf{J}) & \mathbf{0} \end{bmatrix} y \right| = |x_2^* \mathbf{W} y_1|$$

$$\leq \epsilon \cdot \|x_2\| \cdot \|y_1\|$$

$$\leq \frac{\epsilon}{2} \cdot (\|x_2\|^2 + \|y_1\|^2)$$

where the second line uses that $\|\mathbf{W}v\|_2 \leq \epsilon \cdot \|v\|_2$ for all $v \perp \vec{\mathbf{1}}$, and the last line uses the AM-GM inequality. Again since $\mathbf{J}v = \vec{\mathbf{0}}$ for all $v \perp \vec{\mathbf{1}}$, we have that the right-hand side of Equation 5 just becomes

$$\frac{\epsilon}{2} \cdot (\|x\|^2 + \|y\|^2),$$

which is certainly at least $\frac{\epsilon}{2} \cdot (\|x_2\|^2 + \|y_1\|^2)$. $\qquad\square$

Now we show that the derandomized square of a regular digraph yields a unit-circle spectral approximation to the true square.

**Theorem 5.9.** *Let $G = (V, E)$ be a $d$-regular directed multigraph with random walk matrix $\mathbf{W}$. Let $H$ be a $c$-regular expander with $\lambda(H) \leq \epsilon$ and let $\widetilde{\mathbf{W}}$ be the random walk matrix of $G \circledS H$. Then*

$$\widetilde{\mathbf{W}} \overset{\circ}{\approx}_{2 \cdot \epsilon} \mathbf{W}^2.$$

20

*Proof.* Define $\mathbf{M}, \widetilde{\mathbf{M}}, \mathbf{T}^{(v)}, \mathbf{W}^{(v)}, \widetilde{\mathbf{W}}^{(v)}$ as in Lemma 5.7. By Lemma 5.7 we have

$$\frac{1}{d} \cdot \sum_{v \in [n]} \mathbf{W}^{(v)} = \mathbf{W}^2.$$

Also, by Lemma 5.8 we have that $\widetilde{\mathbf{M}} \overset{\circ}{\approx}_\epsilon \mathbf{M}$, which means that for all $x, y \in \mathbb{C}^n$ we can consider $x' = \mathbf{T}^{(v)}x$ and $y' = \mathbf{T}^{(v)}y$ to conclude

$$\left| x^*(\widetilde{\mathbf{W}}^{(v)} - \mathbf{W}^{(v)})y \right| \leq \frac{\epsilon}{2} \cdot \left( x^*(\mathbf{T}^{(v)})^\top \mathbf{T}^{(v)}x + y^*(\mathbf{T}^{(v)})^\top \mathbf{T}^{(v)}y - \left| x^*\mathbf{W}^{(v)}x + y^*\mathbf{W}^{(v)}y \right| \right)$$

Summing both sides of the inequality over all vertices $v \in [n]$, gives

$$\sum_{v \in [n]} \left| x^*(\widetilde{\mathbf{W}}^{(v)} - \mathbf{W}^{(v)})y \right|$$
$$\leq \frac{\epsilon}{2} \cdot \left( \sum_{v \in [n]} \left( x^*(\mathbf{T}^{(v)})^\top \mathbf{T}^{(v)}x + y^*(\mathbf{T}^{(v)})^\top \mathbf{T}^{(v)}y \right) - \left| x^*\mathbf{W}^{(v)}x + y^*\mathbf{W}^{(v)}y \right| \right) \tag{6}$$

Applying the triangle inequality to the left-hand side of Inequality 6 gives

$$\sum_{v \in [n]} \left| x^*(\widetilde{\mathbf{W}}^{(v)} - \mathbf{W}^{(v)})y \right| \geq \left| \sum_{v \in [n]} x^*(\widetilde{\mathbf{W}}^{(v)} - \mathbf{W}^{(v)})y \right|$$
$$= d \cdot \left| x^*(\widetilde{\mathbf{W}} - \mathbf{W}^2)y \right|.$$

Similarly, we can apply the triangle inequality to the sum of complex magnitudes on the right-hand side of Inequality 6 to conclude

$$d \cdot \left| x^*(\widetilde{\mathbf{W}} - \mathbf{W}^2)y \right| \leq \frac{\epsilon}{2} \cdot \left( \sum_{v \in [n]} \left( x^*(\mathbf{T}^{(v)})^\top \mathbf{T}^{(v)}x + y^*(\mathbf{T}^{(v)})^\top \mathbf{T}^{(v)}y \right) - 2 \cdot d \cdot \left| x^*\mathbf{W}^2x + y^*\mathbf{W}^2y \right| \right).$$

Finally, we show that

$$\sum_{v \in [n]} (\mathbf{T}^{(v)})^\top \mathbf{T}^{(v)} = 2 \cdot d \cdot \mathbf{I}_n. \tag{7}$$

This will complete the proof, because our inequality will become

$$d \cdot \left| x^*(\widetilde{\mathbf{W}} - \mathbf{W}^2)y \right| \leq 2 \cdot d \cdot \frac{\epsilon}{2} \cdot \left( \|x\|^2 + \|y\|^2 - \left| x^*\mathbf{W}^2x + y^*\mathbf{W}^2y \right| \right),$$

which says that $\widetilde{\mathbf{W}} \overset{\circ}{\approx}_{2 \cdot \epsilon} \mathbf{W}^2$. Now we prove Equation 7. Fix $v, i, j \in [n]$. We can write

$$\left( \left( \mathbf{T}^{(v)} \right)^\top \mathbf{T}^{(v)} \right)_{ij} = \sum_{k \in [2d]} (\mathbf{T}^{(v)})^\top_{ik} \mathbf{T}^{(v)}_{kj}$$
$$= \sum_{k \in [2d]} \mathbf{T}^{(v)}_{ki} \mathbf{T}^{(v)}_{kj}.$$

From the definition of $\mathbf{T}^{(v)}$, there is exactly one 1 in every row of the matrix, so when $i \neq j$, the above sum is 0. When $i = j$, the sum contributes a 1 for each edge connecting $v$ and $i$ in the graph (in either direction). Therefore $\sum_{v \in [n]} (\mathbf{T}^{(v)})^\top \mathbf{T}^{(v)}$ is always 0 off the diagonal and $2 \cdot d$ on the diagonal because every vertex has exactly $2 \cdot d$ edges incident to it. This confirms Equation 7 and completes the proof. $\qquad\square$

# 6 Approximate Pseudoinverse for Cycle-Lifted Graphs

Let $\mathbf{I} - \mathbf{W}$ be the random-walk Laplacian of a strongly connected, aperiodic, regular digraph $G$. Our goal is to compute an accurate approximation of $(\mathbf{I} - \mathbf{W})^+$. To do this we consider the Laplacian of a cycle-lifted graph $\mathbf{L} = \mathbf{I}_{2^k n} - \mathbf{C}_{2^k} \otimes \mathbf{W}$ for some positive integer $k$, and show how to compute an accurate approximation of $\mathbf{L}^+$, namely $\widetilde{\mathbf{L}^+}$. Then we show that under some conditions, an $n \times n$ projection of $\widetilde{\mathbf{L}^+}$ (specifically, $(\vec{\mathbf{1}}_{2^k} \otimes \mathbf{I}_n)^\top \widetilde{\mathbf{L}^+} (\vec{\mathbf{1}}_{2^k} \otimes \mathbf{I}_n)$) gives an accurate approximation for $(\mathbf{I} - \mathbf{W})^+$ (see Lemma 6.4).

To estimate $\mathbf{L}^+$ we first show how to obtain a weak approximation to it. Then we show how to get an accurate approximation using Richardson iteration (see Lemma 6.2). The following is the main theorem we prove in this section. In this theorem, we only give sufficient conditions for having an approximate pseudo-inverse of the cycle-lifted graphs, and discuss an actual space-efficient algorithm for computing such a matrix in Section 7.

**Theorem 6.1.** *Let $\mathbf{W}$ be the transition matrix of a strongly connected regular digraph with $n$ vertices, $\epsilon \in (0, 1/2)$ and suppose we have a sequence of matrices $\mathbf{W} = \mathbf{W}_0, \ldots, \mathbf{W}_k$, such that*

$$\mathbf{W}_{i+1} \overset{\circ}{\approx}_{\epsilon/k} \mathbf{W}_i^2 \quad \forall 0 < i < k$$

*and each $\mathbf{W}_i$ is a transition matrix of a strongly connected regular digraph. We use $\mathbf{W}_i$'s to define a sequence of $2^k n$ by $2^k n$ matrices $\mathbf{L}^{(i)}$ as in Equation (11). Then for $\mathbf{L} = \mathbf{I}_{2^k n} - \mathbf{C}_{2^k} \otimes \mathbf{W}$ and $\widetilde{\mathbf{L}} \overset{\text{def}}{=} \mathbf{L}^{(k)}$, there exists a PSD matrix $\mathbf{F}$ such that $\|\mathbf{I}_{2^k n} - \widetilde{\mathbf{L}}^+ \mathbf{L}\|_{\mathbf{F}} \le O(k\epsilon)$ and $\mathbf{U_L}/O(k) \preceq \mathbf{F} \preceq O(2^{2k} n^2 k^5) \mathbf{U_L}$.*

In the above theorem $\widetilde{\mathbf{L}}$ is defined in a way so that it has a nice LU factorization. This lets us efficiently compute $\widetilde{\mathbf{L}}^+$. Below we describe how we use Theorem 6.1 in our solver and then we prove the theorem in Section 6.1. We give the characterization of $\widetilde{\mathbf{L}}^+$ in Proposition 6.6. In Section 7, for $k = O(\log n)$, we show how to space efficiently generate the $\mathbf{W}_i$'s and compute $\widetilde{\mathbf{L}}^+$.

The following lemma shows how we can obtain an accurate solver by boosting the precision of an approximate pseudo-inverse through the well-known method of preconditioned Richardson iteration [CKP+1, PS].

**Lemma 6.2.** *Given matrices $\mathbf{A}, \mathbf{B}, \mathbf{F} \in \mathbb{R}^{n \times n}$, such that $\mathbf{F}$ is PSD, and $\|\mathbf{I} - \mathbf{BA}\|_{\mathbf{F}} \le \alpha$ for some constant $\alpha > 0$. Let $\mathbf{P}_m = \sum_{i=0}^{m} (\mathbf{I} - \mathbf{BA})^i \mathbf{B}$. Then*

$$\|\mathbf{I} - \mathbf{P}_m \mathbf{A}\|_{\mathbf{F}} \le \alpha^{m+1}$$

*Proof.* We have $\mathbf{I} - \mathbf{P}_m \mathbf{A} = (\mathbf{I} - \mathbf{BA})^{m+1}$, and then the proof follows by the submultiplicity of $\|\cdot\|_{\mathbf{F}}$. $\qquad\square$

Since we can obtain a reasonably good approximate pseudo-inverse for $\mathbf{L}$ via Theorem 6.1 and boost the quality of that approximation with Lemma 6.2, we can ultimately get a very accurate approximate pseudo-inverse. This is stated rigorously in the following corollary.

**Corollary 6.3.** *Given a transition matrix $\mathbf{W}$ of a regular digraph with $n$ vertices, and $\delta \in (0, 1/2)$. Let $\mathbf{L} = \mathbf{I}_{2^k n} - \mathbf{C}_{2^k} \otimes \mathbf{W}$, and let $\widetilde{\mathbf{L}}^+$ be the approximate pseudo-inverse obtained from Theorem 6.1 by setting $\epsilon = \frac{1}{ck}$ for a large enough constant $c$. For $m = O\left(k + \log n + \log(\frac{1}{\delta})\right)$, and $\mathbf{P}_m = \sum_{i=0}^{m} (\mathbf{I} - \widetilde{\mathbf{L}}^+ \mathbf{L})^i \widetilde{\mathbf{L}}^+$, we have*

$$\|\mathbf{I} - \mathbf{P}_m \mathbf{L}\|_{\mathbf{U_L}} \le \delta$$

*Proof.* Note that with the choice of $\epsilon$, we have $\|\mathbf{I} - \widetilde{\mathbf{L}}^+ \mathbf{L}\|_{\mathbf{F}} \le \frac{1}{2}$. Therefore, by Lemma 6.2 we get $\|\mathbf{I} - \mathbf{P}_m \mathbf{L}\|_{\mathbf{F}} \le \frac{\delta}{poly(2^k, n)}$. Since $\mathbf{U_L}/O(k) \preceq \mathbf{F} \preceq O(2^{2k} n^2 k^5) \mathbf{U_L}$, this implies $\|\mathbf{I} - \mathbf{P}_m \mathbf{L}\|_{\mathbf{U_L}} \le \delta$. $\qquad\square$

22

In the lemma below, we show how to obtain an approximate pseudo-inverse of the original Laplacian system, given an approximate pseudo-inverse for the cycle-lifted graph.

**Lemma 6.4.** *Given a matrix* $\mathbf{W} \in \mathbb{R}_{\geq 0}^{n \times n}$ *with* $\|\mathbf{W}\| \leq 1$ *and an arbitrary integer* $\ell > 0$, *let* $\mathbf{L}_W = \mathbf{I}_n - \mathbf{W}$, *and* $\mathbf{L}_C = \mathbf{I}_{\ell n} - \mathbf{C}_\ell \otimes \mathbf{W}$. *Let* $\mathbf{B}_C \in \mathbb{R}^{\ell n \times \ell n}$ *such that*

$$\|\mathbf{I}_{\ell n} - \mathbf{B}_C \mathbf{L}_C\|_{\mathbf{U}_{\mathbf{L}_C}} \leq \delta.$$

*Let* $\mathbf{B}_W = \frac{1}{\ell}(\vec{\mathbf{1}}_\ell \otimes \mathbf{I}_n)^\top \mathbf{B}_C (\vec{\mathbf{1}}_\ell \otimes \mathbf{I}_n)$; *then*

$$\|\mathbf{I}_n - \mathbf{B}_W \mathbf{L}_W\|_{\mathbf{U}_{\mathbf{L}_W}} \leq \delta.$$

*Proof.* $\|\mathbf{I}_n - \mathbf{B}_W \mathbf{L}_W\|_{\mathbf{U}_{\mathbf{L}_W}} \leq \delta$ is equivalent to

$$(\mathbf{I}_n - \mathbf{B}_W \mathbf{L}_W)^\top \mathbf{U}_{\mathbf{L}_W}(\mathbf{I}_n - \mathbf{B}_W \mathbf{L}_W) \preceq \delta^2 \mathbf{U}_{\mathbf{L}_W},$$

For the RHS we have $\delta^2 \cdot \frac{1}{\ell}(\vec{\mathbf{1}}_\ell \otimes \mathbf{I}_n)^\top \mathbf{U}_{\mathbf{L}_C}(\vec{\mathbf{1}}_\ell \otimes \mathbf{I}_n) = \delta^2 \mathbf{U}_{\mathbf{L}_W}$. For the LHS, let $\Pi = \frac{1}{\ell}(\vec{\mathbf{1}}_\ell \otimes \mathbf{I}_n)(\vec{\mathbf{1}}_\ell \otimes \mathbf{I}_n)^\top = (\frac{\vec{\mathbf{1}}_\ell \vec{\mathbf{1}}_\ell^\top}{\ell}) \otimes \mathbf{I}_n$, then we get

$$\frac{1}{\ell}(\vec{\mathbf{1}}_\ell \otimes \mathbf{I}_n)^\top (\mathbf{I}_{\ell n} - \mathbf{B}_C \Pi \mathbf{L}_C)^\top \Pi \mathbf{U}_{\mathbf{L}_C} \Pi (\mathbf{I}_{\ell n} - \mathbf{B}_C \Pi \mathbf{L}_C)(\vec{\mathbf{1}}_\ell \otimes \mathbf{I}) = (\mathbf{I}_n - \mathbf{B}_W \mathbf{L}_W)^\top \mathbf{U}_{\mathbf{L}_W}(\mathbf{I}_n - \mathbf{B}_W \mathbf{L}_W).$$

Thus it is sufficient to show,

$$\frac{1}{\ell}(\vec{\mathbf{1}}_\ell \otimes \mathbf{I}_n)^\top (\mathbf{I}_{\ell n} - \mathbf{B}_C \Pi \mathbf{L}_C)^\top \Pi \mathbf{U}_{\mathbf{L}_C} \Pi (\mathbf{I}_{\ell n} - \mathbf{B}_C \Pi \mathbf{L}_C)(\vec{\mathbf{1}}_\ell \otimes \mathbf{I}) \preceq \delta^2 \cdot \frac{1}{\ell}(\vec{\mathbf{1}}_\ell \otimes \mathbf{I}_n)^\top \mathbf{U}_{\mathbf{L}_C}(\vec{\mathbf{1}}_\ell \otimes \mathbf{I}_n).$$

Note that $\Pi$ is an orthogonal projection, and $\Pi \mathbf{L}_C = \mathbf{L}_C \Pi$. Thus by the lemma assumption and Lemma D.3, we have,

$$(\mathbf{I}_{\ell n} - \mathbf{B}_C \mathbf{L}_C)^\top \Pi \mathbf{U}_{\mathbf{L}_C}(\mathbf{I}_{\ell n} - \mathbf{B}_C \mathbf{L}_C) \preceq (\mathbf{I}_{\ell n} - \mathbf{B}_C \mathbf{L}_C)^\top \mathbf{U}_{\mathbf{L}_C}(\mathbf{I}_{\ell n} - \mathbf{B}_C \mathbf{L}_C) \preceq \delta^2 \mathbf{U}_{\mathbf{L}_C}.$$

Since $\Pi^2 = \Pi$, and $\Pi$ commutes with $\mathbf{L}_C$ and $\mathbf{U}_{\mathbf{L}_C}$, we get

$$(\mathbf{I}_{\ell n} - \mathbf{B}_C \Pi \mathbf{L}_C)^\top \Pi \mathbf{U}_{\mathbf{L}_C} \Pi (\mathbf{I}_{\ell n} - \mathbf{B}_C \Pi \mathbf{L}_C) \preceq \delta^2 \mathbf{U}_{\mathbf{L}_C}.$$

Now by Proposition 2.1, we get

$$\frac{1}{\ell}(\vec{\mathbf{1}}_\ell \otimes \mathbf{I}_n)^\top (\mathbf{I}_{\ell n} - \mathbf{B}_C \Pi \mathbf{L}_C)^\top \Pi \mathbf{U}_{\mathbf{L}_C} \Pi (\mathbf{I}_{\ell n} - \mathbf{B}_C \Pi \mathbf{L}_C)(\vec{\mathbf{1}}_\ell \otimes \mathbf{I}) \preceq \delta^2 \cdot \frac{1}{\ell}(\vec{\mathbf{1}}_\ell \otimes \mathbf{I}_n)^\top \mathbf{U}_{\mathbf{L}_C}(\vec{\mathbf{1}}_\ell \otimes \mathbf{I}_n). \quad (8)$$

which completes the proof. $\square$

## 6.1 Approximate Pseudoinverse of Cycle-Lifted Graphs

To get an approximate pseudo-inverse of $\mathbf{L}$ we first compute an approximate LU factorization of it. For any matrix $\mathbf{M} \in \mathbb{R}^{n \times n}$, if $F, C$ partition $[n]$ and $\mathbf{M}_{FF}$ is invertible we can write $\mathbf{M}$ as the product of a lower triangular matrix, a block diagonal matrix, and an upper triangular matrix:

$$\mathbf{M} = \begin{bmatrix} \mathbf{M}_{FF} & \mathbf{M}_{FC} \\ \mathbf{M}_{CF} & \mathbf{M}_{CC} \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{I} & 0 \\ \mathbf{M}_{CF}\mathbf{M}_{FF}^{-1} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{M}_{FF} & 0 \\ 0 & \mathbf{M}_{CC} - \mathbf{M}_{CF}\mathbf{M}_{FF}^{-1}\mathbf{M}_{FC} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{M}_{FF}^{-1}\mathbf{M}_{FC} \\ 0 & \mathbf{I} \end{bmatrix} \quad (9)$$

23

Note that $\mathbf{M}_{CC} - \mathbf{M}_{CF}\mathbf{M}_{FF}^{-1}\mathbf{M}_{FC}$ is the Schur complement of $\mathbf{M}$ onto the set $C$. For an invertible matrix $\mathbf{M}$, the above factorization gives a formula to compute $\mathbf{M}^{-1}$:

$$\mathbf{M}^{-1} = \begin{bmatrix} \mathbf{I} & -\mathbf{M}_{FF}^{-1}\mathbf{M}_{FC} \\ 0 & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{M}_{FF}^{-1} & 0 \\ 0 & \mathrm{Sc}(\mathbf{M},C)^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{I} & 0 \\ -\mathbf{M}_{CF}\mathbf{M}_{FF}^{-1} & \mathbf{I} \end{bmatrix}$$

The lower and upper triangular parts are easy to invert. Therefore, the above formula reduces inverting $\mathbf{M}$ to inverting $\mathbf{M}_{FF}$ and $\mathrm{Sc}(\mathbf{M}, C)$. This approach has been used in many recent time efficient algorithms for solving both symmetric and asymmetric diagonally dominant systems [CKK+, KLP+]. We use the same approach of LU factorization to compute an approximate pseudo-inverse of $\mathbf{L}$.

Without loss of generality we use the following ordering of rows and columns for $\mathbf{C}_{2^k}$. Let $\mathbf{C}_1 = [1]$, and for $k > 0$ let

$$\mathbf{C}_{2^k} = \begin{bmatrix} 0 & \mathbf{C}_{2^{k-1}} \\ \mathbf{I}_{2^{k-1}} & 0 \end{bmatrix}.$$

Let $H$ be the set of coordinates $\{2^{k-1}n+1, 2^{k-1}n+2, \ldots, 2^k n\}$, so that the cycle alternates between $H$ and $H^c$. We get the following nice characterization for the Schur complement of $\mathbf{I}_{2^k n} - \mathbf{C}_{2^k} \otimes \mathbf{W}$ on to set $H$.

$$\mathrm{Sc}(\mathbf{I}_{2^k n} - \mathbf{C}_{2^k} \otimes \mathbf{W}, H) = \mathbf{I}_{2^{k-1}n} - (\mathbf{I}_{2^{k-1}} \otimes \mathbf{W})(\mathbf{C}_{2^{k-1}} \otimes \mathbf{W}) = \mathbf{I}_{2^{k-1}n} - (\mathbf{C}_{2^{k-1}} \otimes \mathbf{W}^2).$$

Using the above relation, and (9) we can get the following factorization for $\mathbf{L}$.

$$\mathbf{L} = \begin{bmatrix} \mathbf{I}_{2^{k-1}n} & 0 \\ -\mathbf{I}_{2^{k-1}} \otimes \mathbf{W} & \mathbf{I}_{2^{k-1}n} \end{bmatrix} \begin{bmatrix} \mathbf{I}_{2^{k-1}n} & 0 \\ 0 & \mathbf{I}_{2^{k-1}n} - \mathbf{C}_{2^{k-1}} \otimes \mathbf{W}^2 \end{bmatrix} \begin{bmatrix} \mathbf{I}_{2^{k-1}n} & -\mathbf{C}_{2^{k-1}} \otimes \mathbf{W} \\ 0 & \mathbf{I}_{2^{k-1}n} \end{bmatrix} \quad (10)$$

Note that the lower right block of the middle matrix is the Laplacian of cycle-lifted graph with cycle length $2^{k-1}$ and adjacency matrix $\mathbf{W}^2$. Applying this recursion one more time by short-cutting every other layer in the smaller cycle-lifted graph leads to a cycle-lifted graph of length $2^{k-2}$ and adjacency matrix $\mathbf{W}^4$. To get an approximate LU factorization, we repeatedly apply this recurrence and replace all occurrences of powers of $\mathbf{W}$ with appropriate approximations of them (see Theorem 6.1). For $j \in \{1, \ldots, k\}$ let $\mathbf{W}_j$'s be defined as in Theorem 6.1. With $\mathbf{L}^{(0)} = \mathbf{L}$, we denote by $\mathbf{L}^{(i)}$ the matrix obtained after applying $i$ steps of recursion. For example, we have

$$\mathbf{L}^{(1)} = \begin{bmatrix} \mathbf{I}_{2^{k-1}n} & 0 \\ -\mathbf{I}_{2^{k-1}} \otimes \mathbf{W} & \mathbf{I}_{2^{k-1}n} \end{bmatrix} \begin{bmatrix} \mathbf{I}_{2^{k-1}n} & 0 \\ 0 & \mathbf{I}_{2^{k-1}n} - \mathbf{C}_{2^{k-1}} \otimes \mathbf{W}_1 \end{bmatrix} \begin{bmatrix} \mathbf{I}_{2^{k-1}n} & -\mathbf{C}_{2^{k-1}} \otimes \mathbf{W} \\ 0 & \mathbf{I}_{2^{k-1}n} \end{bmatrix}.$$

More generally, for $1 \leq i \leq k$ we get,

$$\mathbf{L}^{(i)} = \mathbf{X}_1 \cdots \mathbf{X}_i \begin{bmatrix} \mathbf{I}_{(2^k - 2^{k-i})n} & 0 \\ 0 & \mathbf{I}_{2^{k-i}n} - \mathbf{C}_{2^{k-i}} \otimes \mathbf{W}_i \end{bmatrix} \mathbf{Y}_i \cdots \mathbf{Y}_1 \quad (11)$$

where $\mathbf{X}_j$'s and $\mathbf{Y}_j$'s are lower and upper triangular matrices, respectively. Specifically, we have

$$\mathbf{X}_j = \begin{bmatrix} \mathbf{I}_{(2^k - 2^{k-j+1})n} & 0 & 0 \\ 0 & \mathbf{I}_{2^{k-j}n} & 0 \\ 0 & -\mathbf{I}_{2^{k-j}} \otimes \mathbf{W}_{j-1} & \mathbf{I}_{2^{k-j}n} \end{bmatrix}, \mathbf{Y}_j = \begin{bmatrix} \mathbf{I}_{(2^k - 2^{k-j+1})n} & 0 & 0 \\ 0 & \mathbf{I}_{2^{k-j}n} & -\mathbf{C}_{2^{k-j}} \otimes \mathbf{W}_{j-1} \\ 0 & 0 & \mathbf{I}_{2^{k-j}n} \end{bmatrix}.$$
$$(12)$$

Given the above factorization for $\mathbf{L}^{(k)}$ it is easy to get one for $\mathbf{L}^{(k)^+}$. We use the following lemma from [CKK+] to give a factorization of $\mathbf{L}^{(k)^+}$.

**Lemma 6.5** (CKKPPRS18 Lemma C.3). *Consider real matrices $\mathbf{A} \in \mathbb{R}^{m \times m}$, $\mathbf{B} \in \mathbb{R}^{m \times n}$, and $\mathbf{C} \in \mathbb{R}^{n \times n}$, where $\mathbf{A}$ and $\mathbf{C}$ are invertible. Let $\mathbf{M} = \mathbf{ABC}$, then $\mathbf{M}^+ = \mathbf{P_M} C^{-1} B^+ A^{-1} \mathbf{P}_{\mathbf{M}^\top}$, where $\mathbf{P_M}$ defines the orthogonal projection into image of $\mathbf{M}$.*

The following proposition gives the characterization of $\mathbf{L}^{(k)^+}$ which we later use for computing $\mathbf{L}^{(k)^+}$ in small space. Note that inverting the $\mathbf{X}_j$'s, and $\mathbf{Y}_j$'s is easy as they are lower and upper triangular matrices.

**Proposition 6.6.** *Given $\mathbf{L}^{(k)}$ defined by Equation (11), we have*

$$\mathbf{L}^{(k)^+} = \mathbf{P}_{\mathbf{L}^{(k)}} \mathbf{Y}_1^{-1} \cdots \mathbf{Y}_k^{-1} \begin{bmatrix} \mathbf{I}_{(2^k-1)n} & 0 \\ 0 & \mathbf{I}_n - \mathbf{W}_k \end{bmatrix}^+ \mathbf{X}_k^{-1} \cdots \mathbf{X}_1^{-1} \mathbf{P}_{\mathbf{L}^{(k)^\top}} \tag{13}$$

*where*

$$\mathbf{X}_j^{-1} = \begin{bmatrix} \mathbf{I}_{(2^k-2^{k-j+1})n} & 0 & 0 \\ 0 & \mathbf{I}_{2^{k-j}n} & 0 \\ 0 & \mathbf{I}_{2^{k-j}} \otimes \mathbf{W}_{j-1} & \mathbf{I}_{2^{k-j}n} \end{bmatrix}, \mathbf{Y}_j^{-1} = \begin{bmatrix} \mathbf{I}_{(2^k-2^{k-j+1})n} & 0 & 0 \\ 0 & \mathbf{I}_{2^{k-j}n} & \mathbf{C}_{2^{k-j}} \otimes \mathbf{W}_{j-1} \\ 0 & 0 & \mathbf{I}_{2^{k-j}n} \end{bmatrix}.$$

To prove Theorem 6.1, we first show that $\mathbf{L}^{(k)}$ is a good approximation of $\mathbf{L}$. For that, we build a PSD matrix $\mathbf{F}$ such that $\mathbf{L}$ and $\mathbf{L}^{(k)}$ approximate each other with respect to the norm defined by $\mathbf{F}$. For $0 \leq j \leq k$, let

$$\mathbf{S}^{(j)} = \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{I}_{2^{k-j} \cdot n} - \mathbf{C}_{2^{k-j}} \otimes \mathbf{W}_j \end{bmatrix}, \tag{14}$$

where the zeros are used as padding to make the dimension of $\mathbf{S}^{(j)}$'s $2^k n \times 2^k n$. Note that $\mathbf{S}^{(j)}$'s correspond to the approximate Schur complement blocks appear in our recursive algorithm (see Equation 11). Using Lemma 2.3 from [CKK$^+$], we show that the average of $\mathbf{U}_{\mathbf{S}^{(i)}}$'s would be a good choice for $\mathbf{F}$. We defer the statement of the lemma from [CKK$^+$] and the proof of Lemma 6.7 to the appendix (see Appendix D).

**Lemma 6.7.** *Let $\mathbf{S}^{(0)}, \mathbf{S}^{(1)}, \ldots, \mathbf{S}^{(k)}$, and $\mathbf{L}^{(0)}, \mathbf{L}^{(1)}, \ldots, \mathbf{L}^{(k)}$ be defined as Equations (14) and (11) respectively. Then for,*

$$\mathbf{F} = \frac{2}{k} \sum_{i=0}^{k} \mathbf{U}_{\mathbf{S}^{(i)}}$$

*we have:*

1. *for each $0 \leq i \leq k$,*

$$\left\| \mathbf{F}^{+/2} \left( \mathbf{L} - \mathbf{L}^{(i)} \right) \mathbf{F}^{+/2} \right\|_2 \leq \epsilon,$$

2. *The final matrix $\mathbf{L}^{(k)}$ satisfies*

$$\mathbf{L}^{(k)\top} \mathbf{F}^+ \mathbf{L}^{(k)} \succeq \frac{1}{40k^2} \cdot \mathbf{F}.$$

Informally, Item 1 states that all of the $\mathbf{L}^{(i)}$'s and in particular $\mathbf{L}^{(k)}$ are good approximations of $\mathbf{L}$ with respect to $\mathbf{F}$, or equivalently $\mathbf{F}^{+/2} \mathbf{L}^{(i)} \mathbf{F}^{+/2}$ is a good approximation of $\mathbf{F}^{+/2} \mathbf{L} \mathbf{F}^{+/2}$ in spectral norm. Item 2 states that $\mathbf{L}^{(k)}$ is not too small with respect to $\mathbf{F}$, or equivalently that the spectral norm of $(\mathbf{F}^{+/2} \mathbf{L}^{(k)} \mathbf{F}^{+/2})^+ = \mathbf{F}^{1/2} (\mathbf{L}^{(k)})^+ \mathbf{F}^{1/2}$ is not too large. Together they are used to show that for $\widetilde{\mathbf{L}} = \mathbf{L}^{(k)}$, $\widetilde{\mathbf{L}}^+$ is a good preconditioner for $\mathbf{L}$ in the $\mathbf{F}$ norm, i.e. $\|\mathbf{I}_{2^k n} - \widetilde{\mathbf{L}}^+ \mathbf{L}\|_{\mathbf{F}}$ is small (see Lemma D.4). Now we are ready to give a proof of Theorem 6.1.

*Proof of Theorem 6.1.* By Lemma 6.7, there exists matrix $\mathbf{F}$ such that,

$$\|\mathbf{F}^{+/2}(\mathbf{L} - \widetilde{\mathbf{L}})\mathbf{F}^{+/2}\| \leq \epsilon$$

and

$$\widetilde{\mathbf{L}}^\top \mathbf{F}^+ \widetilde{\mathbf{L}} \succeq \frac{1}{40k^2}\mathbf{F}.$$

Therefore, by Lemma D.4, we get $\|\mathbf{I}_{2^k n} - \widetilde{\mathbf{L}}^+\mathbf{L}\|_{\mathbf{F}} \leq \sqrt{40}k\epsilon$. Note that $\mathbf{F} \succeq \frac{\mathbf{U_L}}{k}$ by construction. To get the upper-bound on $\mathbf{F}$, we have

$$\mathbf{F} \preceq O(k^2)\widetilde{\mathbf{L}}^\top \mathbf{F}^+ \widetilde{\mathbf{L}} \preceq O(k^4)\mathbf{L}^\top \mathbf{F}^+ \mathbf{L} \preceq O(k^5)\mathbf{L}^\top \mathbf{U_L}^+ \mathbf{L}$$

where the second inequality holds by Lemma D.5, and the last inequality comes from $\mathbf{F} \succeq \frac{\mathbf{U_L}}{k}$. By applying Lemma D.6 we get,

$$\mathbf{F} \preceq O(k^5)\mathbf{L}^\top \mathbf{U_L}^+ \mathbf{L} \preceq O(k^5 2^{2k} n^2)\mathbf{U_L}.$$

$\square$

# 7  Space-Efficient Eulerian Laplacian Solver

In this section we show that the LU factorization approach to computing the pseudoinverse of an Eulerian Laplacian described in Section 6 can be implemented space-efficiently. In particular, we prove the following theorem.

**Theorem 7.1.** *There is a deterministic algorithm that, given $\epsilon \in (0,1)$ and an Eulerian digraph $G$ with random-walk Laplacian $\mathbf{L} = \mathbf{I} - \mathbf{W}$, computes a matrix $\widetilde{\mathbf{L}}^+$ whose entries differ from $\mathbf{L}^+$ by at most $\epsilon$. The algorithm uses space $O(\log N \cdot \log\log(N/\epsilon))$ where $N$ is the bitlength of the input.*

## 7.1  Model of Space-Bounded Computation

We use a standard model of space bounded computation. The machine has a read-only input tape, a constant number of read/write work tapes, and a write-only output tape. We say the machine runs in space $s$ if throughout the computation, it only uses $s$ total tape cells on the work tapes. The machine may write outputs to the output tape that are larger than $s$ (in fact as large as $2^{O(s)}$) but the output tape is write-only. We use the following fact about the composition of space-bounded algorithms.

**Proposition 7.2** (Composition of Space-Bounded Algorithms)**.** *Let $f_1$ and $f_2$ be functions that can be computed in space $s_1(n), s_2(n) \geq \log n$, respectively, and $f_1$ has output of length $\ell_1(n)$ on inputs of size $n$. Then $f_2 \circ f_1$ can be computed in space*

$$O(s_2(\ell_1(n)) + s_1(n)).$$

## 7.2  Reduction to Regular, Aperiodic, Strongly Connected Case

It is convenient to work with regular, aperiodic, strongly connected digraphs. In this section, we show that this is without loss of generality.

We can find the strongly connected components of an Eulerian graph by ignoring the directionality on the edges and running Reingold's algorithm for UNDIRECTED S-T CONNECTIVITY on each pair of vertices using $O(\log N)$ space. Then we can solve systems in a disconnected graph

by solving systems on each of its strongly connected components separately. So without loss of generality, our graph is strongly connected.

Given an Eulerian digraph $G$ with maximum degree $d_{\max}$ and Laplacian $\mathbf{D} - \mathbf{A}$, we can create a $d$-regular, aperiodic graph of any degree $d > d_{\max}$ by adding $d - \mathrm{degree}(v)$ self loops to each vertex $v$. Notice that self loops do not change the Laplacian $\mathbf{D} - \mathbf{A}$. Our solver works with random-walk Laplacians and so is able to approximate $((\mathbf{D} - \mathbf{A})/d)^+$ and now we want to show how to use this to approximate $((\mathbf{D} - \mathbf{A})\mathbf{D}^{-1})^+$.

Let $\mathbf{P}$ be the orthogonal projection onto the column space of $(\mathbf{D} - \mathbf{A})\mathbf{D}^{-1}$ and let $\mathbf{Q}$ be the orthogonal projection onto the column space of $\mathbf{D}^{-1}(\mathbf{D} - \mathbf{A}^\top)$. By Lemma 6.5 we have

$$((\mathbf{D} - \mathbf{A})\mathbf{D}^{-1})^+ = \mathbf{P}\mathbf{D}(\mathbf{D} - \mathbf{A})^+\mathbf{Q}.$$

Given an $\epsilon$-approximation to $((\mathbf{D} - \mathbf{A})/d)^+ = d \cdot (\mathbf{D} - \mathbf{A})^+$, we can divide this by $d$ to get an $\epsilon$-approximation to $(\mathbf{D} - \mathbf{A})^+$. Plugging this in to the formula above and making $\epsilon$ sufficiently small says that we just need to compute the matrices $\mathbf{P}$ and $\mathbf{Q}$ in order to get an approximation to $((\mathbf{D} - \mathbf{A})\mathbf{D}^{-1})^+$. If $s$ is the stationary distribution of $G$, then $\mathbf{P}$ is the matrix $\mathbf{I} - ss^\top/\|s\|^2$. Since $G$ is Eulerian, the kernel of $\mathbf{D}^{-1}(\mathbf{D} - \mathbf{A}^\top)$ is simply the all ones vector and so $\mathbf{Q} = \mathbf{I} - \vec{\mathbf{1}}\vec{\mathbf{1}}^\top/n$. The stationary distribution of a strongly connected Eulerian graph is proprtional to its vertex degrees and so is easy to compute in logspace. Thus $\mathbf{P}$ and $\mathbf{Q}$ can both be computed in deterministic logspace and we can approximate $((\mathbf{D} - \mathbf{A})\mathbf{D}^{-1})^+$ from $(\mathbf{D} - \mathbf{A})^+$.

## 7.3  Proof of Theorem 7.1

To prove Theorem 7.1, we follow the LU factorization approach discussed in Section 6.1. Just as in Section 6.1, we first show that we can compute a weak approximation to the pseudoinverse space-efficiently and then argue that we can afford to reduce the error using preconditioned Richardson iteration. The analysis in this section is similar to the space complexity analysis from [MRSV1].

Proposition 6.6 says that in order to compute a weak approximation to a pseudoinverse of a random-walk Laplacian $\mathbf{I} - \mathbf{W}$ it suffices to compute a particular polynomial $p$ in the matrices $\mathbf{I}, \mathbf{W}$, and approximations to $\mathbf{W}^{2^k}$ for $k = \{1, \dots, O(\log n)\}$. To compute the approximations to $\mathbf{W}^{2^k}$, we will show that the $k$th derandomized square can be computed in space $O(\log n + k \cdot \log c)$, where $c$ is the degree of the expanders used. To bottom out the recursion, we replace the random-walk Laplacian of the final derandomized power with $(\mathbf{I} - \mathbf{J}) = (\mathbf{I} - \vec{\mathbf{1}}\vec{\mathbf{1}}^\top/n)$, the random-walk Laplacian of the complete graph (with self loops), which is its own pseudoinverse $((\mathbf{I} - \mathbf{J})^+ = \mathbf{I} - \mathbf{J})$. Then to compute $p$, we note that $p$ has degree $O(\log n)$ and the product of $O(\log n)$ matrices can be computed in space $O(\log n \cdot \log \log n)$. Finally, to boost the weak approximation, we use preconditioned Richardson iteration (Lemma 6.2).

The discussion above gives the following informal outline for proving Theorem 7.1:

1. Show that the transition matrix of the $k$th derandomized square using $c$-regular expanders can be computed in space $O(\log n + k \cdot \log c)$

2. Show that the complete graph is a good approximation of the $k$th derandomized power for sufficiently large $k$

3. Show that $k$ arbitrary $n \times n$ matrices can be multiplied in space $O(\log n \cdot \log k)$

4. Show that preconditioned Richardson can be used to boost a weak approximation to a pseudoinverse to an $\epsilon$ approximation in space $O(\log n \cdot \log \log(n/\epsilon))$.

27

5. Show that an extremely good approximation to a pseudoinverse in a spectral sense is also a good approximation in an entrywise sense.

To show that derandomized powers can be computed in small space, we follow the proof techniques of [Rei, RV, Vad]. First we note that neighbors in the sequence of expanders we use for the iterated derandomized square can be explicitly computed space-efficiently.

**Lemma 7.3.** *For every $t \in \mathbb{N}$ and $\mu > 0$, there is a graph $H_{t,\mu}$ with a two-way labeling such that:*

- *$H$ has $2^t$ vertices and is $c$-regular for $c$ being a power of 2 bounded by $\text{poly}(t, 1/\mu)$.*

- *$\lambda(H) \leq \mu$*

- *$\text{Rot}_H$ can be evaluated in linear space in its input length, i.e. space $O(t + \log c)$.*

A short proof sketch for Lemma 7.3 can be found in [MRSV1]. The following lemma says that high derandomized powers can be computed space-efficiently.

**Lemma 7.4.** *Let $G_0$ be a $d$-regular, directed multigraph on $n$ vertices with a two-way labeling and $H_1, \ldots, H_k$ be $c$-regular undirected graphs with two-way labelings where for each $i \in [k]$, $H_i$ has $d \cdot c^{i-1}$ vertices. For each $i \in [k]$ let*

$$G_i = G_{i-1} \text{ⓢ} H_i.$$

*Then given $v_0 \in [n], i_0 \in [d \cdot c^{i-1}], j_0 \in [c]$, $\text{Rot}_{G_i}(v, (i_0, j_0))$ can be computed in space $O(\log(n \cdot d) + k \cdot \log c)$ with oracle queries to $\text{Rot}_{H_1}, \ldots, \text{Rot}_{H_k}$.*

Lemma 7.4 is stated and proven in [MRSV1] for the case of undirected multigraphs, however nothing in that proof required undirectedness. We include the proof in Appendix E.

**Corollary 7.5.** *Let $\mathbf{W}_0, \ldots, \mathbf{W}_k$ be the transition matrices of $G_0, \ldots, G_k$ as defined in Lemma 7.4. For all $\ell \in [k]$, given coordinates $i, j \in [n]$, entry $i, j$ of $\mathbf{W}_\ell$ can be computed in space $O(\log(n \cdot d) + k \cdot \log c)$.*

*Proof.* Lemma 7.4 shows that we can compute neighbors in the graph $G_\ell$ in space $O(\log(n \cdot d) + k \cdot \log c)$. Given coordinates $i, j$ the algorithm initiates a tally $t$ at 0 and computes $\text{Rot}_{G_\ell}(i, q)$ for each $q$ from 1 to $d \cdot c^{\ell-1}$, the degree of $G_\ell$. If the vertex outputted by $\text{Rot}_{G_\ell}$ is $j$, then $t$ is incremented by 1. After the loop finishes, $t$ contains the number of edges from $i$ to $j$ and the algorithm outputs $t/d \cdot c^{\ell-1}$, which is entry $i, j$ of $\mathbf{W}_\ell$. This used space $O(\log(n \cdot d) + k \cdot \log c)$ to compute the rotation map of $G_\ell$ plus space $O(\log(d \cdot c^{\ell-1}))$ to store $q$ and $t$. So the total space usage is $O(\log(n \cdot d) + k \cdot \log c) + O(\log(d \cdot c^{\ell-1})) = O(\log(n \cdot d) + k \cdot \log c)$. $\square$

Next we show that the complete graph is a good approximation of the $k$th derandomized power for sufficiently large $k$. We do this by noting that the derandomized square of a graph $G$, reduces $\lambda(G)$ and that, the smaller $\lambda(G)$, the better approximation it is to the complete graph.

**Lemma 7.6** ([RV]). *Let $G$ be a $d$-regular digraph with a two-way labeling and $\lambda(G) \leq \lambda$ and let $H$ be a $c$-regular graph on $d$ vertices with a two-way labeling and $\lambda(H) \leq \mu$. Then we have*

$$\lambda(G \text{ⓢ} H) \leq 1 - (1 - \lambda^2) \cdot (1 - \mu) \leq \lambda^2 + \mu.$$

Noting that a $d$-regular digraph $G$ on $n$ vertices has $\lambda(G) \leq 1 - 1/(2 \cdot d^2 \cdot n^2)$ [RV], we get that setting $k = O(\log n)$ and setting $\mu = O(1/\text{polylog}(n/\epsilon))$ (and hence $c = \text{polylog}(n/\epsilon)$) in the recurrence from Lemma 7.6 yields that the $k$th derandomized square $G_k$ has $\lambda(G_k) \leq \epsilon$. By Lemma 5.2 this implies that the transition matrix $\mathbf{W}_k$ of $G_k$ is such that $\mathbf{W}_k \overset{\circ}{\approx}_\epsilon \mathbf{J}$.

Next we note the space complexity of multiplying arbitrary matrices.

**Lemma 7.7.** *Given $n \times n$ matrices $\mathbf{M}_1, \ldots, \mathbf{M}_k$, their product $\mathbf{M}_1 \cdot \ldots \cdot \mathbf{M}_k$ can be computed using $O(\log N \cdot \log k)$ space, where $N$ is the size of the input ($N = k \cdot n^2 \cdot$ (bitlength of matrix entries)).*

*Proof.* This uses the natural divide and conquer algorithm and the fact that two matrices can be multiplied in logarithmic space. A detailed proof can be found in [MRSV1]. $\quad\square$

Fourth, we show that we can boost a weak approximation to the pseudoinverse to a strong approximation in small space.

**Lemma 7.8.** *There is a deterministic algorithm that, given matrices $\mathbf{A}, \mathbf{B}, \mathbf{F} \in \mathbb{R}^{n \times n}$, such that $\mathbf{F}$ is PSD, and $\|\mathbf{I} - \mathbf{B}\mathbf{A}\|_{\mathbf{F}} \leq \alpha$ for some constant $\alpha < 1$, computes a matrix $\mathbf{P}$ such that $\|\mathbf{I} - \mathbf{P}\mathbf{A}\|_{\mathbf{F}} \leq \epsilon$. The algorithm uses space $O(\log N \cdot \log(\log_{1/\alpha}(1/\epsilon)))$, where $N$ is the input size.*

*Proof.* From Lemma 6.2 we have that

$$\mathbf{P} = \sum_{i=0}^{O(\log_{1/\alpha}(1/\epsilon))} (\mathbf{I} - \mathbf{B}\mathbf{A})^i \mathbf{B}$$

has the desired property. Since the above polynomial has degree $O(\log(1/\epsilon))$, From Lemma 7.7, we can compute $\mathbf{P}$ in space $O(\log N \cdot \log(\log_{1/\alpha}(1/\epsilon)))$ as desired. $\quad\square$

Finally, we show any extremely good approximation of the pseudoinverse in the "spectral" sense of the above Lemma 7.8 is also a good entrywise approximation of the true pseudoinverse.

**Lemma 7.9.** *Suppose $\|\mathbf{I} - \mathbf{P}\mathbf{A}\|_{\mathbf{F}} \leq \epsilon$ for conformable real square matrices $\mathbf{P}, \mathbf{A}, \mathbf{F}$ where $\mathbf{F}$ is positive semidefinite. Suppose $\mathbf{P}, \mathbf{A}, \mathbf{F}$ all have the same left and right kernels, which are equal to each other. Then every entry of $\mathbf{P}$ differs by at most $\pm\frac{\epsilon}{2} \cdot (\lambda_{max}(\mathbf{A}^{\top +}\mathbf{F}\mathbf{A}^+) + \lambda_{max}(\mathbf{F}^+))$ additively from the corresponding entry of $\mathbf{A}^+$.*

*In particular, if for some parameter $N$, the matrices $\mathbf{F}, \mathbf{A}$ have min and max nonzero singular values between $1/\mathrm{poly}(N)$ and $\mathrm{poly}(N)$, then every entry of $\mathbf{P}$ differs by at most $\pm\epsilon \cdot \mathrm{poly}(N)$ from the corresponding entry of $\mathbf{A}^+$.*

*Even more specifically, if $\mathbf{A}$ is a regular Eulerian directed Laplacian matrix with integer edge weights $\leq N$ of a graph with $\gamma(G) \geq 1/\mathrm{poly}(N)$ for some $N \geq n$ and $F$ is bounded polynomially in $\mathbf{U}_{\mathbf{P}}$ or $\mathbf{I}$, then $\mathbf{P}$ differs by at most $\pm\epsilon \cdot \mathrm{poly}(N)$ from the corresponding entry of $\mathbf{A}^+$.*

*Proof.* The inequality in the lemma statement is equivalent to the statement that for all real conformable vectors $x, y$,

$$y^\top (\mathbf{A}^+ - \mathbf{P})x \leq \epsilon\sqrt{(x^\top \mathbf{A}^{\top +}\mathbf{F}\mathbf{A}^+ x) \cdot (y^\top \mathbf{F}^+ y)}.$$

By the AM-GM inequality,

$$y^\top (\mathbf{A}^+ - \mathbf{P})x \leq \frac{\epsilon}{2}\left(x^\top \mathbf{A}^{\top +}\mathbf{F}\mathbf{A}^+ x + y^\top \mathbf{F}^+ y\right).$$

Since multiplying $x$ by a negative number changes the sign of the LHS but not the RHS,

$$|y^\top (\mathbf{A}^+ - \mathbf{P})x| \leq \frac{\epsilon}{2}\left(x^\top \mathbf{A}^{\top +}\mathbf{F}\mathbf{A}^+ x + y^\top \mathbf{F}^+ y\right).$$

Consider any entry $(i, j)$ of $P^+$. Set $x, y$ to be the $i$th and $j$th standard basis vectors, respectively. Then the inequality says that the $(i, j)$ entries of $P^+$ and $A^+$ differ additvely by at most $\pm$ the RHS.

Thus, we need only bound the RHS. We know that the $x$ and $y$ we have selected are unit vectors. Thus, we may upper bound the RHS as

$$\frac{\epsilon}{2}\left(x^\top \mathbf{A}^{\top +}\mathbf{F}\mathbf{A}^+ x + y^\top \mathbf{F} y\right) \leq \frac{\epsilon}{2}\left(\max_x \frac{x^\top \mathbf{A}^{\top +}\mathbf{F}\mathbf{A}^+ x}{x^\top x} + \max_x \frac{y^\top \mathbf{F}^+ y}{y^\top y}\right).$$

The two terms inside the parentheses are the max eigenvalues of the symmetric matrices $\mathbf{A}^{\top +}\mathbf{F}\mathbf{A}^+$ and $\mathbf{F}^+$, respectively. The second part of the result follows from the fact that the operator norm is submultiplicative.

For the final part of the result, we prove that the max and min nonzero singular values of $\mathbf{A}, \mathbf{F}$ under the additional assumptions stated are between $1/\mathrm{poly}(N)$ and $\mathrm{poly}(N)$.

Note first that if we have a directed Eulerian graph with integer edge weights between 0 and $N$ that the maximum singular value of its Laplacian is at most $\mathrm{poly}(n,N) \leq \mathrm{poly}(N)$ since all entries in the matrix are at most $\mathrm{poly}(N)$. It's minimum nonzero singular value is at least $1/\mathrm{poly}(N)$ by the assumption that the spectral gap is at least $1/\mathrm{poly}(N)$ and the triangle inequality. Thus, the max and min nonzero singular values of $\mathbf{A}$ are between $1/\mathrm{poly}(N)$ and $\mathrm{poly}(N)$.

For the maximum and minimum nonzero singular values of $\mathbf{F}$, or equivalently, max and min nonzero eigenvalues of $\mathbf{F}$, it suffices to bound the min and max nonzero eigenvalues of $\mathbf{U_P}$. By the integrality of the edge weights of $\mathbf{P}$ and the fact that they are at most $N$, we have that the undirected Laplacian matrix $\mathbf{U_P}$ has has nonzero eigenvalues between $1/\mathrm{poly}(N)$ and $\mathrm{poly}(N)$. $\qquad\square$

Now we can prove Theorem 7.1

*Proof of Theorem 7.1.* Set $k = O(\log N)$. Let $\mathbf{W} = \mathbf{W}_0, \ldots, \mathbf{W}_{k-1}$ be the transition matrices of the repeated derandomized square graphs $G_0, \ldots, G_{k-1}$ as defined in Lemma 7.4 using expanders $\{H_i\}$ with degree $c = \mathrm{polylog}(N)$ and hence $\lambda(H_i) \leq 1/\mathrm{polylog}(N))$. Let $\mathbf{W}_k = \mathbf{J} = \vec{\mathbf{1}}\vec{\mathbf{1}}^\top/n$. It follows from Theorem 5.9 that for each $i \in \{0, \ldots, k-2\}$, $\mathbf{W}_{i+1} \overset{\circ}{\approx}_{1/\mathrm{polylog}(N)} \mathbf{W}i^2$ and from Lemmas 7.6 and 5.2 that $\mathbf{W}_{k-1} \overset{\circ}{\approx}_{1/\mathrm{polylog}(N)} \mathbf{W}_k = \mathbf{J}$.

Theorem 6.1 and Proposition 6.6 say that the matrix

$$\mathbf{L}^{(k)+} = \mathbf{P}_{\mathbf{L}^{(k)}}\mathbf{Y}_1^{-1}\cdots\mathbf{Y}_i^{-1}\begin{bmatrix}\mathbf{I}_{(2^k-1)n} & 0 \\ 0 & \mathbf{I}_n - \mathbf{W}_k\end{bmatrix}^+ \mathbf{X}_k^{-1}\cdots\mathbf{X}_1^{-1}\mathbf{P}_{\mathbf{L}^{(k)\top}}$$

where

$$\mathbf{X}_j^{-1} = \begin{bmatrix}\mathbf{I}_{(2^k-2^{k-j+1})n} & 0 & 0 \\ 0 & \mathbf{I}_{2^{k-j}n} & 0 \\ 0 & \mathbf{I}_{2^{k-j}}\otimes\mathbf{W}_{j-1} & \mathbf{I}_{2^{k-j}n}\end{bmatrix}, \mathbf{Y}_j^{-1} = \begin{bmatrix}\mathbf{I}_{(2^k-2^{k-j+1})n} & 0 & 0 \\ 0 & \mathbf{I}_{2^{k-j}n} & \mathbf{C}_{2^{k-j}}\otimes\mathbf{W}_{j-1} \\ 0 & 0 & \mathbf{I}_{2^{k-j}n}\end{bmatrix}.$$

is such that there exists a PSD matrix $\mathbf{F}$ such that $\|\mathbf{I}_{2^k n} - \widetilde{\mathbf{L}}^+\mathbf{L}\|_\mathbf{F} \leq O(k\epsilon)$ and $\mathbf{U_L}/O(k) \preceq \mathbf{F} \preceq O(2^{2k}n^2k^5)\mathbf{U_L}$.

Note that $\mathbf{L}^{(k)+}$ matrix is the product of $O(k) = O(\log N)$ matrices, each of which can be computed in space $O(\log N + k \cdot \log c) = O(\log N \cdot \log\log N)$ by Corollary 7.5. Multiplying them together to compute $\widetilde{L}^+$ adds an additional $O(\log N \cdot \log k) = O(\log N \cdot \log\log N)$ space by Lemma 7.7 and the composition of space-bounded algorithms (Proposition 7.2). Thus the total space for computing $\mathbf{L}^{(k)+}$ is $O(\log N \cdot \log\log N)$.

Applying preconditioned Richardson to $\mathbf{L}^{(k)+}$, Lemma 7.8 says that we can compute a matrix $\widetilde{\mathbf{L}}^+$ such that $\|\mathbf{I}_{\mathrm{im}(\mathbf{L})} - \widetilde{\mathbf{L}}^+\mathbf{L}\|_\mathbf{F} \leq \epsilon$ using an additional $O(\log N \cdot \log\log(1/\epsilon)$ space.

Since $\mathbf{F}$ is bounded polynomially by $\mathbf{U}_{\widetilde{\mathbf{L}}^+}$, we can adjust $\epsilon$ in Lemma 7.8 by a poly$(N)$ factor to get approximation guarantees in $\|\cdot\|_{\mathbf{U_L}}$ (or spectral norm) in space $O(\log N \cdot \log\log(N/\epsilon))$.

Applying Lemma 6.4 to this approximate pseudoinverse yields an approximate pseudoinverse $\mathbf{P}$ for $\mathbf{I} - \mathbf{W}_0$.

Applying Lemma 7.9, we get that if we redefine the value of $\epsilon$ we have been using to $\epsilon \leftarrow \epsilon/\text{poly}(N)$ from the very beginning, then $\mathbf{P}$ is a $\pm\epsilon$ additive entrywise approximation of $(\mathbf{I}-\mathbf{W}_0)^+$. This only changes the space usage by at most a constant factor. $\qquad\square$

# 8 Estimating Random Walk Probabilities

In this section we show that a space-efficient algorithm for Eulerian Laplacian systems can be used to get strong approximations to $k$-step random walk probabilities on Eulerian digraphs, and more generally, to approximate the product of potentially distinct Eulerian transition matrices. We then give a new result for approximating random walk probabilities on arbitrary digraphs.

**Theorem 8.1.** *Let* $\mathbf{W}_1, \ldots, \mathbf{W}_k$ *be the transition matrices of Eulerian digraphs* $G_1, \ldots, G_k$ *on* $n$ *vertices such that for vertex* $v \in [n]$, *the degree of* $v$ *is the same in* $G_1, \ldots, G_k$. *There is a deterministic algorithm that, when given* $u, v \in [n]$ *and* $\epsilon > 0$, *computes*

$$(\mathbf{W}_1\mathbf{W}_2 \ldots \mathbf{W}_k)_{uv}$$

*to within* $\pm\epsilon$. *The algorithm uses space* $O(\log(N) \cdot \log\log(N/\epsilon))$, *where* $N$ *is the size (bitlength) of the input.*

An important special case of Theorem 8.1 is when $\mathbf{W}_1 = \mathbf{W}_2 = \ldots = \mathbf{W}_k$, in which case the algorithm computes an additive $\epsilon$ approximation to the probability that a random walk of length $k$ in the graph starting at $v$ ends at $u$ and uses space $O(\log(k \cdot N) \cdot \log\log(k \cdot N/\epsilon))$ where $N$ is the bitlength of $\mathbf{W}_1$.

To prove Theorem 8.1, we first define escape probabilities.

**Definition 8.2.** In a graph $G = (V, E)$, for two vertices $u$ and $v$, the *escape probability* $p_w(u, v)$ denotes the probability that a random walk starting at vertex $w$ reaches $u$ before first reaching $v$.

In [CKP+2], they give a useful characterization of escape probabilities on arbitrary digraphs in terms of the pseudoinverse, which we restate below for Eulerian graphs.

**Lemma 8.3** ([CKP+2] (Restated for Eulerian Graphs)). *Let* $\mathbf{W}$ *be a random walk matrix associated with a strongly connected Eulerian graph* $G$. *Let* $s$ *be the stationary distribution of* $G$ *and let* $S$ *be the diagonal matrix with* $s$ *on the diagonal. Let* $u, v$ *be two vertices. Let* $p$ *be the vector of escape probabilities, where* $p_w$ *represents the probability that a random walk starting at* $w$ *reaches* $u$ *before* $v$. *Then*

$$p = \beta(\alpha \cdot s + (\mathbf{I} - \mathbf{W})^+(e_u - e_v))$$

*where*

$$\alpha = -e_v^\top S^{-1}(\mathbf{I} - \mathbf{W})^+(e_u - e_v)$$

*and*

$$\beta = \frac{1}{s_u(e_u - e_v)^\top S^{-1}(\mathbf{I} - \mathbf{W})^+(e_u - e_v)}.$$

*Futhermore,* $\alpha = \text{poly}(N)$ *and* $1/\beta = \text{poly}(N)$.

31

It follows from the above lemma that we can use our algorithm to approximate escape probabilities in Eulerian graphs. Now we can prove Theorem 8.1 by reducing computing certain matrix products to computing escape probabilities.

*Proof of Theorem 8.1.* Note that $(\mathbf{W}_1\mathbf{W}_2\ldots\mathbf{W}_k)_{uv}$ equals the probability that a random walk starting at vertex $v$ ends at $u$, if we take the first step on $G_k$, the second step on $G_{k-1}$ etc. and the final step on $G_1$.

We construct a layered graph with $k+2$ layers and $n$ vertices in layers $1,\ldots,k+1$ each and a single vertex in layer $k+2$. All edges in the graph proceed from one layer to the subsequent one. For each $i \in [k]$, we place edges from layer $i$ to layer $i+1$ according to the graph $G_{k-i+1}$ (so layer 1 to 2 contains the edges from $G_k$ and layer $k$ to $k+1$ contains the edges from $G_1$). All vertices from layer $k+1$ send edges to the sole vertex in layer $k+2$ equal to their in-degree and the vertex in layer $k+2$ sends edges back to layer 1 equal to each vertex's out-degree. This ensures that the resulting graph is Eulerian.

Notice that the probability that a random walk starting at vertex $v$ in layer 1 reaches vertex $u$ in layer $k+1$ before reaching the vertex in layer $k+2$ is exactly $(\mathbf{W}_1\mathbf{W}_2\ldots\mathbf{W}_k)_{uv}$ because all walks of length greater than $k$ pass through the vertex in layer $k+2$ and no walks shorter than $k$ beginning in the first layer pass through it. Furthermore, this is an escape probability on an Eulerian graph and so by Lemma 8.3, this can be computed by applying our Eulerian Laplacian solver to the layered graph followed by some vector multiplications and arithmetic.

Applying the solver to the layered graph, uses space $O(\log(N) \cdot \log\log(N/\epsilon))$ to apply the pseudoinverse in the formula for escape probabilities given in Lemma 8.3 to error $\epsilon/\text{poly}(N)$. Since $\beta$ is at most $\text{poly}(N)$ we can compute $\beta$ times the pseudoinverse term to $\epsilon$ error in this same amount of space. Since the stationary distribution of Eulerian graphs can be computed space efficiently, the remaining operations of computing $\alpha$ times the stationary distribution only add logarithmic space overhead. $\qquad\square$

Using perspectives developed for our space-efficient Eulerian Laplacian solver, we also obtain improved space-efficient algorithms for computing high-quality approximations to random walk probabilities in *arbitrary* digraphs or Markov chains, improving both the best known randomized algorithms and deterministic algorithms.

Aleliunas et. al. [AKL$^+$] observed that random walks can be easily simulated in randomized logarithmic space, and this yields the following algorithm for estimating random walk probabilities:

**Theorem 8.4.** *Suppose that $\mathbf{W} \in \mathbb{R}^{n\times n}$ is a substochastic matrix, $k$ is a positive integer, and $\epsilon > 0$. There is a randomized algorithm that, with high probability, computes a matrix $\widetilde{\mathbf{W}}$ such that $\|\widetilde{\mathbf{W}} - \mathbf{W}^k\| \le \epsilon$. The algorithm runs in space $O(\log(k \cdot N/\epsilon))$, where $N$ is the bitlength of the input.*

*Proof sketch.* For each entry $(\mathbf{W}^k)_{uv}$, we simulate $\text{poly}(n/\epsilon)$ random walks of length $k$ started at $v$ and count the fraction of them that end at $u$. $\qquad\square$

Saks and Zhou gave the best known derandomization of Theorem 8.4 [SZ1].

**Theorem 8.5** ([SZ1]). *Suppose that $\mathbf{W} \in \mathbb{R}^{n\times n}$ is a substochastic matrix, $k$ is a positive integer, and $\epsilon > 0$. There is a deterministic algorithm that computes a matrix $\widetilde{\mathbf{W}}$ such that $\|\widetilde{\mathbf{W}} - \mathbf{W}^k\| \le \epsilon$. The algorithm uses $O(\log(k \cdot n/\epsilon) \cdot \log^{1/2}(k))$ space.*[5]

---

[5]Saks and Zhou state their result in terms of the $L_\infty$ matrix norm rather than the spectral norm. These are equivalent up to polynomial changes in $\epsilon$ (and hence constant factors in the space complexity).

We strengthen Theorems 8.4 and 8.5 when $\epsilon = 1/(k \cdot N)^{\omega(1)}$. This range of parameters is useful because $k$-step walk probabilities can be exponentially small in $k$, even on undirected graphs.

**Theorem 8.6.** *Suppose that $\mathbf{W} \in \mathbb{R}^{n \times n}$ is a substochastic matrix, $k$ is a positive integer, and $\epsilon > 0$. Then:*

1. *There is a randomized algorithm that computes a matrix $\widetilde{\mathbf{W}}$ such that each entry of $\widetilde{\mathbf{W}}$ is with $\pm\epsilon$ of that of $\mathbf{W}^k$ additively and runs in space $O(\log(k \cdot N) \cdot \log(\log_{k \cdot N}(1/\epsilon)))$*

2. *There is a deterministic algorithm that computes a matrix $\widetilde{\mathbf{W}}$ such that each entry of $\widetilde{\mathbf{W}}$ is with $\pm\epsilon$ of that of $\mathbf{W}^k$ additively and runs in space $O(\log(k \cdot N) \cdot \log^{1/2}(k) + \log(k \cdot n) \cdot \log(\log_{k \cdot N}(1/\epsilon)))$*

*Proof.* We obtain this result by applying preconditioned Richardson iteration (which is the same technique that gets the doubly logarithmic dependence on $\epsilon$ in Theorem 8.1) to a path-lifted (rather than cycle-lifted) graph. Our plan is to use Theorem 8.4 or Theorem 8.5 to compute an approximation to the inverse of a path-lifted graph with quality $1/\text{poly}(k \cdot n)$. Then we use preconditioned Richardson iteration to boost the error to $\epsilon$ using only additive $O(\log(k \cdot n) \cdot \log(\log_{k \cdot N}(1/\epsilon)))$ space as in Lemma 7.8.

Let $\mathbf{P}_k$ denote the adjacency matrix of the directed path of length $k$ ($\mathbf{P}$ is $(k+1) \times (k+1)$ with all ones just below the diagonal and zeroes elsewhere). Let $\mathbf{M} = \mathbf{P}_k \otimes \mathbf{W}$. Note that $\mathbf{M}^{k+1}$ is the all zeroes matrix and all eigenvalues of $\mathbf{M}$ are 0. Let $\mathbf{I} = \mathbf{I}_{(k+1) \cdot n}$. We will approximate $\mathbf{I} - \mathbf{M}$ using the following identity: for all $\mathbf{A}$ with eigenvalues less than 1 in magnitude, we have

$$(\mathbf{I} - \mathbf{A})^{-1} = \sum_{i=0}^{\infty} \mathbf{A}^i.$$

This means that

$$(\mathbf{I} - \mathbf{M})^{-1} = \sum_{i=0}^{k} \mathbf{M}^i$$
$$= \sum_{i=0}^{k} (\mathbf{P}_k)^i \otimes \mathbf{W}^i$$

because all powers beyond the $k$th are zero. Note that this matrix has $\mathbf{W}^k$ in the lower left block, $\mathbf{W}^{k-1}$ in the second diagonal from the bottom, etc.

From Theorem 8.4, there is a randomized algorithm that can compute matrices $\mathbf{W}_1, \ldots, \mathbf{W}_k$ such that for each $i \in [k]$, we have $\|\mathbf{W}^i - \mathbf{W}_i\| \leq 1/(2 \cdot k \cdot N)$ in space $O(\log(k \cdot N))$. Likewise, by Theorem 8.5, there is a deterministic algorithm that can compute matrices $\mathbf{W}_1, \ldots, \mathbf{W}_k$ such that for each $i \in [k]$, we have $\|\mathbf{W}^i - \mathbf{W}_i\| \leq 1/(2 \cdot k \cdot N)$ in space $O(\log(k \cdot N) \cdot \log^{1/2}(k))$. We can then approximate $(\mathbf{I} - \mathbf{M})^{-1}$ as

$$\mathbf{N} = \mathbf{I} + \sum_{i=1}^{k} (\mathbf{P}_k)^i \otimes \mathbf{W}_i.$$

Note that for each $i \in [k]$ we have $\|\mathbf{P}_k^i\| = 1$. Also, we have

$$\|\mathbf{I} - \mathbf{M}\| = \|\mathbf{I} - \mathbf{P}_k \otimes \mathbf{W}\|$$
$$\leq \|\mathbf{I}\| + \|\mathbf{P}_k\| \cdot \|\mathbf{W}\|$$
$$\leq 2$$

33

because $\|\mathbf{W}\| \leq 1$. Putting this together gives

$$\|\mathbf{I} - \mathbf{N}(\mathbf{I} - \mathbf{M})\| = \|((\mathbf{I} - \mathbf{M})^{-1} - \mathbf{N})(\mathbf{I} - \mathbf{M})\|$$
$$= \left\|\sum_{i=1}^{k}(\mathbf{P}_k)^i \otimes (\mathbf{W}^i - \mathbf{W}_i)(\mathbf{I} - \mathbf{M})\right\|$$
$$\leq \sum_{i=1}^{k} \|\mathbf{P}_k^i\| \cdot \|\mathbf{W}^i - \mathbf{W}_i\| \cdot \|\mathbf{I} - \mathbf{M}\|$$
$$\leq 2 \cdot \sum_{i=1}^{k} 1/(2 \cdot k \cdot N)$$
$$= 1/(k \cdot N)$$

From Lemma 7.8, we can compute a matrix $\widetilde{\mathbf{N}}$ such that $\|\mathbf{I} - \widetilde{\mathbf{N}}(\mathbf{I} - \mathbf{M})\| \leq \epsilon/\|(\mathbf{I} - \mathbf{M})^+\|$ using an additional space

$$O(\log(k \cdot N) \cdot \log(\log_{k \cdot N}(\|(\mathbf{I} - \mathbf{M})^+\|/\epsilon))) = O(\log(k \cdot N) \cdot \log(\log_{k \cdot N}(N/\epsilon)))$$
$$= O(\log(k \cdot N) \cdot \log(\log_{k \cdot N}(1/\epsilon)))$$

for $\epsilon$ sufficiently small.

This implies $\|\widetilde{\mathbf{N}} - (\mathbf{I} - \mathbf{M})^{-1}\| \leq \epsilon$. Since the norm of a matrix is an upper bound on the norm of its submatrices, we have that the lower left $n \times n$ block of $\widetilde{\mathbf{N}}$, is a matrix $\widetilde{\mathbf{W}}$ satisfying $\|\widetilde{\mathbf{W}} - \mathbf{W}^k\| \leq \epsilon$.

To complete the proof, Lemma 7.9 lets us convert this guarantee to entrywise approximation with a blowup of $\text{poly}(N)$ in $\epsilon$. If we reduce $\epsilon$ to compensate for this, it only affects the stated space complexity by a constant factor. $\square$

# 9 Acknowledgements

# References

[AJSS]     AmirMahdi Ahmadinejad, Arun Jambulapati, Amin Saberi, and Aaron Sidford. Perron-frobenius theory in nearly linear time: Positive eigenvectors, m-matrices, graph kernels, and other applications. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2019, San Diego, California, USA, January 6-9, 2019*, pages 1387–1404, 2019.

[AKL+]     Romas Aleliunas, Richard M. Karp, Richard J. Lipton, László Lovász, and Charles Rackoff. Random walks, universal traversal sequences, and the complexity of maze problems. In *20th Annual Symposium on Foundations of Computer Science (San Juan, Puerto Rico, 1979)*, pages 218–223. IEEE, New York, 1979.

[BCP]      A. Borodin, S. Cook, and N. Pippenger. Parallel computation for well-endowed rings and space-bounded probabilistic machines. *Information and Control*, 58(1-3):113–136, 1983.

[BRRY]     Mark Braverman, Anup Rao, Ran Raz, and Amir Yehudayoff. Pseudorandom generators for regular branching programs. In *FOCS*, pages 40–47. IEEE Computer Society, 2010.

[BSST]     Joshua Batson, Daniel A Spielman, Nikhil Srivastava, and Shang-Hua Teng. Spectral sparsification of graphs: theory and algorithms. *Communications of the ACM*, 56(8):87–94, 2013.

[CCL+]     Dehua Cheng, Yu Cheng, Yan Liu, Richard Peng, and Shang-Hua Teng. Spectral sparsification of random-walk matrix polynomials. *arXiv preprint arXiv:1502.03496*, 2015.

[CKK+]     Michael B Cohen, Jonathan Kelner, Rasmus Kyng, John Peebles, Richard Peng, Anup B Rao, and Aaron Sidford. Solving directed laplacian systems in nearly-linear time through sparse lu factorizations. In *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 898–909. IEEE, 2018.

[CKM+]     Michael B. Cohen, Rasmus Kyng, Gary L. Miller, Jakub W. Pachocki, Richard Peng, Anup B. Rao, and Shen Chen Xu. Solving SDD linear systems in nearly $m\log^{1/2}n$ time. In *Symposium on Theory of Computing, STOC 2014, New York, NY, USA, May 31 - June 03, 2014*, pages 343–352, 2014.

[CKP+1]    Michael B Cohen, Jonathan Kelner, John Peebles, Richard Peng, Anup B Rao, Aaron Sidford, and Adrian Vladu. Almost-linear-time algorithms for markov chains and new spectral primitives for directed graphs. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 410–419. ACM, 2017.

[CKP+2]    Michael B Cohen, Jonathan Kelner, John Peebles, Richard Peng, Aaron Sidford, and Adrian Vladu. Faster algorithms for computing the stationary distribution, simulating random walks, and more. In *Foundations of Computer Science (FOCS), 2016 IEEE 57th Annual Symposium on*, pages 583–592. IEEE, 2016.

[De]       Anindya De. Pseudorandomness for permutation and regular branching programs. In *IEEE Conference on Computational Complexity*, pages 221–231. IEEE Computer Society, 2011.

[Gil]      John Gill. Computational complexity of probabilistic Turing machines. *SIAM Journal on Computing*, 6(4):675–695, 1977.

[GMR+] Parikshit Gopalan, Raghu Meka, Omer Reingold, Luca Trevisan, and Salil Vadhan. Better pseudorandom generators via milder pseudorandom restrictions. In *Proceedings of the 53rd Annual IEEE Symposium on Foundations of Computer Science (FOCS '12)*. IEEE, 20–23 October 2012.

[HZ] William Hoza and David Zuckerman. Simple optimal hitting sets for small-success RL. In *59th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2018, Paris, France, October 7-9, 2018*, pages 59–64, 2018.

[Jun] H. Jung. Relationships between probabilistic and deterministic tape complexity. In *Mathematical foundations of computer science, 1981 (Štrbské Pleso, 1981)*, volume 118 of *Lecture Notes in Comput. Sci.*, pages 339–346. Springer, Berlin-New York, 1981.

[KLP+] Rasmus Kyng, Yin Tat Lee, Richard Peng, Sushant Sachdeva, and Daniel A Spielman. Sparsified cholesky and multigrid solvers for connection laplacians. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pages 842–850. ACM, 2016.

[KMP1] Ioannis Koutis, Gary L Miller, and Richard Peng. Approaching optimality for solving sdd systems. march 2010.

[KMP2] Ioannis Koutis, Gary L Miller, and Richard Peng. A nearly-m log n time solver for sdd linear systems. In *Foundations of Computer Science (FOCS), 2011 IEEE 52nd Annual Symposium on*, pages 590–598. IEEE, 2011.

[KNP] Michal Koucký, Prajakta Nimbhorkar, and Pavel Pudlák. Pseudorandom generators for group products: extended abstract. In Lance Fortnow and Salil P. Vadhan, editors, *STOC*, pages 263–272. ACM, 2011.

[KOSZ] Jonathan A Kelner, Lorenzo Orecchia, Aaron Sidford, and Zeyuan Allen Zhu. A simple, combinatorial algorithm for solving sdd systems in nearly-linear time. In *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*, pages 911–920. ACM, 2013.

[KS] Rasmus Kyng and Sushant Sachdeva. Approximate gaussian elimination for laplacians - fast, sparse, and simple. In *IEEE 57th Annual Symposium on Foundations of Computer Science, FOCS 2016, 9-11 October 2016, Hyatt Regency, New Brunswick, New Jersey, USA*, pages 573–582, 2016.

[KvM] Adam R. Klivans and Dieter van Melkebeek. Graph nonisomorphism has subexponential size proofs unless the polynomial-time hierarchy collapses. *SIAM Journal on Computing*, 31(5):1501–1526 (electronic), 2002.

[LS] Yin Tat Lee and Aaron Sidford. Efficient accelerated coordinate descent methods and faster algorithms for solving linear systems. In *Foundations of Computer Science (FOCS), 2013 IEEE 54th Annual Symposium on*, pages 147–156. IEEE, 2013.

[Mih] Milena Mihail. Conductance and convergence of markov chains-a combinatorial treatment of expanders. In *30th Annual Symposium on Foundations of Computer Science (Research Triangle Park, North Carolina)*, pages 526–531. IEEE, 1989.

[MP]      Gary L Miller and Richard Peng. Approximate maximum flow on separable undirected graphs. In *Proceedings of the twenty-fourth annual ACM-SIAM symposium on Discrete algorithms*, pages 1151–1170. Society for Industrial and Applied Mathematics, 2013.

[MPV]     Jack Murtagh, Edward Pyne, and Salil Vadhan. Pseudorandom generators for permutation branching programs of polynomial width. Unpublished manuscript, April 2020.

[MRSV1]   Jack Murtagh, Omer Reingold, Aaron Sidford, and Salil P. Vadhan. Derandomization beyond connectivity: Undirected laplacian systems in nearly logarithmic space. In *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017, Berkeley, CA, USA, October 15-17, 2017*, pages 801–812, 2017.

[MRSV2]   Jack Murtagh, Omer Reingold, Aaron Sidford, and Salil P. Vadhan. Deterministic approximation of random walks in small space. *CoRR*, abs/1903.06361, 2019.

[MRT]     Raghu Meka, Omer Reingold, and Avishay Tal. Pseudorandom generators for width-3 branching programs. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 626–637. ACM, 2019.

[Nis]     Noam Nisan. Pseudorandom bits for constant depth circuits. *Combinatorica*, 11(1):63–70, 1991.

[PS]      Richard Peng and Daniel A. Spielman. An efficient parallel solver for SDD linear systems. *STOC*, 2014.

[Rei]     Omer Reingold. Undirected connectivity in log-space. *Journal of the ACM*, 55(4):Art. 17, 24, 2008.

[RTV]     Omer Reingold, Luca Trevisan, and Salil Vadhan. Pseudorandom walks in regular digraphs and the RL vs. L problem. In *Proceedings of the 38th Annual ACM Symposium on Theory of Computing (STOC '06)*, pages 457–466, 21–23 May 2006. Preliminary version as *ECCC* TR05-22, February 2005.

[RV]      Eyal Rozenman and Salil Vadhan. Derandomized squaring of graphs. In *Proceedings of the 8th International Workshop on Randomization and Computation (RANDOM '05)*, number 3624 in Lecture Notes in Computer Science, pages 436–447, Berkeley, CA, August 2005. Springer.

[RVW]     Omer Reingold, Salil Vadhan, and Avi Wigderson. Entropy waves, the zig-zag graph product, and new constant-degree expanders. *Annals of Mathematics*, 155(1), January 2001.

[Sak]     Michael E. Saks. Randomization and derandomization in space bounded computation. In *IEEE Conference on Computational Complexity*, pages 128–149, 1996.

[Sim]     Janos Simon. On tape-bounded probabilistic Turing machine acceptors. *Theoretical Computer Science*, 16(1):75–91, 1981.

[ST1]     Daniel A Spielman and Shang-Hua Teng. Nearly-linear time algorithms for graph partitioning, graph sparsification, and solving linear systems. In *Proceedings of the thirty-sixth annual ACM symposium on Theory of computing*, pages 81–90. ACM, 2004.

[ST2]     Daniel A Spielman and Shang-Hua Teng. Spectral sparsification of graphs. *SIAM Journal on Computing*, 40(4):981–1025, 2011.

[Ste]     Thomas Steinke. Pseudorandomness for permutation branching programs without the group theory. Technical Report TR12-083, Electronic Colloquium on Computational Complexity (ECCC), July 2012.

[SZ1]     Michael Saks and Shiyu Zhou. $BP_HSPACE(S) \subseteq DSPACE(S^{3/2})$. *Journal of Computer and System Sciences*, 58(2):376–403, 1999.

[SZ2]     Jirí Síma and Stanislav Zák. Almost $k$-wise independent sets establish hitting sets for width-3 1-branching programs. In Alexander S. Kulikov and Nikolay K. Vereshchagin, editors, *CSR*, volume 6651 of *Lecture Notes in Computer Science*, pages 120–133. Springer, 2011.

[Vad]     Salil P Vadhan. Pseudorandomness. *Foundations and Trends® in Theoretical Computer Science*, 7(1–3):1–336, 2012.

# A    Approximation Equivalences Proofs

Here we prove some lemmas from Section 3. We begin with some linear algebraic technical lemmas.

**Lemma A.1** ([CKP+1] Lemma B.2). *For all $\mathbf{A} \in \mathbb{R}^{n \times n}$ and symmetric PSD $\mathbf{M}, \mathbf{N} \in \mathbb{R}^{n \times n}$, the following are equivalent:*

1. *For all $x, y \in \mathbb{R}^n$*

$$|x^\top \mathbf{A} y| \leq \frac{\epsilon}{2} \cdot \left( x^\top \mathbf{M} x + y^\top \mathbf{N} y \right)$$

2. *For all $x, y \in \mathbb{C}^n$*

$$|x^\top \mathbf{A} y| \leq \epsilon \cdot \sqrt{(x^\top \mathbf{M} x) \cdot (y^\top \mathbf{N} y)}$$

3. *$\|\mathbf{M}^{+/2} \mathbf{A} \mathbf{N}^{+/2}\|_{\mathbb{R}^n \to \mathbb{R}^n} \leq \epsilon$ and $\ker(\mathbf{M}) \subseteq \ker(\mathbf{A}^\top)$ and $\ker(\mathbf{N}) \subseteq \ker(\mathbf{A})$*

*where the notation $\| \cdot \|_{\mathbb{R}^n \to \mathbb{R}^n}$ denotes the 2-norm on a linear operator that maps $\mathbb{R}^n$ to $\mathbb{R}^n$.*

Lemma A.1 was used in [CKP+1] to give equivalent formulations of directed spectral approximation. Below, we extend the lemma to complex space, which allows us to give even more equivalent formulations of directed spectral approximation as well as unit-circle spectral approximation.

**Lemma A.2** (Extension of [CKP+1] Lemma B.2). *For all $\mathbf{A} \in \mathbb{C}^{n \times n}$ and Hermitian PSD $\mathbf{M}, \mathbf{N} \in \mathbb{C}^{n \times n}$, the following are equivalent:*

1. *For all $x, y \in \mathbb{C}^n$*

$$|x^* \mathbf{A} y| \leq \frac{\epsilon}{2} \cdot (x^* \mathbf{M} x + y^* \mathbf{N} y)$$

2. *For all $x, y \in \mathbb{C}^n$*

$$|x^* \mathbf{A} y| \leq \epsilon \cdot \sqrt{(x^* \mathbf{M} x) \cdot (y^* \mathbf{N} y)}$$

3. *$\|\mathbf{M}^{+/2} \mathbf{A} \mathbf{N}^{+/2}\|_{\mathbb{C}^n \to \mathbb{C}^n} \leq \epsilon$ and $\ker(\mathbf{M}) \subseteq \ker(\mathbf{A}^*)$ and $\ker(\mathbf{N}) \subseteq \ker(\mathbf{A})$*

*where the notation $\| \cdot \|_{\mathbb{C}^n \to \mathbb{C}^n}$ denotes the 2-norm on a linear operator that maps $\mathbb{C}^n$ to $\mathbb{C}^n$.*

*Proof.* Let $\mathbf{L} = \|\mathbf{M}^{+/2} \mathbf{A} \mathbf{N}^{+/2}\|_2$. Since $\|x\|_2 = \max_{\|y\|_2 = 1} |y^* x|$ we have that

$$\mathbf{L} = \max_{\|x\| = \|y\| = 1} |x^* \mathbf{M}^{+/2} \mathbf{A} \mathbf{N}^{+/2} y|.$$

Performing the mapping $x \colon \mathbf{M}^{+/2} x$ and $y \colon \mathbf{M}^{+/2} y$ we have

$$\mathbf{L} = \max_{x^* \mathbf{M} x = y^* \mathbf{N} y = 1} |x^* \mathbf{A} y| = \max_{x \notin \ker(\mathbf{M}), y \notin \ker(\mathbf{N})} \frac{|x^* \mathbf{A} y|}{\sqrt{x^* \mathbf{M} x \cdot y^* \mathbf{N} y}}.$$

It follows that if $\mathbf{L} \leq \epsilon$ and $\ker(\mathbf{M}) \subseteq \ker(\mathbf{A}^*)$ and $\ker(\mathbf{N}) \subseteq \ker(\mathbf{A})$ then

$$|x^* \mathbf{A} y| \leq \epsilon \cdot \sqrt{(x^* \mathbf{M} x) \cdot (y^* \mathbf{N} y)}.$$

The kernel constraints guarantee that whenever the right-hand side is zero, the left-hand side must also be zero. Finally, by the AM-GM inequality we get $\sqrt{x^* \mathbf{M} x \cdot y^* \mathbf{N} y} \leq (x^* \mathbf{M} x + y^* \mathbf{N} y)/2$ and this inequality is tight when $x^* \mathbf{M} x = y^* \mathbf{N} y = 1$, so for all $x, y \in \mathbb{C}^n$ we have

$$|x^* \mathbf{A} y| \leq \frac{\epsilon}{2} \cdot (x^* \mathbf{M} x + y^* \mathbf{N} y)$$

$\square$

If follows from Lemma 2.2 that in fact all six formulations in Lemmas A.1 and A.2 are equivalent because the spectral norm of a matrix is the same whether one defines it in terms of real or complex vectors.

Now we can prove the equivalences from Section 3.

**Lemma 3.5 restated.** *Let* $\mathbf{W}, \widetilde{\mathbf{W}} \in \mathbb{R}^{n \times n}$ *be (possibly asymmetric) matrices. Then* $\widetilde{\mathbf{W}}$ *is a directed $\epsilon$-approximation of* $\mathbf{W}$ *if and only if* $\widetilde{\mathbf{W}}$ *is a complex $\epsilon$-approximation of* $\mathbf{W}$.

*Proof.* Set $\mathbf{A} = (\mathbf{W} - \widetilde{\mathbf{W}})$ and $\mathbf{M} = \mathbf{N} = \mathbf{U_{I-W}}$. Note that $\mathbf{U_{I-W}}$ is PSD. Observe that for all $v \in \mathbb{C}^n$

$$v^* \mathbf{U_{I-W}} v = \|v\|^2 - v^* \mathbf{U_W} v.$$

Lemmas A.1,A.2, and 2.2 tell us that the following are equivalent:

1. For all $x, y \in \mathbb{R}^n$

$$|x^\top \mathbf{A} y| \leq \frac{\epsilon}{2} \cdot \left( x^\top \mathbf{M} x + y^\top \mathbf{N} y \right)$$

2. For all $x, y \in \mathbb{C}^n$

$$|x^* \mathbf{A} y| \leq \frac{\epsilon}{2} \cdot (x^* \mathbf{M} x + y^* \mathbf{N} y)$$

Plugging in our settings for $\mathbf{A}$, $\mathbf{M}$, and $\mathbf{N}$ completes the proof. $\qquad\square$

**Lemma 3.7 restated.** *Let* $\mathbf{W}, \widetilde{\mathbf{W}} \in \mathbb{R}^{n \times n}$ *be symmetric matrices. Then the following are equivalent:*

1. $\widetilde{\mathbf{W}} \overset{\circ}{\approx}_\epsilon \mathbf{W}$

2. $\widetilde{\mathbf{W}} \approx_\epsilon \mathbf{W}$ *and* $-\widetilde{\mathbf{W}} \approx_\epsilon -\mathbf{W}$

3. *For all $x \in \mathbb{R}^n$ we have*

$$\left| x^\top (\mathbf{W} - \widetilde{\mathbf{W}}) x \right| \leq \epsilon \cdot \left( \|x\|^2 - |x^\top W x| \right).$$

*Proof.* First we show that items 2 and 3 are equivalent. From Definition 3.1, we have $\widetilde{\mathbf{W}} \approx_\epsilon \mathbf{W}$ and $-\widetilde{\mathbf{W}} \approx_\epsilon -\mathbf{W}$ if and only if for all $x \in \mathbb{R}^n$ we have

$$\left| x^\top (\mathbf{W} - \widetilde{\mathbf{W}}) x \right| \leq \epsilon \cdot \left( \|x\|^2 - x^\top W x \right)$$

and for all $x \in \mathbb{R}^n$ we have

$$\left| x^\top (\mathbf{W} - \widetilde{\mathbf{W}}) x \right| \leq \epsilon \cdot \left( \|x\|^2 + x^\top W x \right).$$

This is equivalent to: for all $x \in \mathbb{R}^n$

$$\left| x^\top (\mathbf{W} - \widetilde{\mathbf{W}}) x \right| \leq \min \left\{ \epsilon \cdot \left( \|x\|^2 - x^\top W x \right), \epsilon \cdot \left( \|x\|^2 + x^\top W x \right) \right\}$$
$$= \epsilon \cdot \left( \|x\|^2 - |x^\top W x| \right),$$

which establishes the equivalence between items 2 and 3. Now we show that items 1 and 2 are equivalent, which will complete the proof. From Lemma 3.5, we have that item 2 is equivalent to: for all $x, y \in \mathbb{C}^n$

$$\left| x^* (\mathbf{W} - \widetilde{\mathbf{W}}) y \right| \leq \frac{\epsilon}{2} \cdot \left( \|x\|^2 + \|y\|^2 - x^* \mathbf{U_W} x - y^* \mathbf{U_W} y \right)$$

and for all $x, y \in \mathbb{C}^n$

$$\left| x^*(\mathbf{W} - \widetilde{\mathbf{W}})y \right| \leq \frac{\epsilon}{2} \cdot \left( \|x\|^2 + \|y\|^2 + x^*\mathbf{U_W}x + y^*\mathbf{U_W}y \right).$$

Since $\mathbf{W}$ is real and symmetric we have $\mathbf{U_W} = \mathbf{W}$ and we have that for all $v \in \mathbb{C}^n$, $v^*\mathbf{W}v$ is real. Therefore the above is equivalent to: for all $x, y \in \mathbb{C}^n$

$$\left| x^*(\mathbf{W} - \widetilde{\mathbf{W}})y \right| \leq \min \left\{ \frac{\epsilon}{2} \cdot \left( \|x\|^2 + \|y\|^2 - x^*\mathbf{W}x - y^*\mathbf{W}y \right), \frac{\epsilon}{2} \cdot \left( \|x\|^2 + \|y\|^2 + x^*\mathbf{W}x + y^*\mathbf{W}y \right) \right\}$$
$$= \frac{\epsilon}{2} \cdot \left( \|x\|^2 + \|y\|^2 - |x^*\mathbf{W}x + y^*\mathbf{W}y| \right),$$

which is the requirement for unit-circle spectral approximation. $\qquad \square$

**Lemma 3.8 restated.** *Let* $\mathbf{W}, \widetilde{\mathbf{W}} \in \mathbb{C}^{n \times n}$ *be (possibly asymmetric) matrices. Then the following are equivalent*

1. $\widetilde{\mathbf{W}} \overset{\circ}{\approx}_\epsilon \mathbf{W}$

2. *For all* $z \in \mathbb{C}$ *such that* $|z| = 1$, $z \cdot \widetilde{\mathbf{W}} \approx_\epsilon z \cdot \mathbf{W}$

3. *For all* $z \in \mathbb{C}$ *such that* $|z| = 1$,

   - $\ker(\mathbf{U}_{\mathbf{I}-z\cdot\mathbf{W}}) \subseteq \ker(\widetilde{\mathbf{W}} - \mathbf{W}) \cap \ker((\widetilde{\mathbf{W}} - \mathbf{W})^\top)$ *and*
   - $\left\| \mathbf{U}_{\mathbf{I}-z\cdot\mathbf{W}}^{+/2}(\widetilde{\mathbf{W}} - \mathbf{W})\mathbf{U}_{\mathbf{I}-z\cdot\mathbf{W}}^{+/2} \right\| \leq \epsilon$

*Proof.* Using the definition of directed spectral approximation (Definition 3.2) and its extenstion to complex space from Lemma 3.5 lets us write item 2 as: for all $x, y \in \mathbb{C}^n$

$$\left| x^*(z \cdot \mathbf{W} - z \cdot \widetilde{\mathbf{W}})y \right| \leq \frac{\epsilon}{2} \cdot \left( \|x\|^2 + \|y\|^2 - x^*\mathbf{U}_{z\cdot\mathbf{W}}x - y^*\mathbf{U}_{z\cdot\mathbf{W}}y \right) \tag{15}$$
$$= \frac{\epsilon}{2} \cdot \left( \|x\|^2 + \|y\|^2 - \text{Re}(x^*(z \cdot \mathbf{W})x + y^*(z \cdot \mathbf{W})y) \right). \tag{16}$$

Lemma A.2 immediately implies the equivalence between line 15 and item 3 in the lemma statement by setting $\mathbf{A} = (\widetilde{\mathbf{W}} - \mathbf{W})$ and $\mathbf{M} = \mathbf{N} = \mathbf{U}_{\mathbf{I}-z\cdot\mathbf{W}}$.

Now we will show the equivalence between items 1 and 2. Since $|z| = 1$, it does not affect the left-hand side of Inequality 15 and we can drop it.

Item 2 in the lemma statement says that line 16 holds for all $z \in \mathbb{C}$ such that $|z| = 1$. This is equivalent to it holding for the worst case $z$ (for each pair of vectors, $x, y$). For every complex number $\omega$, we have $\max_{z: |z|=1} \text{Re}(z\omega) = |\omega|$ by setting $\text{Re}(z) = \text{Re}(\omega)/|\omega|$ and $\text{Im}(z) = -\text{Im}(\omega)/|\omega|$. Therefore the inequality becomes

$$\left| x^*(\mathbf{W} - \widetilde{\mathbf{W}})y \right| \leq \frac{\epsilon}{2} \cdot \left( \|x\|^2 + \|y\|^2 - |x^*\mathbf{W}x - y^*\mathbf{W}y| \right),$$

which is exactly the condition we require for $\widetilde{\mathbf{W}} \overset{\circ}{\approx}_\epsilon \mathbf{W}$. $\qquad \square$

# B  Omitted Proofs from Section 4

**Lemma 4.2 restated.** *Fix $\widetilde{\mathbf{W}}, \mathbf{W} \in \mathbb{C}^{n \times n}$. For any matrix $M \in \mathbb{C}^{n \times n}$, let $V_\lambda(M)$ denote the eigenspace of $M$ of eigenvalue $\lambda$.*

1. *If $\widetilde{\mathbf{W}} \approx_c \mathbf{W}$ for a finite $c > 0$, then $V_1(\mathbf{W}) \subseteq V_1(\widetilde{\mathbf{W}})$. If $c < 1$, then $V_1(\mathbf{W}) = V_1(\widetilde{\mathbf{W}})$.*

2. *If $\widetilde{\mathbf{W}} \overset{\circ}{\approx}_c \mathbf{W}$ for a finite $c > 0$, then for all $\lambda \in \mathbb{C}$ such that $|\lambda| = 1$, $V_\lambda(\mathbf{W}) \subseteq V_\lambda(\widetilde{\mathbf{W}})$ and if $c < 1$ then $V_\lambda(\mathbf{W}) = V_\lambda(\widetilde{\mathbf{W}})$.*

*Proof.* Suppose $\widetilde{\mathbf{W}} \approx_c \mathbf{W}$ for a finite $c$. Applying the equivalence of Lemma A.2, this means that for all $x, y \in \mathbb{C}^n$,

$$|x^*(\mathbf{W} - \widetilde{\mathbf{W}})y| \leq c \cdot \sqrt{(x^*(\mathbf{I} - \mathbf{W})x) \cdot (y^*(\mathbf{I} - \mathbf{W})y)}.$$

Let $v$ be an eigenvector of $\mathbf{W}$ with eigenvalue 1. Then setting $y = v$ above makes the right-hand side equal 0. Therefore we have that for all $x \in \mathbb{C}^n$,

$$|x^*(\mathbf{W} - \widetilde{\mathbf{W}})v| = |x^*(\mathbf{I} - \widetilde{\mathbf{W}})v| = 0$$

which can only occur if $(\mathbf{I} - \widetilde{\mathbf{W}})v = \mathbf{0}$. Suppose now that $c < 1$ and let $v$ be an eigenvector of $\widetilde{\mathbf{W}}$ with eigenvalue 1. Then we have

$$|v^*(\mathbf{W} - \widetilde{\mathbf{W}})v| = v^*(\mathbf{I} - \mathbf{W})v \leq \frac{c}{2} \cdot (v^*(\mathbf{I} - \mathbf{W})v + v^*(\mathbf{I} - \mathbf{W})v) = c \cdot v^*(\mathbf{I} - \mathbf{W})v.$$

Since $c < 1$ and $\mathbf{U}_{\mathbf{I} - \mathbf{W}}$ is PSD, the above inequality can only be true when both sides equal zero. Applying the equivalence of Lemma A.2, we have that for all $x \in \mathbb{C}^n$

$$|x^*(\mathbf{W} - \widetilde{\mathbf{W}})v| = |x^*(\mathbf{I} - \mathbf{W})v| \leq c \cdot \sqrt{(x^*(\mathbf{I} - \mathbf{W})x) \cdot (v^*(\mathbf{I} - \mathbf{W})v)} = 0.$$

The above can only happen for all vectors $x$ if $(\mathbf{I} - \mathbf{W})v = \mathbf{0}$.

For the second claim, suppose $\widetilde{\mathbf{W}} \overset{\circ}{\approx}_c \mathbf{W}$ for finite $c$. Then for all $z$ such that $|z| = 1$ we have $z \cdot \widetilde{\mathbf{W}} \approx_c z \cdot \mathbf{W}$. Let $v$ be an eigenvector of $\mathbf{W}$ with eigenvalue $\lambda$ such that $|\lambda| = 1$. Set $z = \lambda^{-1}$ so $z \cdot \mathbf{W}v = v$. From the first part of the lemma, we have $V_1(z \cdot \mathbf{W}) \subseteq V_1(z \cdot \widetilde{\mathbf{W}})$ and hence $z \cdot \widetilde{\mathbf{W}}v = v$, which implies $\widetilde{\mathbf{W}}v = \lambda \cdot v$. A similar argument applies when $c < 1$. $\qquad \square$

**Lemma 4.7 restated.** *Let $\mathbf{M} \colon \mathbb{C}^n \to \mathbb{C}^n$ be a linear operator and $V_1, \ldots, V_\ell \subseteq \mathbb{C}^n$ subspaces such that*

1. *$V_j \perp V_k$ for all $j \neq k$*

2. *$V_1 \oplus \ldots \oplus V_\ell = \mathbb{C}^n$*

3. *$\mathbf{M}V_j \subseteq V_j$ for all $j \in [\ell]$.*

*I.E., M is block diagonal with respect to the subspaces $V_1, \ldots, V_\ell$. Then,*

$$\|\mathbf{M}\| = \max_{j \in [\ell]} \|\mathbf{M}|_{V_j}\|$$

*Proof.* For $v \in \mathbb{C}^n$, by property 2 we can write $v = v_1 + \ldots + v_\ell$ where $v_j \in V_j$ for all $j \in [\ell]$. Using this as well as properties 1 and 3, we get

$$\|\mathbf{M}v\|^2 = \left\| \sum_{j \in [\ell]} \mathbf{M}v_j \right\|^2$$

$$= \sum_{j \in [\ell]} \|\mathbf{M}v_j\|^2$$

$$\leq \sum_{j \in [\ell]} \|\mathbf{M}|_{V_j}\|^2 \cdot \|v_j\|^2$$

$$\leq \left( \max_{j \in [\ell]} \|\mathbf{M}|_{V_j}\|^2 \right) \cdot \|v\|^2.$$

So $\|\mathbf{M}\| \leq \max_{j \in [\ell]}\|\mathbf{M}|_{V_j}\|$. The other direction, $\|\mathbf{M}\| \geq \|\mathbf{M}|_{V_j}\|$ for all $j \in [\ell]$ is immediate. $\square$

**Theorem 4.8 restated.** *Fix $\mathbf{W}, \widetilde{\mathbf{W}} \in \mathbb{C}^n$ and suppose that $\widetilde{\mathbf{W}} \approx_\epsilon \mathbf{W}$ for $\epsilon \in (0, 2/3)$. Let $F \subseteq [n]$ such that $(\mathbf{I}_{|F|} - \mathbf{W}_{FF})$ is invertible and let $C = [n] \setminus F$. Then*

$$\mathbf{I}_{|C|} - \mathrm{Sc}(\mathbf{I}_n - \widetilde{\mathbf{W}}, C) \approx_{\epsilon/(1-3\epsilon/2)} \mathbf{I}_{|C|} - \mathrm{Sc}(\mathbf{I}_n - \mathbf{W}, C)$$

*Proof.* Let $\mathbf{A} = \mathbf{I}_n - \widetilde{\mathbf{W}}$, $\mathbf{B} = \mathbf{I}_n - \mathbf{W}$. Note that since $\widetilde{\mathbf{W}} \approx_\epsilon \mathbf{W}$, we have for all $x, y \in \mathbb{C}^n$

$$|x^*(\mathbf{B} - \mathbf{A})y| \leq \frac{\epsilon}{2} \left( x^* \mathbf{U_B} x + y^* \mathbf{U_B} y \right), \tag{17}$$

our goal is to show that $\mathbf{I}_{|C|} - \mathrm{Sc}(\mathbf{A}, C) \approx_{\epsilon/(1-3\epsilon/2)} \mathbf{I}_{|C|} - \mathrm{Sc}(\mathbf{B}, C)$, which is equivalent to

$$\left| [x^\ell]^* \left( \mathrm{Sc}(\mathbf{A}, C) - \mathrm{Sc}(\mathbf{B}, C) \right) x^r \right| \leq \frac{\epsilon}{2 - 3\epsilon} \left[ [x^\ell]^* \mathbf{U}_{\mathrm{Sc}(\mathbf{B}, C)} x^\ell + [x^r]^* \mathbf{U}_{\mathrm{Sc}(\mathbf{B}, C)} [x^r] \right].$$

for all $x^\ell, x^r \in \mathbb{C}^{|C|}$.

Let $x^r, x^\ell \in \mathbb{C}^C$ be arbitrary. Now define $y^r \in \mathbb{C}^n$ by $y^r_C = x^r$ and $y^r_F = -\mathbf{A}_{FF}^{-1}\mathbf{A}_{FC}x^r$. Further, define $y^\ell \in \mathbb{C}^n$ by $y^\ell_C = x^\ell$, $y^\ell_F = -[\mathbf{B}_{FF}^*]^{-1}[\mathbf{B}_{CF}^*]x^\ell$. Note that $y^r$ and $y^\ell$ are defined so that

$$\mathbf{A}y^r = \begin{pmatrix} \mathbf{A}_{FF} & \mathbf{A}_{FC} \\ \mathbf{A}_{CF} & \mathbf{A}_{CC} \end{pmatrix} \begin{pmatrix} -\mathbf{A}_{FF}^{-1}\mathbf{A}_{FC}x^r \\ x^r \end{pmatrix} = \begin{pmatrix} \vec{0}_F \\ \mathrm{Sc}(\mathbf{A}, C)x^r \end{pmatrix}$$

and

$$[y^\ell]^*\mathbf{B} = \begin{pmatrix} -[\mathbf{B}_{FF}^*]^{-1}[\mathbf{B}_{CF}^*]x^\ell \\ x^\ell \end{pmatrix}^* \begin{pmatrix} \mathbf{B}_{FF} & \mathbf{B}_{FC} \\ \mathbf{B}_{CF} & \mathbf{B}_{CC} \end{pmatrix} = \begin{pmatrix} \vec{0}_F \\ \mathrm{Sc}(\mathbf{B}, C)^* x^\ell \end{pmatrix}^*.$$

Consequently $[y^\ell]^*\mathbf{A}y^r = [x^\ell]^*\mathrm{Sc}(\mathbf{A}, C)x^r$ and $[y^\ell]^*\mathbf{B}y^r = [x^\ell]^*\mathrm{Sc}(\mathbf{B}, C)x^r$. Therefore, by Equation (17) we have that

$$\left| [x^\ell]^*\mathrm{Sc}(\mathbf{A}, C)x^r - [x^\ell]^*\mathrm{Sc}(\mathbf{B}, C)x^r \right| = \left| [y^\ell]^* \left( \mathbf{A} - \mathbf{B} \right) y^r \right| \leq \frac{\epsilon}{2} \left[ [y^\ell]^* \mathbf{U_B} y^\ell + [y^r]^* \mathbf{U_B} y^r \right]. \tag{18}$$

Now,

$$[y^\ell]^* \mathbf{U_B} y^\ell = [y^\ell]^* \left( \frac{\mathbf{B} + \mathbf{B}^*}{2} \right) y^\ell = [x^\ell]^* \mathbf{U}_{\mathrm{Sc}(\mathbf{B}, C)} x^\ell$$

and since we know that $(1 - \epsilon)\mathbf{U_B} \preceq \mathbf{U_A} \preceq (1 + \epsilon)\mathbf{U_B}$ we have that

$$[y^r]^*\mathbf{U}_B[y^r] \preceq \frac{1}{1-\epsilon} \cdot [y^r]^*\mathbf{U_A}[y^r] = \frac{1}{1-\epsilon} \cdot [x^r]^*\mathbf{U}_{\mathrm{Sc}(\mathbf{A},C)}[x^r].$$

Substituting these into Equation 18 gives

$$\left| [x^\ell]^*(\mathrm{Sc}(\mathbf{A},C) - \mathrm{Sc}(\mathbf{B},C))x^r \right| \leq \frac{\epsilon}{2} \left[ [x^\ell]^*\mathbf{U}_{\mathrm{Sc}(\mathbf{B},C)}x^\ell + [x^r]^*\mathbf{U}_{\mathrm{Sc}(\mathbf{A},C)}[x^r]/(1-\epsilon) \right].$$

Plugging in $x^\ell = x^r$ above,

$$[x^r]^* \left( \mathbf{U}_{\mathrm{Sc}(\mathbf{A},C)} - \mathbf{U}_{\mathrm{Sc}(\mathbf{B},C)} \right) x^r \leq |[x^r]^* \left( \mathrm{Sc}(\mathbf{A},C) - \mathrm{Sc}(\mathbf{B},C) \right) x^r|$$
$$\leq \frac{\epsilon}{2} \left[ [x^r]^*\mathbf{U}_{\mathrm{Sc}(\mathbf{B},C)}x^r + [x^r]^*\mathbf{U}_{\mathrm{Sc}(\mathbf{A},C)}[x^r]/(1-\epsilon) \right].$$

and rearranging terms gives

$$[x^r]^*\mathbf{U}_{\mathrm{Sc}(\mathbf{A},C)}[x^r] \leq \frac{1 + \epsilon/2}{1 - \epsilon/(2 \cdot (1-\epsilon))} \cdot [x^r]^*\mathbf{U}_{\mathrm{Sc}(\mathbf{B},C)}[x^r],$$

which implies

$$\left| [x^\ell]^* \left( \mathrm{Sc}(\mathbf{A},C) - \mathrm{Sc}(\mathbf{B},C) \right) x^r \right| \leq \frac{\epsilon}{2} \left[ [x^\ell]^*\mathbf{U}_{\mathrm{Sc}(\mathbf{B},C)}x^\ell + \frac{1 + \epsilon/2}{1 - \epsilon/(2 \cdot (1-\epsilon))} \cdot \frac{1}{1-\epsilon} \cdot [x^r]^*\mathbf{U}_{\mathrm{Sc}(\mathbf{B},C)}[x^r] \right]$$
$$= \frac{\epsilon}{2} \left[ [x^\ell]^*\mathbf{U}_{\mathrm{Sc}(\mathbf{B},C)}x^\ell + \frac{1 + \epsilon/2}{1 - 3\epsilon/2} \cdot [x^r]^*\mathbf{U}_{\mathrm{Sc}(\mathbf{B},C)}[x^r] \right]$$

By symmetry, i.e. repeating the above argument on the conjugate transposes of the matrices we also have that

$$\left| [x^\ell]^* \left( \mathrm{Sc}(\mathbf{A},C) - \mathrm{Sc}(\mathbf{B},C) \right) x^r \right| \leq \frac{\epsilon}{2} \cdot \frac{1 + \epsilon/2}{1 - 3\epsilon/2} \cdot \left[ [x^\ell]^*\mathbf{U}_{\mathrm{Sc}(\mathbf{B},C)}x^\ell + [x^r]^*\mathbf{U}_{\mathrm{Sc}(\mathbf{B},C)}[x^r] \right].$$

Taking the average of these two equations then yields that

$$\left| [x^\ell]^* \left( \mathrm{Sc}(\mathbf{A},C) - \mathrm{Sc}(\mathbf{B},C) \right) x^r \right| \leq \frac{\epsilon}{4} \cdot \frac{2 - \epsilon}{1 - 3\epsilon/2} \cdot \left[ [x^\ell]^*\mathbf{U}_{\mathrm{Sc}(\mathbf{B},C)}x^\ell + [x^r]^*\mathbf{U}_{\mathrm{Sc}(\mathbf{B},C)}[x^r] \right].$$

Since

$$\frac{\epsilon}{4} \cdot \frac{2 - \epsilon}{1 - 3\epsilon/2} \leq \frac{1}{2} \cdot \frac{\epsilon}{2 - 3\epsilon}$$

and $x^\ell$ and $x^r$ were arbitrary, the result follows. $\qquad \square$

The loss of $1/(1 - 3\epsilon/2)$ in the approximation quality in Theorem 4.8 is avoidable with a small change to the notion of approximation.

**Definition B.1** (Min Spectral Approximation). Let $\mathbf{W}, \widetilde{\mathbf{W}} \in \mathbb{C}^{n \times n}$ be (possibly asymmetric) matrices. We say that $\widetilde{\mathbf{W}}$ is a *min $\epsilon$-approximation* of $\mathbf{W}$ (written $\widetilde{\mathbf{W}} \overset{\min}{\approx}_\epsilon \mathbf{W}$) if for all $x, y \in \mathbb{C}^n$

$$\left| x^*(\mathbf{W} - \widetilde{\mathbf{W}})y \right| \leq \frac{\epsilon}{2} \cdot \min \left\{ x^*\mathbf{U}_{\mathbf{I}-\mathbf{W}}x + y^*\mathbf{U}_{\mathbf{I}-\widetilde{\mathbf{W}}}y, y^*\mathbf{U}_{\mathbf{I}-\mathbf{W}}y + x^*\mathbf{U}_{\mathbf{I}-\widetilde{\mathbf{W}}}x \right\}$$

In min spectral approximation, the error on the right-hand side is measured with respect to both matrices rather than just one of them. Min spectral approximation is equivalent to the original notion of spectral approximation up to a small loss in approximation quality, as seen in the following lemma.

**Lemma B.2.** *Let* $\mathbf{W}, \widetilde{\mathbf{W}} \in \mathbb{C}^{n \times n}$ *be (possibly asymmetric) matrices and suppose* $\epsilon \in (0, 1)$.

1. *If* $\widetilde{\mathbf{W}} \stackrel{\min}{\approx}_{\epsilon} \mathbf{W}$ *then* $\widetilde{\mathbf{W}} \approx_{\epsilon/(1-\epsilon/2)} \mathbf{W}$

2. *If* $\widetilde{\mathbf{W}} \approx_{\epsilon} \mathbf{W}$ *then* $\widetilde{\mathbf{W}} \stackrel{\min}{\approx}_{\epsilon/(1-\epsilon)} \mathbf{W}$.

*Proof.* Suppose $\widetilde{\mathbf{W}} \stackrel{\min}{\approx}_{\epsilon} \mathbf{W}$ and fix $x \in \mathbb{C}^n$. Then we can write

$$
\begin{aligned}
x^* \left( \mathbf{U}_{\mathbf{I}-\widetilde{\mathbf{W}}} - \mathbf{U}_{\mathbf{I}-\mathbf{W}} \right) x = \mathrm{Re}\left( x^*(\mathbf{W} - \widetilde{\mathbf{W}})x \right) \\
\leq \left| x^*(\mathbf{W} - \widetilde{\mathbf{W}})x \right| \\
\leq \frac{\epsilon}{2} \cdot \left( x^* \mathbf{U}_{\mathbf{I}-\mathbf{W}} x + x^* \mathbf{U}_{\mathbf{I}-\widetilde{\mathbf{W}}} x \right).
\end{aligned}
$$

Rearranging the above gives

$$
x^* \mathbf{U}_{\mathbf{I}-\widetilde{\mathbf{W}}} x \leq \frac{1 + \epsilon/2}{1 - \epsilon/2} \cdot x^* \mathbf{U}_{\mathbf{I}-\mathbf{W}} x
$$

which implies

$$
\begin{aligned}
\left| x^*(\mathbf{W} - \widetilde{\mathbf{W}})y \right| \leq \frac{\epsilon}{2} \cdot \left( y^* \mathbf{U}_{\mathbf{I}-\mathbf{W}} y + x^* \mathbf{U}_{\mathbf{I}-\widetilde{\mathbf{W}}} x \right) \\
\leq \frac{\epsilon}{2} \cdot \left( y^* \mathbf{U}_{\mathbf{I}-\mathbf{W}} y + \frac{1 + \epsilon/2}{1 - \epsilon/2} \cdot x^* \mathbf{U}_{\mathbf{I}-\mathbf{W}} x \right).
\end{aligned}
$$

A similar calculation gives

$$
\left| x^*(\mathbf{W} - \widetilde{\mathbf{W}})y \right| \leq \frac{\epsilon}{2} \cdot \left( \frac{1 + \epsilon/2}{1 - \epsilon/2} \cdot y^* \mathbf{U}_{\mathbf{I}-\mathbf{W}} y + x^* \mathbf{U}_{\mathbf{I}-\mathbf{W}} x \right).
$$

Averaging these two inequalities and noting that

$$
\frac{1}{2} \cdot \frac{\epsilon}{2} \cdot \left( 1 + \frac{1 + \epsilon/2}{1 - \epsilon/2} \right) = \frac{\epsilon}{2 \cdot (1 - \epsilon/2)}
$$

gives

$$
\left| x^*(\mathbf{W} - \widetilde{\mathbf{W}})y \right| \leq \frac{\epsilon}{2 \cdot (1 - \epsilon/2)} \cdot \left( x^* \mathbf{U}_{\mathbf{I}-\mathbf{W}} x + y^* \mathbf{U}_{\mathbf{I}-\mathbf{W}} y \right),
$$

or equivalently $\widetilde{\mathbf{W}} \approx_{\epsilon/(1-\epsilon/2)} \mathbf{W}$.

Now suppose $\widetilde{\mathbf{W}} \approx_{\epsilon} \mathbf{W}$. It follows that $\mathbf{U}_{\widetilde{\mathbf{W}}} \approx_{\epsilon} \mathbf{U}_{\mathbf{W}}$ and hence

$$
(1 - \epsilon) \cdot \mathbf{U}_{\mathbf{I}-\mathbf{W}} \preceq \mathbf{U}_{\mathbf{I}-\widetilde{\mathbf{W}}} \preceq (1 + \epsilon) \cdot \mathbf{U}_{\mathbf{I}-\mathbf{W}},
$$

which implies that $\mathbf{U}_{\mathbf{I}-\mathbf{W}} \preceq (1/(1 - \epsilon)) \cdot \mathbf{U}_{\mathbf{I}-\widetilde{\mathbf{W}}}$. Now we can write that for all $x, y \in \mathbb{C}^n$

$$
\begin{aligned}
\left| x^*(\mathbf{W} - \widetilde{\mathbf{W}})y \right| \leq \frac{\epsilon}{2} \cdot \left( x^* \mathbf{U}_{\mathbf{I}-\mathbf{W}} x + y^* \mathbf{U}_{\mathbf{I}-\mathbf{W}} y \right) \\
\leq \frac{\epsilon}{2} \cdot \left( \frac{1}{1 - \epsilon} \cdot x^* \mathbf{U}_{\mathbf{I}-\widetilde{\mathbf{W}}} x + y^* \mathbf{U}_{\mathbf{I}-\mathbf{W}} y \right) \\
\leq \frac{\epsilon}{2 \cdot (1 - \epsilon)} \cdot \left( x^* \mathbf{U}_{\mathbf{I}-\widetilde{\mathbf{W}}} x + y^* \mathbf{U}_{\mathbf{I}-\mathbf{W}} y \right).
\end{aligned}
$$

A similar calculation shows

$$\left|x^*(\mathbf{W} - \widetilde{\mathbf{W}})y\right| \leq \frac{\epsilon}{2 \cdot (1 - \epsilon)} \cdot \left(x^*\mathbf{U_{I-W}}x + y^*\mathbf{U_{I-\widetilde{W}}}y\right)$$

and hence $\widetilde{\mathbf{W}} \overset{\min}{\approx}_{\epsilon/(1-\epsilon)} \mathbf{W}$. □

Now we show that min spectral approximation is preserved under schur complements with no loss in approximation quality.

**Theorem B.3.** *Let* $\mathbf{W}, \widetilde{\mathbf{W}} \in \mathbb{C}^n$ *and suppose that* $\widetilde{\mathbf{W}} \overset{\min}{\approx}_{\epsilon} \mathbf{W}$ *for* $\epsilon \in (0,1)$. *Let* $F \subseteq [n]$ *such that* $(\mathbf{I}_{|F|} - \mathbf{W}_{FF})$ *is invertible and let* $C = [n] \setminus F$. *Then*

$$\mathbf{I}_{|C|} - \mathrm{Sc}(\mathbf{I}_n - \widetilde{\mathbf{W}}, C) \overset{\min}{\approx}_{\epsilon} \mathbf{I}_{|C|} - \mathrm{Sc}(\mathbf{I}_n - \mathbf{W}, C)$$

*Proof.* Let $\mathbf{A} = \mathbf{I}_n - \widetilde{\mathbf{W}}$, $\mathbf{B} = \mathbf{I}_n - \mathbf{W}$. Note that since $\widetilde{\mathbf{W}} \overset{\min}{\approx}_{\epsilon} \mathbf{W}$, we have for all $x, y \in \mathbb{C}^n$

$$|x^*(\mathbf{B} - \mathbf{A})y| \leq \frac{\epsilon}{2} \min\left\{x^*\mathbf{U_B}x + y^*\mathbf{U_A}y, y^*\mathbf{U_B}y + x^*\mathbf{U_A}x\right\}, \tag{19}$$

Our goal is to show that $\mathbf{I}_{|C|} - \mathrm{Sc}(\mathbf{A}, C) \overset{\min}{\approx}_{\epsilon} \mathbf{I}_{|C|} - \mathrm{Sc}(\mathbf{B}, C)$, which is equivalent to

$$\left|[x^\ell]^* \left(\mathrm{Sc}(\mathbf{A}, C) - \mathrm{Sc}(\mathbf{B}, C)\right)x^r\right|$$
$$\leq \frac{\epsilon}{2} \cdot \min\left\{[x^\ell]^*\mathbf{U}_{\mathrm{Sc}(\mathbf{A},C)}x^\ell + [x^r]^*\mathbf{U}_{\mathrm{Sc}(\mathbf{B},C)}[x^r], [x^\ell]^*\mathbf{U}_{\mathrm{Sc}(\mathbf{B},C)}x^\ell + [x^r]^*\mathbf{U}_{\mathrm{Sc}(\mathbf{A},C)}[x^r]\right\}.$$

for all $x^\ell, x^r \in \mathbb{C}^{|C|}$.

Let $x^r, x^\ell \in \mathbb{C}^C$ be arbitrary. Now define $y^r \in \mathbb{C}^n$ by $y^r_C = x^r$ and $y^r_F = -\mathbf{A}_{FF}^{-1}\mathbf{A}_{FC}x^r$. Further, define $y^\ell \in \mathbb{C}^n$ by $y^\ell_C = x^\ell$, $y^\ell_F = -[\mathbf{B}^*_{FF}]^{-1}[\mathbf{B}^*_{CF}]x^\ell$. Note that $y^r$ and $y^\ell$ are defined so that

$$\mathbf{A}y^r = \left(\begin{array}{cc} \mathbf{A}_{FF} & \mathbf{A}_{FC} \\ \mathbf{A}_{CF} & \mathbf{A}_{CC} \end{array}\right)\left(\begin{array}{c} -\mathbf{A}_{FF}^{-1}\mathbf{A}_{FC}x^r \\ x^r \end{array}\right) = \left(\begin{array}{c} \vec{0}_F \\ \mathrm{Sc}(\mathbf{A}, C)x^r \end{array}\right)$$

and

$$[y^\ell]^*\mathbf{B} = \left(\begin{array}{c} -[\mathbf{B}^*_{FF}]^{-1}[\mathbf{B}^*_{CF}]x^\ell \\ x^\ell \end{array}\right)^*\left(\begin{array}{cc} \mathbf{B}_{FF} & \mathbf{B}_{FC} \\ \mathbf{B}_{CF} & \mathbf{B}_{CC} \end{array}\right) = \left(\begin{array}{c} \vec{0}_F \\ \mathrm{Sc}(\mathbf{B}, C)^*x^\ell \end{array}\right)^*.$$

Consequently $[y^\ell]^*\mathbf{A}y^r = [x^\ell]^*\mathrm{Sc}(\mathbf{A}, C)x^r$ and $[y^\ell]^*\mathbf{B}y^r = [x^\ell]^*\mathrm{Sc}(\mathbf{B}, C)x^r$. We also have $[y^\ell]^*\mathbf{U_B}y^\ell = [x^\ell]^*\mathbf{U}_{\mathrm{Sc}(\mathbf{B},C)}x^\ell$ and $[y^r]^*\mathbf{U_A}y^r = [x^r]^*\mathbf{U}_{\mathrm{Sc}(\mathbf{A},C)}x^r$. Therefore, by Equation (19) we get

$$\left|[x^\ell]^*\mathrm{Sc}(\mathbf{A}, C)x^r - [x^\ell]^*\mathrm{Sc}(\mathbf{B}, C)x^r\right| = \left|[y^\ell]^*\left(\mathbf{A} - \mathbf{B}\right)y^r\right|$$
$$\leq \frac{\epsilon}{2} \cdot \left([y^\ell]^*\mathbf{U_B}y^\ell + [y^r]^*\mathbf{U_A}y^r\right)$$
$$= \frac{\epsilon}{2} \cdot \left([x^\ell]^*\mathbf{U}_{\mathrm{Sc}(\mathbf{B},C)}x^\ell + [x^r]^*\mathbf{U}_{\mathrm{Sc}(\mathbf{A},C)}x^r\right).$$

Redefining $y^\ell$ and $y^r$, we can get the analogous inequality

$$\left|[x^\ell]^*\mathrm{Sc}(\mathbf{A}, C)x^r - [x^\ell]^*\mathrm{Sc}(\mathbf{B}, C)x^r\right| \leq \frac{\epsilon}{2} \cdot \left([x^\ell]^*\mathbf{U}_{\mathrm{Sc}(\mathbf{A},C)}x^\ell + [x^r]^*\mathbf{U}_{\mathrm{Sc}(\mathbf{B},C)}x^r\right).$$

Therefore $\mathbf{I}_{|C|} - \mathrm{Sc}(\mathbf{A}, C) \overset{\min}{\approx}_{\epsilon} \mathbf{I}_{|C|} - \mathrm{Sc}(\mathbf{B}, C)$. □

## C    Proof of Lemma 5.2

**Lemma 5.2 restated.** *Let $G$ be a strongly connected, regular directed multigraph on $n$ vertices with transition matrix $\mathbf{W}$ and let $\mathbf{J} \in \mathbb{R}^{n \times n}$ be a matrix with $1/n$ in every entry (i.e. $\mathbf{J}$ is the transition matrix of the complete graph with a self loop on every vertex). Then $\lambda(G) \le \lambda$ if and only if $\mathbf{W} \overset{\circ}{\approx}_{\lambda} \mathbf{J}$.*

*Proof.* First we will show that if $\mathbf{W} \overset{\circ}{\approx}_{\lambda} \mathbf{J}$ then $\lambda(G) \le \lambda$. By the equivalence of items 1 and 2 in Lemma A.1, we have that for all $x, y \in \mathbb{C}^n$

$$x^*(\mathbf{W} - \mathbf{J})y \le \lambda \cdot \sqrt{x^*(\mathbf{I} - \mathbf{J})x \cdot y^*(\mathbf{I} - \mathbf{J})y}.$$

Note that we can write

$$\lambda(G) = \max_{u \perp \mathbf{1}, v \perp \mathbf{1}} \frac{u^*\mathbf{W}v}{\|u\|\|v\|}.$$

Let $\bar{u}$ and $\bar{v}$ be vectors that achieve the maximization above. Setting $x = \bar{u}$ and $y = \bar{v}$, and noting that $\mathbf{J}\bar{v} = \mathbf{J}\bar{u} = \mathbf{0}$ because $\bar{u}$ and $\bar{v}$ are perpendicular to $\mathbf{1}$, our inequality becomes

$$\bar{u}^*\mathbf{W}\bar{v} \le \lambda \cdot \sqrt{\|\bar{u}\|^2 \cdot \|\bar{v}\|^2} = \|\bar{u}\| \cdot \|\bar{v}\|.$$

Dividing by $\|\bar{u}\| \cdot \|\bar{v}\|$ completes the proof.

For the other direction, assume $\lambda(G) \le \lambda$ and we will show that for all $x, y \in \mathbb{C}^n$ and all $z \in \mathbb{C}$ such that $|z| = 1$,

$$|x^*(\mathbf{W} - \mathbf{J})y| \le \lambda \cdot \sqrt{|x^*(\mathbf{I} - z\mathbf{J})x| \cdot |y^*(\mathbf{I} - z\mathbf{J})y|} \tag{20}$$

Note that for all constant vectors $v$ (i.e. $v$ has the same entry in every coordinate) we have by the regularity of $\mathbf{W}$ and $\mathbf{J}$,

$$\mathbf{W}v = \mathbf{W}^\top v = \mathbf{J}v = \mathbf{J}^\top v = v.$$

So when $x$ or $y$ is constant, the left-hand side of Inequality (20) is zero and the inequality is true (because the right-hand side is always non-negative). Furthermore, orthogonality to $\vec{\mathbf{1}}$ is preserved under $\mathbf{I}, \mathbf{J}, \mathbf{W}$, and $\mathbf{W}^\top$ so it suffices to consider vectors $x, y \perp \vec{\mathbf{1}}$. Fix two such vectors. For $v \perp \vec{\mathbf{1}}$, we have $\mathbf{J}v = \vec{\mathbf{0}}$. So Inequality (20) becomes

$$|x^*\mathbf{W}y| \le \lambda \cdot \sqrt{x^*x \cdot y^*y}$$
$$= \lambda \cdot \|x\| \cdot \|y\|.$$

Applying Cauchy-Schwarz, we have

$$|x^\top\mathbf{W}y| \le \|x\| \cdot \|\mathbf{W}y\|$$
$$\le \lambda \cdot \|x\| \cdot \|y\|$$

where the last line follows from the assumption that $\lambda(G) \le \lambda$. $\qquad\square$

## D    Omitted Proofs from Section 6

Below is the statement of Lemma 2.3 from [CKK$^+$]. we give a proof of Lemma 6.7 using this lemma.

**Lemma D.1** (CKKPPRS18 Lemma 2.3)**.** *Consider a sequence of $m$-by-$m$ matrices $\mathbf{S}^{(0)}, \mathbf{S}^{(1)}, \ldots, \mathbf{S}^{(m)}$ such that*

1. $\mathbf{S}^{(i)}$ has non-zero entries only on the indices $[i+1, m]$,

2. The left/right kernels of $\mathbf{S}^{(i)}$ are equal, and after restricting $\mathbf{S}^{(i)}$ to the indices $[i+1, m]$, the kernel of the resulting matrix equals the coordinate restriction of the vectors in the kernel of $\mathbf{S}$. Formally, $\ker\left(\mathbf{S}^{(i)}_{[i+1,m],[i+1,m]}\right) = \{b_{[i+1,m]} : b \in \ker(\mathbf{S}^{(0)})\}$.

3. The symmetrization of each $\mathbf{S}^{(i)}$, $\mathbf{U}_{\mathbf{S}^{(i)}} = \frac{1}{2}(\mathbf{S}^{(i)} + (\mathbf{S}^{(i)})^\top)$, is positive semi-definite.

Let $\mathbf{M} = \mathbf{M}^{(0)} = \mathbf{S}^{(0)}$, and define matrices $\mathbf{M}^{(1)}, \mathbf{M}^{(2)}, \ldots, \mathbf{M}^{(m)}$ iteratively by

$$\mathbf{M}^{(i+1)} \stackrel{\text{def}}{=} \mathbf{M}^{(i)} + \left(\mathbf{S}^{(i+1)} - \mathrm{Sc}\left(\mathbf{M}^{(i)}, [i+1, m]\right)\right) \qquad \forall\, 0 \le i < p_{\max}.$$

If for a subsequence of indices $1 = i_0 < i_1 < i_2 < \ldots < i_{p_{\max}}$ associated scaling parameters $0 < \theta_0, \theta_1, \ldots, \theta_{p_{\max}-1} < 1/2$ such that $\sum_{p=0}^{p_{\max}-1} \theta_p = 1$, and some global error $0 < \epsilon < 1/2$, we have for every $0 \le p < p_{\max}$:

$$\left\| \mathbf{U}^{+/2}_{\mathbf{S}^{(i_p)}} \left(\mathbf{M}^{(i_p)} - \mathbf{M}^{(i_{p+1})}\right) \mathbf{U}^{+/2}_{\mathbf{S}^{(i_p)}} \right\| \le \theta_p \epsilon,$$

then for a matrix-norm defined from the symmetrization of the $\mathbf{S}^{(i_p)}$ matrices and the scaling parameters:

$$\mathbf{F} = \sum_{0 \le p < p_{\max}} \theta_p \mathbf{U}_{\mathbf{S}^{(i_p)}},$$

we have:

1. for each $0 \le i \le p_{\max}$,

$$\left\| \mathbf{F}^{+/2} \left(\mathbf{M} - \mathbf{M}^{(i)}\right) \mathbf{F}^{+/2} \right\|_2 \le \epsilon,$$

2. The final matrix $\mathbf{M}^{(p_{\max})}$ satisfies

$$\mathbf{M}^{(p_{\max})\top} \mathbf{F}^+ \mathbf{M}^{(p_{\max})} \succeq \frac{1}{10 p_{\max}^2} \cdot \mathbf{F}.$$

**Lemma D.2.** Let $\mathbf{L}^{(i)}$'s be $2^k n$ by $2^k n$ matrices defined as Equation (11), then

$$\mathbf{L}^{(i+1)} - \mathbf{L}^{(i)} = \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{I}_{2^{k-i-1}n} - \mathbf{C}_{2^{k-i-1}} \otimes \mathbf{W}_{i+1} \end{bmatrix} - \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{I}_{2^{k-i-1}n} - \mathbf{C}_{2^{k-i-1}} \otimes \mathbf{W}_i^2 \end{bmatrix}.$$

*Proof.* From Equation (11), we have

$$\mathbf{L}^{(i+1)} = \mathbf{X}_1 \cdots \mathbf{X}_{i+1} \begin{bmatrix} \mathbf{I}_{(2^k - 2^{k-i-1})n} & 0 \\ 0 & \mathbf{I}_{2^{k-i-1}n} - \mathbf{C}_{2^{k-i-1}} \otimes \mathbf{W}_{i+1} \end{bmatrix} \mathbf{Y}_{i+1} \cdots \mathbf{Y}_1,$$

applying $\mathbf{X}_{i+1}$, and $\mathbf{Y}_{i+1}$ to the middle matrix we get,

$$\mathbf{L}^{(i+1)} = \mathbf{X}_1 \cdots \mathbf{X}_i \begin{bmatrix} \mathbf{I}_{(2^k - 2^{k-i})n} & 0 & 0 \\ 0 & \mathbf{I}_{\frac{2^k n}{2^{i+1}}} & -\mathbf{C}_{\frac{2^k n}{2^{i+1}}} \otimes \mathbf{W}_{i+1} \\ 0 & -\mathbf{I}_{\frac{2^k n}{2^{i+1}}} \otimes \mathbf{W}_{i+1} & \mathbf{I}_{\frac{2^k n}{2^{i+1}}} - \mathbf{C}_{\frac{2^k n}{2^{i+1}}} \otimes \mathbf{W}_{i+1} + \mathbf{C}_{\frac{2^k n}{2^{i+1}}} \otimes \mathbf{W}_i^2 \end{bmatrix} \mathbf{Y}_i \cdots \mathbf{Y}_1.$$

Therefore,

$$\mathbf{L}^{(i+1)} - \mathbf{L}^{(i)} = \mathbf{X}_1 \cdots \mathbf{X}_i \begin{bmatrix} 0 & 0 \\ 0 & -\mathbf{C}_{2^{k-i-1}} \otimes \mathbf{W}_{i+1} + \mathbf{C}_{2^{k-i-1}} \otimes \mathbf{W}_i^2 \end{bmatrix} \mathbf{Y}_i \cdots \mathbf{Y}_1.$$

Note that the error only resides on the lower right block of the middle matrix. It is easy to verify that hitting the middle matrix by $\mathbf{X}^{(i)}$'s from left and $\mathbf{Y}^{(i)}$'s from right does not change the matrix (as the non-zero part is only multiplied by identity blocks). Therefore,

$$\mathbf{L}^{(i+1)} - \mathbf{L}^{(i)} = \begin{bmatrix} 0 & 0 \\ 0 & -\mathbf{C}_{2^{k-i-1}} \otimes \mathbf{W}_{i+1} + \mathbf{C}_{2^{k-i-1}} \otimes \mathbf{W}_i^2 \end{bmatrix}.$$

$\square$

Now we are ready to give a proof of Lemma 6.7.

*Proof of Lemma 6.7.* From $\mathbf{S}^{(i)}$'s and $\mathbf{L}^{(i)}$'s, we build a sequence of $\hat{\mathbf{S}}^{(j)}$'s and $\hat{\mathbf{M}}^{(j)}$'s that satisfy the conditions of Lemma D.1, and using that we derive the statement of Lemma 6.7. For $0 \leq i < k$, and $0 \leq j < 2^{k-i-1}n$, let $a_i \overset{\text{def}}{=} (2^k - 2^{k-i})n$, $\hat{\mathbf{S}}^{(a_k)} = \mathbf{S}^{(k)}$, and

$$\hat{\mathbf{S}}^{(a_i+j)} = \begin{cases} \mathbf{S}^{(i)}, & \text{if } j = 0 \\ \mathrm{Sc}(\hat{\mathbf{M}}^{a_i+j-1}, [a_i + j, 2^k n]) & \text{otherwise.} \end{cases}$$

and

$$\hat{\mathbf{M}}^{(h+1)} \overset{\text{def}}{=} \hat{\mathbf{M}}^{(h)} + \left( \hat{\mathbf{S}}^{(h+1)} - \mathrm{Sc}\left( \hat{\mathbf{M}}^{(h)}, \left[ h + 1, 2^k n \right] \right) \right) \qquad \forall\, 0 \leq h < (2^k - 1)n.$$

Note that $\hat{\mathbf{S}}$'s satisfy all the three premises in Lemma D.1. First $\hat{\mathbf{S}}^{(i)}$ has non-zero entries only on the indices $[i + 1, 2^k n]$. Further as all $\hat{\mathbf{S}}^{(i)}$'s are random-walk Laplacian of Eulerian connected aperiodic regular digraphs, the left and right kernels are the same, all have the same kernel up to the restriction to non-zero entries, and all $\mathbf{U}_{\mathbf{S}^{(i)}}$'s are PSD. Therefore all the three premises of the lemma are satisfied.

Next we show that for all $i$'s $\mathbf{L}^{(i+1)}$ approximates $\mathbf{L}^{(i)}$ in the norm defined by $\mathbf{U}_{\mathbf{S}^{(i)}}$. By Lemma D.2,

$$\mathbf{L}^{(i+1)} - \mathbf{L}^{(i)} = \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{C}_{2^{k-i-1}} \otimes \mathbf{W}_i^2 - \mathbf{C}_{2^{k-i-1}} \otimes \mathbf{W}_{i+1}. \end{bmatrix}$$

Now, given $\mathbf{W}_{i+1} \overset{\circ}{\approx}_{\epsilon/k} \mathbf{W}_i^2$, by Theorem 4.5 we get

$$\|\mathbf{U}_{\mathrm{Sc}(\mathbf{S}_i, H_i)}^{+/2} \left( \mathbf{L}^{(i+1)} - \mathbf{L}^{(i)} \right) \mathbf{U}_{\mathrm{Sc}(\mathbf{S}_i, H_i)}^{+/2} \| \leq \epsilon/k.$$

Since $\mathbf{U}_{\mathrm{Sc}(\mathbf{S}_i, H_i)} \preceq 2 \mathbf{U}_{\mathbf{S}^{(i)}}$,

$$\|\mathbf{U}_{\mathbf{S}^{(i)}}^{+/2} \left( \mathbf{L}^{(i+1)} - \mathbf{L}^{(i)} \right) \mathbf{U}_{\mathbf{S}^{(i)}}^{+/2} \| \leq 2\epsilon/k.$$

By construction, we have $\hat{\mathbf{S}}^{(a_i)} = \mathbf{S}^{(i)}$ and $\hat{\mathbf{M}}^{(a_i)} = \mathbf{L}^{(i)}$ for all $0 \leq i \leq k$. Therefore, we get

$$\left\| \mathbf{U}_{\hat{\mathbf{S}}^{(a_i)}}^{+/2} \left( \hat{\mathbf{M}}^{(a_i)} - \hat{\mathbf{M}}^{(a_{i+1})} \right) \mathbf{U}_{\hat{\mathbf{S}}^{(a_i)}}^{+/2} \right\| \leq 2\epsilon/k,$$

Thus by Lemma D.1, for $\mathbf{F} = \frac{2}{k} \sum_{i=0}^{k} \mathbf{U}_{\mathbf{S}^{(i)}}$,

$$\|\mathbf{F}^{+/2}(\mathbf{L} - \mathbf{L}^{(i)})\mathbf{F}^{+/2}\| \leq \epsilon \quad \forall 0 \leq i \leq k$$

and

$${\mathbf{L}^{(k)}}^\top \mathbf{F}^+ \mathbf{L}^{(k)} \succeq \frac{1}{40k^2} \mathbf{F}.$$

$\square$

**Lemma D.3.** *Given a PSD matrix* $\mathbf{U} \in \mathbb{R}^{n \times n}$, *and an orthogonal projection* $\Pi \in \mathbb{R}^{n \times n}$ *such that* $\mathbf{U}\Pi = \Pi\mathbf{U}$ *we have*

$$\mathbf{U}\Pi \preceq \mathbf{U}$$

*Proof.* Let $x \in \mathbb{R}^n$ be an arbitrary vector, and let $x_1$ denote the projection of $x$ into $\mathrm{im}(\Pi)$ and $x_2$ be the projection to the subspace orthogonal to $\mathrm{im}(\Pi)$, such that $x = x_1 + x_2$. Note that $\Pi x_1 = x_1$, and $\Pi x_2 = 0$. Therefore we have,

$$(x_1 + x_2)^\top \mathbf{U}(x_1 + x_2) = (\Pi x_1 + x_2)^\top \mathbf{U}(\Pi x_1 + x_2) = x_1^\top \mathbf{U} x_1 + x_2^\top \mathbf{U} x_2,$$

and

$$(x_1 + x_2)^\top \mathbf{U}\Pi(x_1 + x_2) = (x_1 + x_2)^\top \mathbf{U}\Pi x_1 = (x_1 + x_2)^\top \Pi\mathbf{U} x_1 = x_1^\top \Pi\mathbf{U}\Pi x_1 = x_1^\top \mathbf{U} x_1.$$

Therefore,

$$x^\top \mathbf{U}\Pi x = x_1^\top \mathbf{U} x_1 \leq x_1^\top \mathbf{U} x_1 + x_2^\top \mathbf{U} x_2 = x^\top \mathbf{U} x.$$

Since the choice of $x$ was arbitrary this completes the proof. $\qquad\square$

**Lemma D.4** (Lemma 2.6 in [CKK$^+$]). *Suppose we are given matrices* $\mathbf{L}, \widetilde{\mathbf{L}}$ *and a positive semi-definite matrix* $\mathbf{F}$ *such that* $\ker(\mathbf{F}) \subseteq \ker(\mathbf{L}) = \ker(\mathbf{L}^\top) = \ker(\widetilde{\mathbf{L}}) = \ker(\widetilde{\mathbf{L}}^\top)$ *and*

1. $\|\mathbf{F}^{+/2}(\mathbf{L} - \widetilde{\mathbf{L}})\mathbf{F}^{+/2}\| \leq \epsilon$

2. $\widetilde{\mathbf{L}}^\top \mathbf{F}^+ \widetilde{\mathbf{L}} \succeq \gamma \mathbf{F}$,

*then* $\|\mathbf{I}_{\mathrm{im}(\mathbf{L})} - \widetilde{\mathbf{L}}^+ \mathbf{L}\|_{\mathbf{F}} \leq \epsilon\sqrt{\gamma^{-1}}$.

**Lemma D.5** (Lemma B.3 in [CKK$^+$]). *Let* $\mathbf{L}, \widetilde{\mathbf{L}}, \mathbf{F}$ *be arbitrary matrices with* $\ker(\mathbf{F}) = \ker(\mathbf{F}^\top) = \ker(\mathbf{L}) = \ker(\mathbf{L}^\top) = \ker(\widetilde{\mathbf{L}}) = \ker(\widetilde{\mathbf{L}}^\top)$. *If* $\|\mathbf{F}^{+/2}(\mathbf{L} - \widetilde{\mathbf{L}})\mathbf{F}^{+/2}\| \leq \epsilon$ *and* $\widetilde{\mathbf{L}}^\top \mathbf{F}^+ \widetilde{\mathbf{L}} \succeq \gamma \mathbf{F}$, *then* $\mathbf{L}^\top F^+ \mathbf{L} \approx_{O(\frac{\epsilon}{\sqrt{\gamma}} + \frac{\epsilon^2}{\gamma})} \widetilde{\mathbf{L}}^\top F^+ \widetilde{\mathbf{L}}$.

**Lemma D.6** (Lemma 13 in [CKP$^+$2]). *Let* $\mathbf{L}$ *be an Eulerian directed Laplacian,* $\mathrm{tr}(\mathbf{U}^{+/2}\mathbf{L}^\top \mathbf{U}^+ \mathbf{L}\mathbf{U}^{+/2}) \leq 2(n-1)^2$.

# E   Proof of Lemma 7.4

**Lemma 7.4 restated.** *Let* $G_0$ *be a d-regular, directed multigraph on n vertices with a two-way labeling and* $H_1, \ldots, H_k$ *be c-regular undirected graphs with two-way labelings where for each* $i \in [k]$, $H_i$ *has* $d \cdot c^{i-1}$ *vertices. For each* $i \in [k]$ *let*

$$G_i = G_{i-1} \text{\textcircled{s}} H_i.$$

*Then given* $v_0 \in [n], i_0 \in [d \cdot c^{i-1}], j_0 \in [c]$, $\mathrm{Rot}_{G_i}(v, (i_0, j_0))$ *can be computed in space* $O(\log(n \cdot d) + k \cdot \log c)$ *with oracle queries to* $\mathrm{Rot}_{H_1}, \ldots, \mathrm{Rot}_{H_k}$.

*Proof.* When $c$ is subpolynomial we are reasoning about sublogarithmic space complexity, which can depend on the model. So we will be explicit about the model we are using. We compute the rotation map of $G_i$ on a multi-tape Turing machine with the following input/output conventions:

- Input Description:

- Tape 1 (read-only): Contains the initial input graph $G_0$, with the head at the leftmost position of the tape.
- Tape 2 (read-write): Contains the input triple $(v_0, (i_0, j_0))$, where $v_0$ is a vertex of $G_i$, $i_0 \in [d \cdot c^{i-1}]$ is an edge label in $G_i$, and $j_0 \in [c]$ is an edge label in $H_i$ on a *read-write* tape, with the head at the *rightmost* position of $j_0$. The rest of the tape may contain additional data.
- Tapes 3+ (read-write): Blank worktapes with the head at the leftmost position.

- Output Description:

  - Tape 1: The head should be returned to the leftmost position.
  - Tape 2: In place of $(v_0, (i_0, j_0))$, it should contain the output $(v_2, (i_3, j_1)) = \mathrm{Rot}_{G_i}(v_0, (i_0, j_0))$ as described in Definition 5.5. The head should be at the rightmost position of $j_1$ and the rest of the tape should remain unchanged from its state at the beginning of the computation.
  - Tapes 3+ (read-write): Are returned to the blank state with the heads at the leftmost position.

Let $\mathrm{Space}(G_i)$ be the amount of space required to compute the rotation map of graph $G_i$. We will show that for all $i \in [k]$, $\mathrm{Space}(G_i) = \mathrm{Space}(G_{i-1}) + O(\log c)$. Note that $\mathrm{Space}(G_0) = O(\log(nd))$.

Fix $i \in [k]$. We begin with $v_0 \in [n], i_0 \in [d \cdot c^{i-1}]$ and $j_0 \in [c]$ on tape 2 with the head on the rightmost position of $j_0$ and we want to compute $\mathrm{Rot}_{G_i}(v_0, (i_0, j_0))$. We move the head left to the rightmost position of $i_0$, recursively compute $\mathrm{Rot}_{G_{i-1}}(v_0, i_0)$ so that tape 2 now contains $(v_1, i_1, j_0)$. Then we move the head to the rightmost position of $j_0$ and compute $\mathrm{Rot}_{H_i}(i_1, j_0)$ so that tape 2 now contains $(v_1, i_2, j_1)$. Finally, we move the head to the rightmost position of $i_2$ and compute $\mathrm{Rot}_{G_{i-1}}(v_1, i_2)$ so that tape 2 contains $(v_2, i_3, j_1)$.

This requires 2 evaluations of the rotation map of $G_{i-1}$ and one evaluation of the rotation map of $H_i$. Note that we can reuse the same space for each of these evaluations because they happen in succession. The space needed on top of $\mathrm{Space}(G_{i-1})$ is the space to store edge label $j_0$, which uses $O(\log c)$ space. So $\mathrm{Space}(G_i) = \mathrm{Space}(G_{i-1}) + O(\log c)$. Since $i$ can be as large as $k$ and $\mathrm{Space}(G_0) = O(\log(n \cdot d))$ we get that for all $i \in [k]$, $\mathrm{Space}(G_i) = O(\log n \cdot d + k \cdot \log c)$. $\qquad \square$