

# Faster Maxflow via Improved Dynamic Spectral Vertex Sparsifiers

Jan van den Brand\*   Yu Gao†   Arun Jambulapati‡   Yin Tat Lee§   Yang P. Liu¶  
 Richard Peng||   Aaron Sidford\*\*

December 2, 2021

## Abstract

We make several advances broadly related to the maintenance of electrical flows in weighted graphs undergoing dynamic resistance updates, including:

1. More efficient dynamic spectral vertex sparsification, achieved by faster length estimation of random walks in weighted graphs using Morris counters [Morris 1978, Nelson-Yu 2020].
2. A direct reduction from detecting edges with large energy in dynamic electric flows to dynamic spectral vertex sparsifiers.
3. A procedure for turning algorithms for estimating a sequence of vectors under updates from an oblivious adversary to one that tolerates adaptive adversaries via the Gaussian-mechanism from differential privacy.

Combining these pieces with modifications to prior robust interior point frameworks gives an algorithm that on graphs with  $m$  edges computes a mincost flow with edge costs and capacities in  $[1, U]$  in time  $\tilde{O}(m^{3/2-1/58} \log^2 U)$ . In prior and independent work, [Axiotis-Mądry-Vladu FOCS 2021] also obtained an improved algorithm for sparse mincost flows on capacitated graphs. Our algorithm implies a  $\tilde{O}(m^{3/2-1/58} \log U)$  time maxflow algorithm, improving over the  $\tilde{O}(m^{3/2-1/328} \log U)$  time maxflow algorithm of [Gao-Liu-Peng FOCS 2021].

---

\*vdbbrand@berkeley.edu. Simons Institute & UC Berkeley, USA.

†ygao380@gatech.edu. Georgia Institute of Technology, USA

‡jmbulapati@stanford.edu. Stanford University, USA.

§yintat@uw.edu. University of Washington and Microsoft Research Redmond, USA.

¶yangpliu@stanford.edu. Stanford University, USA.

||rpeng@cc.gatech.edu. Georgia Institute of Technology, USA & University of Waterloo, Canada.

\*\*sidford@stanford.edu. Stanford University, USA.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Key Algorithmic Pieces . . . . .	2
1.2	Related Work . . . . .	3
1.3	General Notation . . . . .	4
1.4	Organization . . . . .	4
<b>2</b>	<b>Overview</b>	<b>5</b>
2.1	Overview of Faster Schur Complements via the Morris walk . . . . .	5
2.2	Overview of Operator-based Electric Flow Heavy Hitters . . . . .	6
2.3	Overview of Reduction from Adaptive to Oblivious Adversaries . . . . .	7
2.4	Overview of IPM Outer Loop . . . . .	9
<b>3</b>	<b>Preliminaries</b>	<b>10</b>
<b>4</b>	<b>Improved Dynamic Schur Complements</b>	<b>11</b>
4.1	Approximate Random Walks with Morris Counters . . . . .	12
4.2	Improved Dynamic Schur Complement . . . . .	19
<b>5</b>	<b>Data Structures for Dynamic Electrical Flows</b>	<b>21</b>
5.1	Harmonic Extension . . . . .	21
5.2	Dynamic Potential Maintenance . . . . .	27
5.3	Dynamic Evaluator . . . . .	28
5.4	Dynamic Locator . . . . .	30
<b>6</b>	<b>Reducing Adaptive to Oblivious Adversaries</b>	<b>33</b>
6.1	Simulating Gaussian Error . . . . .	34
6.2	Recursive Simulation . . . . .	37
6.3	Proof of Theorem <a href="#">6.2</a> . . . . .	38
<b>7</b>	<b>Interior Point Method</b>	<b>39</b>
7.1	Robust IPM Framework . . . . .	40
7.2	Robust IPM Tools . . . . .	41
7.3	Robust IPM Implementation . . . . .	43
7.4	Efficient Solution Approximation . . . . .	45
7.5	Robust IPM Stability Bound . . . . .	48
<b>8</b>	<b>Final Runtime Bound</b>	<b>52</b>
8.1	Efficient Solution Maintenance . . . . .	53
8.2	Initial Point, Final Point, and Proof of Main Theorem . . . . .	55

# 1 Introduction

The maximum flow (maxflow) problem asks to route the maximum amount of flow between two vertices  $s$  and  $t$  in a directed graph  $G$  such that the amount of flow on every edge is at most its capacity. The more general minimum cost flow (mincost flow) problem asks to route a fixed demands in a directed graph  $G$  without sending more flow on any edge than its capacity, while minimizing a linear cost. Together these well-studied problems cover a wide range of combinatorial and numerical problems, including maximum cardinality bipartite matching, minimum  $s$ - $t$  cut, shortest paths in graphs with negative edge length, and optimal transport (see e.g. [BLN<sup>+</sup>20, BLL<sup>+</sup>21]).

While classical algorithms for these problems revolved around using augmenting paths or cycle primitives (such as blocking flows) [Kar73, ET75, GT90, GR98], the last decade has seen significant runtime improvements for maxflow and mincost flow in various settings based on *electrical flows*. For a graph  $G$  with vertex set  $V$  and edge set  $E$ , with edge resistances  $\mathbf{r} \in \mathbb{R}_{\geq 0}^E$ , and a demand vector  $\mathbf{d} \in \mathbb{R}^E$ , the electric flow on  $G$  is the flow that routes a fixed demand while minimizing the energy  $\sum_{e \in E} \mathbf{r}_e \mathbf{f}_e^2$ . Better maxflow algorithms have been given in several regimes using electrical flows and stronger primitives [AKPS19, KPSW19], including unit capacity graphs [Mađ13, Mađ16, CMSV17, LS20, KLS20, AMV20], approximate maxflow on undirected graphs [CKM<sup>+</sup>11, LRS13, She13, KLOS14, Pen16, She17, ST18], and dense graphs [LS19, BLN<sup>+</sup>20, BLL<sup>+</sup>21].

However, it has been particularly challenging to obtain running time improvements for solving mincost flow and maxflow to high precision in sparse capacitated graphs. Recently, [GLP21] gave an  $\tilde{O}(m^{1.5-1/328} \log U)$  time algorithm for sparse graphs with capacities bounded by  $U$ , the first improvement over  $\tilde{O}(m^{1.5} \log U)$  for maxflow on sparse graphs with arbitrary, polynomially bounded capacities. Their improvement involved an intricate interaction of dynamic data structures for electrical flows, a modification of the standard interior point method (IPM) outer loop which builds an maxflow via  $\sqrt{m}$  approximate electric flows [Kar84, Vai89], and sketching techniques. Additionally, issues relating to randomness in data structures and combinatorial reasoning about errors resulting from random walks required careful analysis that significantly increase the runtime and resulted in the small improvement over  $m^{1.5}$ .

Our main result is an algorithm that while has a similar high-level picture as [GLP21], significantly simplifies the major pieces described previously and their interactions. Specifically, we give a general purpose sketching tool for electrical flows, a graph theoretic (instead of algebraic) way of constructing the random walks at the core of the data structure, and handle randomness dependencies using ideas from differential privacy. Further, we provide modifications to prior robust interior point frameworks to do  $\ell_2$ -based recentering (Definition 7.6) within a robust IPM via additional spectral vertex sparsification techniques. As a result, we achieve a faster runtime than in [GLP21]. Also, as a result of our simplified electric flow data structure, our algorithm and IPM seamlessly extend to mincost flow. In contrast, the data structure complications in [GLP21] restricted their algorithm to be applied to maxflow. <sup>1</sup>

**Theorem 1.1.** *There is an algorithm which given any  $m$ -edge directed graph  $G$  with integral capacities in  $[1, U]$ , feasible demand vector  $\mathbf{d} \in \mathbb{Z}^E$ , and an integral cost vector  $\mathbf{c} \in [-U, U]^E$ , computes a flow  $\mathbf{f}$  that routes demand  $\mathbf{d}$ , satisfies the capacity constraints, and minimizes  $\mathbf{c}^\top \mathbf{f}$ . The algorithm succeeds whp. and runs in time  $\tilde{O}(m^{3/2-1/58} \log^2 U)$ .*

---

<sup>1</sup> [AMV21], in FOCS 2021, claims an improvement to sparse capacitated mincost flow in the title. A preprint was recently made available at <https://arxiv.org/abs/2111.10368v1>. The results in this paper were derived independently: we defer detailed comparisons to a future version.

Overall, this paper simplifies the key pieces of [GLP21] and, as a result, clarifies the important components used to achieve faster maxflow algorithms via dynamic electric flows. Further, we consider each of these pieces to be interesting in their own right: dynamic maintenance of Schur complements, sketching and maintenance of high energy edges in dynamic electric flows, and understanding reductions between adaptive and oblivious adversaries. Ultimately, this paper can be read independently of [GLP21] and the proofs are simpler and more natural in many cases.

## 1.1 Key Algorithmic Pieces

Here we cover the key algorithmic pieces underlying our algorithm. The first major piece is a faster algorithm for generating random walks of a fixed length from a vertex, a core primitive in all random-walk based approaches to dynamic electric flows [DGGP19, GLP21]. Our second key contribution is an algorithm for detecting large energy edges in electric flows for graphs with dynamically changing resistances and demands based on a direct reduction to dynamic spectral vertex sparsifiers (Schur complements). As our data structures for maintaining dynamic electric flows naturally work only against oblivious adversaries, we develop an approach that black-box reduces dynamic electric flows against adaptive adversaries to the same problem against oblivious adversaries, at the cost of a small runtime increase.

**Faster generation of random walks and Schur complements.** All previous algorithms for dynamic electric flows [DGGP19, GLP21] require dynamic maintenance of spectral vertex sparsifiers or Schur complements, which approximate the electric flow and potentials onto a smaller set of terminal vertices. The Schur complement is generated by sampling several random walks from vertices  $v$  with exit probabilities proportional to inverse resistances until the walk visited a fixed number  $L$  of distinct vertices, and by estimating the sum of resistances of edges along the walk.

Because there may be edges with very large or small resistances, a naïve simulation may get stuck for polynomially many steps. Consequently, [DGGP19] gave an algorithm for this based on taking high powers of the random walk matrix (which [GLP21] applied in a black-box fashion). This generated large factors in the runtime of the data structures. We give an approach to significantly speed up the sampling of vertices and length estimation by applying a *Morris counter* from the streaming/sketching literature [NY20], and reducing the problem to solving a sequence of electric flow computations (Laplacian systems) as opposed to the more expensive matrix multiplication operations of [DGGP19]. This significant runtime improvement immediately translates to our dynamic electric flow data structure described above, which directly uses dynamic Schur complements.

**Simplified electric flow heavy hitter.** To design our dynamic algorithm for detecting edges with large energy in dynamic electric flows, we maintain an  $\ell_2$  heavy hitter sketch of the electric flow vector. The algorithm of [GLP21] maintained this sketch by using a dynamic spectral vertex sparsifier or Schur complement, which approximates the electric flow and potentials on a smaller set of terminal vertices, and several random walks for “moving” the heavy-hitter sketch vector to the terminal set. This latter piece (maintaining random walks for moving the heavy hitter vector) introduced several complications into the analysis and generated a large overall running time for the data structure. On the other hand, our algorithm is more directly based on spectral approximations. In particular, we show how to dynamically maintain the result of moving the heavy hitter vector onto the terminal set by simply calling another dynamic Schur complement data structure, and carefully reasoning about spectral approximations to bound how that affects the resulting error.

**Simplified IPM outer loop.** As a result of our more linear algebraic approach to maintaining electric flows, our data structure for detecting large energy edges in electric flows works for both dynamic resistance changes and dynamic demands, while the algorithm of [GLP21] required restricting to only  $s$ - $t$  flows. Our generalization also allows us to use a more standard and efficient robust IPM (from [DLY21]) to implement the outer loop utilizing the data structure, while [GLP21] had to redesign the IPM to carefully only use  $s$ - $t$  electric flows to interact with their data structure. Our robust IPM implements an additional batching, or  $\ell_2$ -based recentering step, by computing the changes on a small subset of edges to higher accuracy by using spectral vertex sparsifiers.

**Black-box reduction from adaptive to oblivious adversaries.** As we are applying randomized data structure inside an algorithmic outer loop, their previous responses may affect future updates. This is referred to an *adaptive adversary* in the literature. On the other hand, our data structures which are based on random walks naturally only work against *oblivious adversaries*, where the input sequence does not depend on the outputs and randomness of the data structure. The algorithm of [GLP21] handled this issue in their data structures by carefully controlling the total number of adaptive phases of their algorithm before snapping back to a deterministic state.

Our approach on the other hand is more black-box, and gives a more general approach for converting data structures against oblivious adversaries to handle adaptive queries. We build a LOCATOR which returns a superset of edges with large energies, and several EVALUATORS with differing accuracy parameters which separately estimate the energies of the edges. By leveraging ideas from the Gaussian-mechanism from differential privacy [DR<sup>+</sup>14] we show how to apply the EVALUATOR data structures to simulate estimating adding Gaussian noise to the true energy vector that we wish to output. We simulate this by making several queries to the EVALUATORS, where we query the least accurate EVALUATORS most often, and only query more accurate EVALUATORS when the estimate of the energy vector is close to certain thresholds and we require finer estimates to decide how to round. Because we are simulating adding noise to the true output, the algorithm succeeds against an adaptive adversary.

## 1.2 Related Work

We briefly survey the lines of work most relevant to our results, and refer the reader to [GLP21] for more comprehensive discussion. Recently, [DGG<sup>+</sup>22] gave a mincost flow algorithm on planar graphs running in nearly linear time. Similar to our paper, it is based on the robust IPM framework of [DLY21] and dynamic Schur complements. However, [DGG<sup>+</sup>22] relies on the fact that the terminal set  $C$  is small due to the existence of planar separators, while our paper relies on the fact that  $C$  is slowly changing.

**Data structures for IPMs.** IPMs are a powerful framework which reduces linear programming with  $m$  variables to a sequence of  $\tilde{O}(\sqrt{m})$  linear system solutions [Ren88]. For maxflow and mincost flow, these linear systems correspond to computing electrical flows, and Daich-Spielman [DS08] leveraged this observation to give a  $\tilde{O}(m^{1.5} \log U)$  mincost flow algorithm. Recently, several works have leveraged the key fact dating back to early works of Karmarkar [Kar84] and Vaidya [Vai89] that the linear systems change slowly and only need to be solved approximately, both in the context of linear programs [LS15, CLS19, LSZ19, Bra20, BLSS20, JSWZ20, BLL<sup>+</sup>21, Bra21] and mincost flows [BLN<sup>+</sup>20, BLL<sup>+</sup>21, GLP21].

**Dynamic electrical flows.** Recent works applying dynamic data structures to IPMs for maxflow require maintaining various properties of electrical flows on dynamically changing graphs. The improvements on dense graphs [BLN<sup>+</sup>20, BLL<sup>+</sup>21] required dynamically maintaining spectral sparsifiers of the Laplacian in  $\tilde{O}(1)$  time per edge update and  $\tilde{O}(n)$  per query, as well as detecting edges with large electrical energies in  $\tilde{O}(n)$  time per query. Both of these pieces were done using dynamic expander decompositions [NS17, Wul17, NSW17, SW19, CGL<sup>+</sup>20, BBG<sup>+</sup>20]. The work of [GLP21] desired *sublinear* time per query and hence required dynamically maintaining Schur complements, whose study was initiated in [DGGP19] to dynamically maintain approximate effective resistances.

**Adaptivity and differential privacy.** There has been significant work towards building techniques to apply oblivious data structures in the context of an algorithmic outer loop, which requires adaptivity. To date, most approaches to this problem involve either making the algorithm deterministic [BC16, BC17, CK19, GWN20, Chu21, BGS21], or resparsifying [BBG<sup>+</sup>20], both of which heavily leverage properties provided by dynamic expander decompositions [NS17, Wul17, NSW17, SW19, CGL<sup>+</sup>20, BBG<sup>+</sup>20]. Our work takes a different perspective and instead more carefully analyzes whether the adversary can learn any randomness leaked from the distribution of our output vector. This perspective is motivated by ideas from differential privacy, and in fact our key result is an adaptation of the Gaussian mechanism [DR<sup>+</sup>14] which simulates adding unbiased Gaussian noise to the true output vector by using a sequence of oblivious estimates. Our recursive scheme is also broadly related to the idea of multilevel Monte Carlo [Gil15, BG15] and its recent applications in leveraging approximate optimization procedures to obtain nearly unbiased estimates of minimizers [ACJ<sup>+</sup>21].

### 1.3 General Notation

We use plaintext to denote scalars, bold lowercase for vectors, and bold uppercase for matrices. For resistances  $\mathbf{r}$  and conductances  $\mathbf{w} \stackrel{\text{def}}{=} \mathbf{r}^{-1}$ , the corresponding capital matrices are diagonal matrices with the vector entries on the diagonal, i.e.  $\mathbf{R} \stackrel{\text{def}}{=} \text{diag}(\mathbf{r})$  and  $\mathbf{W} \stackrel{\text{def}}{=} \text{diag}(\mathbf{w})$ . As our algorithm heavily use approximations, we will use  $\tilde{\cdot}$  to denote the approximate versions of true variables.

We use  $\tilde{O}(\cdot)$  to suppress logarithmic factors in  $m$  and  $\tilde{\Omega}(\cdot)$  to suppress inverse logarithmic factors in  $m$ . For vectors  $\mathbf{x}, \mathbf{y}$  we sometimes let  $\mathbf{x}\mathbf{y}$  denote the entry-wise product of  $\mathbf{x}, \mathbf{y}$ , so  $(\mathbf{x}\mathbf{y})_i \stackrel{\text{def}}{=} x_i y_i$ . Similarly, we let  $(\mathbf{x}/\mathbf{y})_i \stackrel{\text{def}}{=} x_i/y_i$ . We say that an event holds with high probability (whp.) if for any constant  $C > 0$ , the event succeeds with probability at least  $1 - n^{-C}$  by adjusting parameters. We let  $[n] = \{1, 2, \dots, n\}$ . We denote the (unweighted) degree of a vertex  $v$  as  $\deg(v)$ .

We say that a symmetric matrix  $\mathbf{M} \in \mathbb{R}^{n \times n}$  is positive semidefinite (PSD) if  $\mathbf{x}^\top \mathbf{A} \mathbf{x} \geq 0$  for all  $\mathbf{x} \in \mathbb{R}^n$ . For PSD matrices  $\mathbf{A}, \mathbf{B}$  we write  $\mathbf{A} \preceq \mathbf{B}$  if  $\mathbf{B} - \mathbf{A}$  is PSD. For positive real numbers  $a, b$  we write  $a \approx_\gamma b$  to denote  $\exp(-\gamma)b \leq a \leq \exp(\gamma)b$ . For PSD matrices  $\mathbf{A}, \mathbf{B}$  we write  $\mathbf{A} \approx_\gamma \mathbf{B}$  if  $\exp(-\gamma)\mathbf{B} \preceq \mathbf{A} \preceq \exp(\gamma)\mathbf{B}$ .

### 1.4 Organization

In Section 2, we give a technical overview of each of our improvements to each of the key components of [GLP21]: faster sampling of Schur complements, operator-based electric flow heavy hitters, and black-box reduction of adaptive to oblivious adversaries. We also overview the robust IPM we use. In Section 3 we give preliminaries for maxflow, mincost flow, and electric flows that we require for the remainder of our paper. In Section 4 we give our algorithm for faster sampling of random walks

and Schur complements, and we combine this with an operator-based heavy hitter in Section 5 to give a faster algorithm for detecting edges with large energy. In Section 6 we show how to black-box reduce adaptive to oblivious adversaries for the problem of estimating dynamic vectors. We give our robust IPM in Section 7, which is an adaptation of that in [DLY21], and additional tools to apply it. Finally, we combine all pieces and compute the final runtime in Section 8.

## 2 Overview

Here we provide a technical overview of our contributions.

### 2.1 Overview of Faster Schur Complements via the Morris walk

Our data structures, as in [GLP21], heavily rely on dynamically maintaining spectral vertex sparsifiers (Schur complements) of  $G$ , which approximate the inverse spectral form of  $G$  onto a subset of the vertices. This was achieved using the algorithm of [DGGP19], which showed how to dynamically maintain an approximate Schur complement under edge resistance updates. The main primitive behind the dynamic Schur complement data structure was a procedure to sample random walks from a vertex with exit probability proportional to inverse resistances, i.e. the probability of going from a vertex  $v$  to a neighbor  $u$  is given by

$$\frac{r_{vu}^{-1}}{\sum_{w \text{ neighbor of } v} r_{vw}^{-1}}.$$

For this walk and a parameter  $L$ , we must run the walk until the total degree of visited vertices is  $L$ , and to estimate the total *resistive length* of the walk up to a  $(1 + \epsilon)$ -factor, where resistive length refers to the sum of resistances of edges on the walk. Directly simulating the random walk is not efficient enough, because there may be polynomially large and small resistances, which cause the walks to “get stuck” on a small set of edges, without visiting new vertices. Thus it can take a long time to visit  $L$  distinct vertices. Despite this, [DGGP19] showed how to sample the walk and resistive length in  $\tilde{O}(L^4 \epsilon^{-2})$  time per vertex, and this large runtime directly led to the fact that [GLP21] only achieved a small 1/328 improvement in the exponent.

Interestingly, if one is only interested in obtaining distinct vertices on the walk (and not the resistive length) until the total degree is  $L$ , this can be done in  $\tilde{O}(L^2)$  time by solving a Laplacian system (corresponding to computing an electric flow) for each of at most  $L$  steps to compute the next exit vertex. However, the approach of [DGGP19] which also computed the resistive length, i.e. the sum of resistances of edges on the walk, was based on matrix-powering/matrix multiplication, and instead had a larger  $\tilde{O}(L^4 \epsilon^{-2})$  runtime. Our main idea is to resolve this runtime discrepancy between sampling the distinct vertices and computing a  $(1 + \epsilon)$ -resistive length estimate by giving an algorithm that computes both quantities by solving a sequence of Laplacian systems. In total, we solve  $\tilde{O}(L + \epsilon^{-2})$  systems, for a runtime of  $\tilde{O}(L^2 + L \epsilon^{-2})$ . In our settings,  $L$  will generally be  $\Omega(\epsilon^{-2})$ , so our runtime is  $\tilde{O}(L^2)$ , matching the time to generate the first  $L$  vertices using a sequence of Laplacian systems, and significantly improving over the  $\tilde{O}(L^4 \epsilon^{-2})$  runtime of [DGGP19, GLP21].

Our algorithm for this task is derived from the Morris counter [MS78, NY20], a probabilistic algorithm for maintaining low-space approximations to a counter  $N$  undergoing increments. For simplicity of exposition, we assume our input graph has integer, polynomially-bounded edge weights. Our algorithm intuitively begins by running a random walk in  $G$ . However, we replace the naive



procedure for computing the resistive length with a Morris counter. More precisely, assume we have run a random walk starting from a vertex  $u$  for  $k$  steps and have estimated the resistive length of the walk via the Morris counter. To estimate the resistive length of this walk after a further step, we simply sample one new step of the walk: if we sampled an edge of resistance length  $w$ , we increment the Morris counter  $w$  times. In this way, the Morris counter enables us to maintain estimates of the resistive length of a random walk.

We use the following properties of Morris counters as shown in [NY20]. First, they take discrete values of  $\frac{1}{a}((1+a)^i - 1)$  for some real number  $a > 0$  and integers  $i \geq 0$ , and if  $a = \epsilon^2/\text{poly log } n$ , the value of the Morris counter is always a  $(1+\epsilon)$ -approximation of the true value with high probability. In particular, for graphs with polynomially bounded weights, the Morris counter takes at most  $\tilde{O}(a^{-1}) = \tilde{O}(\epsilon^{-2})$  distinct values.

Our key insight we can simulate incrementing of the Morris counter and the sequence of vertices visited as a random walk on a “lifted” graph with  $\tilde{O}(\epsilon^{-1})$  layers. Each time we explore a “new” neighbor in this lifted space, we either find a new unexplored vertex along the random walk or increment the Morris counter. However, there are only  $\tilde{O}(\epsilon^{-2})$  distinct values of the counter: thus we must explore only an additional  $\tilde{O}(\epsilon^{-2})$  distinct vertices in this lifted space to obtain the desired guarantee on the number of new vertices seen. We obtain our final algorithm by replacing the explicit random walks with a subroutine based on Laplacian linear system solvers: in this way our final complexity of  $\tilde{O}(L^2 + L\epsilon^{-2})$  follows. Overall, this provides a graph-theoretic approach for estimating lengths of random walks on graphs, as opposed to the previous algorithm in [DGGP19] which was based on matrix multiplication.

## 2.2 Overview of Operator-based Electric Flow Heavy Hitters

Our next major improvement over [GLP21] is a data structure that detects large energy edges in electric flows on graphs with dynamic resistances and demands by direct reduction to maintaining dynamic Schur complements. To be precise, we give a data structure that on a graph  $G$  with dynamically changing resistances and demands, solves a *electric flow heavy hitter* problem, by returning a set  $S$  of  $O(\epsilon^{-2})$  edges containing all edges  $e$  with at least  $\epsilon^2$  fraction of the electric energy, i.e.  $\mathbf{r}_e \mathbf{f}_e^2 \geq \epsilon^2 \sum_{e \in E} \mathbf{r}_e \mathbf{f}_e^2$  where  $\mathbf{f}$  is the electric flow vector. [GLP21] gave a data structure that solved this problem in sublinear time per resistance update and query as a core piece of their algorithm. We give improved runtimes for solving this problem and shed light on its complexity by directly reducing to dynamically maintaining Schur complements.

Note that the dynamic electric flow heavy hitter problem is equivalent to detecting large coordinates of the vector  $\mathbf{R}^{1/2} \mathbf{f}$  compared to its  $\ell_2$  norm. Hence, it is natural to apply an  $\ell_2$  heavy-hitter sketch [KNPW11], which at a high-level consists of  $\tilde{O}(\epsilon^{-2})$  Johnson-Lindenstrauss  $\ell_2$  sketches. In total, this consists of maintaining the value of  $\mathbf{q}^\top \mathbf{R}^{1/2} \mathbf{f}$  for  $\tilde{O}(\epsilon^{-2})$  random sketch vectors  $\mathbf{q} \in \{-1, 0, 1\}^E$ . The flow  $\mathbf{f}$  can be represented as  $\mathbf{f} = \mathbf{R}^{-1} \mathbf{B} \phi$  for electric potentials  $\phi$  and edge-vertex incidence matrix  $\mathbf{B}$ , so

$$\mathbf{q}^\top \mathbf{R}^{1/2} \mathbf{f} = \langle \mathbf{B}^\top \mathbf{R}^{-1/2} \mathbf{q}, \phi \rangle.$$

Let  $\mathbf{y} = \mathbf{B}^\top \mathbf{R}^{-1/2} \mathbf{q}$ , so that we focus on maintaining  $\mathbf{y}^\top \phi$ . However,  $\phi$  is still a  $|V|$ -dimensional vector, so in order to achieve sublinear time [GLP21] used a smaller terminal set  $C$  to estimate  $\mathbf{y}^\top \phi$ . In particular, they write  $\phi = \mathcal{H}_C \phi_C$  where  $\phi_C$  is the restriction of  $\phi$  to  $C$ , and  $\mathcal{H}_C \in \mathbb{R}^{V(G) \times C}$  is the *harmonic extension* (Definition 5.1) operator which extends  $\phi_C$  to  $\phi$  using that for any



vertex  $v$ ,  $\phi_v$  is the average of its neighbors, weighted proportional to inverse resistances. This way,  $\mathbf{y}^\top \phi = \langle \mathcal{H}_C^\top \mathbf{y}, \phi_C \rangle$ . Assuming that we can approximate maintain  $\phi_C$  (which we discuss towards the end of this section’s overview), it suffices to maintain  $\mathcal{H}_C^\top \mathbf{y}$ .

Our major difference from [GLP21] is in how we maintain  $\mathcal{H}_C^\top \mathbf{y}$ . While [GLP21] used the combinatorial interpretation of the operator  $\mathcal{H}_C^\top$  as using random walks to “move” the mass from vector  $\mathbf{y}$  onto  $C$ , we use the spectral fact (Lemma 5.2) that

$$\mathcal{H}_C = \mathbf{L}^\dagger \begin{bmatrix} 0 & \mathbf{SC}(\mathbf{L}, C) \end{bmatrix},$$

where  $\mathbf{L}$  is the graph Laplacian and  $\mathbf{SC}(\mathbf{L}, C)$  is the Schur complement of  $\mathbf{L}$  onto  $C$ . Thus, we get

$$\mathcal{H}_C^\top \mathbf{y} = \begin{bmatrix} 0 \\ \mathbf{SC}(\mathbf{L}, C) \end{bmatrix} \mathbf{L}^\dagger \mathbf{y}.$$

Thus, we could optimistically precompute  $\mathbf{L}^\dagger \mathbf{y}$  and then compute  $\mathcal{H}_C^\top \mathbf{y}$  as long as we can dynamically maintain the Schur complement  $\mathbf{SC}(\mathbf{L}, C)$ , which is a size  $C$  object. The remaining issue is that the Laplacian  $\mathbf{L}$  may change because the resistances change. However, the operator  $\mathcal{H}_C^\top$  does not depend on the resistances of edges completely inside  $C$  (by definition), so we may actually let  $\mathbf{L}$  be the Laplacian of the *original graph* as long as all endpoints of edges with resistance changes are added to  $C$ . Finally we are able to show that using an approximate Schur complement in place of  $\mathbf{SC}(\mathbf{L}, C)$  still suffices for our data structures (Lemma 5.7).

Finally we discuss the (approximate) maintenance of the potential  $\phi_C$ . For an electric flow  $\mathbf{f}$  routing demand  $\mathbf{d}$ , i.e.  $\mathbf{B}^\top \mathbf{f} = \mathbf{d}$ , the potentials  $\phi_C$  are given by

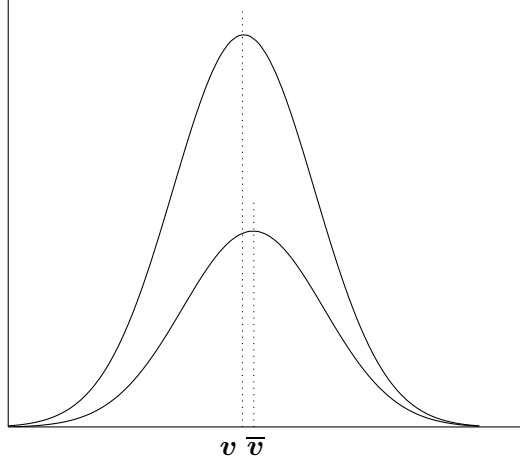
$$\phi_C = \mathbf{SC}(\mathbf{L}, C)^\dagger \mathcal{H}_C^\top \mathbf{d}.$$

In other words, we first “move” the demands to the terminal set using  $\mathcal{H}_C^\top$  just as above, and then invert the Schur complement on it. Therefore we can maintain  $\phi_C$  as follows: maintain  $\mathcal{H}_C^\top \mathbf{d}$  approximately as above, and then also approximate maintain  $\mathbf{SC}(\mathbf{L}, C)$  using an approximate Schur complement data structure. In all, this reduces the maintenance of the heavy hitter vector to three calls to an approximate Schur complement oracle.

### 2.3 Overview of Reduction from Adaptive to Oblivious Adversaries

The dynamic electric flow data structures built in Section 2.2 naïvely only work against oblivious adversaries, i.e. the inputs must be independent of the outputs and randomness of the data structure. In [GLP21] this was handled by carefully designing the data structures to utilize the fact that the IPM central path is a deterministic object. However, we take a more general approach, by applying ideas from the Gaussian-mechanism from differential privacy [DR<sup>+</sup>14] to build versions of these data structures that work directly against adaptive adversaries, allowing them to be applied within the interior point outer loop. In fact, we give a generic reduction for estimating vectors against adaptive adversaries to oblivious adversaries.

Consider an oblivious data structure that outputs vectors  $\bar{\mathbf{v}} \in \mathbb{R}^m$  that are supposed to approximate a true underlying vector  $\mathbf{v} \in \mathbb{R}^m$ . In our dynamic electric flow setting, this corresponds to a data structure which detects edges with large electric energy, and approximates their flow values. Consider what would happen if instead of  $\bar{\mathbf{v}}$ , our algorithm uses  $\mathbf{z} \sim \mathcal{N}(\mathbf{v}, \sigma^2)$  for small enough  $\sigma$  (i.e. the vector  $\mathbf{v}$  with some Gaussian noise added to it). If  $\sigma$  is small enough, then  $\mathbf{z}$  would be



**Figure 1:** Density function  $d$  of  $\mathcal{N}(\mathbf{v}, \sigma^2)$ , and density function  $\bar{d}$  of  $\mathcal{N}(\bar{\mathbf{v}}, \sigma^2)$  scaled by some  $\exp(-\alpha)$ ,  $\alpha > 0$  so that  $\bar{d}(\mathbf{x}) \exp(-\alpha) \leq d(\mathbf{x})$ .

an accurate approximation for our algorithm to work. Additionally, the vector  $\mathbf{z}$  obviously does not depend on the internal randomness of the data structure, since it is defined with respect to  $\mathbf{v}$ , not the approximation  $\bar{\mathbf{v}}$ . Unfortunately, computing  $\mathbf{z}$  by computing  $\mathbf{v}$  and adding noise is rather inefficient since  $\mathbf{v}$  is the exact solution, not an approximation. We now explain how to obtain vector  $\mathbf{z}$  more efficiently from oblivious estimate vectors  $\bar{\mathbf{v}}$  by using the Gaussian-mechanism from differential privacy [DR<sup>+</sup>14].

Specifically, it is known that for any  $\sigma > 0$  there is small enough  $\alpha > 0$  such that if  $d$  is the density function of  $\mathcal{N}(\mathbf{v}, \sigma^2)$  and  $\bar{d}$  is the density function of  $\mathcal{N}(\bar{\mathbf{v}}, \sigma^2)$ , then  $\bar{d}(\mathbf{x}) \leq \exp(\alpha)d(\mathbf{x})$  for all  $\mathbf{x}$ .<sup>2</sup> For example, Figure 1 shows density function  $d(\mathbf{x})$  and the scaled density function  $\exp(-\alpha) \cdot \bar{d}(\mathbf{x})$  for the 1-dimensional case. Note that we can pick a random  $\mathbf{z} \sim \mathcal{N}(\mathbf{v}, \sigma^2)$  by picking uniformly at random a point below the curve of  $d(\bar{\mathbf{x}})$  and returning the  $x$ -coordinate. We can also split this sampling scheme into two phases: (i) With probability  $1 - \exp(-\alpha)$ , sample from the area between the two curves. (ii) Alternatively, with probability  $\exp(-\alpha)$  sample from the area below the bottom curve  $\exp(-\alpha)\bar{d}(\mathbf{x})$  in Figure 1.

When case (i) happens, we handle it directly by computing  $\mathbf{v}$  exactly (which is expensive), which gives us the distributions  $d$  and  $\bar{d}$  explicitly. However, note that if  $\alpha$  is close to 0, then this case only occurs infrequently: with probability  $1 - \exp(-\alpha) = O(\alpha)$ , which balances out the expensive cost of computing  $\mathbf{v}$ . On the other hand, case (ii), which occurs with probability  $\exp(-\alpha)$ , corresponds to flipping an unbalanced coin and with probability  $\exp(-\alpha)$  we sample a  $\mathbf{z}' \sim \mathcal{N}(\bar{\mathbf{v}}, \sigma^2)$ . So with probability  $\exp(-\alpha)$  we do not need to know/compute the exact vector  $\mathbf{v}$  in order to obtain a sample with distribution  $\mathcal{N}(\mathbf{v}, \sigma^2)$  and just knowing the approximate result  $\bar{\mathbf{v}}$  already suffices.

Now, this scheme can be extended recursively to handle case (ii), i.e. sampling from  $\mathbf{z} \sim \mathcal{N}(\bar{\mathbf{v}}, \sigma^2)$ . We can use the same scheme again via some  $\bar{\mathbf{v}}'$ , i.e. sampling from  $\mathcal{N}(\bar{\mathbf{v}}', \sigma^2)$  with probability  $\exp(-\alpha)$  instead of  $\mathcal{N}(\bar{\mathbf{v}}, \sigma^2)$ . This leads to another speed-up because of the following reason: the probability  $\exp(-\alpha)$  depends on the approximation quality of  $\bar{\mathbf{v}}'$  compared to  $\mathbf{v}$ . We want to use a large  $\alpha$  in order to reduce the probability of computing  $\mathbf{v}$ , but this requires  $\bar{\mathbf{v}}'$  to be a better approximation. Thus, we are able to compute higher accuracy approximations (which take

<sup>2</sup>This is actually only true for  $\mathbf{x} \in D$  for some event  $D$  that holds whp. We ignore this here for simplicity.

more runtime) less frequently, and this leads to a speedup. Overall, by using this scheme, our data structures will work against an adaptive adversary because the output has distribution  $\mathcal{N}(\mathbf{v}, \sigma^2)$ , i.e. a distribution that is independent of the internal randomness of the data structures.

## 2.4 Overview of IPM Outer Loop

Here we overview how we apply the above primitives in a robust IPM to give an algorithm for algorithm, which reduces solving maxflow to computing a sequence of  $\tilde{O}(\sqrt{m})$  approximate electric flows. The IPM of [GLP21] required several nonstandard modifications, including restricting to using  $s$ - $t$  flows, which resulted in using more than  $\tilde{O}(\sqrt{m})$  steps, and overall higher runtime. On the other hand, our algorithm is based on the more standard robust IPM of [DLY21], with an additional procedure that allows for recentering in the context of a robust IPM that allows us to control errors that accumulate over longer periods of time.

We start by briefly introducing a standard robust IPM setup for the mincost flow problem based on [DLY21] (in Section 7 we change notation slightly to work with general linear programs)

$$\min_{\mathbf{f} \in \mathbb{R}^m: \mathbf{B}^\top \mathbf{f} = \mathbf{d} \text{ and } \ell \leq \mathbf{f} \leq \mathbf{u}} \mathbf{c}^\top \mathbf{f}, \quad (1)$$

where  $\mathbf{c} \in \mathbb{R}^E$  is the cost vector, and  $\ell, \mathbf{u} \in \mathbb{R}^E$  are lower/upper capacities on edges. For  $e \in E$  and real number  $f \in \mathbb{R}$ , define the logarithmic barrier function  $\phi_e(f) \stackrel{\text{def}}{=} -\log(f - \ell_e) - \log(\mathbf{u}_e - f)$ , and for flow  $\mathbf{f} \in \mathbb{R}^E$  define  $\phi(\mathbf{f}) \stackrel{\text{def}}{=} \sum_{e \in E} \phi_e(\mathbf{f}_e)$ . For a path parameter  $\mu$  that decreases towards 0 over the course of  $\tilde{O}(\sqrt{m})$  steps, the robust IPM maintains an approximate minimizer to the expression

$$\mathbf{f}_\mu \stackrel{\text{def}}{=} \min_{\mathbf{f} \in \mathbb{R}^m: \mathbf{B}^\top \mathbf{f} = \mathbf{d} \text{ and } \ell \leq \mathbf{f} \leq \mathbf{u}} \mathbf{c}^\top \mathbf{f} + \mu \phi(\mathbf{f}). \quad (2)$$

Since  $\phi$  is convex, the KKT conditions for (2) give that there is a vector  $\mathbf{y}$  such that  $\mathbf{c} + \mu \nabla \phi(\mathbf{f}_\mu) = \mu \mathbf{B} \mathbf{y}$ . Thus there is a vector  $\mathbf{s}_\mu \in \mathbb{R}^E$  such that  $\mathbf{B} \mathbf{y} + \mathbf{s}_\mu = \mathbf{c}/\mu$  and  $\mathbf{s}_\mu + \nabla \phi(\mathbf{f}_\mu) = 0$ . In this way, we define a  $\mu$ -centered point as a pair  $(\mathbf{f}, \mathbf{s})$  such that  $\mathbf{B} \mathbf{y} + \mathbf{s} = \mathbf{c}/\mu$  for some  $\mathbf{y} \in \mathbb{R}^V$  and  $\|\nabla^2 \phi(\mathbf{f})^{-1/2}(\mathbf{s} + \nabla \phi(\mathbf{f}))\|_\infty \leq 1/64$ . The robust IPM maintains  $\mu$ -centered points throughout by tracking the potential function

$$\sum_{e \in E} \cosh \left( \lambda \phi_e''(\mathbf{f}_e)^{-1/2} (\mathbf{s}_e + \phi_e'(\mathbf{f}_e)) \right)$$

for  $\lambda = 128 \log(16m)$ . Now IPM steps are taken to simulate gradient descent steps on the potential to keep it small, and hence maintain  $\mu$ -centered points at all times.

Because our data structures work in time sublinear in the number of vertices, the flows we maintain during the robust IPM are stored implicitly, even without the ability to query in  $\tilde{O}(1)$  time the “true flow” on an edge  $e$ . Further, the error of our flow estimate from the true value accumulates over the steps of our method. Hence we require a procedure to recompute a feasible  $\mu$ -centered flow in a robust IPM every  $k$  steps in  $\tilde{O}(m)$  time for some  $k = m^{\Omega(1)}$  (Theorem 7.7). To see why this could be possible, note that the true step per iteration is an electric flow with some resistances and demands. Additionally, over the course of  $k$  steps, these resistances and demands will only change at most  $\text{poly}(k)$  total times. Thus, we can put all edges whose resistance or demand changed into a terminal set  $C$  and compute an  $\epsilon$ -approximate Schur complement onto  $C$ . [LS18] shows that such a Schur complement (onto a slightly larger set) can be computed in time

$\tilde{O}(m + |C|/\epsilon^2) = \tilde{O}(m + \text{poly}(k)/\epsilon^2) = \tilde{O}(m)$  for some  $k = m^{\Omega(1)}$ . Leveraging this, we show that we can recover a centered point in the context of a robust IPM in  $\tilde{O}(m)$  time every  $k$  steps.

Overall, our algorithm splits the  $\tilde{O}(\sqrt{m})$  robust IPM steps in  $\tilde{O}(\sqrt{m}/k)$  batches of  $k$  steps. Within each batch, we ensure that at most  $\text{poly}(k)$  edges have their resistances change in the graph  $G$  (but there may be more resistance updates in between batches). Each step in the batch is maintained using the dynamic electric flow heavy hitter data structure we built, as described in Sections 2.1 to 2.3. At the end of each batch, we use the approximate recentering procedure described in the previous paragraph. Combining these pieces along with the standard bound that over  $T$  IPM steps, at most  $\tilde{O}(T^2)$  resistances change by a constant factor, gives our final runtimes.

### 3 Preliminaries

We give preliminaries on maxflow, mincost flow, electric flows, and Schur complements.

**Maxflow and mincost flow.** Throughout, we let  $G = (V, E)$  be our graph with  $n = |V|$  vertices and  $m = |E|$  edges. We let  $\mathbf{B} \in \mathbb{R}^{E \times V}$  denote the edge-vertex incidence matrix of  $G$ . Additionally, we let  $\ell, \mathbf{u} \in \mathbb{Z}^E$  denote the lower/upper capacities on edges in  $G$ . We assume that  $\|\ell\|_\infty, \|\mathbf{u}\|_\infty \leq U$ . A flow  $\mathbf{f} \in \mathbb{R}^E$  is any assignment of real numbers of edges of  $G$ . We say that a flow  $\mathbf{f}$  is *feasible* if  $\ell_e \leq \mathbf{f}_e \leq \mathbf{u}_e$  for all  $e \in E$ . We say that  $\mathbf{f}$  routes the demand  $\mathbf{d} \in \mathbb{R}^V$  if  $\mathbf{B}^\top \mathbf{f} = \mathbf{d}$ .

The maximum flow problem asks to find a feasible flow routing the maximum multiple of a demand  $\mathbf{d}$  (generally assumed to be  $s$ - $t$ ). Written linear algebraically, this asks to find the largest  $F^*$  such that there is a flow  $\mathbf{f}$  satisfying  $\mathbf{B}^\top \mathbf{f} = F^* \mathbf{d}$  and  $\ell_e \leq \mathbf{f}_e \leq \mathbf{u}_e$  for all  $e \in E$ . The minimum cost flow problem asks to minimize a linear cost  $\mathbf{c}$  over flows routing a fixed demand  $\mathbf{d}$ . Linear algebraically, this can be written as

$$\min_{\substack{\mathbf{B}^\top \mathbf{f} = \mathbf{d} \\ \ell_e \leq \mathbf{f}_e \leq \mathbf{u}_e \text{ for all } e \in E}} \mathbf{c}^\top \mathbf{f}.$$

We work with mincost flow throughout, as it is known to generalize maxflow. We also focus on finding high-accuracy solutions in runtime depending logarithmically on  $U$  and  $\|\mathbf{c}\|_\infty$ , as it is known that this suffices to get an exact solution with linear time overhead [DS08, BLL<sup>+</sup>21].

**Electric flows and Schur complements.** Electric flows are  $\ell_2$ -minimization analogues of maxflow on undirected graphs, and are used in all current state-of-the-art high accuracy maxflow algorithms [KLS20, AMV20, BLN<sup>+</sup>20, BLL<sup>+</sup>21, GLP21] based on IPMs. On a graph  $G$  with resistances  $\mathbf{r}$ , the electric flow routing demand  $\mathbf{d}$  is given by

$$\arg \min_{\mathbf{B}^\top \mathbf{f} = \mathbf{d}} \sum_{e \in E} \mathbf{r}_e \mathbf{f}_e^2. \quad (3)$$

The minimizer in (3) is given by the solution to a linear system:  $\mathbf{f} = \mathbf{R}^{-1} \mathbf{B} (\mathbf{B}^\top \mathbf{R}^{-1} \mathbf{B})^\dagger \mathbf{d}$ . The matrix  $\mathbf{B}^\top \mathbf{R}^{-1} \mathbf{B}$  is known as the *Laplacian* of  $G$ , which can be solved in nearly-linear time [ST04, KMP10, KMP11, KOSA13, LS13, CKM<sup>+</sup>14, KLP<sup>+</sup>16, KS16, JS21]. Precisely, solving a Laplacian system gives high accuracy *vertex potentials*, defined as  $\phi = (\mathbf{B}^\top \mathbf{R}^{-1} \mathbf{B})^\dagger \mathbf{d}$ .

**Theorem 3.1.** *Let  $G$  be a graph with  $n$  vertices and  $m$  edges. Let  $\mathbf{r} \in \mathbb{R}_{>0}^E$  denote edge resistances. For any demand vector  $\mathbf{d}$  and  $\epsilon > 0$  there is an algorithm which computes in  $\tilde{O}(m \log \epsilon^{-1})$  time*

potentials  $\phi$  such that  $\|\phi - \phi^*\|_{\mathbf{L}} \leq \epsilon \|\phi^*\|_{\mathbf{L}}$ , where  $\mathbf{L} = \mathbf{B}^\top \mathbf{R}^{-1} \mathbf{B}$  is the Laplacian of  $G$ , and  $\phi^* = \mathbf{L}^\dagger \mathbf{d}$  are the true potentials determined by the resistances  $\mathbf{r}$ .

For notational convenience, we define the *conductances*  $\mathbf{w} \stackrel{\text{def}}{=} \mathbf{r}^{-1}$ , and let  $\mathbf{L}(\mathbf{w}) \stackrel{\text{def}}{=} \mathbf{B}^\top \mathbf{W} \mathbf{B}$ .

Several of our algorithms want to solve Laplacian systems in *sublinear time*. This can be done in the following natural sense: instead of returning the full potential vector  $\phi$ , we only wish to determine  $\phi$  restricted to a subset of vertices  $C \subseteq V$ . This is captured by a *Schur complement*, which is defined as  $\mathbf{SC}(\mathbf{L}, C) \stackrel{\text{def}}{=} \mathbf{L}_{CC} - \mathbf{L}_{CF} \mathbf{L}_{FF}^{-1} \mathbf{L}_{FC}$ , where  $F = V \setminus C$  and  $\mathbf{L}_{FF}, \mathbf{L}_{CF}, \mathbf{L}_{FC}, \mathbf{L}_{CC}$  are blocks of the Laplacian  $\mathbf{L}$  corresponding to rows/columns in  $F, C$ . Schur complements satisfy two key properties which are essential for our algorithm: they are also graph Laplacians, and they are directly related to  $\mathbf{L}^\dagger$  via the Cholesky factorization.

**Lemma 3.2** (Cholesky factorization). *For a connected graph  $G$  with Laplacian  $\mathbf{L} \in \mathbb{R}^{V \times V}$ , subset  $C \subseteq V$ , and  $F \stackrel{\text{def}}{=} V \setminus C$ ,*

$$\mathbf{L}^\dagger = \begin{bmatrix} \mathbf{I} & -\mathbf{L}_{FF}^{-1} \mathbf{L}_{FC} \\ 0 & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{L}_{FF}^{-1} & 0 \\ 0 & \mathbf{SC}(\mathbf{L}, C)^\dagger \end{bmatrix} \begin{bmatrix} \mathbf{I} & 0 \\ -\mathbf{L}_{CF} \mathbf{L}_{FF}^{-1} & \mathbf{I} \end{bmatrix}.$$

The matrix  $\begin{bmatrix} -\mathbf{L}_{FF}^{-1} \mathbf{L}_{FC} \\ \mathbf{I} \end{bmatrix}$  appearing in the Cholesky factorization corresponds to mapping the potentials on  $C$  back to the whole graph via a *harmonic extension*. In other words, a random walk on  $G$ , with exit probabilities proportional to conductances is a martingale on potentials. We give a more formal definition and properties later in Section 5.

Finally, it is very useful intuition that electric flows are inherently connected with the following random walk on  $G$ : a vertex  $v$  goes to a neighbor  $u$  with probability proportional to conductance (inverse resistances), i.e.  $\frac{w_{uv}}{\sum_{w \in N(v)} w_{vw}}$ , where  $N(v)$  are the neighbors of  $v$  in  $G$ . This random walk is the one used to define the harmonic extension, and also is used more directly in our algorithm for sampling Schur complements (see Lemma 4.11). Throughout, any mention of random walks refers to this random walk.

## 4 Improved Dynamic Schur Complements

In this section, we give our main algorithm for maintaining a Schur complement in a dynamic graph. Our main result is the following (see Theorem 4.10 for a more precise statement):

**Theorem 4.1** (Dynamic Schur complement (informal)). *There is a data structure that supports the following operations against oblivious adversaries given a graph  $G = (V, E)$  with dynamic edge conductances  $\mathbf{w} \in \mathbb{R}^{E(G)}$  and parameters  $\beta < \epsilon^2 < 1$ .*

- *INITIALIZE( $G, \mathbf{w}, \epsilon, \beta$ ). Initializes the data structure with accuracy parameter  $\epsilon$ , and chooses a set of  $O(\beta m)$  terminals  $C$ .  $\overline{\mathbf{w}}$  is initialized as  $\mathbf{w}$ . Runtime:  $\tilde{O}(m\beta^{-2}\epsilon^{-2})$ .*
- *ADDTERMINAL( $v$ ). Makes  $v$  a terminal, i.e.  $C \leftarrow C \cup \{v\}$ . Runtime: amortized  $\tilde{O}(\beta^{-2}\epsilon^{-2})$ .*
- *UPDATE( $e, \overline{\mathbf{w}}^{\text{new}}$ ). Under the guarantee that both endpoints of  $e$  are terminals in  $C$ , updates  $\overline{\mathbf{w}}_e \leftarrow \overline{\mathbf{w}}^{\text{new}}$ . Runtime: amortized  $\tilde{O}(1)$ .*

- $SC()$ . Returns a Laplacian  $\widetilde{\mathbf{SC}}$  with  $\widetilde{O}(\beta m \epsilon^{-2})$  edges which  $(1 + \epsilon)$ -spectrally approximates the Schur complement of  $\mathbf{L}$  with terminal set  $C$  in time  $\widetilde{O}(\beta m \epsilon^{-2})$ .

All outputs and runtimes are correct with high probability if  $|C| = O(\beta m)$  at all times and there are at most  $O(\beta m)$  total calls to  $UPDATE$ .

Our proof is organized in two parts. In Section 4.1 we give an algorithm to efficiently generate useful attributes of a random walk in graphs with polynomially bounded edge weights. Next, in Section 4.2 we describe how to use these walk attributes to maintain a Schur complement under modifications to the terminal set and edge weights.

#### 4.1 Approximate Random Walks with Morris Counters

Our main contribution in this section is an improved algorithm to sample random walks in weighted graphs, based on the Morris counter of [MS78, NY20]. The main technical result of this section is the following:

**Theorem 4.2** (Morris Walk). *Let  $G = (V, E, w, \ell)$  be a graph with edge weights  $w$  and edge lengths  $\ell$  bounded between 1 and  $n^{O(1)}$ . For any vertex  $u$ , and parameters  $L, \epsilon \geq 0$ , Algorithm 3 with high probability runs in  $\widetilde{O}(L^2 + L\epsilon^{-2})$  time and generates the following attributes of a random walk  $\mathcal{W}_u$  in  $G$  which starts from  $u$ , samples the edges it traverses with probabilities proportional to  $w_e$ , and stops when  $\sum_{v \in \mathcal{W}_u} \deg(v) > L$ <sup>3</sup>:*

- $u_1, u_2, \dots$ , the  $O(L)$  distinct vertices of  $\mathcal{W}_u$  in order of their encounter.
- For each  $u_i$ ,  $\delta_{u_i}$  is a  $(1 + \epsilon)$ -approximation of

$$\sum_{k=1}^{f_i-1} \ell_{(u_k, u_{k+1})},$$

where  $f_i$  is the index of the first visit of  $u_i$  in  $\mathcal{W}_u$ .

Our algorithm is based on simulating random walks in a graph by repeatedly solving linear systems, a technique that has been used in prior work on sampling random spanning trees and dynamically maintaining Schur complements [KM09, MST14, DKP<sup>+</sup>17, Sch18, DGGP19, GLP21]. However, a difficulty in applying this approach to our setting is the need to estimate the length of the resulting random walk. We address this issue by appealing to an approximate counter algorithm to estimate the length of prefixes of the walk by simulating random walks on a larger graph.

To aid our exposition, we begin by recalling a variant of the Morris counter algorithm and an improved analysis of such from [NY20], which we present in Algorithm 1.

**Theorem 4.3** (Modification of Theorem 1.2 from [NY20]). *Consider an instantiation of Algorithm 1 for parameters  $\epsilon, \delta$ , where  $INCREMENT()$  has been called  $N$  times after one call to  $INIT-COUNTER()$ . Then  $APPROXVAL()$  returns a value  $\widehat{N}$  satisfying  $\mathbb{E}[\widehat{N}] = N$  and*

$$(1 - \epsilon)N \leq \widehat{N} \leq (1 + \epsilon)N$$

with probability  $1 - \delta$ .

---

<sup>3</sup>Here,  $\deg$  denotes the *unweighted* degree of a vertex in  $G$ .



---

**Algorithm 1:** Morris Counter MORRIS

---

```

1 global variables
2   |  $X$ : current counter value
3   |  $a \geq 0$ : accuracy parameter
4 procedure INITCOUNTER() ( $\epsilon, \delta$ )
5   |  $X = 0$ .  $a \leftarrow \frac{\epsilon^2}{8 \log(1/\delta)}$ .
6 procedure INCREMENT()
7   | // Probabilistically updates the Morris counter  $X$ 
8   | with probability  $(1 + a)^{-X}$  do
9   |   |  $X = X + 1$ 
10  | end
11 procedure APPROXVAL()
12   | // Returns unbiased estimator for number of times INCREMENT() was
13   |   called.
14 return  $\frac{1}{a} \left( (1 + a)^X - 1 \right)$ 

```

---

The above theorem may be recovered directly from the analysis of Section 2.2 in [NY20]. We will employ this algorithm in a white-box fashion to estimate the length of a random walk in a graph. To do this, we condense the behavior of the counter over a collection of  $w$  increments into an explicit probability distribution:

**Definition 4.4** (Morris Increment Probabilities). Given a parameter  $a \geq 0$ , for integers  $Y, Z$  we define the Morris increment probabilities

$$p_{a,Y}^Z(\ell) = \Pr(\text{MORRIS}.X = Z \text{ after processing } \ell \text{ INCREMENT() calls} \mid \text{MORRIS}.X = Y \text{ originally})$$

We remark that these probabilities may be nontrivial to compute. However, in our algorithms we only require the ability to sample a  $Z$  with probability proportional to  $p_{a,Y}^Z(\ell)$ : we will later show how this may be done efficiently.

**Definition 4.5** (Layer Graph). For weighted graph  $G = (V, E, w, \ell)$  and parameter  $a \geq 0$ , the  $a$ -layer graph is an (infinite) weighted directed graph  $\hat{G}$  with vertex set  $\hat{V} = V \otimes \{0, 1, \dots\}$ <sup>4</sup> and edge set  $\hat{E}$  constructed in the following fashion: For each edge  $(u, v) \in E$  of weight  $w$  and length  $\ell$ , each  $0 \leq i$ , and each  $i \leq j$ , add a directed edge  $(u, i) \rightarrow (v, j)$  of weight  $w \cdot p_{a,i}^j\left(\left\lfloor \frac{\ell}{a} \right\rfloor\right)$  to  $\hat{E}$ .

We remark that although the layer graph as defined is infinite, we only access finite subgraphs of it in our algorithms. Our proof strategy in this section is in two parts. First, we describe an idealized algorithm (Algorithm 2) that directly runs a random walk in a graph and generates an output matching the requirements of Theorem 4.2. We then provide an efficient variant (Algorithm 3) which with high probability returns an output matching that of Algorithm 2 in distribution: the result follows.

---

<sup>4</sup>We use the notation  $A \otimes B$  to denote the Cartesian product of  $A$  and  $B$ : it consists of all tuples  $(i, j)$  for  $i \in A$  and  $j \in B$ .

---

**Algorithm 2:** Conceptual Morris Walk
 

---

```

1 procedure CONCEPTUALMORRISWALK( $G, L, \epsilon, u, c$ )
2    $a = \frac{\epsilon^2}{8 \log(n^{1+c})}$ 
3    $\hat{G} = a$ -layer graph of  $G$ ,  $u_{rw} = (u, 0)$ 
4    $S_{\text{visited}} = [u]$ 
5   while  $\sum_{v \in S_{\text{visited}}} \deg_G(v) \leq L$  do
6      $(v_{rw}, i_{rw}) \leftarrow$  random neighbor of  $u_{rw}$  in  $\hat{G}$ 
7     if  $v_{rw} \notin S_{\text{visited}}$  then
8        $S_{\text{visited}} = [S_{\text{visited}}; v_{rw}]$ 
9        $\delta_{v_{rw}} = \frac{1}{8} ((1+a)^{i_{rw}} - 1)$ 
10    end
11     $u_{rw} = (v_{rw}, i_{rw})$ 
12  end
13  return  $S_{\text{visited}}, \{\delta\}$ 

```

---

**Theorem 4.6.** Let  $G = (V, E, w, \ell)$  be a graph with edge weights  $w$  and edge lengths  $\ell$  bounded between 1 and  $n^{O(1)}$ . For parameters  $L, \epsilon, c \geq 0$  and starting vertex  $u$ , Algorithm 2 with high probability returns the following attributes of  $\mathcal{W}_u$ , a random walk in  $G$  which starts from  $u$ , samples the edges it traverses with probabilities proportional to  $w_e$ , and stops once  $\sum_{v \in \mathcal{W}_u} \deg_G(v) \geq L$ :

- A set  $S$  of the first  $O(L)$  distinct vertices in  $\mathcal{W}_u$ , in order of encounter
- With probability  $1 - n^{-c}$ , values  $\{\delta\}$  such that for each  $v \in S$ ,  $\delta_v (1 + \epsilon)$ -approximates the length in  $\mathcal{W}_u$  (measured with respect to  $\ell$ ) from  $u$  to the first encounter of  $v$ .

*Proof.* Let  $a = \frac{\epsilon^2}{8 \log(n^{1+c})}$ , and let  $\hat{\mathcal{W}}$  be the ordered collection of vertices  $(v_{rw}, i_{rw}) \in \hat{G}$  encountered on Line 6: note that these vertices form a random walk on  $\hat{G}$  by construction. We define an auxiliary random walk  $\mathcal{W}$  in  $G$  as follows: if the  $k^{\text{th}}$  in  $\hat{\mathcal{W}}$  is  $(v, i)$ , the  $k^{\text{th}}$  node in  $\mathcal{W}$  is  $v$ . We will show the following two facts:

- $\mathcal{W}$  is a random walk in  $G$  which starts from  $u$ , samples its edges with probabilities proportional to  $w_e$ , and stops once  $\sum_{v \in \mathcal{W}} \deg_G(v) \geq L$ .
- For any  $k$ , let the  $k^{\text{th}}$  node in  $\hat{\mathcal{W}}$  be  $(v_k, i_k)$ , and let  $R_k$  be inductively defined by  $R_1 = 0$ ,  $R_{i+1} = R_i + \left\lfloor \frac{8}{a} \ell_{(v_i, v_{i+1})} \right\rfloor$ . Then  $i_k$  is distributed as  $\text{MORRIS}(a).X$  after processing  $R_k$  increments.

The first of these claims follows immediately: if we sample a random neighbor of  $(v, i)$  in  $\hat{G}$ , the probability that it is of the form  $(v', j)$  for some  $j$  is simply

$$\frac{\sum_{j=0}^{\infty} w_{(v', v)} p_{a, i}^j \left( \left\lfloor \frac{8}{a} \ell_{(v', v)} \right\rfloor \right)}{\sum_{x \in V} \sum_{j=0}^{\infty} w_{(x, v)} p_{a, i}^j \left( \left\lfloor \frac{8}{a} \ell_{(x, v)} \right\rfloor \right)} = \frac{w_{(v', v)}}{\sum_{w \in V} w_{(x, v)}}.$$

Thus the  $k^{\text{th}}$  element of  $\mathcal{W}$  is a neighbor of the  $(k-1)^{\text{st}}$  sampled proportional to  $w$ : since  $\mathcal{W}$  starts from  $u$  the claim follows.

For the second claim, we proceed by induction on  $k$ . The claim is trivially true for  $k = 1$  (as the first node in  $\widehat{\mathcal{W}}$  is  $(u, 0)$ ). It remains to show the induction step. Let the  $k^{th}$  node of  $\widehat{\mathcal{W}}$  be  $(v_k, i_k)$ : by the induction hypothesis  $i_k$  is distributed as  $\text{MORRIS}(a).X$  after  $R_k$  increments. Now conditioned on the value of  $v_{k+1}$ , we have

$$\begin{aligned} \Pr(i_{k+1} = x) &= \sum_{y=0}^x \Pr(i_{k+1} = x | i_k = y) \Pr(i_k = y) \\ &= \sum_{y=0}^x \frac{w_{(v_k, v_{k+1})} p_{a,y}^x \left( \left\lfloor \frac{8}{a} \ell_{(v_k, v_{k+1})} \right\rfloor \right)}{\sum_{z=0}^{\infty} w_{(v_k, v_{k+1})} p_{a,y}^z \left( \left\lfloor \frac{8}{a} \ell_{(v_k, v_{k+1})} \right\rfloor \right)} \Pr(i_k = y) = \sum_{y=0}^x p_{a,y}^x \left( \left\lfloor \frac{8}{a} \ell_{(v_k, v_{k+1})} \right\rfloor \right) \Pr(i_k = y). \end{aligned}$$

But by the induction hypothesis, each term of the expression is the probability that  $\text{MORRIS}(a).X$  equals  $y$  after  $R_k$  increments and also equals  $y$  after a further  $\left\lfloor \frac{8}{a} \ell_{(v_k, v_{k+1})} \right\rfloor$  increments. Since  $R_{k+1} = R_k + \left\lfloor \frac{8}{a} \ell_{(v_k, v_{k+1})} \right\rfloor$  conditioned on the value of  $v_{k+1}$ , the claim follows by the law of total probability.

We finally show how these claims imply the theorem. First, note that  $S$  consists of the vertices in  $\mathcal{W}$  in the order of their encounter: since  $\mathcal{W}$  is a random walk in  $G$  the correctness of  $S$  follows. Next, for each  $v \in S$  let  $u_v = (v, i_v)$  be the value of  $u_{rw}$  set on Line 6 where  $v$  was first encountered. Observe that each edge in  $G$  has weight at least 1: thus

$$\left(1 - \frac{a}{8}\right) \ell_{(v_k, v_{k+1})} \leq \ell_{(v_k, v_{k+1})} - \frac{a}{8} \leq \frac{a}{8} \left\lfloor \frac{8}{a} \ell_{(v_k, v_{k+1})} \right\rfloor \leq \ell_{(v_k, v_{k+1})}.$$

Thus for any  $k$ ,

$$\left(1 - \frac{a}{8}\right) \frac{8}{a} \sum_{i=0}^k \ell_{(v_i, v_{i+1})} \leq R_k \leq \frac{8}{a} \sum_{i=0}^k \ell_{(v_i, v_{i+1})}.$$

By the second claim, we see that  $i_v$  is distributed as  $\text{MORRIS}(a).X$  after processing  $N_{uv}$  increments, where  $N_{uv} = R_k$  if  $k$  is the smallest index where  $v$  appears in  $\mathcal{W}$ . This number of increments is larger than  $\frac{8}{a}$ : by Theorem 4.3 we thus have

$$\Pr\left(\left|\frac{1}{a} \left((1+a)^{i_v} - 1\right) - \frac{8}{a} N_{uv}\right| \geq \frac{8\epsilon}{a} N_{uv}\right) \leq 1 - n^{-3}.$$

But now,  $\frac{a}{8} N_{uv}$  is within a  $1 + \frac{a}{4} \leq 1 + \epsilon$  factor of  $\mathcal{L}_{uv}$ , the length of  $\mathcal{W}$  from  $u$  to the first visit of  $v$ . Thus,

$$\Pr(|\delta_v - \mathcal{L}_{uv}| \geq 2\epsilon \mathcal{L}_{uv}) \leq 1 - n^{-C-1}.$$

As there are at most  $n$  vertices in  $S$ , the claim follows by scaling down  $\epsilon$  and union bounding over these failure probabilities.  $\square$

With Theorem 4.6 in hand, we prove the main result of this section by giving an efficient implementation of Algorithm 2. Our algorithm works by simulating a random walk over the  $a$ -layer graph  $\widehat{G}$  using a Laplacian linear system solver. We will employ the following (standard) lemma on the hitting probabilities of a random walk in undirected graphs:

**Lemma 4.7** (Corollary of Lemma 5.6 from [GLP21]). *Let  $G = (V, E, w)$  be a weighted undirected graph, and let  $x$  be any vertex in  $V$ . For any  $C \subseteq V$ , the probability that a random walk starting from  $x$  first enters  $C$  at a vertex  $y$  is given by*

$$- \left[ \mathbf{L}_{C, V \setminus C} \mathbf{L}_{C, C}^{-1} \boldsymbol{\chi}_x \right]_y.$$

*Thus, in  $\tilde{O}(|E|)$  time we may sample a vertex  $y \in C$  with probability equal to a random walk starting from  $x$  first entering  $C$  at  $y$ .*

We will use this fact within the framework of CONCEPTUALMORRISWALK to replace the explicit sampling of random walk (which may take  $\text{poly}(n, W)$  time) with a computationally efficient subroutine. We now describe the graphs on which we apply Lemma 4.7: In the below,  $N_G(S)$  denotes the vertices which are neighbors of  $S$  but do not themselves belong to  $S$ .

**Definition 4.8** ( $(a, \iota, S)$ -shortcut graph). Let  $G = (V, E, w, \ell)$  be an undirected graph with edge weights  $w$  and edge lengths  $\ell$ . Let  $a \geq 0$  be a parameter, and let  $\iota \geq 0$  be an integer. For  $S \subseteq V$ , we define the  $(a, \iota, S)$ -shortcut graph  $H = (V_H, E_H, w_H)$  as follows:

- For each  $v \in S \cup N_G(S)$ , add  $v$  to  $V_H$ .
- For each edge  $e = (u, v) \in E$  of weight  $w$  and length  $\ell$  with  $u, v \in S$ , add vertices  $v_u^+, v_v^+$  to  $V_H$ , an edge  $(u, v)$  of weight  $w \cdot p_{a, \iota}^t \left( \left\lfloor \frac{8}{a} \ell \right\rfloor \right)$  to  $E_H$ , and edges  $(u, v_u^+), (v_v^+, v)$  of weight  $w \cdot \left( 1 - p_{a, \iota}^t \left( \left\lfloor \frac{8}{a} \ell \right\rfloor \right) \right)$  to  $E_H$ .
- For each edge  $(u, v) \in E$  of weight  $w$  and length  $\ell$  with  $u \in S, v \in N(S)$ , add a vertex  $v_u^+$  to  $V_H$ , an edge  $(u, v)$  of weight  $w \cdot p_{a, \iota}^t \left( \left\lfloor \frac{8}{a} \ell \right\rfloor \right)$  to  $E_H$ , and an edge  $(u, v_u^+)$  of weight  $w \cdot \left( 1 - p_{a, \iota}^t \left( \left\lfloor \frac{8}{a} \ell \right\rfloor \right) \right)$  to  $E_H$ .

Let  $V_S^+$  denote the set of vertices of the form  $v_u^+ \in H$  for  $v \in S$ ,  $N_S$  denote vertices of the form  $v \in H$  for  $v \in N_G(S)$ , and  $N_S^+$  denote vertices of the form  $v_u^+ \in H$  for  $v \in N_G(S)$ .

Note that computing the shortcut graph defined above only requires computing Morris increment probabilities of the form  $p_{a, \iota}^t(s)$ . We will show that this admits a simple closed form, and that we may sample a variable proportional to the increment probabilities efficiently.

**Lemma 4.9.** *Given a parameter  $a \geq 0$  and integers  $Y, \ell$ , the Morris increment probabilities (Definition 4.4) satisfy*

$$p_{a, Y}^Y(\ell) = \left( 1 - (1 + a)^{-Y} \right)^\ell.$$

*In addition, we may sample an integer  $Z \geq Y + 1$  such that*

$$\Pr(Z = \Gamma) = \frac{p_{a, Y}^\Gamma(\ell)}{1 - p_{a, Y}^Y(\ell)}$$

*in time  $\tilde{O}(Z - Y)$ .*

*Proof.* For the first claim, note that each call to `INCREMENT()` increments `MORRIS.X` with probability  $(1 + a)^{-Y}$ . The probability that  $\ell$  such increments fails to increase `MORRIS.X` is therefore  $(1 - (1 + a)^{-Y})^\ell$  as claimed.

For the second claim, we describe an algorithm to sample from the claimed distribution. We first observe that the desired distribution is precisely the value of `MORRIS.X` after processing  $\ell$  increments, conditioned on

- The initial value of `MORRIS.X` was  $Y$ .
- The final value of `MORRIS.X` is strictly larger than  $Y$ .

We will sample from this distribution by implicitly simulating the Morris counter algorithm itself: we repeatedly sample from the distribution over the number of `INCREMENT()` calls required to increase `MORRIS.X`, and return the final value of `MORRIS.X` after  $\ell$  simulated increments were processed. For the below, we let  $Geom(p)$  denote the geometric random variable over  $\{1, 2, \dots\}$  with failure probability  $p$  and let  $Geom^k(p)$  denote  $Geom(p)$  conditioned on the output being at most  $k$ : note that both distributions may be sampled from in  $\tilde{O}(1)$  time.

Assume that  $Y' = \text{MORRIS.X}$  at some point. Let  $p_{Y'}$  be a random variable representing the number of `INCREMENT()` calls required to increase `MORRIS.X`: note that

$$\Pr(p_{Y'} > s) = p_{a, Y'}^{Y'}(s)$$

by definition. By the closed-form representation of these probabilities, we may therefore conclude that  $p_{Y'}$  is distributed as  $Geom((1 + a)^{-Y'})$ .

By the definition of `MORRIS`, it is therefore clear that we may sample  $Z \sim p_{a, Y}^Z(\ell)$  by repeating the following operations:

- Initialize a running increment counter  $\ell' = 0$  and a counter value  $Z = Y$ .
- Generate a sample  $k_Z \sim Geom((1 + a)^{-Z})$  and set  $\ell' = \ell' + k_Z$ .
- If  $\ell' > \ell$ , return  $Z$ . Else, increment  $Z$  by 1 and go back to the previous line.

To sample  $Z$  conditioned on  $Z \neq Y$ , it is thus sufficient to sample the first  $k_Y \sim Geom^\ell((1 + a)^{-Z})$  to ensure  $Z$  is not incremented 0 times. To bound the running time, we additionally observe that the number of geometric and truncated geometric random variables sampled is proportional to  $Z - Y$ : as the total work performed is  $\tilde{O}(1)$  times this the claim follows.  $\square$

*Proof of Theorem 4.2.* Our proof proceeds in two steps. We will first show that the vertices added to  $S$  and the values  $\{\delta\}$  have the same distribution as the output of Algorithm 2. We will then bound the runtime of the algorithm.

Let  $\hat{G}$  be the  $a$ -layer graph of  $G$ , and fix a parameter  $\iota$  and visited set  $S$  during a single iteration of Algorithm 3. We consider the subgraph  $\hat{G}_{\iota, S}$  consisting of all directed edges with tail of the form  $(v, \iota)$  for  $v \in S$ . Consider the process of running a random walk from  $(v, \iota) \in \hat{G}_{\iota, S}$  until a vertex not of the form  $(u, \iota)$  with  $u \in S$  is reached. It is self-evident that the only such vertices in  $\hat{G}_{\iota, S}$  belong to three classes:

- $(v, \iota)$  where  $v \in N_G(S)$

---

**Algorithm 3:** Morris Walk
 

---

```

1 procedure MORRISWALK( $G, L, \epsilon, u$ )
2    $a = \frac{\epsilon^2}{8 \log(n^3)}$ 
3    $S = [u], \iota = 0, u_{rw} = u$ 
4   while  $\sum_{v \in S} \deg_G(v) \leq L$  do
5      $G_S^\iota \leftarrow (a, \iota, S)$ -shortcut graph for  $G$  (Definition 4.8)
6      $C = V_S^+ \cup N_S \cup N_S^+$ 
7      $x \leftarrow$  vertex sampled with probability a random walk starting from  $u_{rw}$  in  $G_S^\iota$  first
       enters  $C$  at  $x$  (Lemma 4.7)
8     if  $x \in N_S$   $\triangleright$  Added new vertex to  $S$  then
9        $S = [S; x]$   $\triangleright$  Interpret  $x$  as a vertex in  $G$ 
10       $\delta_x = \frac{1}{8} ((1 + a)^\iota - 1)$ 
11       $u_{rw} = x$ 
12    end
13    if  $x \in V_S^+$   $\triangleright$  Incremented  $\iota$  then
14       $v_s^+ \stackrel{\text{def}}{=} x$ 
15       $\ell \stackrel{\text{def}}{=} \text{length of edge } (s, v) \in G$ 
16       $\iota \leftarrow \iota'$  sampled with probability  $\propto p'_{a,\iota}(\lfloor \frac{8}{a} \ell \rfloor)$ , conditioned on  $\iota' > \iota$  (Lemma 4.9)
17       $u_{rw} = v$ 
18    end
19    if  $v \in N_S^+$   $\triangleright$  Incremented  $\iota$  and added vertex to  $S$  then
20       $v_s^+ \stackrel{\text{def}}{=} x$ 
21       $\ell \stackrel{\text{def}}{=} \text{length of edge } (s, v) \in G$ 
22       $S = [S, v]$ 
23       $\iota \leftarrow \iota'$  sampled with probability  $\propto p'_{a,\iota}(\lfloor \frac{8}{a} \ell \rfloor)$ , conditioned on  $\iota' > \iota$  (Lemma 4.9)
24       $\delta_v = \frac{1}{8} ((1 + a)^\iota - 1)$ 
25       $u_{rw} = v$ 
26    end
27  end
28  return  $S, \{\delta\}$ 

```

---

- $(v, \iota')$  where  $v \in S$  and  $\iota' > \iota$
- $(v, \iota')$  where  $v \in N_G(S)$  and  $\iota' > \iota$ .

Let  $C$  denote the collection of vertices of this type. Note that the subgraph of  $\widehat{G}_{\iota,S}$  induced on vertices of the form  $(v, \iota)$  for  $v \in S$  is essentially undirected (since each directed edge  $(x, y)$  is matched by a directed edge  $(y, x)$  of the same weight). Let  $G_{\iota,S}$  be the graph obtained by replacing these parallel directed edges with an undirected edge of the same weight, and by removing edge directions from all other edges. It is clear that the probability distribution over vertices that a random walk starting from  $(u, \iota)$  for  $u \in S$  enters  $C$  at is induced by a Laplacian linear system solve via Lemma 4.7. By direct calculation, it may be verified that these sampling probabilities are equivalent to the sampling performed on Line 7: when sampling the number of increments to the



counter, Algorithm 3 simply samples the event that the counter is incremented at least once and then samples from the appropriate conditional distribution for the true number of increments to apply.

We now bound the running time of our algorithm. We observe that the termination condition of the while loop ensures that  $G_S^\iota$  never contains more than  $O(L)$  edges: thus the call to Lemma 4.7 on on Line 7 can be implemented in  $\tilde{O}(L)$  time. We additionally see via the remaining operations in the loop that each linear system we solve ensures that we either add a new vertex to  $S$  or increase the value of  $\iota$ . Next, we note that since  $G$ 's weights and lengths are polynomially-bounded, the total length of a random walk which covers the entirety of  $G$  is bounded by  $\text{poly}(n)$ . Thus for any  $v$  in the returned set  $S$ ,  $\delta_v$  is a  $(1 + \epsilon)$ -approximation to a quantity which is also bounded by  $\text{poly}(n)$ . But this implies that the variable  $\iota$  satisfies

$$(1 + a)^\iota \leq \text{poly}(n) \implies \iota \leq \tilde{O}(\epsilon^{-2})$$

with high probability. Thus at most  $\tilde{O}(\epsilon^{-2})$  calls to Lemma 4.7 can increase the value of  $\iota$ : as the while loop must terminate after adding  $L$  vertices to  $S$  it follows that Algorithm 3 solves at most  $\tilde{O}(L + \epsilon^{-2})$  linear systems with high probability. Finally, the only remaining nontrivial computation of the algorithm is performed on Line 16 and Line 23. But as  $\iota \leq \tilde{O}(\epsilon^{-2})$  by Lemma 4.9 these lines cost  $\tilde{O}(\epsilon^{-2})$  amortized over the whole algorithm. The claimed runtime follows.  $\square$

## 4.2 Improved Dynamic Schur Complement

Here we provide our main result regarding the dynamic maintenance of Schur complements under edge resistance changes in  $G$ . We achieve this by plugging in our improved algorithm Theorem 4.2 for estimating lengths of random walks visiting a fixed number of vertices into previous frameworks [DGGP19, GLP21]. Below, the additional operation INITIALSC maintains the approximate Schur complement ignoring edge updates, but still tracking terminal additions. It is useful for our dynamic EVALUATOR and LOCATOR data structures in Section 5. We use the notation  $\mathbf{SC}_\mathcal{H}$  for the approximation as it eventually gets used to approximately compute a harmonic extension  $\mathcal{H}$ .

**Theorem 4.10** (Dynamic Schur complement). *There is a data structure DYNAMICSC that supports the following operations against oblivious adversaries given a graph  $G = (V, E)$  with dynamic edge conductances  $\mathbf{w} \in \mathbb{R}^{E(G)}$  and parameters  $\beta < \epsilon^2 < 1$ .*

- **INITIALIZE**( $G, \mathbf{w}, \epsilon, \beta$ ). *Initializes the data structure with accuracy parameter  $\epsilon$ , and chooses a set of  $O(\beta m)$  terminals  $C$ .  $\overline{\mathbf{w}}$  is initialized as  $\mathbf{w}$ . Runtime:  $\tilde{O}(m\beta^{-2}\epsilon^{-2})$ .*
- **ADDTERMINAL**( $v$ ). *Makes  $v$  a terminal, i.e.  $C \leftarrow C \cup \{v\}$ . Runtime: amortized  $\tilde{O}(\beta^{-2}\epsilon^{-2})$ .*
- **UPDATE**( $e, \overline{\mathbf{w}}^{\text{new}}$ ). *Under the guarantee that both endpoints of  $e$  are terminals in  $C$ , updates  $\overline{\mathbf{w}}_e \leftarrow \overline{\mathbf{w}}^{\text{new}}$ . Runtime: amortized  $\tilde{O}(1)$ .*
- **SC**( $\cdot$ ). *Returns a Laplacian  $\widetilde{\mathbf{SC}} \approx_\epsilon \mathbf{SC}(\mathbf{L}(\overline{\mathbf{w}}), C)$  with  $\tilde{O}(\beta m \epsilon^{-2})$  edges in time  $\tilde{O}(\beta m \epsilon^{-2})$ .*
- **INITIALSC**( $\cdot$ ). *Returns a Laplacian  $\widetilde{\mathbf{SC}}_\mathcal{H}$  with  $\tilde{O}(\beta m \epsilon^{-2})$  edges in time  $\tilde{O}(\beta m \epsilon^{-2})$ . Let  $Z$  be the set of edges which were input to UPDATE after initialization. Define  $\mathbf{w}_Z$  as  $(\mathbf{w}_Z)_e = 0$  for  $e \in Z$  and  $(\mathbf{w}_Z)_e = \mathbf{w}_e$  otherwise. Then  $\widetilde{\mathbf{SC}}_\mathcal{H}$  satisfies*

$$\mathbf{SC}(\mathbf{L}(\mathbf{w}), C) - \epsilon \mathbf{SC}(\mathbf{L}(\mathbf{w}_Z), C) \preceq \widetilde{\mathbf{SC}}_\mathcal{H} \preceq \mathbf{SC}(\mathbf{L}(\mathbf{w}), C) + \epsilon \mathbf{SC}(\mathbf{L}(\mathbf{w}_Z), C). \quad (4)$$

All outputs and runtimes are correct whp. if  $|C| = O(\beta m)$  at all times and there are at most  $O(\beta m)$  total calls to `UPDATE`.

We note that we could achieve the tighter approximation guarantee in (4) for the operation `SC()`. However, we do not need use it in this paper (eg. Section 5) and therefore, do not state it.

We require the following process which samples Schur complements using random walks.

**Lemma 4.11** (Schur complement approximation, [DGGP19] Theorem 3.1). *Let  $G = (V, E, \mathbf{r})$  be an undirected, weighted multigraph with a subset of vertices  $C$ . Furthermore, let  $\epsilon \in (0, 1)$  and let  $\rho = 1000\epsilon^{-2} \log n$ . Let  $H$  be an initially empty graph with vertices  $C$ , and for each edge  $e = (u, v) \in E(G)$  repeat the following procedure  $\rho$  times.*

1. *Simulate a random walk from  $u$  until it hits  $C$  at  $c_1$ .*
2. *Simulate a random walk from  $v$  until it hits  $C$  at  $c_2$ .*
3. *Combine these random walks (along with edge  $e = (u, v)$ ) to form a walk  $W$ .*
4. *Add edge  $(c_1, c_2)$  to  $H$  with resistance  $\rho \sum_{e \in W} \mathbf{r}_e$ .*

*The resulting graph  $H$  satisfies  $\mathbf{L}(H) \approx_\epsilon \mathbf{SC}(\mathbf{L}(\mathbf{w}), C)$  with probability at least  $1 - n^{-10}$ .*

Finally, we require a dynamic spectral sparsification procedure.

**Lemma 4.12** ([GLP21, Lemma 4.10]). *There is a data structure that supports insertions and deletions of edges on a graph  $G$  which have underlying conductances/resistances in amortized  $\tilde{O}(\log U)$  time per operation. Additionally, it can output a  $(1 + \epsilon)$ -spectral sparsifier of  $G$  in  $\tilde{O}(n\epsilon^{-2} \log U)$  time.*

Now, we can show Theorem 4.10 exactly as done in [DGGP19, GLP21] by sampling random walks using Lemma 4.11 and shortcutting them as terminals get added.

*Proof of Theorem 4.10.* We explain how to implement each operation in Theorem 4.10.

**INITIALIZE:** Randomly sample an initial terminal set  $C$  of size  $O(\beta m)$ . From each edge  $e = (u, v) \in G$ , sample  $\rho = \tilde{O}(\epsilon^{-2})$  random walks from  $u, v$  to  $C$  as in Lemma 4.11, and record  $(1 + \epsilon)$  approximations of the sums of resistances of all prefixes. Note that these walks visit  $\tilde{O}(\beta^{-1})$  distinct vertices whp. Initialize the data structure  $D^{(s)}$  in Lemma 4.12. Based on these random walks, add edges to  $C$  using the data structure  $D^{(s)}$ . Additionally, maintain a set  $Z$  of updated edges, whose original and final conductances we track explicitly.

The runtime of **INITIALIZE** is dominated by the time to sample the random walks, which is  $\tilde{O}(m\epsilon^{-2}(\beta^{-2} + \beta^{-1}\epsilon^{-2})) = \tilde{O}(m\epsilon^{-2}\beta^{-2})$  by Theorem 4.2 (the length  $L = \tilde{O}(\beta^{-1})$ ) and  $\beta < \epsilon^2$ .

**ADDTERMINAL( $v$ ):** Update  $C \leftarrow C \cup \{v\}$  and shortcut all walks passing through  $v$ . The total length of all walks is  $\tilde{O}(m\epsilon^{-2}\beta^{-1})$ , so over the course of  $O(\beta m)$  terminal insertions, the amortized runtime is  $\tilde{O}(m\epsilon^{-2}\beta^{-1}/(\beta m)) = \tilde{O}(\beta^{-2}\epsilon^{-2})$ . Finally, pass all edge insertions/deletions in  $C$  to  $D^{(s)}$ .

**UPDATE( $e, \bar{\mathbf{w}}^{\text{new}}$ ):** Delete the edge  $e$  (do not insert an edge with conductance  $\bar{\mathbf{w}}^{\text{new}}$ ), and pass the deletion to  $D^{(s)}$ . Update  $Z \leftarrow Z \cup \{e\}$ . From now on, the algorithm explicitly stores in memory the original and current conductances of edge  $e$ . Clearly, the update time is  $\tilde{O}(1)$ .

**SC():** Call  $D^{(s)}$  to output a  $(1 + \epsilon)$ -approximation of  $\mathbf{SC}(\mathbf{L}(\mathbf{w}_Z), C)$  with high probability. The approximation guarantee follows from Lemma 4.11 and the guarantee of Theorem 4.2 that the total

resistive length of each random walk is correct up to  $(1 + \epsilon/10)$  with high probability. Finally, add the edges  $e \in Z$  back in with the current conductances. The runtime is  $\tilde{O}(\beta m \epsilon^{-2})$  by Lemma 4.12 as  $|C| = O(\beta m)$ .

INITIALSC(): Same as SC(), except we add back edges in  $Z$  with their original conductances. The tighter approximation holds because the algorithm is returning a  $(1 + \epsilon)$ -approximation of  $\text{SC}(\mathbf{L}(\mathbf{w}_{\overline{Z}}), C)$  and the edges  $e \in Z$  that are added in contribute no error.  $\square$

## 5 Data Structures for Dynamic Electrical Flows

The goal of this section is to apply the dynamic Schur complement data structure of Theorem 4.10 to give algorithms that dynamically maintain electric potentials and edges with large electric energies in dynamic electrical flows. In Section 5.1, we will introduce the harmonic extension and use it to decompose the energy vector we need to maintain for the outer IPM. In Section 5.2, we show how to maintain a potential vector which is a key component for the following subsections. In Section 5.3, we build the EVALUATOR that estimates the energy of any edge. In Section 5.4, we build the LOCATOR that returns a superset of edges with large energies.

### 5.1 Harmonic Extension

A key notion we use throughout is the harmonic extension, which is a linear operator that maps the potentials restricted to a terminal set to the full electric potentials  $\phi$ . We use  $\mathbf{T}$  to denote the projection orthogonal to the all-ones vector.

**Definition 5.1** (Harmonic extension). For a graph  $G = (V, E)$  with edge conductances  $\mathbf{w} \in \mathbb{R}_{>0}^E$  and  $C \subseteq V(G)$ , define the *harmonic extension* operator  $\mathcal{H}_C \in \mathbb{R}^{V(G) \times C}$  as

$$\mathcal{H}_C \stackrel{\text{def}}{=} \begin{bmatrix} -\mathbf{L}(\mathbf{w})_{FF}^{-1} \mathbf{L}(\mathbf{w})_{FC} \mathbf{T} \\ \mathbf{T} \end{bmatrix}.$$

Note that the harmonic extension does not depend on edges with both endpoints in  $C$ . Leveraging this yields the following alternative characterization of the harmonic extension. These properties are crucial for our data structures as they maintain a growing terminal set where resistance changes are on edges completely inside the terminal set. In this section, we use  $\tilde{\mathbf{w}}$  to denote modified conductances and  $\mathbf{w}$  to denote initial conductances.

**Lemma 5.2** (Alternate definition of harmonic extension). *For a graph  $G = (V, E)$  with edge conductances  $\mathbf{w} \in \mathbb{R}_{>0}^E$  and  $C \subseteq V$ , let  $\tilde{G}$  be a graph with the same edge set as  $G$  whose conductances  $\tilde{\mathbf{w}}$  agree with  $\mathbf{w}$  except potentially on edges with both endpoints inside  $C$ . Then*

$$\mathcal{H}_C = \mathbf{L}(\tilde{\mathbf{w}})^\dagger \begin{bmatrix} 0 \\ \text{SC}(\mathbf{L}(\tilde{\mathbf{w}}), C) \end{bmatrix}. \quad (5)$$

*Proof.* By Definition 5.1, the harmonic extension does not depend on the edges inside  $C$ . Hence, we can simply show the lemma for the Laplacian  $\mathbf{L} = \mathbf{L}(\tilde{\mathbf{w}})$ . By the Cholesky factorization

(Lemma 3.2), we have

$$\begin{aligned}
\mathbf{L}^\dagger \begin{bmatrix} 0 \\ \mathbf{SC}(\mathbf{L}, C) \end{bmatrix} &= \begin{bmatrix} \mathbf{I} & -\mathbf{L}_{FF}^{-1} \mathbf{L}_{FC} \\ 0 & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{L}_{FF}^{-1} & 0 \\ 0 & \mathbf{SC}(\mathbf{L}, C)^\dagger \end{bmatrix} \begin{bmatrix} \mathbf{I} & 0 \\ -\mathbf{L}_{CF} \mathbf{L}_{FF}^{-1} & \mathbf{I} \end{bmatrix} \begin{bmatrix} 0 \\ \mathbf{SC}(\mathbf{L}, C) \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{I} & -\mathbf{L}_{FF}^{-1} \mathbf{L}_{FC} \\ 0 & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{L}_{FF}^{-1} & 0 \\ 0 & \mathbf{SC}(\mathbf{L}, C)^\dagger \end{bmatrix} \begin{bmatrix} 0 \\ \mathbf{SC}(\mathbf{L}, C) \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{I} & -\mathbf{L}_{FF}^{-1} \mathbf{L}_{FC} \\ 0 & \mathbf{T} \end{bmatrix} \begin{bmatrix} 0 \\ \mathbf{T} \end{bmatrix} = \mathcal{H}_C.
\end{aligned}$$

□

This is why we use the notation  $\mathcal{H}_C$  without reference to  $G$  – when we use  $\mathcal{H}_C$  in our dynamic data structures all changed edges will lie inside  $C$ . Consequently, the actual graph (beyond initialization) does not affect  $\mathcal{H}_C$ !

The inverse of the Laplacian can be represented by a contribution from the Schur complement, plus  $\mathbf{L}_{FF}^{-1}$ . This is essentially just a restatement of the Cholesky factorization (Lemma 3.2).

**Lemma 5.3.** *Let  $G = (V, E, \mathbf{w})$  be a graph. Then*

$$\mathbf{L}(\mathbf{w})^\dagger = \mathcal{H}_C \mathbf{SC}(\mathbf{L}(\mathbf{w}), C)^\dagger \mathcal{H}_C^\top + \begin{bmatrix} \mathbf{L}(\mathbf{w})_{F,F}^{-1} & 0 \\ 0 & 0 \end{bmatrix}.$$

*Proof.* The Cholesky factorization (Lemma 3.2) says that

$$\mathbf{L}(\mathbf{w})^\dagger = \begin{bmatrix} \mathbf{I} & -\mathbf{L}_{FF}^{-1} \mathbf{L}_{FC} \\ 0 & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{L}_{FF}^{-1} & 0 \\ 0 & \mathbf{SC}(\mathbf{L}, C)^\dagger \end{bmatrix} \begin{bmatrix} \mathbf{I} & 0 \\ -\mathbf{L}_{CF} \mathbf{L}_{FF}^{-1} & \mathbf{I} \end{bmatrix}.$$

As  $\mathbf{TSC}(\mathbf{L}, C)^\dagger \mathbf{T} = \mathbf{SC}(\mathbf{L}, C)^\dagger$ , the equation above is equal to

$$\begin{aligned}
&\begin{bmatrix} \begin{bmatrix} \mathbf{I} \\ 0 \end{bmatrix} & \mathcal{H}_C \end{bmatrix} \begin{bmatrix} \mathbf{L}_{FF}^{-1} & 0 \\ 0 & \mathbf{SC}(\mathbf{L}, C)^\dagger \end{bmatrix} \begin{bmatrix} \begin{bmatrix} \mathbf{I} & 0 \end{bmatrix} \\ \mathcal{H}_C^\top \end{bmatrix} \\
&= \mathcal{H}_C \mathbf{SC}(\mathbf{L}(\overline{\mathbf{w}}), C)^\dagger \mathcal{H}_C^\top + \begin{bmatrix} \mathbf{I} \\ 0 \end{bmatrix} \mathbf{L}(\mathbf{w})_{F,F}^{-1} \begin{bmatrix} \mathbf{I} & 0 \end{bmatrix} \\
&= \mathcal{H}_C \mathbf{SC}(\mathbf{L}(\overline{\mathbf{w}}), C)^\dagger \mathcal{H}_C^\top + \begin{bmatrix} \mathbf{L}(\mathbf{w})_{F,F}^{-1} & 0 \\ 0 & 0 \end{bmatrix}
\end{aligned}$$

□

Let  $\mathbf{v} \in \mathbb{R}^E$  be a (dynamic) vector. To implement the outer IPM, we must be able to maintain a heavy-hitter sketch on the following vector

$$\mathbf{\Pi} \mathbf{v} \stackrel{\text{def}}{=} \mathbf{\Pi}(\overline{\mathbf{w}}) \mathbf{v} = \overline{\mathbf{W}}^{1/2} \mathbf{B} \mathbf{L}(\overline{\mathbf{w}})^\dagger \mathbf{B}^\top \overline{\mathbf{W}}^{1/2} \mathbf{v}.$$

For this, we decompose  $\Pi \mathbf{v}$  into three terms. Let  $\hat{\mathbf{v}}$  be any vector that agrees with  $\mathbf{v}$  in  $F = V \setminus C$ . Then We first decompose  $\Pi(\mathbf{v})$  by Lemma 5.3.

$$\begin{aligned}
\Pi(\mathbf{v}) &= \bar{\mathbf{W}}^{1/2} \mathbf{B} \mathbf{L}(\bar{\mathbf{w}})^\dagger \mathbf{B}^\top \bar{\mathbf{W}}^{1/2} \mathbf{v} \\
&= \bar{\mathbf{W}}^{1/2} \mathbf{B} \left( \mathcal{H}_C \mathbf{S} \mathbf{C}(\mathbf{L}(\bar{\mathbf{w}}), C)^\dagger \mathcal{H}_C^\top + \begin{bmatrix} \mathbf{L}(\bar{\mathbf{w}})_{F,F}^{-1} & 0 \\ 0 & 0 \end{bmatrix} \right) \mathbf{B}^\top \bar{\mathbf{W}}^{1/2} \mathbf{v} \quad (\text{by Lemma 5.3}) \\
&= \bar{\mathbf{W}}^{1/2} \mathbf{B} \mathcal{H}_C \mathbf{S} \mathbf{C}(\mathbf{L}(\bar{\mathbf{w}}), C)^\dagger \mathcal{H}_C^\top \mathbf{B}^\top \bar{\mathbf{W}}^{1/2} \mathbf{v} + \bar{\mathbf{W}}^{1/2} \mathbf{B} \begin{bmatrix} \mathbf{L}(\bar{\mathbf{w}})_{F,F}^{-1} & 0 \\ 0 & 0 \end{bmatrix} \mathbf{B}^\top \bar{\mathbf{W}}^{1/2} \hat{\mathbf{v}}.
\end{aligned}$$

Since  $\mathbf{w}_C$  does not affect the value of the second term, we have

$$\begin{aligned}
&\bar{\mathbf{W}}^{1/2} \mathbf{B} \mathcal{H}_C \mathbf{S} \mathbf{C}(\mathbf{L}(\bar{\mathbf{w}}), C)^\dagger \mathcal{H}_C^\top \mathbf{B}^\top \bar{\mathbf{W}}^{1/2} \mathbf{v} + \bar{\mathbf{W}}^{1/2} \mathbf{B} \begin{bmatrix} \mathbf{L}(\bar{\mathbf{w}})_{F,F}^{-1} & 0 \\ 0 & 0 \end{bmatrix} \mathbf{B}^\top \bar{\mathbf{W}}^{1/2} \hat{\mathbf{v}} \\
&= \bar{\mathbf{W}}^{1/2} \mathbf{B} \mathcal{H}_C \mathbf{S} \mathbf{C}(\mathbf{L}(\bar{\mathbf{w}}), C)^\dagger \mathcal{H}_C^\top \mathbf{B}^\top \bar{\mathbf{W}}^{1/2} \mathbf{v} + \mathbf{W}^{1/2} \mathbf{B} \begin{bmatrix} \mathbf{L}(\mathbf{w})_{F,F}^{-1} & 0 \\ 0 & 0 \end{bmatrix} \mathbf{B}^\top \mathbf{W}^{1/2} \hat{\mathbf{v}}.
\end{aligned}$$

Then we use Lemma 5.3 in the other direction to get

$$\begin{aligned}
&\bar{\mathbf{W}}^{1/2} \mathbf{B} \mathcal{H}_C \mathbf{S} \mathbf{C}(\mathbf{L}(\bar{\mathbf{w}}), C)^\dagger \mathcal{H}_C^\top \mathbf{B}^\top \bar{\mathbf{W}}^{1/2} \mathbf{v} + \mathbf{W}^{1/2} \mathbf{B} \begin{bmatrix} \mathbf{L}(\mathbf{w})_{F,F}^{-1} & 0 \\ 0 & 0 \end{bmatrix} \mathbf{B}^\top \mathbf{W}^{1/2} \hat{\mathbf{v}} \\
&= \bar{\mathbf{W}}^{1/2} \mathbf{B} \mathcal{H}_C \mathbf{S} \mathbf{C}(\mathbf{L}(\bar{\mathbf{w}}), C)^\dagger \mathcal{H}_C^\top \mathbf{B}^\top \bar{\mathbf{W}}^{1/2} \mathbf{v} + \mathbf{W}^{1/2} \mathbf{B} \left( \mathbf{L}(\mathbf{w})^\dagger - \mathcal{H}_C \mathbf{S} \mathbf{C}(\mathbf{L}(\mathbf{w}), C)^\dagger \mathcal{H}_C^\top \right) \mathbf{B}^\top \mathbf{W}^{1/2} \hat{\mathbf{v}} \\
&\quad (\text{by Lemma 5.3}) \\
&= \bar{\mathbf{W}}^{1/2} \mathbf{B} \mathcal{H}_C \mathbf{S} \mathbf{C}(\mathbf{L}(\bar{\mathbf{w}}), C)^\dagger \mathcal{H}_C^\top \mathbf{B}^\top \bar{\mathbf{W}}^{1/2} \mathbf{v} - \mathbf{W}^{1/2} \mathbf{B} \mathcal{H}_C \mathbf{S} \mathbf{C}(\mathbf{L}(\mathbf{w}), C)^\dagger \mathcal{H}_C^\top \mathbf{B}^\top \mathbf{W}^{1/2} \hat{\mathbf{v}} \\
&\quad + \mathbf{W}^{1/2} \mathbf{B} \mathbf{L}(\mathbf{w})^\dagger \mathbf{B}^\top \mathbf{W}^{1/2} \hat{\mathbf{v}}.
\end{aligned}$$

We will use

$$\begin{aligned}
\Pi \mathbf{v} &= \bar{\mathbf{W}}^{1/2} \mathbf{B} \mathcal{H}_C \mathbf{S} \mathbf{C}(\mathbf{L}(\bar{\mathbf{w}}), C)^\dagger \mathcal{H}_C^\top \mathbf{B}^\top \bar{\mathbf{W}}^{1/2} \mathbf{v} - \mathbf{W}^{1/2} \mathbf{B} \mathcal{H}_C \mathbf{S} \mathbf{C}(\mathbf{L}(\mathbf{w}), C)^\dagger \mathcal{H}_C^\top \mathbf{B}^\top \mathbf{W}^{1/2} \hat{\mathbf{v}} \\
&\quad + \mathbf{W}^{1/2} \mathbf{B} \mathbf{L}(\mathbf{w})^\dagger \mathbf{B}^\top \mathbf{W}^{1/2} \hat{\mathbf{v}} \quad (6)
\end{aligned}$$

in two cases

- where  $\hat{\mathbf{v}} = \mathbf{v}$ , and
- where  $\mathbf{v}$  being the current vector and  $\hat{\mathbf{v}}$  being the initial  $\mathbf{v}$ .

At a high level, our approach will use several spectral approximations of the RHS of (6). We will replace  $\mathbf{S} \mathbf{C}(\mathbf{L}(\bar{\mathbf{w}}), C)$  with an approximate Schur complement using Theorem 4.10. Additionally, we will replace  $\mathcal{H}_C$  and  $\mathcal{H}_C^\top$  by replacing the Schur complements in (5) with approximate Schur complements given by Theorem 4.10.

We now focus on approximating the “right” of the first two terms of the RHS (6), i.e. the induced potentials on  $C$

$$\phi = \mathbf{S} \mathbf{C}(\mathbf{L}(\bar{\mathbf{w}}), C)^\dagger \mathcal{H}_C^\top \mathbf{B}^\top \bar{\mathbf{W}}^{1/2} \mathbf{v} \quad (7)$$

and

$$\psi = \mathbf{SC}(\mathbf{L}(\mathbf{w}), C)^\dagger \mathcal{H}_C^\top \mathbf{B}^\top \mathbf{W}^{1/2} \hat{\mathbf{v}}. \quad (8)$$

Lemma 5.5 below defines the approximation of  $\phi$  that our data structures maintain. To analyze the quality of the approximation, we will need a standard spectral approximation inequality, proven for completeness.

**Lemma 5.4** (Spectral approximation of differences). *For PSD matrices  $\mathbf{X} \approx_\epsilon \mathbf{Y}$ , we have that*

$$(\mathbf{X} - \mathbf{Y})\mathbf{X}^\dagger(\mathbf{X} - \mathbf{Y}) \preceq \epsilon^2 \mathbf{X}.$$

*Proof.* The desired inequality follows from  $\|\mathbf{I} - \mathbf{X}^{\dagger/2} \mathbf{Y} \mathbf{X}^{\dagger/2}\|_2 \leq \epsilon$  and multiplying the LHS and RHS by  $\mathbf{X}^{1/2}$  on the left and right. Now, this follows because  $\mathbf{X} \approx_\epsilon \mathbf{Y}$  implies that  $(1 - \epsilon)\mathbf{I} \preceq \mathbf{X}^{\dagger/2} \mathbf{Y} \mathbf{X}^{\dagger/2} \preceq (1 + \epsilon)\mathbf{I}$  as desired.  $\square$

**Lemma 5.5** (Approximate potential). *Let  $G$  be a graph with weights  $\bar{\mathbf{w}} \in \mathbb{R}^E$  which differ from weights  $\mathbf{w} \in \mathbb{R}^E$  except on an edge subset  $Z \subseteq E(G)$ . Let  $C \subseteq V(G)$  contain all endpoints of edges in  $Z$ . Let  $\mathbf{w}_{\bar{Z}} \in \mathbb{R}^E$  be defined as  $(\mathbf{w}_{\bar{Z}})_e = 0$  for  $e \in Z$  and  $(\mathbf{w}_{\bar{Z}})_e = \mathbf{w}_e$  otherwise. Let  $\widetilde{\mathbf{SC}} \approx_\epsilon \mathbf{SC}(\mathbf{L}(\bar{\mathbf{w}}), C)$  and let  $\tilde{\mathcal{H}} = \mathbf{L}(\mathbf{w})^\dagger \begin{bmatrix} 0 \\ \widetilde{\mathbf{SC}}_{\mathcal{H}} \end{bmatrix}$  for some  $\widetilde{\mathbf{SC}}_{\mathcal{H}}$  satisfying*

$$\mathbf{SC}(\mathbf{L}(\mathbf{w}), C) - \epsilon \mathbf{SC}(\mathbf{L}(\mathbf{w}_{\bar{Z}}), C) \preceq \widetilde{\mathbf{SC}}_{\mathcal{H}} \preceq \mathbf{SC}(\mathbf{L}(\mathbf{w}), C) + \epsilon \mathbf{SC}(\mathbf{L}(\mathbf{w}_{\bar{Z}}), C). \quad (9)$$

*Then the vectors  $\phi = \mathbf{SC}(\mathbf{L}(\bar{\mathbf{w}}), C)^\dagger \mathcal{H}_C^\top \mathbf{B}^\top \bar{\mathbf{W}}^{1/2} \mathbf{v}$  ((7)) and  $\tilde{\phi} = \widetilde{\mathbf{SC}}^\dagger \mathbf{B}^\top (\bar{\mathbf{W}}^{1/2} - \mathbf{W}^{1/2}) \mathbf{v} + \widetilde{\mathbf{SC}}^\dagger \tilde{\mathcal{H}}^\top \mathbf{B}^\top \mathbf{W}^{1/2} \mathbf{v}$  in  $\mathbb{R}^C$  satisfy  $\|\phi - \tilde{\phi}\|_{\mathbf{SC}(\mathbf{L}(\bar{\mathbf{w}}), C)} \leq 3\epsilon \|\mathbf{v}\|_2$ .*

*Proof.* We first calculate that

$$\begin{aligned} \phi &= \left( \mathbf{SC}(\mathbf{L}(\bar{\mathbf{w}}), C)^\dagger - \widetilde{\mathbf{SC}}^\dagger \right) \mathcal{H}_C^\top \mathbf{B}^\top \bar{\mathbf{W}}^{1/2} \mathbf{v} + \widetilde{\mathbf{SC}}^\dagger \mathcal{H}_C^\top \mathbf{B}^\top \bar{\mathbf{W}}^{1/2} \mathbf{v} \\ &= \left( \mathbf{SC}(\mathbf{L}(\bar{\mathbf{w}}), C)^\dagger - \widetilde{\mathbf{SC}}^\dagger \right) \mathcal{H}_C^\top \mathbf{B}^\top \bar{\mathbf{W}}^{1/2} \mathbf{v} + \widetilde{\mathbf{SC}}^\dagger \mathcal{H}_C^\top \mathbf{B}^\top (\bar{\mathbf{W}}^{1/2} - \mathbf{W}^{1/2}) \mathbf{v} + \widetilde{\mathbf{SC}}^\dagger \mathcal{H}_C^\top \mathbf{B}^\top \mathbf{W}^{1/2} \mathbf{v} \\ &\stackrel{(i)}{=} \left( \mathbf{SC}(\mathbf{L}(\bar{\mathbf{w}}), C)^\dagger - \widetilde{\mathbf{SC}}^\dagger \right) \mathcal{H}_C^\top \mathbf{B}^\top \bar{\mathbf{W}}^{1/2} \mathbf{v} + \widetilde{\mathbf{SC}}^\dagger \mathbf{B}_{:,C}^\top (\bar{\mathbf{W}}^{1/2} - \mathbf{W}^{1/2}) \mathbf{v} + \widetilde{\mathbf{SC}}^\dagger \mathcal{H}_C^\top \mathbf{B}^\top \mathbf{W}^{1/2} \mathbf{v} \end{aligned}$$

(i) is because  $\mathcal{H}_C^\top \mathbf{B}^\top (\bar{\mathbf{W}}^{1/2} - \mathbf{W}^{1/2}) = \mathbf{B}_{:,C}^\top (\bar{\mathbf{W}}^{1/2} - \mathbf{W}^{1/2})$  as  $\bar{\mathbf{w}} = \mathbf{w}$  except on  $C$ . We extract  $\tilde{\phi}$  by swapping the  $\tilde{\mathcal{H}}$  by  $\tilde{\mathcal{H}}_C$  in the last term:

$$\begin{aligned} \phi &= \left( \mathbf{SC}(\mathbf{L}(\bar{\mathbf{w}}), C)^\dagger - \widetilde{\mathbf{SC}}^\dagger \right) \mathcal{H}_C^\top \mathbf{B}^\top \bar{\mathbf{W}}^{1/2} \mathbf{v} + \widetilde{\mathbf{SC}}^\dagger \mathbf{B}_{:,C}^\top (\bar{\mathbf{W}}^{1/2} - \mathbf{W}^{1/2}) \mathbf{v} + \widetilde{\mathbf{SC}}^\dagger \mathcal{H}_C^\top \mathbf{B}^\top \mathbf{W}^{1/2} \mathbf{v} \\ &= \tilde{\phi} + \left( \mathbf{SC}(\mathbf{L}(\bar{\mathbf{w}}), C)^\dagger - \widetilde{\mathbf{SC}}^\dagger \right) \mathcal{H}_C^\top \mathbf{B}^\top \bar{\mathbf{W}}^{1/2} \mathbf{v} + \widetilde{\mathbf{SC}}^\dagger (\mathcal{H}_C^\top - \tilde{\mathcal{H}}^\top) \mathbf{B}^\top \mathbf{W}^{1/2} \mathbf{v}. \end{aligned}$$

Hence

$$\phi - \tilde{\phi} = \left( \mathbf{SC}(\mathbf{L}(\bar{\mathbf{w}}), C)^\dagger - \widetilde{\mathbf{SC}}^\dagger \right) \mathcal{H}_C^\top \mathbf{B}^\top \bar{\mathbf{W}}^{1/2} \mathbf{v} + \widetilde{\mathbf{SC}}^\dagger (\mathcal{H}_C^\top - \tilde{\mathcal{H}}^\top) \mathbf{B}^\top \mathbf{W}^{1/2} \mathbf{v}. \quad (10)$$



We bound both terms separately. For the first term,

$$\begin{aligned} & \left\| \left( \mathbf{SC}(\mathbf{L}(\overline{\mathbf{w}}), C)^\dagger - \widetilde{\mathbf{SC}}^\dagger \right) \mathcal{H}_C^\top \mathbf{B}^\top \overline{\mathbf{W}}^{1/2} \mathbf{v} \right\|_{\mathbf{SC}(\mathbf{L}(\overline{\mathbf{w}}), C)} \stackrel{(i)}{\leq} \epsilon \left\| \mathcal{H}_C^\top \mathbf{B}^\top \overline{\mathbf{W}}^{1/2} \mathbf{v} \right\|_{\mathbf{SC}(\mathbf{L}(\overline{\mathbf{w}}), C)^\dagger} \\ & \stackrel{(ii)}{\leq} \epsilon \left\| \mathbf{B}^\top \overline{\mathbf{W}}^{1/2} \mathbf{v} \right\|_{\mathbf{L}(\overline{\mathbf{w}})^\dagger} \leq \epsilon \|\mathbf{v}\|_2. \end{aligned}$$

where (i) follows from  $\widetilde{\mathbf{SC}} \approx_\epsilon \mathbf{SC}(\mathbf{L}(\overline{\mathbf{w}}), C)$  and Lemma 5.4 for  $\mathbf{X} = \mathbf{SC}(\mathbf{L}(\overline{\mathbf{w}}), C)^\dagger$  and  $\mathbf{Y} = \widetilde{\mathbf{SC}}^\dagger$ , and (ii) follows from Lemma 5.3 and the fact that  $\mathbf{L}_{F,F}^{-1}$  is positive definite. For the second term,

$$\begin{aligned} & \left\| \widetilde{\mathbf{SC}}^\dagger (\mathcal{H}_C^\top - \widetilde{\mathcal{H}}^\top) \mathbf{B}^\top \mathbf{W}^{1/2} \mathbf{v} \right\|_{\mathbf{SC}(\mathbf{L}(\overline{\mathbf{w}}), C)} \leq 2 \left\| (\mathcal{H}_C^\top - \widetilde{\mathcal{H}}^\top) \mathbf{B}^\top \mathbf{W}^{1/2} \mathbf{v} \right\|_{\mathbf{SC}(\mathbf{L}(\overline{\mathbf{w}}), C)^\dagger} \\ & \leq 2 \left\| (\mathcal{H}_C^\top - \widetilde{\mathcal{H}}^\top) \mathbf{B}^\top \mathbf{W}^{1/2} \mathbf{v} \right\|_{\mathbf{SC}(\mathbf{L}(\mathbf{w}_{\overline{Z}}), C)^\dagger} \stackrel{(i)}{\leq} 2\epsilon \left\| [\mathbf{L}(\mathbf{w})^\dagger \mathbf{B}^\top \mathbf{W}^{1/2} \mathbf{v}]_C \right\|_{\mathbf{SC}(\mathbf{L}(\mathbf{w}_{\overline{Z}}), C)} \\ & \stackrel{(ii)}{\leq} 2\epsilon \left\| [\mathbf{L}(\mathbf{w})^\dagger \mathbf{B}^\top \mathbf{W}^{1/2} \mathbf{v}]_C \right\|_{\mathbf{SC}(\mathbf{L}(\mathbf{w}), C)} \stackrel{(iii)}{\leq} 2\epsilon \|\mathbf{v}\|_2, \end{aligned}$$

where (i) follows from Lemma 5.2 and (9), (ii) follows from  $\mathbf{SC}(\mathbf{L}(\mathbf{w}), C) - \mathbf{SC}(\mathbf{L}(\mathbf{w}_{\overline{Z}}), C) = \mathbf{L}(G[C])$  being positive semidefinite, and (iii) follows from the fact that the Schur complement is spectrally smaller than the Laplacian:  $\mathbf{SC}(\mathbf{L}(\mathbf{w}), C) \preceq \mathbf{L}(\mathbf{w})$ .  $\square$

By Lemma 5.5 with  $\overline{\mathbf{w}} = \mathbf{w}$ , we can approximate the other potential vector in the RHS of (6).

**Corollary 5.6.** *Let  $\psi = \mathbf{SC}(\mathbf{L}(\mathbf{w}), C)^\dagger \mathcal{H}_C^\top \mathbf{B} \mathbf{W}^{1/2} \mathbf{v}$  ((8)) and  $\tilde{\psi} = \widetilde{\mathbf{SC}}(\mathbf{L}(\mathbf{w}), C)^\dagger \widetilde{\mathcal{H}}^\top \mathbf{B} \mathbf{W}^{1/2} \mathbf{v}$  in  $\mathbb{R}^C$  where  $\widetilde{\mathbf{SC}}(\mathbf{L}(\mathbf{w}), C)$  satisfies  $\widetilde{\mathbf{SC}}(\mathbf{L}(\mathbf{w}), C) \approx_\epsilon \mathbf{SC}(\mathbf{L}(\mathbf{w}), C)$ . We have  $\|\psi - \tilde{\psi}\|_{\mathbf{SC}(\mathbf{L}(\mathbf{w}), C)} \leq 3\epsilon \|\mathbf{v}\|_2$ .*

*Proof.* Apply Lemma 5.5 with  $\overline{\mathbf{w}} = \mathbf{w}$ ,  $\psi = \phi$  and  $\tilde{\psi} = \tilde{\phi}$ . The first term of  $\psi$

$$\widetilde{\mathbf{SC}}^\dagger \mathbf{B}(\overline{\mathbf{W}}^{1/2} - \mathbf{W}^{1/2}) \mathbf{v}$$

equals 0 because  $\overline{\mathbf{W}} = \mathbf{W}$ .  $\square$

We can use our approximate potential  $\tilde{\phi}$ ,  $\tilde{\psi}$  in Lemma 5.5 and Corollary 5.6 to define a full approximate projection of  $\Pi \mathbf{v}$ . Our starting point is that  $\Pi \mathbf{v} = \overline{\mathbf{W}}^{1/2} \mathbf{B} \mathcal{H}_C \phi + \mathbf{W}^{1/2} \mathbf{B} \mathcal{H}_C \psi$  for  $\phi$ ,  $\psi$  as in Lemma 5.5 and Corollary 5.6.

**Lemma 5.7** (Approximate projection). *Let  $\mathbf{w}, \overline{\mathbf{w}}, Z, \mathbf{w}_{\overline{Z}}, \widetilde{\mathcal{H}}, \phi, \tilde{\phi}$  be as in Lemma 5.5, let  $\psi, \tilde{\psi}$  be as in Corollary 5.6, and let*

$$\begin{aligned} \tilde{\Pi} \mathbf{v} &= (\overline{\mathbf{W}}^{1/2} - \mathbf{W}^{1/2}) \mathbf{B} \tilde{\phi} + \mathbf{W}^{1/2} \mathbf{B} \widetilde{\mathcal{H}} \tilde{\phi} \\ &\quad + \mathbf{W}^{1/2} \mathbf{B} \widetilde{\mathcal{H}} \tilde{\psi} \\ &\quad + \mathbf{W}^{1/2} \mathbf{B} \mathbf{L}(\mathbf{w})^\dagger \mathbf{B}^\top \mathbf{W}^{1/2} \mathbf{v} \end{aligned}$$

where  $\tilde{\phi}$  is padded with zeroes for computing  $\mathbf{B} \tilde{\phi}$ . Then

$$\left\| \Pi \mathbf{v} - \tilde{\Pi} \mathbf{v} \right\|_2 \leq 2\epsilon \|\mathbf{v}\|_2 + (1 + \epsilon) \left\| \phi - \tilde{\phi} \right\|_{\mathbf{SC}(\mathbf{L}(\overline{\mathbf{w}}), C)} + (1 + \epsilon) \left\| \psi - \tilde{\psi} \right\|_{\mathbf{SC}(\mathbf{L}(\mathbf{w}), C)}.$$

*Proof.* We first prove the first two terms of  $\tilde{\Pi}\mathbf{v}$

$$\tilde{\Pi}_\phi(\mathbf{v}) \stackrel{\text{def}}{=} (\bar{\mathbf{W}}^{1/2} - \mathbf{W}^{1/2}) \mathbf{B}\tilde{\phi} + \mathbf{W}^{1/2} \mathbf{B}\tilde{\mathcal{H}}\tilde{\phi}$$

approximates  $\Pi_\phi(\mathbf{v}) \stackrel{\text{def}}{=} \bar{\mathbf{W}}^{1/2} \mathbf{B}\mathcal{H}_C \phi$ . Specifically,

$$\left\| \Pi_\phi(\mathbf{v}) - \tilde{\Pi}_\phi(\mathbf{v}) \right\|_2 \leq \epsilon \|\mathbf{v}\|_2 + (1 + \epsilon) \left\| \phi - \tilde{\phi} \right\|_{\text{SC}(\mathbf{L}(\bar{\mathbf{w}}), C)}.$$

We start by calculating

$$\begin{aligned} \Pi_\phi(\mathbf{v}) &= \bar{\mathbf{W}}^{1/2} \mathbf{B}\mathcal{H}_C \phi = \bar{\mathbf{W}}^{1/2} \mathbf{B}\mathcal{H}_C (\phi - \tilde{\phi}) + \bar{\mathbf{W}}^{1/2} \mathbf{B}\mathcal{H}_C \tilde{\phi} \\ &= \bar{\mathbf{W}}^{1/2} \mathbf{B}\mathcal{H}_C (\phi - \tilde{\phi}) + (\bar{\mathbf{W}}^{1/2} - \mathbf{W}^{1/2}) \mathbf{B}\mathcal{H}_C \tilde{\phi} + \mathbf{W}^{1/2} \mathbf{B}\mathcal{H}_C \tilde{\phi} \\ &\stackrel{(i)}{=} \bar{\mathbf{W}}^{1/2} \mathbf{B}\mathcal{H}_C (\phi - \tilde{\phi}) + (\bar{\mathbf{W}}^{1/2} - \mathbf{W}^{1/2}) \mathbf{B}\tilde{\phi} + \mathbf{W}^{1/2} \mathbf{B}\mathcal{H}_C \tilde{\phi} \\ &= \bar{\mathbf{W}}^{1/2} \mathbf{B}\mathcal{H}_C (\phi - \tilde{\phi}) + \mathbf{W}^{1/2} \mathbf{B}(\mathcal{H}_C - \tilde{\mathcal{H}}) \tilde{\phi} + \tilde{\Pi}_\phi(\mathbf{v}), \end{aligned}$$

where (i) follows from  $(\bar{\mathbf{W}}^{1/2} - \mathbf{W}^{1/2}) \mathbf{B}\mathcal{H}_C = (\bar{\mathbf{W}}^{1/2} - \mathbf{W}^{1/2}) \mathbf{B}$  as  $\bar{\mathbf{w}} = \mathbf{w}$  outside  $C$ . Hence

$$\Pi_\phi(\mathbf{v}) - \tilde{\Pi}_\phi(\mathbf{v}) = \bar{\mathbf{W}}^{1/2} \mathbf{B}\mathcal{H}_C (\phi - \tilde{\phi}) + \mathbf{W}^{1/2} \mathbf{B}(\mathcal{H}_C - \tilde{\mathcal{H}}) \tilde{\phi}. \quad (11)$$

We bound both terms of (11) separately. For the first term, note that

$$\left\| \bar{\mathbf{W}}^{1/2} \mathbf{B}\mathcal{H}_C (\phi - \tilde{\phi}) \right\|_2 \stackrel{(i)}{\leq} \left\| \phi - \tilde{\phi} \right\|_{\text{SC}(\mathbf{L}(\bar{\mathbf{w}}), C)},$$

where (i) follows from properties of  $\mathcal{H}_C$ . For the second term of (11),

$$\left\| \mathbf{W}^{1/2} \mathbf{B}(\mathcal{H}_C - \tilde{\mathcal{H}}) \tilde{\phi} \right\|_2 \stackrel{(i)}{\leq} \epsilon \left\| \tilde{\phi} \right\|_{\text{SC}(\mathbf{L}(\bar{\mathbf{w}}), C)} \leq \epsilon \left\| \tilde{\phi} \right\|_{\text{SC}(\mathbf{L}(\bar{\mathbf{w}}), C)} \stackrel{(ii)}{\leq} \epsilon \|\mathbf{v}\|_2 + \left\| \phi - \tilde{\phi} \right\|_{\text{SC}(\mathbf{L}(\bar{\mathbf{w}}), C)}$$

where (i) follows from (9), and (ii) is because

$$\left\| \tilde{\phi} \right\|_{\text{SC}(\mathbf{L}(\bar{\mathbf{w}}), C)} \leq \left\| \tilde{\phi} - \phi \right\|_{\text{SC}(\mathbf{L}(\bar{\mathbf{w}}), C)} + \left\| \phi \right\|_{\text{SC}(\mathbf{L}(\bar{\mathbf{w}}), C)} \leq \left\| \phi - \tilde{\phi} \right\|_{\text{SC}(\mathbf{L}(\bar{\mathbf{w}}), C)} + \|\mathbf{v}\|_{\mathbf{P}_{\bar{\mathbf{w}}}} \leq \|\mathbf{v}\|_2$$

because  $\mathbf{P}_{\bar{\mathbf{w}}}$  is an orthogonal projection matrix. Summing these errors completes the proof for

$$\left\| \Pi_\phi(\mathbf{v}) - \tilde{\Pi}_\phi(\mathbf{v}) \right\|_2 \leq \epsilon \|\mathbf{v}\|_2 + (1 + \epsilon) \left\| \phi - \tilde{\phi} \right\|_{\text{SC}(\mathbf{L}(\bar{\mathbf{w}}), C)}. \quad (12)$$

We then define

$$\tilde{\Pi}_\psi(\mathbf{v}) \stackrel{\text{def}}{=} (\mathbf{W}^{1/2} - \bar{\mathbf{W}}^{1/2}) \mathbf{B}\tilde{\phi} + \mathbf{W}^{1/2} \mathbf{B}\tilde{\mathcal{H}}\tilde{\phi}$$

which is the third term of  $\Pi\mathbf{v}$  and

$$\Pi_\psi(\mathbf{v}) \stackrel{\text{def}}{=} \bar{\mathbf{W}}^{1/2} \mathbf{B}\mathcal{H}_C \psi.$$

Then, by the proof above with  $\bar{\mathbf{w}}$  replaced by  $\mathbf{w}$  (and  $\bar{\mathbf{W}}$  replaced by  $\mathbf{W}$ ), we get

$$\left\| \Pi_\psi(\mathbf{v}) - \tilde{\Pi}_\psi(\mathbf{v}) \right\|_2 \leq \epsilon \|\mathbf{v}\|_2 + (1 + \epsilon) \left\| \psi - \tilde{\psi} \right\|_{\text{SC}(\mathbf{L}(\mathbf{w}), C)}. \quad (13)$$

Recall that

$$\begin{aligned} \Pi \mathbf{v} = & \bar{\mathbf{W}}^{1/2} \mathbf{B} \mathcal{H}_C \mathbf{S}\mathbf{C}(\mathbf{L}(\bar{\mathbf{w}}), C)^\dagger \mathcal{H}_C^\top \mathbf{B}^\top \bar{\mathbf{W}}^{1/2} \mathbf{v} - \mathbf{W}^{1/2} \mathbf{B} \mathcal{H}_C \mathbf{S}\mathbf{C}(\mathbf{L}(\mathbf{w}), C)^\dagger \mathcal{H}_C^\top \mathbf{B}^\top \mathbf{W}^{1/2} \mathbf{v} \\ & + \mathbf{W}^{1/2} \mathbf{B} \mathbf{L}(\mathbf{w})^\dagger \mathbf{B}^\top \mathbf{W}^{1/2} \mathbf{v}. \end{aligned}$$

Its first two terms are approximated respectively by  $\tilde{\Pi}_\phi(\mathbf{v})$  and  $\tilde{\Pi}_\psi(\mathbf{v})$ , the first two terms in  $\tilde{\Pi} \mathbf{v}$ . Its last term is exactly the last term of  $\tilde{\Pi} \mathbf{v}$ . By triangle inequality and (12), (13), we have

$$\|\Pi \mathbf{v} - \tilde{\Pi} \mathbf{v}\|_2 \leq 2\epsilon \|\mathbf{v}\|_2 + (1 + \epsilon) \|\phi - \tilde{\phi}\|_{\mathbf{S}\mathbf{C}(\mathbf{L}(\bar{\mathbf{w}}), C)} + (1 + \epsilon) \|\psi - \tilde{\psi}\|_{\mathbf{S}\mathbf{C}(\mathbf{L}(\mathbf{w}), C)}.$$

□

In the following sections, we will use  $\text{SOLVE}(\mathbf{L}, \mathbf{b})$  to denote a high accuracy Laplacian solver that returns  $\mathbf{x}$  such that  $\mathbf{L}\mathbf{x} = \mathbf{b}$  and runs in nearly linear time. We will overload notation to extend any dimension of a matrix from a subset of  $V$  to  $V$ , or from a subset of  $E$  to  $E$  by padding zeroes.

## 5.2 Dynamic Potential Maintenance

In this section, we show how to maintain the vector  $\tilde{\phi}$  (Lemma 5.5) that approximates the potential vector  $\phi$ . This data structure can also be used for  $\psi$  (Corollary 5.6).

**Lemma 5.8** (Dynamic Potential). *For a graph  $G = (V, E)$  with dynamic edge conductances  $\bar{\mathbf{w}} \in \mathbb{R}_{\geq 0}^{E(G)}$  and a dynamic vector  $\bar{\mathbf{v}} \in \mathbb{R}^{E(G)}$  for some constant  $C$ , there is a data structure (Algorithm 4) that supports the following operations against an oblivious adversary for parameters  $\beta < \epsilon^2 < 1$ .*

- *INITIALIZE( $G, \mathbf{w}, \mathbf{v}^{(\text{init})}, \beta, \epsilon$ ). Initializes the data structure in time  $\tilde{O}(m\beta^{-2}\epsilon^{-2})$  with an empty set  $Z \leftarrow \emptyset$  of marked edges. Initialize  $\bar{\mathbf{w}}$  as  $\mathbf{w}$  and  $\bar{\mathbf{v}}$  as  $\mathbf{v}^{(\text{init})}$ .*
- *UPDATEV( $e, \bar{\mathbf{v}}^{\text{new}}$ ). Updates  $\bar{\mathbf{v}}_e \leftarrow \bar{\mathbf{v}}^{\text{new}}$  in  $\tilde{O}(\beta^{-2}\epsilon^{-2})$  time.*
- *UPDATEW( $e, \bar{\mathbf{w}}^{\text{new}}$ ). Updates  $\bar{\mathbf{w}}_e \leftarrow \bar{\mathbf{w}}^{\text{new}}$  in  $\tilde{O}(\beta^{-2}\epsilon^{-2})$  time.*
- *QUERYPOTENTIAL(). For  $C \subseteq V$  with  $|C| = O(\beta m)$  and  $Z \subseteq E(C)$ , returns in  $\tilde{O}(\beta m \epsilon^{-2})$  time a vector  $\tilde{\phi}$  satisfying  $\|\phi - \tilde{\phi}\|_{\mathbf{S}\mathbf{C}(\mathbf{L}(\bar{\mathbf{w}}), C)} \leq \epsilon \|\bar{\mathbf{v}}\|_2$  where  $\phi = \mathbf{S}\mathbf{C}(\mathbf{L}(\bar{\mathbf{w}}), C)^\dagger \mathcal{H}_C^\top \mathbf{B}^\top \bar{\mathbf{W}}^{1/2} \bar{\mathbf{v}}$ .*

*Runtimes and output correctness hold w.h.p. if there are at most  $O(\beta m)$  calls to UPDATEV and UPDATEW in total.*

*Proof.* The pseudocode for the proof of Lemma 5.8 is in Algorithm 4. At a high-level, it simply maintains  $\tilde{\phi}$  as in Lemma 5.5. The one difference is that it handles changes to  $\mathbf{v}^{(\text{init})}$  directly because both endpoints of all edges changed in  $\mathbf{v}^{(\text{init})}$  are in the marked set  $Z$ . We start by analyzing the correctness of the algorithm, then move the runtime.

**Correctness.** We only need to analyze the `QUERYPOTENTIAL()` operation. In this proof, we show the weaker bound  $\|\tilde{\phi} - \phi\|_{\text{SC}(\mathbf{L}(\bar{\mathbf{w}}), C)} \leq 4\epsilon \left( \|\bar{\mathbf{v}}\|_2 + \|\mathbf{v}^{(\text{init})}\|_2 \right)$ . However, this suffices because we can build  $\tilde{O}(1)$  copies of the data structure. For  $-\tilde{O}(1) \leq j \leq \tilde{O}(1)$ , the  $j$ -th instance initializes and answers queries only when  $\|\bar{\mathbf{v}}\|_2 \in (2^j, 2^{j+1}]$ . Updates are passed to all instances. When the number of updates exceeds  $O(\beta m)$  for an instance but it cannot be initialized because  $\|\bar{\mathbf{v}}\|_2$  does not fall in its range, it simply ignore following updates until it can be initialized. This only increases the runtime by  $\tilde{O}(1)$  factors.

Let  $\widetilde{\text{SC}}_{\mathcal{H}}$  be the value of  $D^{(\text{sc})}.\text{INITIALSC}()$  returned in line 23 of Algorithm 4. It satisfies condition (9) by the guarantees of Theorem 4.10. Also, by inspection of the procedure `QUERYPOTENTIAL()` in Algorithm 4, the returned vector  $\tilde{\phi}$  is defined as

$$\begin{aligned} \tilde{\phi} &= \widetilde{\text{SC}}^\dagger \begin{bmatrix} \widetilde{\text{SC}}_{\mathcal{H}} & 0 \end{bmatrix} \mathbf{d}^{(\text{init})} + \widetilde{\text{SC}}^\dagger \mathbf{B} \mathbf{W}^{1/2} (\bar{\mathbf{v}}_Z - \mathbf{v}_Z^{(\text{init})}) + \widetilde{\text{SC}}^\dagger \mathbf{B} (\bar{\mathbf{W}}^{1/2} - \mathbf{W}^{1/2}) \bar{\mathbf{v}}_Z \\ &= \widetilde{\text{SC}}^\dagger \begin{bmatrix} \widetilde{\text{SC}}_{\mathcal{H}} & 0 \end{bmatrix} \mathbf{L}(\mathbf{w})^\dagger \mathbf{B} \mathbf{W}^{1/2} \mathbf{v}^{(\text{init})} + \widetilde{\text{SC}}^\dagger \mathbf{B} \mathbf{W}^{1/2} (\bar{\mathbf{v}}_Z - \mathbf{v}_Z^{(\text{init})}) + \widetilde{\text{SC}}^\dagger \mathbf{B} (\bar{\mathbf{W}}^{1/2} - \mathbf{W}^{1/2}) \bar{\mathbf{v}}_Z \\ &= \widetilde{\text{SC}}^\dagger \tilde{\mathcal{H}}_C \mathbf{B} \mathbf{W}^{1/2} \mathbf{v}^{(\text{init})} + \widetilde{\text{SC}}^\dagger \mathbf{B} (\bar{\mathbf{W}}^{1/2} - \mathbf{W}^{1/2}) \mathbf{v}^{(\text{init})} + \widetilde{\text{SC}}^\dagger \mathbf{B} \bar{\mathbf{W}}^{1/2} (\bar{\mathbf{v}}_Z - \mathbf{v}_Z^{(\text{init})}). \end{aligned}$$

Additionally, because  $\mathbf{v}^{(\text{init})} - \bar{\mathbf{v}}$  is supported on  $Z$ , the true  $\phi$  can be written as

$$\begin{aligned} \phi &= \text{SC}(\mathbf{L}(\bar{\mathbf{w}}), C)^\dagger \mathcal{H}_C^\top \bar{\mathbf{B}} \bar{\mathbf{W}}^{1/2} \bar{\mathbf{v}} \\ &= \text{SC}(\mathbf{L}(\bar{\mathbf{w}}), C)^\dagger \mathcal{H}_C^\top \bar{\mathbf{B}} \bar{\mathbf{W}}^{1/2} \mathbf{v}^{(\text{init})} + \text{SC}(\mathbf{L}(\bar{\mathbf{w}}), C)^\dagger \bar{\mathbf{B}} \bar{\mathbf{W}}^{1/2} (\bar{\mathbf{v}} - \mathbf{v}^{(\text{init})}). \end{aligned}$$

Hence we get that

$$\begin{aligned} &\|\tilde{\phi} - \phi\|_{\text{SC}(\mathbf{L}(\bar{\mathbf{w}}), C)} \\ &\leq \left\| \widetilde{\text{SC}}^\dagger \tilde{\mathcal{H}}_C \mathbf{B} \mathbf{W}^{1/2} \mathbf{v}^{(\text{init})} + \widetilde{\text{SC}}^\dagger \mathbf{B} (\bar{\mathbf{W}}^{1/2} - \mathbf{W}^{1/2}) \mathbf{v}^{(\text{init})} - \text{SC}(\mathbf{L}(\bar{\mathbf{w}}), C)^\dagger \mathcal{H}_C^\top \bar{\mathbf{B}} \bar{\mathbf{W}}^{1/2} \mathbf{v}^{(\text{init})} \right\|_{\text{SC}(\mathbf{L}(\bar{\mathbf{w}}), C)} \\ &\quad + \left\| \widetilde{\text{SC}}^\dagger \bar{\mathbf{B}} \bar{\mathbf{W}}^{1/2} (\bar{\mathbf{v}}_Z - \mathbf{v}_Z^{(\text{init})}) - \text{SC}(\mathbf{L}(\bar{\mathbf{w}}), C)^\dagger \bar{\mathbf{B}} \bar{\mathbf{W}}^{1/2} (\bar{\mathbf{v}} - \mathbf{v}^{(\text{init})}) \right\| \\ &\stackrel{(i)}{\leq} 3\epsilon \|\mathbf{v}^{(\text{init})}\|_2 + \epsilon \|\bar{\mathbf{v}} - \mathbf{v}^{(\text{init})}\|_2 \leq 4\epsilon \left( \|\bar{\mathbf{v}}\|_2 + \|\mathbf{v}^{(\text{init})}\|_2 \right), \end{aligned}$$

where (i) follows from Lemma 5.5 for the first term, and  $\widetilde{\text{SC}} \approx_\epsilon \text{SC}(\mathbf{L}(\bar{\mathbf{w}}), C)$  from the guarantee of Theorem 4.10 for the second term. This suffices because Algorithm 4 set  $\epsilon \leftarrow \epsilon/10$  in line 2.

**Runtime.** The runtimes of `UPDATEV`, `UPDATEW` are trivially the same as the runtime of `MARK`. The runtimes of `MARK` and `INITIALIZE` follows from the `ADDTerminal` and `INITIALIZE` operations respectively of Theorem 4.10. The runtime of `QUERYPOTENTIAL` is  $\tilde{O}(\beta m \epsilon^{-2})$  by the runtime guarantees of `SC` and `INITIALSC` of Theorem 4.10, and the fact that  $\widetilde{\text{SC}}$  and  $\widetilde{\text{SC}}_{\mathcal{H}}$  all have  $\tilde{O}(\beta m \epsilon^{-2})$  edges, and hence solving or multiplying by them costs  $\tilde{O}(\beta m \epsilon^{-2})$  time.  $\square$

### 5.3 Dynamic Evaluator

**Theorem 5.9** (Dynamic Evaluator). *For a graph  $G = (V, E)$  with dynamic edge conductances  $\bar{\mathbf{w}} \in \mathbb{R}_{\geq 0}^{E(G)}$  and a dynamic vector  $\bar{\mathbf{v}} \in \mathbb{R}^{E(G)}$ , there is a data structure `EVALUATOR` that supports the following operations against an oblivious adversary for parameters  $\beta < \epsilon^2 < 1$ .*

---

**Algorithm 4: Dynamic Potential**


---

```

// This implementation assumes that  $\|\bar{v}\|_2 \approx \|v^{(\text{init})}\|_2$  always. However, this
// can be achieved by duplicating the data structure  $\tilde{O}(1)$  times, one
// handling each range  $\|\bar{v}\|_2 \in [2^j, 2^{j+1}]$  for  $-\tilde{O}(1) \leq j \leq \tilde{O}(1)$ .
1 procedure INITIALIZE( $G, w, v^{(\text{init})}, \beta, \epsilon$ )
2    $\epsilon \leftarrow \epsilon/10$ .
3   Let  $D^{(\text{sc})}$  be a instance of the dynamic Schur complement data structure of
   Theorem 4.10.
4    $D^{(\text{sc})}.\text{INITIALIZE}(G, w, \epsilon, \beta)$ .
5    $\bar{v} \leftarrow v^{(\text{init})}$ .
   //  $v^{(\text{init})}$  is the initial vector and  $\bar{v}$  to denote the current vector  $v$ 
   // throughout the algorithm.
6    $\bar{w} \leftarrow w$ .
   // We use  $w$  to denote the initial vector and  $\bar{w}$  to denote the current
   // vector  $w$  throughout the algorithm.
7    $d^{(\text{init})} \leftarrow \text{SOLVE}(\mathbf{L}(w), \mathbf{B}\mathbf{W}^{1/2}v^{(\text{init})})$ .
8    $Z \leftarrow \emptyset$ . // Marked edges.
9 procedure MARK( $e$ )
10   $D^{(\text{sc})}.\text{ADDTERMINAL}(u)$ .
11   $D^{(\text{sc})}.\text{ADDTERMINAL}(v)$ .
12   $Z \leftarrow Z \cup \{e\}$ .
13   $D^{(\text{sc})}.\text{UPDATE}(e, w_e)$ . // Make sure the  $D^{(\text{sc})}$  puts edge  $e$  in  $Z$ .
14 procedure UPDATEV( $e, v^{\text{new}}$ )
15  MARK( $e$ ).
16   $\bar{v}_e \leftarrow v^{\text{new}}$ .
17 procedure UPDATEW( $e, w^{\text{new}}$ )
18  MARK( $e$ ).
19   $D^{(\text{sc})}.\text{UPDATE}(e, w^{\text{new}})$ .
20   $\bar{w}_e \leftarrow w^{\text{new}}$ .
21 procedure QUERYPOTENTIAL()
22   $\widetilde{\mathbf{SC}} \leftarrow D^{(\text{sc})}.\text{SC}()$ .
23   $\tilde{\phi} \leftarrow \text{SOLVE}(\widetilde{\mathbf{SC}}, \begin{bmatrix} D^{(\text{sc})}.\text{INITIALSC}() & 0 \end{bmatrix} d^{(\text{init})})$ .
24   $\tilde{\phi} \leftarrow \tilde{\phi} + \text{SOLVE}(\widetilde{\mathbf{SC}}, \mathbf{B}\mathbf{W}^{1/2}(\bar{v}_Z - v_Z^{(\text{init})}))$ .
25   $\tilde{\phi} \leftarrow \tilde{\phi} + \text{SOLVE}(\widetilde{\mathbf{SC}}, \mathbf{B}(\bar{\mathbf{W}}^{1/2} - \mathbf{W}^{1/2})\bar{v}_Z)$ .
26  return  $\tilde{\phi}$ .

```

---

- $\text{INITIALIZE}(G, w, v^{(\text{init})}, \beta, \epsilon)$ . Initializes the data structure in time  $\tilde{O}(m\beta^{-2}\epsilon^{-2})$  with an empty set  $Z \leftarrow \emptyset$  of marked edges. Initializes  $\bar{w}$  as  $w$ ,  $\bar{v}$  as  $v^{(\text{init})}$ .
- $\text{UPDATEV}(e, v^{\text{new}})$ . Updates  $\bar{v}_e \leftarrow v^{\text{new}}$  in  $\tilde{O}(\beta^{-2}\epsilon^{-2})$  time.
- $\text{UPDATEW}(e, w^{\text{new}})$ . Updates  $\bar{w}_e \leftarrow w^{\text{new}}$  in  $\tilde{O}(\beta^{-2}\epsilon^{-2})$  time.

- $\text{QUERY}()$ . Returns a vector  $\mathbf{u} \in \mathbb{R}^Z$  satisfying  $\|\mathbf{u} - [\mathbf{P}_{\bar{\mathbf{w}}}\bar{\mathbf{v}}]_Z\|_2 \leq \epsilon\|\bar{\mathbf{v}}\|_2 + \epsilon\|\mathbf{v}\|_2$  in time  $\tilde{O}(\beta m \epsilon^{-2})$ .

Runtimes and output correctness hold w.h.p. if there are at most  $O(\beta m)$  calls to  $\text{UPDATEV}$ ,  $\text{UPDATEW}$  in total.

*Proof.* We decompose  $\mathbf{P}_{\bar{\mathbf{w}}}\bar{\mathbf{v}}$  by (6). The  $\mathbf{v}$  in (6) is the current vector  $\bar{\mathbf{v}}$  and the  $\hat{\mathbf{v}}$  in (6) is the initial vector  $\mathbf{v}$  here. We create two instances of Algorithm 4  $D^{(\phi)}$  and  $D^{(\psi)}$  maintaining

$$\tilde{\phi} = \widetilde{\mathbf{SC}}^\dagger \mathbf{B}(\bar{\mathbf{W}}^{1/2} - \mathbf{W}^{1/2})\bar{\mathbf{v}} + \widetilde{\mathbf{SC}}^\dagger \tilde{\mathcal{H}}^\top \mathbf{B}\mathbf{W}^{1/2}\bar{\mathbf{v}}$$

(Lemma 5.5) and

$$\tilde{\psi} = \widetilde{\mathbf{SC}}(\mathbf{L}(\mathbf{w}), C)^\dagger \tilde{\mathcal{H}}^\top \mathbf{B}\mathbf{W}^{1/2}\mathbf{v}$$

(Corollary 5.6) respectively.  $\text{INITIALIZE}$ ,  $\text{MARK}$  are forwarded to both  $D^{(\phi)}$  and  $D^{(\psi)}$ . The operations  $\text{UPDATEV}$ ,  $\text{UPDATEW}$  are forwarded only to  $D^{(\phi)}$  as  $D^{(\psi)}$  maintains  $\tilde{\psi}$  where  $\mathbf{w}$  and  $\mathbf{v}$  do not change. We also compute the exact value of the last term  $\mathbf{W}^{1/2}\mathbf{B}\mathbf{L}(\mathbf{w})^\dagger \mathbf{B}^\top \mathbf{W}^{1/2}\mathbf{v}$  of  $\mathbf{P}_{\bar{\mathbf{w}}}\bar{\mathbf{v}}$  by

$$\mathbf{x} = \mathbf{W}^{1/2}\mathbf{B}\text{SOLVE}(\mathbf{L}(\mathbf{w}), \mathbf{B}^\top \mathbf{W}^{1/2}\mathbf{v}).$$

For  $\text{QUERY}()$ , let  $\tilde{\phi} = D^{(\phi)}.\text{QUERYPOTENTIAL}()$ ,  $\tilde{\psi} = D^{(\psi)}.\text{QUERYPOTENTIAL}()$ . The EVALUATOR returns

$$\mathbf{u} = [\bar{\mathbf{W}}^{1/2}\mathbf{B}\tilde{\phi}]_Z + [\mathbf{W}^{1/2}\mathbf{B}\tilde{\psi}]_Z + \mathbf{x}_Z.$$

Clearly all runtimes transfer exactly from Lemma 5.8. It suffices to show the correctness of  $\text{QUERY}()$ .

For the true potentials  $\phi = \mathbf{SC}(\mathbf{L}(\bar{\mathbf{w}}), C)^\dagger \mathcal{H}_C^\top \mathbf{B}\bar{\mathbf{W}}^{1/2}\bar{\mathbf{v}}$  and  $\psi = \mathbf{SC}(\mathbf{L}(\mathbf{w}), C)^\dagger \mathcal{H}_C^\top \mathbf{B}\mathbf{W}^{1/2}\mathbf{v}$  we have  $[\mathbf{P}_{\bar{\mathbf{w}}}\bar{\mathbf{v}}]_Z = [\bar{\mathbf{W}}^{1/2}\mathbf{B}\phi + \mathbf{W}^{1/2}\mathbf{B}\psi]_Z + \mathbf{x}_Z$ . Thus,

$$\begin{aligned} \|\mathbf{u} - [\mathbf{P}_{\bar{\mathbf{w}}}\bar{\mathbf{v}}]_Z\|_2 &= \|\bar{\mathbf{W}}_Z^{1/2}\mathbf{B}_Z(\tilde{\phi} - \phi) + \mathbf{W}_Z^{1/2}\mathbf{B}_Z(\tilde{\psi} - \psi)\|_2 \\ &\leq \|\bar{\mathbf{W}}_Z^{1/2}\mathbf{B}_Z(\tilde{\phi} - \phi)\|_2 + \|\mathbf{W}_Z^{1/2}\mathbf{B}_Z(\tilde{\psi} - \psi)\|_2 \\ &= \|\tilde{\phi} - \phi\|_{\mathbf{B}_Z^\top \bar{\mathbf{W}}_Z \mathbf{B}_Z} + \|\tilde{\psi} - \psi\|_{\mathbf{B}_Z^\top \mathbf{W}_Z \mathbf{B}_Z} \\ &\stackrel{(i)}{\leq} \|\tilde{\phi} - \phi\|_{\mathbf{SC}(\mathbf{L}(\bar{\mathbf{w}}), C)} + \|\tilde{\psi} - \psi\|_{\mathbf{SC}(\mathbf{L}(\mathbf{w}), C)} \\ &\stackrel{(ii)}{\leq} \epsilon\|\bar{\mathbf{v}}\|_2 + \epsilon\|\mathbf{v}\|_2 \end{aligned}$$

where (i) follows from the fact that  $\mathbf{L}(\bar{\mathbf{w}}_Z) \preceq \mathbf{SC}(\mathbf{L}(\bar{\mathbf{w}}), C)$  (and  $\mathbf{L}(\mathbf{w}_Z) \preceq \mathbf{SC}(\mathbf{L}(\mathbf{w}), C)$ ) as  $Z$  is completely inside  $C$ , and (ii) follows from the guarantee of  $\text{QUERYPOTENTIAL}()$  of Lemma 5.8. This completes the proof.  $\square$

## 5.4 Dynamic Locator

**Theorem 5.10** (Dynamic Locator). *For a graph  $G = (V, E)$  with dynamic edge conductances  $\bar{\mathbf{w}} \in \mathbb{R}_{\geq 0}^{E(G)}$  and a dynamic vector  $\bar{\mathbf{v}} \in \mathbb{R}^{E(G)}$ , there is a data structure  $\text{LOCATOR}$  (given in Algorithm 5) that supports the following operations against an oblivious adversary for parameters  $\beta < \epsilon^2 < 1$ .*



- $\text{INITIALIZE}(G, \mathbf{w}, \mathbf{v}^{(\text{init})}, \beta, \epsilon)$ . Initializes the data structure in time  $\tilde{O}(m\beta^{-2}\epsilon^{-2})$  and sets  $\overline{\mathbf{w}} \leftarrow \mathbf{w}$  and  $\overline{\mathbf{v}} \leftarrow \mathbf{v}^{(\text{init})}$ .
- $\text{UPDATEV}(e, \overline{\mathbf{v}}^{\text{new}})$ . Updates  $\overline{\mathbf{v}}_e \leftarrow \overline{\mathbf{v}}^{\text{new}}$  in  $\tilde{O}(\beta^{-2}\epsilon^{-2})$  time.
- $\text{UPDATEW}(e, \overline{\mathbf{w}}^{\text{new}})$ . Updates  $\overline{\mathbf{w}}_e \leftarrow \overline{\mathbf{w}}^{\text{new}}$  in  $\tilde{O}(\beta^{-2}\epsilon^{-2})$  time.
- $\text{LOCATE}()$ . Returns in time  $\tilde{O}(\beta m \epsilon^{-2})$  a set  $S \subseteq E(G)$  with  $|S| \leq O(\epsilon^{-2})$  containing all edges  $e$  with  $|\llbracket \mathbf{P}_{\mathbf{w}} \overline{\mathbf{v}} \rrbracket_e| \geq \epsilon \|\overline{\mathbf{v}}\|_2$  whp.

Runtimes and output correctness hold w.h.p. if there are at most  $O(\beta m)$  calls to  $\text{UPDATEV}$ ,  $\text{UPDATEW}$  in total.

The following lemma is implicit in [KNPW11] and allows us to recover the large entries of  $\mathbf{x}$  by a low-dimensional projection of it.

**Lemma 5.11** ( $\ell_2$ -heavy hitter, [KNPW11]). *There exists a function  $\text{SKETCH}(\epsilon, n)$  that given  $\epsilon > 0$  explicitly returns a random matrix  $\mathbf{Q} \in \mathbb{R}^{N \times m}$  with  $N = O(\epsilon^{-2} \log^3 m)$  and column sparsity  $c = O(\log^3 m)$  in  $\tilde{O}(N + m)$  time, and uses  $\tilde{O}(N + m)$  spaces to store the matrix  $\mathbf{Q}$ . There further exists a function  $\text{RECOVER}(\mathbf{Q}\mathbf{x})$  that in time  $O(\epsilon^{-2} \log^3 m)$  reports a list  $S \subset [m]$  of size  $O(\epsilon^{-2})$ . For any fixed  $\mathbf{x}$ , the list includes all  $i$  with  $|\mathbf{x}_i| \geq \epsilon \|\mathbf{x}\|_2$  with high probability over the randomness of  $\mathbf{Q}$ .*

*Proof of Theorem 5.10.* At a high level, Algorithm 5 simply maintains the formula given by Lemma 5.7 for  $\tilde{\phi}$  and  $\tilde{\psi}$  given by the output of the dynamic potential maintenance data structure in Lemma 5.8. Let us first show correctness and then analyze runtime. We only have to check correctness of  $\text{LOCATE}()$ . We follow the decomposition (6) with both  $\mathbf{v}$  and  $\hat{\mathbf{v}}$  being the current vector  $\overline{\mathbf{v}}$  here. The  $\tilde{\phi}$  and  $\tilde{\psi}$  maintained by Algorithm 5 satisfy

$$\tilde{\phi} = \widetilde{\mathbf{S}\mathbf{C}}^\dagger \mathbf{B}(\overline{\mathbf{W}}^{1/2} - \mathbf{W}^{1/2})\mathbf{v} + \widetilde{\mathbf{S}\mathbf{C}}^\dagger \tilde{\mathcal{H}}^\top \mathbf{B}\mathbf{W}^{1/2}\overline{\mathbf{v}}$$

and

$$\tilde{\psi} = \widetilde{\mathbf{S}\mathbf{C}}^\dagger \tilde{\mathcal{H}}^\top \mathbf{B}\mathbf{W}^{1/2}\overline{\mathbf{v}}.$$

(Note that  $\tilde{\psi}$  is defined differently from Theorem 5.9.)

Thus, for  $\widetilde{\mathbf{S}\mathbf{C}}_{\mathcal{H}} = D^{(\text{sc})}.\text{INITIALSC}()$ ,  $\mathbf{p}$  as defined in  $\text{LOCATE}()$  of Algorithm 5 satisfies

$$\begin{aligned} \mathbf{p} &= \mathbf{Q}(\overline{\mathbf{W}}^{1/2} - \mathbf{W}^{1/2})\mathbf{B}\tilde{\phi} + \Gamma \begin{bmatrix} 0 \\ D^{(\text{sc})}.\text{INITIALSC}() \end{bmatrix} \tilde{\phi} + \Gamma \begin{bmatrix} 0 \\ D^{(\text{sc})}.\text{INITIALSC}() \end{bmatrix} \tilde{\psi} + \Gamma \mathbf{y} \\ &= \mathbf{Q}(\overline{\mathbf{W}}^{1/2} - \mathbf{W}^{1/2})\mathbf{B}\tilde{\phi} + \begin{bmatrix} 0 \\ \mathbf{Q}\mathbf{W}^{1/2}\mathbf{B}\mathbf{L}(\mathbf{w})^\dagger \widetilde{\mathbf{S}\mathbf{C}}_{\mathcal{H}} \end{bmatrix} \tilde{\phi} + \begin{bmatrix} 0 \\ \mathbf{Q}\mathbf{W}^{1/2}\mathbf{B}\mathbf{L}(\mathbf{w})^\dagger \widetilde{\mathbf{S}\mathbf{C}}_{\mathcal{H}} \end{bmatrix} \tilde{\psi} \\ &\quad + \mathbf{Q}\mathbf{W}^{1/2}\mathbf{B}\mathbf{L}(\mathbf{w})^\dagger \mathbf{B}^\top \mathbf{W}^{1/2}\overline{\mathbf{v}} \\ &= \mathbf{Q}\tilde{\Pi}\overline{\mathbf{v}} \end{aligned}$$

for  $\tilde{\Pi}\overline{\mathbf{v}}$  as defined in Lemma 5.7. Note that  $\tilde{\phi}$  is padded with zeroes for computing  $\mathbf{B}\tilde{\phi}$  as in Algorithm 5. Because

$$\|\phi - \tilde{\phi}\|_{\text{SC}(\mathbf{L}(\overline{\mathbf{w}}), C)} \leq \epsilon \|\overline{\mathbf{v}}\|_2 \quad (14)$$

---

**Algorithm 5: Dynamic Locator**


---

```

1 procedure INITIALIZE( $G, \mathbf{w}, \mathbf{v}^{(\text{init})}, \beta, \epsilon$ )
2    $\epsilon \leftarrow \epsilon/10$ .
3   Let  $D^{(\phi)}$  be an instance of the dynamic potential data structure of Lemma 5.8.
4   Let  $D^{(\psi)}$  be an instance of the dynamic potential data structure of Lemma 5.8.
5   Let  $D^{(\text{sc})}$  be an instance of the dynamic Schur complement data structure of
      Theorem 4.10.
6    $D^{(\phi)}$ .INITIALIZE( $G, \mathbf{w}, \mathbf{v}^{(\text{init})}, \epsilon, \beta$ ).
7    $D^{(\psi)}$ .INITIALIZE( $G, \mathbf{w}, \mathbf{v}^{(\text{init})}, \epsilon, \beta$ ).
8    $D^{(\text{sc})}$ .INITIALIZE( $G, \mathbf{w}, \epsilon, \beta$ ).
9   Initialize an  $N = O(\epsilon^{-2} \log^3 m)$  by  $m$  matrix  $\mathbf{Q}$  with rows  $\mathbf{q}^{(1)}, \mathbf{q}^{(2)}, \dots, \mathbf{q}^{(N)} \in \mathbb{R}^m$  by
      Lemma 5.11.
10  for  $i \in [N]$  do
11     $\gamma^{(i)} \leftarrow \text{SOLVE}(\mathbf{L}(\mathbf{w}), \mathbf{B}^\top \mathbf{W}^{1/2} \mathbf{q}^{(i)})$ . //  $\gamma^{(i)}$  are rows of  $\mathbf{\Gamma} \stackrel{\text{def}}{=} \mathbf{Q} \mathbf{W}^{1/2} \mathbf{B} \mathbf{L}(\mathbf{w})^\dagger$ .
12  end
13   $\bar{\mathbf{w}} \leftarrow \mathbf{w}, \bar{\mathbf{v}} \leftarrow \mathbf{v}, \mathbf{y} \leftarrow \mathbf{B}^\top \mathbf{W} \mathbf{v}$ .
14  procedure UPDATEV( $e, \mathbf{v}^{\text{new}}$ )
15     $D^{(\phi)}$ .UPDATEV( $e, \mathbf{v}^{\text{new}}$ ).
16     $D^{(\psi)}$ .UPDATEV( $e, \mathbf{v}^{\text{new}}$ ).
17     $D^{(\text{sc})}$ .UPDATEV( $e, \mathbf{v}^{\text{new}}$ ).
18     $\bar{\mathbf{v}}_e \leftarrow \mathbf{v}^{\text{new}}, \mathbf{y} \leftarrow \mathbf{y} + \mathbf{B}^\top \mathbf{W}(\mathbf{v}^{\text{new}} - \mathbf{v}_e)$ .
19  procedure UPDATEW( $e, \mathbf{w}^{\text{new}}$ )
20     $D^{(\phi)}$ .UPDATEW( $e, \mathbf{w}^{\text{new}}$ ).
21    //  $D^{(\psi)}$  does not update  $\bar{\mathbf{w}}_e$  to  $\mathbf{w}^{\text{new}}$ .
22     $D^{(\text{sc})}$ .UPDATEW( $e, \mathbf{w}^{\text{new}}$ ).
23     $\bar{\mathbf{w}}_e \leftarrow \mathbf{w}^{\text{new}}$ .
24  procedure LOCATE()
25     $\tilde{\phi} \leftarrow D^{(\phi)}$ .QUERYPOTENTIAL().
26    //  $\tilde{\phi}$  is padded with zeroes for computing  $\mathbf{B}\tilde{\phi}$ 
27     $\tilde{\psi} \leftarrow D^{(\psi)}$ .QUERYPOTENTIAL().
28     $\mathbf{p} \leftarrow \mathbf{Q} \left( \bar{\mathbf{W}}^{1/2} - \mathbf{W}^{1/2} \right) \mathbf{B}\tilde{\phi} + \mathbf{\Gamma} \begin{bmatrix} 0 \\ D^{(\text{sc})}.\text{INITIALSC}() \end{bmatrix} \tilde{\phi} + \mathbf{\Gamma} \begin{bmatrix} 0 \\ D^{(\text{sc})}.\text{INITIALSC}() \end{bmatrix} \tilde{\psi} + \mathbf{\Gamma} \mathbf{y}$ .
29    Return the set  $S$  returned by calling RECOVER( $\mathbf{p}$ ) of Lemma 5.11.

```

---

and

$$\left\| \boldsymbol{\psi} - \tilde{\boldsymbol{\psi}} \right\|_{\text{SC}(\mathbf{L}(\mathbf{w}), C)} \leq \epsilon \|\bar{\mathbf{v}}\|_2, \quad (15)$$

by the guarantee of Lemma 5.8 we have that

$$\begin{aligned}
\left\| \tilde{\mathbf{\Pi}} \bar{\mathbf{v}} \right\|_2 &\leq \left\| \mathbf{\Pi} \bar{\mathbf{v}} \right\|_2 + \left\| \tilde{\mathbf{\Pi}} \bar{\mathbf{v}} - \mathbf{\Pi} \bar{\mathbf{v}} \right\|_2 \\
&\leq \|\bar{\mathbf{v}}\|_2 + 2\epsilon \|\bar{\mathbf{v}}\|_2 + (1 + \epsilon) \left\| \boldsymbol{\phi} - \tilde{\boldsymbol{\phi}} \right\|_{\text{SC}(\mathbf{L}(\bar{\mathbf{w}}), C)} + (1 + \epsilon) \left\| \boldsymbol{\psi} - \tilde{\boldsymbol{\psi}} \right\|_{\text{SC}(\mathbf{L}(\mathbf{w}), C)}
\end{aligned}$$

$$\leq 2\|\bar{\mathbf{v}}\|_2.$$

By Lemma 5.11, the set  $S \leftarrow \text{RECOVER}(\mathbf{p})$  contains all  $e$  such that  $\left| \left[ \tilde{\mathbf{\Pi}}\bar{\mathbf{v}} \right]_e \right|$  is at least

$$\epsilon \left\| \tilde{\mathbf{\Pi}}\bar{\mathbf{v}} \right\|_2 \leq 2\epsilon\|\bar{\mathbf{v}}\|_2.$$

Finally, if  $e$  satisfies  $|\left[ \mathbf{\Pi}\bar{\mathbf{v}} \right]_e| \geq 10\epsilon\|\bar{\mathbf{v}}\|_2$  then

$$\begin{aligned} \left| \left[ \tilde{\mathbf{\Pi}}\bar{\mathbf{v}} \right]_e \right| &\geq |\left[ \mathbf{\Pi}\bar{\mathbf{v}} \right]_e| - \left\| \tilde{\mathbf{\Pi}}\bar{\mathbf{v}} - \mathbf{\Pi}\bar{\mathbf{v}} \right\|_2 \\ &\geq 10\epsilon\|\bar{\mathbf{v}}\|_2 - 2\epsilon\|\bar{\mathbf{v}}\|_2 - (1 + \epsilon) \left\| \phi - \tilde{\phi} \right\|_{\text{SC}(\mathbf{L}(\bar{\mathbf{w}}), C)} - (1 + \epsilon) \left\| \psi - \tilde{\psi} \right\|_{\text{SC}(\mathbf{L}(\mathbf{w}), C)} \geq 2\epsilon\|\bar{\mathbf{v}}\|_2 \end{aligned}$$

where the final step follows from Lemma 5.7 with (14). Thus  $e \in S$  as desired.

Now we bound the runtimes. The runtimes of `UPDATEV` and `UPDATEW` follow directly from Theorem 4.10 and Lemma 5.8. The cost of `INITIALIZE` is the cost of `INITIALIZE` in Theorem 4.10 and Lemma 5.8 plus the cost of computing  $\mathbf{\Gamma}$ . This involves solving  $N$  Laplacian systems, which costs  $\tilde{O}(Nm) = \tilde{O}(m\epsilon^{-2})$  time. This is dominated by  $\tilde{O}(m\beta^{-2}\epsilon^{-2})$ . Finally, the cost of `LOCATE()` is  $\tilde{O}(\beta m\epsilon^{-2})$  for computing  $\tilde{\phi}, \tilde{\psi}$  by Lemma 5.8, and the cost of computing  $\mathbf{p}$  in line 26 of Algorithm 5. The first term in line 26 can be computed in time  $O(N\beta m) = O(\beta m\epsilon^{-2})$  as  $\bar{\mathbf{W}}^{1/2} - \mathbf{W}^{1/2}$  is supported on  $O(\beta m)$  entries and  $\mathbf{Q}$  has  $N$  rows. The second and third terms in line 26 can be computed by first multiplying  $D^{(\text{sc})}.\text{INITIALSC}()$  times  $\tilde{\phi}$  (or  $\tilde{\psi}$ ) in time  $\tilde{O}(\beta m\epsilon^{-2})$ , as  $D^{(\text{sc})}.\text{INITIALSC}()$  has  $\tilde{O}(\beta m\epsilon^{-2})$  edges, and then multiplying by  $\mathbf{\Gamma}$  which is a  $N$ -by- $O(\beta m)$  size matrix in time  $O(N\beta m) = \tilde{O}(\beta m\epsilon^{-2})$  time. Thus the total runtime of `LOCATE()` is  $\tilde{O}(\beta m\epsilon^{-2})$  as desired.  $\square$

## 6 Reducing Adaptive to Oblivious Adversaries

In this section we show a blackbox reduction that is able to transform any dynamic algorithm that maintains some sequence of vectors  $(\mathbf{v}^t)_{t \geq 1}$  against oblivious adversaries to one that can maintain the vectors against adaptive adversaries. We formalize the requirements of the dynamic algorithm via Definition 6.1. Roughly, Definition 6.1 states that the dynamic algorithm must support two operations: (i) find the entries of the current vector  $\mathbf{v}^t$  with large absolute value, and (ii) query some set of the entries approximately.

**Definition 6.1.** We call a dynamic algorithm an  $\epsilon$ -approximate  $(L, S)$ -locator for an online<sup>5</sup> sequence of vectors  $(\mathbf{v}^t)_{t \geq 1}$ , if in each iteration  $t \geq 1$  the dynamic algorithm returns a set  $I \subset [n]$  of size at most  $S$  containing all  $i$  with  $|\mathbf{v}_i^t| > \epsilon$  in  $L$  time.

We call a dynamic algorithm an  $\epsilon$ -approximate  $C$ -evaluator, if it supports a query operation that, given some  $I \subset [n]$ , returns all  $\bar{\mathbf{v}}_i$  for  $i \in I$  in  $C(|I|)$  time for some  $\bar{\mathbf{v}} \in \mathbb{R}^n$  with  $\|\bar{\mathbf{v}} - \mathbf{v}^t\|_2 \leq \epsilon$ .

We show that, given a locator (a dynamic algorithm that can tell us the large entries), and locators (dynamic algorithms that tells us the entries of the vectors) with different accuracies, we can combine these dynamic algorithms to work against an adaptive adversary. The more accurate locators will be used less frequently, resulting in an expected time complexity faster than the most accurate locator.

---

<sup>5</sup>The sequence is may depend on outputs of the data structures.

**Theorem 6.2.** *Assume we have  $\epsilon$ -accurate  $(L, S)$ -locator and  $(\epsilon/2^i)$ -accurate  $C_i$ -evaluators for  $i = 0, \dots, K$  for an online sequence of  $n$ -dimensional vectors  $(\mathbf{v}^t)_{t \geq 1}$ . Both dynamic algorithms hold against an oblivious adversary. Also assume there is an  $\epsilon/2^K$ -accurate  $T$ -evaluator against an adaptive adversary.*

*Then there exists a dynamic algorithm against an adaptive adversary that in each iteration returns whp. some  $\bar{\mathbf{v}}^t$  with  $\|\bar{\mathbf{v}}^t - \mathbf{v}^t\|_\infty \leq O(\epsilon \log^2 n)$ . Each iteration takes expected time*

$$O\left(SK + \frac{T(S)}{2^K} + L + \sum_{i=0}^K \frac{C_i(S)}{2^i}\right).$$

Note that the  $T$ -evaluator against an adaptive adversary could just be a method to compute the exact solution statically. Alternatively, one could run several dynamic algorithms against an oblivious adversary in parallel, but use each data structure only once to answer a query.

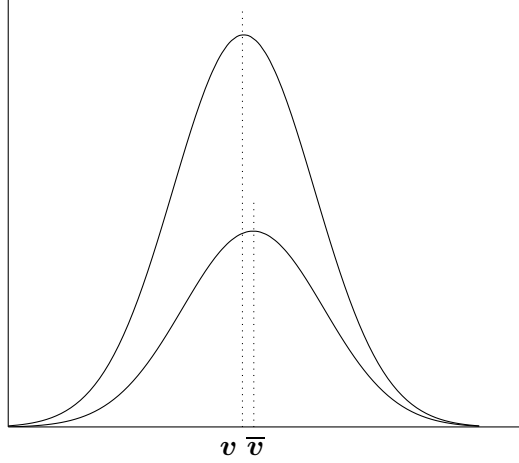
In the overview of Section 2.3 we outlined how Theorem 6.2 is obtained. We here give a quick recap. Let  $\mathbf{w}$  be the result of the  $T$ -evaluator and  $\mathbf{w}'$  be the result of one of the other evaluators against an oblivious adversary. We want to construct an output  $\bar{\mathbf{w}}$  whose distribution is similar to  $\mathcal{N}(\mathbf{w}, \sigma^2)$  for some variance  $\sigma = O(\epsilon \log n)$ . Note that w.h.p  $\|\bar{\mathbf{w}} - \mathbf{v}^t\|_\infty \leq O(\epsilon \log^2 n)$  because of the random Gaussian noise we added. We wish to improve upon the naive time of explicitly computing  $\bar{\mathbf{w}}$  by directly adding Gaussian noise to  $\mathbf{w}$ . We achieve this by performing this sampling in a different way which guarantees that we compute  $\mathbf{w}$  explicitly only with some small probability.

Let  $d$  be the density function of  $\mathcal{N}(\mathbf{w}, \sigma^2)$  and  $d'$  be the density function of  $\mathcal{N}(\mathbf{w}', \sigma^2)$ . Then there is some small  $\alpha > 0$  and very unlikely event  $D$  with  $d'(\mathbf{x}) \leq \exp(\alpha)d(\mathbf{x})$  for all  $\mathbf{x} \notin D$ . For example, Figure 2 shows density function  $d(\mathbf{x})$  and the scaled density function  $\exp(-\alpha) \cdot d'(\mathbf{x})$  for the 1-dimensional case. If one were to pick uniformly at random a point below the top curve in Figure 2 and return its  $\mathbf{x}$ -coordinate, then this corresponds to sampling from  $\mathcal{N}(\mathbf{w}, \sigma^2)$ . The same distribution can be obtained by first flipping an unbalanced coin, and with probability  $\exp(-\alpha)$  we sample from the area below the bottom curve  $\exp(-\alpha)d'(\mathbf{x})$  in Figure 2 (i.e. sample according to  $\mathcal{N}(\mathbf{w}', \sigma^2)$ ). Otherwise, with probability  $1 - \exp(-\alpha)$ , we sample from the area between the two curves. This way we are able to sample from  $\mathcal{N}(\mathbf{w}, \sigma^2)$  more efficiently because only with probability  $1 - \exp(-\alpha)$  must we compute  $\mathbf{w}$ . As computing  $\mathbf{w}'$  is faster than computing  $\mathbf{w}$ , the expected time complexity improves.

This scheme is proven formally in Section 6.1 for the general case where the vectors are  $n$ -dimensional. The scheme can be extended recursively: note that in order to sample from  $\mathcal{N}(\mathbf{w}', \sigma^2)$ , we can use the same scheme again via some  $\mathbf{w}''$ , i.e. when sampling from  $\mathcal{N}(\mathbf{w}', \sigma^2)$  we can sample from  $\mathcal{N}(\mathbf{w}'', \sigma^2)$  instead with probability  $\exp(-\alpha)$ . This is why Theorem 6.2 has  $K$  many different evaluators with increasing accuracy. The evaluators with higher accuracy are used with smaller probability, thus the expected time complexity improves. This recursive scheme is proven in Section 6.2 and we use it in Section 6.3 to prove Theorem 6.2.

## 6.1 Simulating Gaussian Error

Here we prove the algorithm outlined in the previous subsection. We want to construct a variable with distribution  $\mathcal{N}(\mathbf{v}, \sigma^2)$ . This is done by flipping a biased coin: with probability  $\exp(-\alpha)$  we return a vector according to  $\mathcal{N}(\mathbf{u}, \sigma^2)$ . Alternatively, with probability  $1 - \exp(-\alpha)$  we must return a random vector whose distribution we pick in such a way, that the result of our algorithm has distribution  $\mathcal{N}(\mathbf{v}, \sigma^2)$ . The exact algorithm is given in Algorithm 6 and Lemma 6.3 stated the guarantees of that algorithm.



**Figure 2:** Density function  $d$  of  $\mathcal{N}(\mathbf{v}, \sigma^2)$ , and density function  $\bar{d}$  of  $\mathcal{N}(\bar{\mathbf{v}}, \sigma^2)$  scaled by some  $\exp(-\alpha)$ ,  $\alpha > 0$  so that  $\bar{d}(\mathbf{x}) \exp(-\alpha) \leq d(\mathbf{x})$ .

---

**Algorithm 6:** Basic Simulation Algorithm

---

```

1 procedure SIMULATE( $\mathbf{v} \in \mathbb{R}^n, \mathbf{u} \in \mathbb{R}^n, \alpha \geq 0, \sigma > 0$ )
2   // Simulates  $\mathbf{v} + \mathbf{x}$  for  $\mathbf{x} \sim \mathcal{N}(0, \sigma^2)$ .
3   with probability  $\exp(-\alpha)$  do
4     | Sample  $\mathbf{x} \sim \mathcal{N}(0, \sigma^2)$ 
5     | return  $\mathbf{u} + \mathbf{x}$ 
6   end
7   while true do
8     | Sample  $\mathbf{x} \sim \mathcal{N}(0, \sigma^2)$  conditioned on  $\frac{||\mathbf{x}||^2 - ||\mathbf{x} - \mathbf{u} + \mathbf{v}||^2}{2\sigma^2} \leq \alpha$ 
9     | with probability  $1 - \exp\left(\frac{||\mathbf{x}||^2 - ||\mathbf{x} - \mathbf{u} + \mathbf{v}||^2}{2\sigma^2} - \alpha\right)$  do
10    | | return  $\mathbf{v} + \mathbf{x}$ 
11  | end
12 end

```

---

**Lemma 6.3.** *Let  $\mathbf{z}$  be the result of a call to *SIMULATE*( $\mathbf{v}, \mathbf{u}, \alpha, \sigma$ ) (Algorithm 6) with  $\sigma \geq 2 \ln(1.25/\delta) \epsilon / \alpha$  for any  $\delta > 0$  and  $\epsilon \geq \|\mathbf{v} - \mathbf{u}\|_2$ . Then the distribution of  $\mathbf{z}$  has total variation distance at most  $\delta$  compared to  $\mathcal{N}(\mathbf{v}, \sigma^2)$ . Further, the expected time complexity is bounded by  $O(n)$ .*

To prove Lemma 6.3, we first consider the distribution of the result returned by Line 6 to Line 9 of Algorithm 6.

**Lemma 6.4.** *Consider executing Line 6 to Line 11 of Algorithm 6 and let  $\mathbf{z}$  be the returned vector, i.e.  $\mathbf{z}$  is the output of Algorithm 6 conditioned on being returned in Line 9. Then the distribution of  $\mathbf{z}$  under this condition has density function*

$$d(\mathbf{z}) = \frac{\exp(-\frac{\|\mathbf{z} - \mathbf{v}\|^2}{2\sigma^2}) - \exp(-\frac{\|\mathbf{z} - \mathbf{u}\|^2}{2\sigma^2} - \alpha)}{\sqrt{2\pi}\sigma \mathbb{P}[||\mathbf{x}||^2 - ||\mathbf{x} - \mathbf{u} + \mathbf{v}||^2 \leq 2\sigma^2\alpha] (1 - \exp(-\alpha))}$$

*Proof.* Up to normalization, the density function of the distribution of  $\mathbf{z}$  is

$$\begin{aligned} & \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\|\mathbf{z} - \mathbf{v}\|^2}{2\sigma^2}\right) \left(1 - \exp\left(\frac{\|\mathbf{z} - \mathbf{v}\|^2 - \|\mathbf{z} - \mathbf{u}\|^2}{2\sigma^2} - \alpha\right)\right) \\ &= \frac{1}{\sqrt{2\pi}\sigma} \left(\exp\left(-\frac{\|\mathbf{z} - \mathbf{v}\|^2}{2\sigma^2}\right) - \exp\left(\frac{-\|\mathbf{z} - \mathbf{u}\|^2}{2\sigma^2} - \alpha\right)\right) \end{aligned}$$

We now compute the normalization factor. For that note the set of possibly returned vectors is  $S = \{\mathbf{z} \mid |\|\mathbf{z} - \mathbf{v}\|_2 - \|\mathbf{z} - \mathbf{u}\|_2| \leq 2\sigma^2\alpha\}$  and

$$\begin{aligned} \int_{\mathbf{z} \in S} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\|\mathbf{z} - \mathbf{u}\|^2}{2\sigma^2}\right) d\mathbf{z} &= \int_{\mathbf{z} \in S} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\|\mathbf{z} - \mathbf{v}\|^2}{2\sigma^2}\right) d\mathbf{z} \\ &= \mathbb{P}[|\|\mathbf{x}\|^2 - \|\mathbf{x} - \mathbf{u} + \mathbf{v}\|^2| \leq 2\sigma^2\alpha] \end{aligned}$$

for some  $\mathbf{x} \sim \mathcal{N}(0, \sigma^2)$ . So the normalization factor is

$$\mathbb{P}[|\|\mathbf{x}\|^2 - \|\mathbf{x} - \mathbf{u} + \mathbf{v}\|^2| \leq 2\sigma^2\alpha] (1 - \exp(-\alpha))$$

and the density function is

$$\frac{\exp\left(-\frac{\|\mathbf{z} - \mathbf{v}\|^2}{2\sigma^2}\right) - \exp\left(\frac{-\|\mathbf{z} - \mathbf{u}\|^2}{2\sigma^2} - \alpha\right)}{\sqrt{2\pi}\sigma \mathbb{P}[|\|\mathbf{x}\|^2 - \|\mathbf{x} - \mathbf{u} + \mathbf{v}\|^2| \leq 2\sigma^2\alpha] (1 - \exp(-\alpha))}$$

□

Our algorithm relies on the fact that the density function of  $\mathcal{N}(\mathbf{u}, \sigma^2)$  is smaller than the density function of  $\mathcal{N}(\mathbf{v}, \sigma^2)$  when scaled by  $\exp(\alpha)$ . This is generally not true, unless we restrict the two density function on to some event  $E \subset \mathbb{R}^n$ . Using the following result from differential privacy, we show that this event occurs with high probability, if the variance  $\sigma^2$  of the added noise and the scaling-parameter  $\alpha$  are sufficiently large.

**Lemma 6.5** ([DR<sup>+</sup>14, Appendix A]). *Let  $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ ,  $\epsilon \geq \|\mathbf{u} - \mathbf{v}\|$ ,  $c^2 > 2 \ln(1.25/\delta)$ ,  $\sigma \geq c\epsilon/\alpha$  and  $\mathbf{x} \sim \mathcal{N}(0, \sigma^2)$ . Then  $\mathbb{P}[|\|\mathbf{x}\|^2 - \|\mathbf{x} - \mathbf{u} + \mathbf{v}\|^2| > 2\alpha\sigma^2] \leq \delta$ .*

We now have all tools available to prove Lemma 6.3.

*Proof of Lemma 6.3.* The density function of  $\mathbf{z}$  conditioned on  $|\|\mathbf{z} - \mathbf{v}\|^2 - \|\mathbf{z} - \mathbf{u}\|^2| \leq 2\sigma^2\alpha$  is

$$\begin{aligned} & \exp(-\alpha) \cdot \frac{\exp\left(-\frac{\|\mathbf{z} - \mathbf{u}\|^2}{2\sigma^2}\right)}{\sqrt{2\pi}\sigma \mathbb{P}[|\|\mathbf{x}\|^2 - \|\mathbf{x} - (\mathbf{u} - \mathbf{v})\|^2| \leq 2\sigma^2\alpha]} \\ &+ (1 - \exp(-\alpha)) \cdot \frac{\exp\left(-\frac{\|\mathbf{z} - \mathbf{v}\|^2}{2\sigma^2}\right) - \exp\left(\frac{-\|\mathbf{z} - \mathbf{u}\|^2}{2\sigma^2} - \alpha\right)}{\sqrt{2\pi}\sigma \mathbb{P}[|\|\mathbf{x}\|^2 - \|\mathbf{x} - (\mathbf{u} - \mathbf{v})\|^2| \leq 2\sigma^2\alpha] (1 - \exp(-\alpha))} \\ &= \frac{\exp\left(-\frac{\|\mathbf{z} - \mathbf{v}\|^2}{2\sigma^2}\right)}{\sqrt{2\pi}\sigma \mathbb{P}[|\|\mathbf{x}\|^2 - \|\mathbf{x} - (\mathbf{u} - \mathbf{v})\|^2| \leq 2\sigma^2\alpha]} \end{aligned}$$

---

**Algorithm 7:** Recursive Simulation Algorithm

---

```

1 procedure SIMULATE( $\mathbf{v}_1, \dots, \mathbf{v}_k \in \mathbb{R}^n, \alpha \geq 0, \sigma > 0$ )
  // If  $k = 2$  we call Algorithm 6 instead.
2   with probability  $\exp(-\alpha)$  do
3     return SIMULATE( $\mathbf{v}_2, \dots, \mathbf{v}_k, 2\alpha, \sigma$ ) // Call Alg. 6 if  $k = 2$ .
4   end
5   while true do
6     Sample  $\mathbf{x} \sim \mathcal{N}(0, \sigma^2)$  conditioned on  $\frac{|\|\mathbf{x}\|^2 - \|\mathbf{x} - \mathbf{v}_2 + \mathbf{v}_1\|^2|}{2\sigma^2} \leq \alpha$ 
7     with probability  $1 - \exp(\frac{\|\mathbf{x}\|^2 - \|\mathbf{x} - \mathbf{v}_2 + \mathbf{v}_1\|^2}{2\sigma^2} - \alpha)$  do
8       return  $\mathbf{v}_1 + \mathbf{x}$ 
9     end
10  end

```

---

which is also the density function of  $\mathbf{x} \sim \mathcal{N}(0, \sigma^2)$  conditioned on  $|\|\mathbf{x}\|^2 - \|\mathbf{x} - (\mathbf{u} - \mathbf{v})\|^2| \leq 2\sigma^2\alpha$ . Thus the total variation distance is bounded by  $\delta$  via Lemma 6.5.

For the time complexity, note that the probability of returning the vector during an iteration of Line 6 is  $1 - \exp(-\alpha)$ . Consequently, if we reach Line 6, it is invoked  $(1 - \exp(-\alpha))^{-1}$  times in expectation. The probability of reaching Line 6 is  $1 - \exp(-\alpha)$ , so Line 6 is invoked 1 time in expectation. As each iteration needs  $O(n)$  time to process the  $n$ -dimensional vectors, the expected runtime of the procedure is  $O(n)$ .  $\square$

## 6.2 Recursive Simulation

In this subsection we provide and analyze a recursive variant of Algorithm 6. This variant replaces Line 3 of Algorithm 6, which samples some  $\mathcal{N}(\mathbf{u}, \sigma^2)$ , by a recursive invocation of Algorithm 6. We first analyze the distribution of the returned vector in Lemma 6.6 and then bound the expected time complexity in Lemma 6.7.

**Lemma 6.6.** *Consider a call to Algorithm 7 with inputs  $\mathbf{v}_1, \dots, \mathbf{v}_k \in \mathbb{R}^n$ ,  $k \geq 2$ ,  $\epsilon \geq \|\mathbf{v}_i - \mathbf{v}_{i+1}\|_2/2^{i-1}$  for all  $i = 1, \dots, k-1$ ,  $\sigma \geq 2\ln(1.25/\delta)\epsilon/\alpha$ . Then the returned value has total variation distance at most  $(k-1)\delta$  compared to  $\mathcal{N}(\mathbf{v}_1, \sigma^2)$ .*

*Proof.* We prove this by induction over  $k$ , the number of vectors.

**Base Case** For  $k = 2$  we call Algorithm 6 instead, so the claim is true by Lemma 6.3.

**Induction** Now assume Lemma 6.6 holds for some  $k-1$  and consider a call to SIMULATE with vectors  $\mathbf{v}_1, \dots, \mathbf{v}_k$ . Let  $\mathbf{v}'_1, \dots, \mathbf{v}'_{k-1}$ ,  $\alpha' = 2\alpha$  be the parameters of the recursive call in Line 3 and let  $\epsilon' = 2\epsilon$ . Then

$$\begin{aligned} \epsilon' &= \epsilon/2 \geq \|\mathbf{v}_i - \mathbf{v}_{i+1}\|_2/2^i = \|\mathbf{v}_{i+1} - \mathbf{v}_i\|_2/2^{i-1} = \|\mathbf{v}'_i - \mathbf{v}'_{i+1}\|_2/2^{i-1} \text{ and} \\ \sigma &\geq 2\ln(1.25/\delta)\epsilon/\alpha = 2\ln(1.25/\delta)\epsilon'/\alpha' \end{aligned}$$

so the conditions to apply the induction hypothesis are satisfied. Thus, the vector returned in Line 3 has the same distribution as  $\mathcal{N}(\mathbf{v}'_1, \sigma^2) = \mathcal{N}(\mathbf{v}_2, \sigma^2)$  up to total variation distance  $(k-2)\delta$ .



Note that Algorithm 7 is the same as Algorithm 6, except for Line 3, so by the same proof as in Lemma 6.3 we return a vector that is distributed like  $\mathcal{N}(\mathbf{v}_1, \sigma^2)$  up to total variation distance  $(k-2)\delta + \delta = (k-1)\delta$ .  $\square$

For computational efficiency, note that Algorithm 7 does not need to read vector  $\mathbf{v}_1$  when performing the branch of Line 3. The following lemma bounds the probability of accessing any  $\mathbf{v}_i$  for  $i < k$ .

**Lemma 6.7.** *Consider a call to Algorithm 7 with inputs  $\mathbf{v}_1, \dots, \mathbf{v}_k \in \mathbb{R}^n, \alpha \geq 0$ . The probability that vector  $\mathbf{v}_i$  is accessed is at most  $2^i \alpha$  for all  $i < k$ . Further, the expected time complexity (ignoring the time for accessing any  $\mathbf{v}_i$ ) is bounded by  $O(kn)$ .*

*Proof.* Vector  $\mathbf{v}_1$  is accessed with probability  $1 - \exp(-\alpha) \leq \alpha \leq \alpha 2^1$ . Vector  $\mathbf{v}_i$  for  $i > 1$  is accessed with probability  $1 - \exp(-\alpha 2^{i-2})$  when calling  $\text{SIMULATE}(\mathbf{v}_{i-1}, \dots, \mathbf{v}_k, \alpha, \sigma)$ , or with probability  $1 - \exp(-\alpha 2^{i-1})$  when calling  $\text{SIMULATE}(\mathbf{v}_i, \dots, \mathbf{v}_k, \alpha, \sigma)$ . Thus the overall probability after a call  $\text{SIMULATE}(\mathbf{v}_1, \dots, \mathbf{v}_k, \alpha, \sigma)$  is

$$\begin{aligned}
& \underbrace{\prod_{t=1}^{i-1} \exp(-\alpha 2^{t-1})}_{\text{Probability of recursing to SIMULATE}(\mathbf{v}_i, \dots, \mathbf{v}_k, \alpha, \sigma)} \left(1 - \exp(-\alpha 2^{i-1})\right) + \underbrace{\prod_{t=1}^{i-2} \exp(-\alpha 2^{t-1})}_{\text{Probability of recursing to SIMULATE}(\mathbf{v}_{i-1}, \dots, \mathbf{v}_k, \alpha, \sigma)} \left(1 - \exp(-\alpha 2^{i-2})\right) \\
&= \exp\left(-\alpha \sum_{t=0}^{i-2} 2^t\right) \left(1 - \exp(-\alpha 2^{i-1})\right) + \exp\left(-\alpha \sum_{t=0}^{i-3} 2^t\right) \left(1 - \exp(-\alpha 2^{i-2})\right) \\
&= \exp\left(-\alpha(2^{i-1} - 1)\right) \left(1 - \exp(-\alpha 2^{i-1})\right) + \exp\left(-\alpha(2^{i-2} - 1)\right) \left(1 - \exp(-\alpha 2^{i-2})\right) \\
&= \exp\left(-\alpha(2^{i-1} - 1)\right) - \exp\left(-\alpha(2^i - 1)\right) + \exp\left(-\alpha(2^{i-2} - 1)\right) - \exp\left(-\alpha(2^{i-1} - 1)\right) \\
&= \exp\left(-\alpha(2^{i-2} - 1)\right) - \exp\left(-\alpha(2^i - 1)\right) \\
&= \left(1 - \exp\left(-\alpha(2^i - 2^{i-2})\right)\right) \cdot \exp\left(-\alpha(2^{i-2} - 1)\right) \\
&\leq 1 - \exp\left(-\alpha(2^i - 2^{i-2})\right) \\
&\leq \alpha(2^i - 2^{i-2}) \leq \alpha 2^i
\end{aligned}$$

The time expected time complexity is at most  $O(kn)$  because each recursion has expected time  $O(n)$  by Lemma 6.3.  $\square$

### 6.3 Proof of Theorem 6.2

We can now prove Theorem 6.2 by applying Algorithm 7 to the vectors returned by the evaluator data structures.

*Proof of Theorem 6.2.* Given the locator and evaluators, we construct a new dynamic algorithm  $\mathcal{A}$  against an adaptive adversary. The construction is done in a paragraph further below. For now, we claim that the output of the new dynamic algorithm  $\mathcal{A}$  has the following distribution.

Let  $\mathbf{w}^1$  be the output of the  $\epsilon/2^K$ -accurate oracle against an adaptive adversary. Then sample  $\mathbf{x} \sim \mathcal{N}(\mathbf{w}^1, (c_1 \epsilon \log n)^2)$  for some sufficiently large constant  $c_1$ . At last, set all entries of  $\mathbf{x}$  with

absolute value smaller than  $c_2\epsilon\log^2 n$  to 0. Call the resulting vector  $\mathbf{u}$ . We claim the dynamic algorithm  $\mathcal{A}$  will have the distribution of this vector  $\mathbf{z}$ .

Vector  $\mathbf{z}$  satisfies w.h.p.  $\|\mathbf{z} - \mathbf{v}^t\|_\infty = O(\epsilon\log^2 n)$ , so returning  $\mathbf{z}$  would satisfy the promised approximation guarantees of Theorem 6.2 and the algorithm would work against an adaptive adversary because the output does not depend on any of the oracles that use the oblivious adversary assumption.

We now describe the new dynamic algorithm  $\mathcal{A}$  and how it constructs this vector  $\mathbf{z}$  more efficiently than the procedure described above.

**Algorithm** Let  $I$  be the set returned by the  $\epsilon$ -accurate locator. For  $i > 1$  let  $\mathbf{w}_I^i$  be result of the  $\epsilon/2^{K-i+2}$ -accurate oracles against oblivious adversaries when querying only entries from  $I$ . Let  $\mathbf{x}'_I = \text{SIMULATE}(\mathbf{w}_S^1, \dots, \mathbf{w}_S^{k+1}, 2^{-K}, c_1\epsilon\log n)$  (Algorithm 7). Then set all entries of  $\mathbf{x}$  with absolute value smaller than  $c_2\epsilon\log^2 n$  to 0 and let  $\mathbf{z}'$  be the resulting vector. Here  $c_2 > c_1$  is picked such that w.h.p.  $|\mathbf{x}_i - \mathbf{w}_I^1| < c_2/2 \cdot \epsilon\log^2 n$ . Our algorithm returns this vector  $\mathbf{z}'$ .

**Correctness** We claim  $\mathbf{z}'$  has the same distribution as  $\mathbf{z}$  up to total variation distance  $1/\text{poly}(n)$ . For  $\epsilon' = 2\epsilon/2^K$  we have  $\|\mathbf{w}^i - \mathbf{w}^{i+1}\|_2 \leq \epsilon/2^{K-i} \leq \epsilon'2^{i-1}$ . So  $\mathbf{x}'_I$  has distribution  $\mathcal{N}(\mathbf{w}_I^1, (c_1\sigma\log n)^2)$  up to total variation distance  $1/\text{poly}(n)$  by Lemma 6.6 for some large enough constant  $c_1$ . Thus if  $I$  only contained indices  $i$  where w.h.p.  $\mathbf{z}_i$  would be 0 anyway, then  $\mathbf{z}'$  has same distribution as  $\mathbf{z}$  up to total variation distance  $1/\text{poly}(n)$ .

Note that by  $\|\mathbf{w}^1 - \mathbf{v}^t\|_2 < \epsilon$  we have that  $I$  (which by definition contains all indices with  $|\mathbf{v}_i^t| > \epsilon$ ) also contains all indices  $i$  with  $|\mathbf{w}_I^1| > 2\epsilon$ . Further, w.h.p. we have  $\|\mathbf{w}^1 - \mathbf{x}\|_\infty < c_2/2\epsilon\log^2 n$  by choice of  $c_2 > c_1$ . So  $i \in I$  this would imply  $|\mathbf{x}_i| \leq |\mathbf{w}_I^1| + |\mathbf{w}_I^1 - \mathbf{x}_i| \leq 2\epsilon + c_2/2\epsilon\log^2 n < c_2\epsilon\log^2 n$  so w.h.p.  $\mathbf{z}_i$  will be set to 0. Thus the total variation distance of  $\mathbf{z}$  and  $\mathbf{z}'$  is at most some  $1/\text{poly}(n)$ .

**Complexity** By Lemma 6.7, we use each  $\mathbf{w}_i$  with probability at most  $2^i/2^K = 2^{i-K}$  for  $i \leq K$  and running SIMULATE on  $K+1$  many  $|I|$ -dimensional vectors needs  $O(|I|K) = O(SK)$  time. We can delay the query to  $\mathbf{w}^i$  until the vectors actually need to be used. As  $\mathbf{w}_i$  is obtained from evaluator with complexity  $C_{K-i+2}$  for  $i > 1$ , we obtain time complexity

$$O(SK + L + C_0(S) + T(S)/2^K + \sum_{i=1}^K \frac{C_i(S)}{2^i}).$$

□

## 7 Interior Point Method

In this section we provide the machinery we use to reduce minimum cost flow to dynamic graph data structure problems. First, in Section 7.1 we provide the general IPM framework for linear programming from [DLY21] that we use. Then, in Section 7.2 we introduce the data structures, subroutines, and bounds that we develop in this paper to implement this framework efficiently and in Section 7.3 we combine these pieces to give the efficient IPM. The proofs for the tools we introduce are provided in Section 7.4, Section 7.5, and Section 8 (for the runtime bound for the graph solution maintainer (Definition 7.4) in Theorem 7.7).

## 7.1 Robust IPM Framework

Here we provide the the linear programming setup that we use to model minimum cost flow and the IPM framework provided by [DLY21] for solving them. In particular, throughout the section, we consider the general linear programming problem. Given  $\mathbf{B} \in \mathbb{R}^{m \times n}$ ,  $\mathbf{c}, \boldsymbol{\ell}, \mathbf{u} \in \mathbb{R}^m$ , and  $\mathbf{d} \in \mathbb{R}^n$  where  $\boldsymbol{\ell} < \mathbf{u}$  entrywise, we wish to solve.

$$\min_{\mathbf{x} \in \mathbb{R}^m \mid \mathbf{B}^\top \mathbf{x} = \mathbf{d} \text{ and } \boldsymbol{\ell} \leq \mathbf{x} \leq \mathbf{u}} \mathbf{c}^\top \mathbf{x}. \quad (16)$$

In the special case where  $\mathbf{B}$  is the incidence matrix of graph and  $\boldsymbol{\ell} = \mathbf{0}$ , this problem directly corresponds to the minimum cost flow problem. Many of the reductions we provide in this section apply to this general linear program and we will explicitly state in which cases we instead assume that  $\mathbf{B}$  is the incidence matrix of graph.

To solve (16) we leverage the general robust IPM framework of [DLY21]. This method crudely follows a central path by maintaining centered points defined as follows.

**Definition 7.1** (Centered Point). For  $\mathcal{X} \stackrel{\text{def}}{=} \{\mathbf{x} \in \mathbb{R}^m \mid \mathbf{x}_i \in (\boldsymbol{\ell}_i, \mathbf{u}_i)\}$  we say  $(\mathbf{x}, \mathbf{s}) \in \mathbb{R}^m \times \mathbb{R}^m$  is  $\mu$ -feasible for  $\mu > 0$  if  $\mathbf{x} \in \mathcal{X}$ ,  $\mathbf{B}^\top \mathbf{x} = \mathbf{d}$ ,  $\mathbf{B}\mathbf{y} + \mathbf{s} = \mathbf{c}/\mu$  for some  $\mathbf{y} \in \mathbb{R}^n$ . We say  $(\mathbf{x}, \mathbf{s})$  is  $\mu$ -centered<sup>6</sup> if  $(\mathbf{x}, \mathbf{s})$  is  $\mu$ -feasible and  $\left\| \nabla^2 \phi(\mathbf{x})^{-1/2} (\mathbf{s} + \nabla \phi(\mathbf{x})) \right\|_\infty \leq \frac{1}{64}$  where  $\phi(\mathbf{x}) \stackrel{\text{def}}{=} \sum_{i \in [m]} \phi_i(\mathbf{x})$  with  $\phi_i(\mathbf{x}) \stackrel{\text{def}}{=} -\log(\mathbf{u}_i - \mathbf{x}_i) - \log(\mathbf{x}_i - \boldsymbol{\ell}_i)$  for  $\mathbf{x} \in \mathcal{X}$ .

This definition is motivated for the fact that,  $\mu$ -central path point, defined as

$$\mathbf{x}_\mu \stackrel{\text{def}}{=} \arg \min_{\mathbf{x} \in \mathcal{X} \mid \mathbf{B}^\top \mathbf{x} = \mathbf{d}} \mu \cdot \mathbf{c}^\top \mathbf{x} + \phi(\mathbf{x})$$

is the unique  $\mu$ -centered point with  $\left\| \nabla^2 \phi(\mathbf{x})^{-1/2} (\mathbf{s} + \nabla \phi(\mathbf{x})) \right\|_\infty = 0$ . To see this, note that  $\phi$  is convex on  $\mathcal{X}$  and the optimality conditions for  $\mathbf{x}_\mu$  are that

$$\mathbf{x} \in \mathcal{X}, \mathbf{B}^\top \mathbf{x}_\mu = \mathbf{d} \text{ and } \mathbf{c} + \mu \nabla \phi(\mathbf{x}_\mu) \perp \ker(\mathbf{B}^\top).$$

However,  $\mathbf{c} + \mu \nabla \phi(\mathbf{x}_\mu) \perp \ker(\mathbf{B}^\top)$  if and only if  $\mathbf{c} + \mu \nabla \phi(\mathbf{x}_\mu) \in \text{im}(\mathbf{B})$  which we can write equivalently as  $\mathbf{c} + \mu \nabla \phi(\mathbf{x}_\mu) = \mu \mathbf{B}\mathbf{y}_\mu$  for some  $\mathbf{y}_\mu$ . Finally, the condition  $\mathbf{c} + \mu \nabla \phi(\mathbf{x}_\mu) = \mu \mathbf{B}\mathbf{y}_\mu$  is equivalent to  $\mathbf{B}\mathbf{y}_\mu + \mathbf{s}_\mu = \mathbf{c}/\mu$  for  $\mathbf{s}_\mu = -\nabla \phi(\mathbf{x}_\mu)$ , i.e.  $\left\| \nabla^2 \phi(\mathbf{x}_\mu)^{-1/2} (\mathbf{s}_\mu + \nabla \phi(\mathbf{x}_\mu)) \right\|_\infty = 0$  as  $\nabla^2 \phi(\mathbf{x}_\mu)$  is positive definite. Consequently, a  $\mu$ -centered point is a point which maintains an approximate notion of the optimality of  $\mathbf{x}_\mu$ .

The IPM framework works by maintaining  $\mu$ -centered points by controlling centrality measures as potentials. The definition of these quantities (Definition 7.2), the framework (Algorithm 8), and the result from [DLY21] that we use about this framework (Theorem 7.3) are all given below.

**Definition 7.2** (Centrality). For  $\mu$ -feasible  $(\mathbf{x}, \mathbf{s})$  we define *centrality measure*  $\gamma(\mathbf{x}, \mathbf{s}) \in \mathbb{R}^m$  where  $\gamma_i(\mathbf{x}, \mathbf{s}) \stackrel{\text{def}}{=} \phi_i''(\mathbf{x})^{-1/2} (\mathbf{s}_i + \phi_i'(\mathbf{x}))$  and  $\phi_i'(\mathbf{x}) \stackrel{\text{def}}{=} [\nabla \phi(\mathbf{x})]_i$  and  $\phi_i''(\mathbf{x}) \stackrel{\text{def}}{=} [\nabla^2 \phi(\mathbf{x})]_{ii}$ . Further, we define *centrality potential*  $\Psi(\mathbf{x}, \mathbf{s}) \stackrel{\text{def}}{=} \sum_{i \in [m]} \cosh(\lambda \cdot \gamma_i(\mathbf{x}, \mathbf{s}))$  where  $\lambda \stackrel{\text{def}}{=} 128 \log(16m)$  and  $\cosh(z) \stackrel{\text{def}}{=} \frac{1}{2}[\exp(z) + \exp(-z)]$  for all  $z \in \mathbb{R}$ .

<sup>6</sup>In [DLY21], the condition is  $\mathbf{B}\mathbf{y} + \mathbf{s} = \mathbf{c}$  and  $\left\| \nabla^2 \phi(\mathbf{x})^{-1/2} (\mathbf{s}/\mu + \nabla \phi(\mathbf{x})) \right\|_\infty \leq \frac{1}{64}$  instead. We do the replacement from  $\mathbf{s}/\mu$  to  $\mathbf{s}$  to simplify the algorithm description and notations in the data structures. Further, the choice of variable names is different in the two papers with variable names chosen here for the application of minimum cost flow.

---

**Algorithm 8:** A robust interior point method in [DLY21]

---

```

1 procedure CENTERING( $\mathbf{B}, \mathbf{x}, \mathbf{s}, \ell, \mathbf{u}, \mu_{\text{start}}, \mu_{\text{end}}$ )
    // Invariant:  $(\mathbf{x}, \mathbf{s})$  is  $\mu$ -centered with  $\Psi(\mathbf{x}, \mathbf{s}) \leq \cosh(\lambda/64)$ 
    // (See Definitions 7.1 and 7.2)
2
3   Define step size  $\alpha \stackrel{\text{def}}{=} \frac{1}{2^{15}\lambda}$ .
4    $\bar{\mu} = \mu = \mu_{\text{start}}, \bar{\mathbf{x}} = \mathbf{x}, \bar{\mathbf{s}} = \mathbf{s}$ 
5   while  $\mu \geq \mu_{\text{end}}$  do
6       Set weight matrix  $\mathbf{W} \leftarrow \nabla^2 \phi(\bar{\mathbf{x}})^{-1}$ .
7       Set iterate approximation  $(\bar{\mathbf{x}}, \bar{\mathbf{s}}) \in \mathbb{R}^m \times \mathbb{R}^m$  such that
           
$$\|\mathbf{W}^{-1/2}(\bar{\mathbf{x}} - \mathbf{x})\|_{\infty} \leq \alpha \text{ and } \|\mathbf{W}^{1/2}(\bar{\mathbf{s}} - \mathbf{s})\|_{\infty} \leq \alpha.$$

8       Set step direction  $\mathbf{v} \in \mathbb{R}^m$  where  $v_i \leftarrow \sinh(\lambda\gamma_i(\bar{\mathbf{x}}, \bar{\mathbf{s}}))$  for all  $i \in [m]$  and
            $\sinh(z) \stackrel{\text{def}}{=} \frac{1}{2}(\exp(z) - \exp(-z)).$ 
9       Set step size  $h \leftarrow -\alpha / \|\cosh(\lambda\gamma(\bar{\mathbf{x}}, \bar{\mathbf{s}}))\|_2$ .
10      Set  $\mathbf{v}^{\parallel}, \mathbf{v}^{\perp}$  such that  $\mathbf{W}^{-1/2}\mathbf{v}^{\parallel} \in \text{Im}\mathbf{B}, \mathbf{B}^{\top}\mathbf{W}^{1/2}\mathbf{v}^{\perp} = 0$ , and
           
$$\|\mathbf{v}^{\parallel} - \mathbf{P}_{\mathbf{W}}\mathbf{v}\|_2 \leq \alpha\|\mathbf{v}\|_2 \text{ and } \|\mathbf{v}^{\perp} - (\mathbf{I} - \mathbf{P}_{\mathbf{W}})\mathbf{v}\|_2 \leq \alpha\|\mathbf{v}\|_2$$

           where  $\mathbf{P} \stackrel{\text{def}}{=} \mathbf{W}^{1/2}\mathbf{B}(\mathbf{B}^{\top}\mathbf{W}\mathbf{B})^{-1}\mathbf{B}^{\top}\mathbf{W}^{1/2}$ 
11      Set  $\mathbf{x} \leftarrow \mathbf{x} + h\mathbf{W}^{1/2}\mathbf{v}^{\parallel}, \mathbf{s} \leftarrow \mathbf{s} + h\mathbf{W}^{-1/2}\mathbf{v}^{\parallel}, \mu \leftarrow \max\{(1 - \frac{\alpha}{64\sqrt{m}})\mu, \mu_{\text{end}}\}$ 
12      If  $|\bar{\mu} - \mu| \geq \alpha\bar{\mu}$ , then  $\mathbf{s} \leftarrow \frac{\mu}{\bar{\mu}}\mathbf{s}, \bar{\mu} \leftarrow \mu$ 
13  end
14  Return  $(\mathbf{x}, \mathbf{s})$ 

```

---

**Theorem 7.3** (Theorem A.16 in [DLY21]). *Using the notation in Algorithm 8, let  $(\mathbf{x}^{(0)}, \mathbf{s}^{(0)})$  be the value of  $(\mathbf{x}, \mathbf{s})$  before the step (Line 11) and let  $(\mathbf{x}^{(1)}, \mathbf{s}^{(1)})$  be the value  $(\mathbf{x}, \mathbf{s})$  after the step. If  $(\mathbf{x}^{(0)}, \mathbf{s}^{(0)})$  is  $\bar{\mu}$ -feasible and  $\Psi(\mathbf{x}^{(0)}, \mathbf{s}^{(0)}) \leq \cosh(\lambda/64)$ , then  $(\mathbf{x}^{(1)}, \mathbf{s}^{(1)})$  is  $\bar{\mu}$ -feasible and*

$$\Psi(\mathbf{x}^{(1)}, \mathbf{s}^{(1)}) \leq \left(1 - \frac{\alpha\lambda}{8\sqrt{m}}\right) \Psi(\mathbf{x}^{(0)}, \mathbf{s}^{(0)}) + \alpha\lambda\sqrt{m} \leq \cosh(\lambda/64).$$

*Proof.* The proof of [DLY21, Theorem A.16] shows  $\Psi^{\mu'}(\mathbf{x}^{(1)}, \mathbf{s}^{(1)}) \leq \left(1 - \frac{\alpha\lambda}{8\sqrt{m}}\right) \Psi^{\mu'}(\mathbf{x}^{(0)}, \mathbf{s}^{(0)}) + \alpha\lambda\sqrt{m}$  for any  $|\mu' - \mu| \leq \alpha\mu$  where  $\Psi^{\mu}(\mathbf{x}, \mathbf{s}) \stackrel{\text{def}}{=} \left\| \nabla^2 \phi(\mathbf{x})^{-1/2}(\mathbf{s}/\mu + \nabla \phi(\mathbf{x})) \right\|_{\infty}$ . We picked  $\mu' = \bar{\mu}$  and replaced  $\mathbf{s}/\bar{\mu}$  by  $\mathbf{s}$ .  $\square$

## 7.2 Robust IPM Tools

Here we discuss the key tools we develop in this paper to efficiently implement the robust IPM (Algorithm 8) of [DLY21] discussed in the previous Section 7.1.

First, as discussed in Section 1.1, a key advance of this paper is efficient procedures for approximately maintaining the iterates of Algorithm 8, i.e. approximating the result of approximate projection steps. We formalize this maintenance problem as a data structure problem defined below.

**Definition 7.4** (Solution Maintainer). We call a data structure a  $(\mathcal{T}_{\text{init}}, \mathcal{T}_{\text{phase}})$ -*solution maintainer* if it supports the following operations against an adaptive adversary with high probability:

- **INITIALIZE**( $\mathbf{B} \in \mathbb{R}^{m \times n}$ ,  $\mathbf{w}^{(0)} \in \mathbb{R}_+^m$ ,  $\mathbf{x}^{(0)} \in \mathbb{R}^m$ ,  $\mathbf{s}^{(0)} \in \mathbb{R}^m$ ,  $\alpha, C_r, k, C_z$ ): Given input constraint matrix  $\mathbf{B}$ , weight vector  $\mathbf{w}^{(0)}$ , iterate  $(\mathbf{x}^{(0)}, \mathbf{s}^{(0)})$ , accuracy parameter  $\alpha$ , weight range  $r$ , phase length  $k$ , sparsity of changes  $z$ , initialize the data structure with  $\mathbf{w} := \mathbf{w}^{(0)}$ ,  $\mathbf{x} := \mathbf{x}^{(0)}$ , and  $\mathbf{s} := \mathbf{s}^{(0)}$  in time  $O(\mathcal{T}_{\text{init}})$  with  $\mathcal{T}_{\text{init}} = \Omega(m)$ .
- **STARTPHASE**( $\tilde{\mathbf{x}} \in \mathbb{R}^m, \tilde{\mathbf{s}} \in \mathbb{R}^m$ ): Given input iterate  $(\tilde{\mathbf{x}}, \tilde{\mathbf{s}})$  with  $\|\mathbf{W}^{-1/2}(\tilde{\mathbf{x}} - \mathbf{x})\|_2 \leq 1$  and  $\|\mathbf{W}^{1/2}(\tilde{\mathbf{s}} - \mathbf{s})\|_2 \leq 1$ , update  $\mathbf{x} \leftarrow \tilde{\mathbf{x}}$  and  $\mathbf{s} \leftarrow \tilde{\mathbf{s}}$  in amortized  $O(\mathcal{T}_{\text{phase}})$  time with  $\mathcal{T}_{\text{phase}} = \Omega(m)$ .
- **MOVE**( $\mathbf{w}^{(j)} \in \mathbb{R}_+^m, \mathbf{v}^{(j)} \in \mathbb{R}^m, h^{(j)} \in \mathbb{R}$ )  $\rightarrow \mathbb{R}^m \times \mathbb{R}^m$ : In the  $j$ -th call to **MOVE**, given input weights  $\mathbf{w}^{(j)}$ , direction  $\mathbf{v}^{(j)}$ , and step size  $h^{(j)}$  with  $h^{(j)}\|\mathbf{v}^{(j)}\|_2 \leq 1$ , **MOVE** updates  $\mathbf{w} \leftarrow \mathbf{w}^{(j)}$ ,

$$\mathbf{x} \leftarrow \mathbf{x} + h^{(j)} \mathbf{W}_j^{1/2} (\mathbf{I} - \mathbf{P}_j) \mathbf{v}^{(j)}, \text{ and } \mathbf{s} \leftarrow \mathbf{s} + h^{(j)} \mathbf{W}_j^{-1/2} \mathbf{P}_j \mathbf{v}^{(j)},$$

where  $\mathbf{W}_j \stackrel{\text{def}}{=} \text{diag}(\mathbf{w}^{(j)})$  and  $\mathbf{P}_j \stackrel{\text{def}}{=} \mathbf{W}_j^{1/2} \mathbf{B} (\mathbf{B}^\top \mathbf{W}_j \mathbf{B})^{-1} \mathbf{B}^\top \mathbf{W}_j^{1/2}$  and **MOVE** outputs  $(\bar{\mathbf{x}}^{(j)}, \bar{\mathbf{s}}^{(j)}) \in \mathbb{R}^m \times \mathbb{R}^m$  with  $\|(\mathbf{W}^{(j)})^{-1/2}(\bar{\mathbf{x}}^{(j)} - \mathbf{x})\|_\infty \leq \alpha$ ,  $\|(\mathbf{W}^{(j)})^{1/2}(\bar{\mathbf{s}}^{(j)} - \mathbf{s})\|_\infty \leq \alpha$ , and the number of coordinates changed from the previous output bounded by  $O(2^{2\ell_j+1} \alpha^{-2} \log^3 m + S_j)$  where

$$S_j \stackrel{\text{def}}{=} \left| \left\{ i \in [m] : \mathbf{w}_i^{(j)} \neq \mathbf{w}_i^{(j-1)}, \bar{\mathbf{x}}_i^{(j-1)} = \bar{\mathbf{x}}_i^{(j-2)}, \text{ and } \bar{\mathbf{s}}_i^{(j-1)} = \bar{\mathbf{s}}_i^{(j-2)} \right\} \right|.$$

The input  $\mathbf{w}^{(j)}$  and  $\mathbf{v}^{(j)}$  and output  $(\bar{\mathbf{x}}^{(j)}, \bar{\mathbf{s}}^{(j)})$  are given implicitly as a list of changes to the previous input and output of **MOVE**.

Furthermore, the above operations need only be supported under the following assumptions:

1. *Phase length*: **STARTPHASE** is called at least every  $k$  calls to **MOVE** and at most twice in a row.
2. *Number of changes*: for all  $j \geq 1$  there are at most  $\min\{C_z 2^{2\ell_j}, m\}$  coordinates changed in  $\mathbf{w}^{(j)}, \mathbf{v}^{(j)}$  from  $\mathbf{w}^{(j-1)}, \mathbf{v}^{(j-1)}$  where  $\ell_j$  is the largest integer with  $\ell$  with  $j \equiv 0 \pmod{2^\ell}$ .
3. *Magnitude of changes*: for any  $|j_2 - j_1| \leq L$ , we have  $\sqrt{\mathbf{w}_i^{(j_2)}/\mathbf{w}_i^{(j_1)}} \leq C_r L^2$  for all  $i \in [m]$ .

Our algorithm actually always has  $S_j = 0$ , but we state Definition 7.4 with possibly nonzero  $S_j$  for more generality.

In the particular case of graphs, one of the key results of this paper is the following efficient solution maintenance data structure in the particular case of graphs (shown in Section 8.1).

**Theorem 7.5** (Graph Solution Maintenance). *In the special case that  $\mathbf{B}$  is the incidence matrix of a  $m$ -edge,  $n$ -node graph, if  $C_r, C_z = \tilde{O}(1)$  and  $\alpha = \tilde{O}(1)$ , there is a  $(\mathcal{T}_{\text{init}}, \mathcal{T}_{\text{phase}})$ -solution maintainer (Definition 7.4) with  $\mathcal{T}_{\text{init}} = \tilde{O}(m)$  and  $\mathcal{T}_{\text{phase}} = \tilde{O}(m + m^{15/16} k^{29/8})$ .*

Note that in the solution maintenance data structure problem it is required that **STARTPHASE** be called at least every  $k$  calls to **MOVE**. Consequently, to apply this data structure to implement the robust IPM framework the input  $\hat{\mathbf{x}}$  and  $\hat{\mathbf{s}}$  to **STARTPHASE**, i.e. weighted  $\ell_2$  approximations to  $(\mathbf{x}, \mathbf{s})$ , need to be computed efficiently. We formalize this problem below.

**Definition 7.6** (Solution Approximation). We call a procedure  $\mathcal{T}_{\text{approx-approximator}}$  if given  $\mu$ -feasible  $(\mathbf{x}, \mathbf{s})$ , weights  $\mathbf{w}^{(1)}, \dots, \mathbf{w}^{(k)} \in \mathbb{R}_+^m$ , directions  $\mathbf{v}^{(1)}, \dots, \mathbf{v}^{(k)} \in \mathbb{R}^m$ , and step sizes  $h^{(1)}, \dots, h^{(k)}$  such that

- $h^{(i)} \|\mathbf{v}^{(i)}\|_2 \leq 1$ ,
- all the changes in  $\mathbf{w}$  and  $\mathbf{v}$  are supported on  $z$  many edges and the input is given as these changes,
- $\frac{1}{r} \leq \sqrt{\mathbf{w}_\ell^{(i)} / \mathbf{w}_\ell^{(j)}} \leq r$  for all  $i, j \in [k]$  and  $\ell \in [m]$ ,

with high probability, we can compute  $\mu$ -feasible  $(\tilde{\mathbf{x}}, \tilde{\mathbf{s}})$  such that

$$\left\| \tilde{\mathbf{x}} - \mathbf{x} - \sum_{i \in [k]} h^{(i)} \mathbf{W}_i^{1/2} (\mathbf{I} - \mathbf{P}_{\mathbf{W}_i}) \mathbf{v}^{(i)} \right\|_{\mathbf{W}_k^{-1}} \leq \epsilon \text{ and } \left\| \tilde{\mathbf{s}} - \mathbf{s} - \sum_{i \in [k]} h^{(i)} \mathbf{W}_i^{-1/2} \mathbf{P}_{\mathbf{W}_i} \mathbf{v}^{(i)} \right\|_{\mathbf{W}_k} \leq \epsilon$$

in  $O(\mathcal{T}_{\text{approx}})$  time where  $\mathbf{W}_i \stackrel{\text{def}}{=} \mathbf{diag}(\mathbf{w}^{(i)})$ .

In the particular case of graphs, in Section 7.4 we provide the following theorem on efficient solution approximation.

**Theorem 7.7.** *In the special case that  $\mathbf{B}$  is the incidence matrix of a  $m$ -edge,  $n$ -node graph there is  $\mathcal{T}_{\text{approx-approximator}}$  with  $\mathcal{T}_{\text{approx}} = \tilde{O}(m + zk^3r^2\epsilon^{-2})$ .*

Finally, to apply these results, we need to prove that the weights which in turn are induced by  $\nabla^2 \phi(\mathbf{x})$  do not change by too much. For this, in Appendix 7.5 we prove the following. The statement is similar to the bound in [GLP21, Lemma 6.5] generalized to our setting. Note that this bound applies to the IPM framework regardless of whether or not  $\mathbf{B}$  is the incidence matrix of a graph.

**Lemma 7.8.** *For  $\mu^{(0)}$ -centered  $(\mathbf{x}^{(0)}, \mathbf{s}^{(0)})$  and  $\mu^{(1)}$ -centered  $(\mathbf{x}^{(1)}, \mathbf{s}^{(1)})$  with  $\mu^{(0)} \approx_{1/32} \mu^{(1)}$ , if  $\boldsymbol{\eta}^{(j)} \stackrel{\text{def}}{=} \mathbf{s}^{(j)} + \nabla \phi(\mathbf{x}^{(j)})$  for  $j \in \{0, 1\}$  it follows that*

$$\sum_{\phi_i''(\mathbf{x}^{(0)})^{1/2} \geq 3\phi_i''(\mathbf{x}^{(1)})^{1/2}} \sqrt{\frac{\phi_i''(\mathbf{x}^{(0)})}{\phi_i''(\mathbf{x}^{(1)})}} \leq 2^{10} \sum_{i \in [m]} \frac{(\boldsymbol{\eta}_i^{(1)} - \boldsymbol{\eta}_i^{(0)})^2}{\phi_i''(\mathbf{x}^{(0)}) + \phi_i''(\mathbf{x}^{(1)})} + 2^4 m \left( \frac{\mu^{(1)} - \mu^{(0)}}{\mu^{(0)}} \right)^2.$$

### 7.3 Robust IPM Implementation

Here we show how to use the tools of Section 7.2 to efficiently implement the Algorithm 8. The algorithm, Algorithm 8, and its analysis, Lemma 7.9, are given below.

**Theorem 7.9.** *For any  $k \geq 1$ ,  $\mu_{\text{end}} \leq \mu_{\text{start}}$ , and  $\mu_{\text{start}}$ -centered  $(\mathbf{x}, \mathbf{s})$  with  $\Psi(\mathbf{x}, \mathbf{s}) \leq \cosh(\lambda/64)$ , Algorithm 9 outputs a  $\mu_{\text{end}}$ -centered  $(\mathbf{x}', \mathbf{s}')$  with  $\Psi(\mathbf{x}', \mathbf{s}') \leq \cosh(\lambda/64)$  in time  $\tilde{O}((\mathcal{T}_{\text{init}} + \frac{\sqrt{m}}{k}(\mathcal{T}_{\text{phase}} + \mathcal{T}_{\text{approx}})) \log(\mu_{\text{start}}/\mu_{\text{end}}))$ .*

*Proof.* First, we verify that the conditions of the solution maintenance data structure (Definition 7.4) are satisfied with  $C_r, C_z$  defined as in the Algorithm 8.

---

**Algorithm 9:** Algorithm 8 Implementation with Solution Maintenance and Estimation

---

```

1 procedure CENTERINGIMPL( $\mathbf{B}, \mathbf{x}, \mathbf{s}, \ell, \mathbf{u}, \mu_{\text{start}}, \mu_{\text{end}}, k$ )
2   Define step size  $\alpha \stackrel{\text{def}}{=} \frac{1}{2^{15}\lambda}$ , weight range  $C_r = \Theta(1)$ , and sparsity parameter
    $C_z = \Theta(\log^5 m)$ 
3   Set  $\bar{\mu} = \mu = \mu_{\text{start}}$ ,  $\bar{\mathbf{x}} = \mathbf{x}$ ,  $\bar{\mathbf{s}} = \mathbf{s}$ ,  $\mathbf{w} = \text{diag}(\nabla^2 \phi(\bar{\mathbf{x}})^{-1})$ ,  $j = 0$ 
4   SOL.INITIALIZE( $\mathbf{B}, \mathbf{w}, \mathbf{x}, \mathbf{s}, \alpha, C_r, k, C_z$ ) where SOL is a  $(\mathcal{T}_{\text{init}}, \mathcal{T}_{\text{phase}})$ -solution maintainer
   (Definition 7.4)
5   while  $\mu \geq \mu_{\text{end}}$  do
6     if  $k$  divides  $j$  or  $\mu = \mu_{\text{end}}$  then // Reset every  $k$  iterations
7       Let  $(\mathbf{x}, \mathbf{s})$  be the solution SOL implicitly maintained.
8       Find  $\bar{\mu}$ -feasible  $(\tilde{\mathbf{x}}, \tilde{\mathbf{s}})$  with  $\|\mathbf{W}^{-1/2}(\mathbf{x} - \tilde{\mathbf{x}})\|_2 \leq \frac{\alpha}{100}$  and  $\|\mathbf{W}^{1/2}(\mathbf{s} - \tilde{\mathbf{s}})\|_2 \leq \frac{\alpha}{100}$  by
       using a  $\mathcal{T}_{\text{approx}}$ -approximator (Definition 7.6) with  $k = k, r = O(k^4), z = \tilde{O}(k^2)$ .
9       if  $|\bar{\mu} - \mu| \geq \alpha \bar{\mu}$  then
10         SOL.INITIALIZE( $\mathbf{B}, \mathbf{w}, \mathbf{x}, \mathbf{s}, \alpha, r, k, z$ )
11          $\tilde{\mathbf{s}} \leftarrow \frac{\mu}{\bar{\mu}} \tilde{\mathbf{s}}, \bar{\mu} \leftarrow \mu$  // Reinitialize. All coordinates may have changed
12       end
13       SOL.STARTPHASE( $\tilde{\mathbf{x}}, \tilde{\mathbf{s}}$ )
14     end
15
16     // Step:  $\mathbf{x} \leftarrow \mathbf{x} + h \mathbf{W}^{1/2}(\mathbf{I} - \mathbf{P}_{\mathbf{W}})\mathbf{v}, \mathbf{s} \leftarrow \mathbf{s} + h \mathbf{W}^{-1/2} \mathbf{P}_{\mathbf{W}} \mathbf{v}$ 
17     Set the direction  $\mathbf{v}_i = \sinh(\lambda \gamma_i(\bar{\mathbf{x}}, \bar{\mathbf{s}}))$  and the step size  $h = -\alpha / \|\cosh(\lambda \gamma(\bar{\mathbf{x}}, \bar{\mathbf{s}}))\|_2$ .
18      $(\bar{\mathbf{x}}, \bar{\mathbf{s}}) \leftarrow \text{SOL.MOVE}(\mathbf{w}, \mathbf{v}, h)$ 
19      $\mu \leftarrow \max((1 - \frac{\alpha}{64\sqrt{m}})\mu, \mu_{\text{end}})$ ,  $\mathbf{w} \leftarrow \text{diag}(\nabla^2 \phi(\bar{\mathbf{x}})^{-1})$ , and  $j \leftarrow j + 1$ 
20   end
21   Return  $(\mathbf{x}, \mathbf{s})$ 

```

---

- $C_z$ : Note that both  $\mathbf{w}, \mathbf{v}$  are entrywise functions of  $\bar{\mathbf{x}}, \bar{\mathbf{s}}$  and Definition 7.4 promises that  $\bar{\mathbf{x}}, \bar{\mathbf{s}}$  changes in at most  $O(2^{2\ell_{j+1}} \epsilon^{-2} \log^3 m + S_j)$  coordinates. Since we only change  $\mathbf{w}$  when  $\bar{\mathbf{x}}$  or  $\bar{\mathbf{s}}$  changes, we have  $S_j = 0$ . Using the parameter choice  $\epsilon = \Theta(1/\log m)$ , the number of changes is bounded by  $O(2^{2\ell_{j+1}} \log^5 m)$ . This verifies the condition  $C_z = O(\log^5 m)$ .
- $C_r$ : For any two iterations  $\mathbf{x}^{(j_1)}$  and  $\mathbf{x}^{(j_2)}$  associated with path parameters  $\mu^{(j_1)}$  and  $\mu^{(j_2)}$ , Lemma 7.8 shows that

$$\sqrt{\frac{\phi_i''(\mathbf{x}^{(j_1)})}{\phi_i''(\mathbf{x}^{(j_2)})}} \leq 3 + 2^{10} \sum_{i \in [m]} \frac{(\eta_i^{(j_2)} - \eta_i^{(j_1)})^2}{\phi_i''(\mathbf{x}^{(j_2)}) + \phi_i''(\mathbf{x}^{(j_1)})} + 2^4 m \left( \frac{\mu^{(j_2)} - \mu^{(j_1)}}{\mu^{(j_1)}} \right)^2$$

where  $\eta^{(j)} \stackrel{\text{def}}{=} \mathbf{s}^{(j)} + \nabla \phi(\mathbf{x}^{(j)})$ . Note that to apply this lemma, we used that all iterations are centered (which we will show later) and that  $\mu^{(j_1)} \approx_{1/32} \mu^{(j_2)}$  (since we reinitialize the data structure every  $\tilde{\Theta}(\sqrt{m})$  steps). For  $L = |j_1 - j_2|$ , we have  $|\mu^{(j_2)} - \mu^{(j_1)}| \leq \frac{\alpha L}{32\sqrt{m}} \mu^{(j_1)}$ , so

$$\sqrt{\frac{\phi_i''(\mathbf{x}^{(j_1)})}{\phi_i''(\mathbf{x}^{(j_2)})}} \leq 3 + 2^{10} \sum_{i \in [m]} \frac{(\eta_i^{(j_2)} - \eta_i^{(j_1)})^2}{\phi_i''(\mathbf{x}^{(j_2)}) + \phi_i''(\mathbf{x}^{(j_1)})} + O(\alpha^2 L^2).$$



To bound the first term, we note that every term in the summation is bounded by  $O(1)$ . We split the sum into two cases. The first case is when  $\phi_i''(\mathbf{x})$  does not change by more than a  $O(1)$  factor. In this case, one can prove that  $\|\boldsymbol{\eta}^{(j+1)} - \boldsymbol{\eta}^{(j)}\|_{\phi''(\mathbf{x}^{(j)})} = O(\alpha)$  because  $\|\mathbf{x}^{(j+1)} - \mathbf{x}^{(j)}\|_{(\mathbf{W}^{(j)})^{-1}} \leq \alpha$ ,  $\|\mathbf{s}^{(j+1)} - \mathbf{s}^{(j)}\|_{\mathbf{W}^{(j)}} \leq \alpha$  and  $\mathbf{W}^{(j)} \approx_{O(1)} \nabla^2 \phi(\mathbf{x}^{(j)})^{-1}$ . Therefore, after  $L$  steps, the sum for the first case is bounded by  $O(\alpha^2 L^2) = O(L^2)$ . For the second case, we can use  $\|\mathbf{x}^{(j+1)} - \mathbf{x}^{(j)}\|_{(\mathbf{W}^{(j)})^{-1}} \leq \alpha$  to show that there are at most  $O(L^2 \alpha^2)$  coordinates where  $\phi''$  changes by more than a constant multiplicative factor. Hence, this shows that

$$\sqrt{\frac{\phi_i''(\mathbf{x}^{(j_1)})}{\phi_i''(\mathbf{x}^{(j_2)})}} = O(L^2). \quad (17)$$

This verifies the condition  $C_r = O(1)$ .

Now, we bound the potential. Theorem 7.3 shows that

$$\Psi(\mathbf{x}^{\text{new}}, \mathbf{x}^{\text{new}}) \leq \left(1 - \frac{\alpha\lambda}{8\sqrt{m}}\right) \Psi(\mathbf{x}, \mathbf{x}) + \alpha\lambda\sqrt{m}$$

for every step (excluding the effect of STARTPHASE). For STARTPHASE, we have that  $\|\mathbf{W}^{-1/2}(\mathbf{x} - \tilde{\mathbf{x}})\|_2 \leq \frac{\alpha}{100}$  and  $\|\mathbf{W}^{1/2}(\mathbf{s} - \tilde{\mathbf{s}})\|_2 \leq \frac{\alpha}{100}$ . This increases  $\Psi$  by at most  $\frac{\alpha\lambda}{16\sqrt{m}}\Psi(\mathbf{x}, \mathbf{s})$  additively. Finally, for the change of  $\bar{\mu}$ , it would increase  $\Psi$  by at most  $2\alpha\lambda\Psi(\mathbf{x}, \mathbf{s})$ , but this happens every  $32\sqrt{m}$  steps. Therefore  $\Psi$  is decreasing on average and stays polynomially bounded.

Next, we discuss the parameters for the  $\mathcal{T}_{\text{approx}}$ -approximator. The number of terms is exactly given by  $k$ . For the number of coordinate changes  $z$ , Definition 7.4 promised that  $\bar{\mathbf{x}}, \bar{\mathbf{s}}$  changes by  $\tilde{O}(2^{2\ell_{j+1}})$  coordinates at the  $j$ -th step. Since we restart every  $k$  iterations by calling STARTPHASE, by aligning our steps numbers appropriately, we have that  $\sum_{j \text{ in a phase}} \tilde{O}(2^{2\ell_j}) = \tilde{O}(\max_{j \text{ in a phase}} 2^{2\ell_j}) = \tilde{O}(k^2)$ . Finally, the weight ratio is due to (17) with  $L = k$ .

Finally, for the runtime, note that there are  $\tilde{O}(\sqrt{m} \log(\mu_{\text{start}}/\mu_{\text{end}}))$  steps. For every  $k$  steps, we use a  $\mathcal{T}_{\text{approx}}$ -approximator and call SOL.STARTPHASE and they cost  $\mathcal{T}_{\text{phase}}$  and  $\mathcal{T}_{\text{approx}}$  respectively. All other costs are linear in the output size of the data structure and are not bottlenecks. Therefore, the total cost is  $\tilde{O}((\mathcal{T}_{\text{init}} + \frac{\sqrt{m}}{k}(\mathcal{T}_{\text{phase}} + \mathcal{T}_{\text{approx}})) \log(\mu_{\text{start}}/\mu_{\text{end}}))$ .  $\square$

## 7.4 Efficient Solution Approximation

In this section, we prove Theorem 7.7. Our algorithms leverage two powerful tools from algorithmic graph theory, in particular nearly linear time algorithms for subspace sparsification [LS18].

*Proof of Theorem 7.7.* Our algorithm for approximating  $\tilde{\mathbf{x}}$  involves two steps, we first find a  $\mathbf{x}'$  such that it is close to the true vector  $\mathbf{x}^* \stackrel{\text{def}}{=} \mathbf{x} + \sum_{i \in [k]} h^{(i)} \mathbf{W}_i^{1/2} (\mathbf{I} - \mathbf{P}_{\mathbf{W}_i}) \mathbf{v}^{(i)}$ , but may not satisfies  $\mathbf{B}^\top \mathbf{x}' = \mathbf{d}$ . Then, we show how to use  $\mathbf{x}'$  to find  $\tilde{\mathbf{x}}$  that is close to  $\mathbf{x}^*$  and satisfies  $\mathbf{B}^\top \mathbf{x}' = \mathbf{d}$ .

Let  $S \subseteq [m]$  be the set of at most  $z$  coordinates of  $\mathbf{w}$  and  $\mathbf{v}$  that change and let  $C \subseteq [n]$  be an arbitrary subset (that we set later) such that every edge in  $S$  has both endpoints in  $C$ . Further, let  $\Delta_1 \stackrel{\text{def}}{=} \sum_{i \in [k]} h^{(i)} \mathbf{W}_i^{1/2} \mathbf{P}_{\mathbf{W}_i} \mathbf{v}^{(i)}$  and  $\Delta_2 \stackrel{\text{def}}{=} \sum_{i \in [k]} h^{(i)} \mathbf{W}_i^{1/2} \mathbf{v}^{(i)}$  so that  $\sum_{i \in [k]} h^{(i)} \mathbf{W}_i^{1/2} (\mathbf{I} - \mathbf{P}_{\mathbf{W}_i}) \mathbf{v}^{(i)} = \Delta_2 - \Delta_1$ . Note that  $\Delta_2$  can be computed in  $O(m + zk)$  time by first computing  $\sum_{i \in [k]} h^{(i)}$  and with this computing  $[\Delta_2]_j$  for  $j \notin S$  in  $O(1)$  time and for  $j \in S$  in  $O(k)$  time. Consequently, to compute  $\tilde{\mathbf{x}}$  in the given time bound, it suffices to approximately compute  $\Delta_1$ .

Next, let  $\mathbf{L}^{(i)} \stackrel{\text{def}}{=} \mathbf{B}^\top \mathbf{W}_i \mathbf{B}$  and  $F \stackrel{\text{def}}{=} V \setminus C$  and note that

$$\Delta_1 = \sum_{i \in [k]} h^{(i)} \mathbf{W}_i \mathbf{B} \begin{bmatrix} \mathbf{I} & -\mathbf{L}_{FF}^{-1} \mathbf{L}_{FC} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{L}_{FF}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{SC}(\mathbf{L}_i, C)^\dagger \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{L}_{CF} \mathbf{L}_{FF}^{-1} & \mathbf{I} \end{bmatrix} \mathbf{B}^\top \mathbf{W}_i^{1/2} \mathbf{v}^{(i)}$$

since  $\mathbf{L}_{FF} = [\mathbf{L}^{(i)}]_{FF}$  and  $\mathbf{L}_{FC} = [\mathbf{L}^{(i)}]_{FC}$  for all  $i$  by the definition of  $C$ . Further, let  $\mathbf{B}_C$  be the incidence matrix of edges with both endpoints in  $C$  and  $\mathbf{B}_{-C}$  be the incidence matrix of the remaining edges so that  $\mathbf{B}^\top \mathbf{W}_i^{1/2} \mathbf{v}^{(i)} = \mathbf{B}_C^\top \mathbf{W}_i^{1/2} \mathbf{v}^{(i)} + \mathbf{B}_{-C}^\top \mathbf{W}^{1/2} \mathbf{v}$  for all  $i \in [k]$ . Combining yields that,  $\Delta_1 = \mathbf{a}_1 + \mathbf{a}_2 + \mathbf{a}_3 + \mathbf{a}_4$  where

$$\mathbf{a}_1 \stackrel{\text{def}}{=} \sum_{i \in [k]} h^{(i)} \mathbf{W}_i \mathbf{B}_C \mathbf{SC}(\mathbf{L}_i, C)^\dagger \mathbf{B}_C^\top \mathbf{W}_i^{1/2} \mathbf{v}^{(i)} \quad (18)$$

$$\mathbf{a}_2 \stackrel{\text{def}}{=} \sum_{i \in [k]} h^{(i)} \mathbf{W} \mathbf{B}_{-C} \begin{bmatrix} -\mathbf{L}_{FF}^{-1} \mathbf{L}_{FC} \\ \mathbf{I} \end{bmatrix} \mathbf{SC}(\mathbf{L}_i, C)^{-1} \mathbf{B}_C^\top \mathbf{W}_i^{1/2} \mathbf{v}^{(i)}$$

$$\mathbf{a}_3 \stackrel{\text{def}}{=} \sum_{i \in [k]} h^{(i)} \mathbf{W}_i \mathbf{B}_C \mathbf{SC}(\mathbf{L}_i, C)^\dagger \begin{bmatrix} -\mathbf{L}_{CF} \mathbf{L}_{FF}^{-1} & \mathbf{I} \end{bmatrix} \mathbf{B}_{-C}^\top \mathbf{W}^{1/2} \mathbf{v}$$

$$\mathbf{a}_4 \stackrel{\text{def}}{=} \sum_{i \in [k]} h^{(i)} \mathbf{W} \mathbf{B}_{-C} \begin{bmatrix} \mathbf{I} & -\mathbf{L}_{FF}^{-1} \mathbf{L}_{FC} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{L}_{FF}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{SC}(\mathbf{L}_i, C)^\dagger \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{L}_{CF} \mathbf{L}_{FF}^{-1} & \mathbf{I} \end{bmatrix} \mathbf{B}_{-C}^\top \mathbf{W}^{1/2} \mathbf{v}. \quad (19)$$

Our algorithm simply computes  $\Delta_1$  through the above formula where every instance of  $\mathbf{SC}(\mathbf{L}_i, C)$  is replaced with some efficiently computed  $\widetilde{\mathbf{SC}}_i \approx_\delta \mathbf{SC}(\mathbf{L}_i, C)$  for  $\delta$  we set later.

To compute the  $\widetilde{\mathbf{SC}}_i$ , first for each  $i$ , we define  $\mathbf{L}_i(S) \stackrel{\text{def}}{=} \mathbf{B}_S^\top \mathbf{W}_i \mathbf{B}_S$  where  $\mathbf{B}_S$  is the incidence matrix of edges  $S$  and let  $\mathbf{L}_{\text{ext}} \stackrel{\text{def}}{=} \mathbf{L}^{(i)} - \mathbf{L}_i(S)$  for any  $i \in [k]$ . Note that this definition does not depend on  $i$  by the definition of  $S$ . Using [LS18, Theorem 1.3], we can compute  $C \subseteq V$  such that every edge in  $S$  has both endpoints in  $C$  and a Laplacian  $\widetilde{\mathbf{SC}} \in \mathbb{R}^{C \times C}$  such that  $\widetilde{\mathbf{SC}} \approx_\delta \mathbf{SC}(\mathbf{L}_{\text{ext}}, C)$  and  $|C| \leq \text{nnz}(\widetilde{\mathbf{SC}}) = \tilde{O}(|S|\delta^{-2}) = \tilde{O}(k\delta^{-2})$  in  $\tilde{O}(m)$  time with high probability. We use this procedure to determine  $C$  and compute  $\widetilde{\mathbf{SC}}$ . Further, we define  $\widetilde{\mathbf{SC}}_i = \widetilde{\mathbf{SC}} + \mathbf{L}_i(S)$  and note that  $\widetilde{\mathbf{SC}}_i \approx_\delta \mathbf{SC}(\mathbf{L}_i, C)$  for all  $i \in [k]$  and  $\text{nnz}(\widetilde{\mathbf{SC}}_i) = \text{nnz}(\widetilde{\mathbf{SC}}) + |S| = \tilde{O}(k\delta^{-2})$ .

Now, we let  $\tilde{\mathbf{a}}_1, \tilde{\mathbf{a}}_2, \tilde{\mathbf{a}}_3, \tilde{\mathbf{a}}_4$  be the result of computing  $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4$  respectively where each  $\mathbf{SC}(\mathbf{L}_i, C)$  is replaced with  $\widetilde{\mathbf{SC}}_i$  and each matrix inversion is computed to high precision using nearly linear time SDD-solvers for  $\mathbf{L}_{FF}^{-1}$  and  $\mathbf{SC}(\mathbf{L}_i, C)^{-1}$  (Theorem 3.1). Further, we let  $\mathbf{x}' = \mathbf{x} + \sum_{i \in [4]} \tilde{\mathbf{a}}_i + \Delta_2$ . Note that  $h^{(i)} \mathbf{B}_C^\top \mathbf{W}_i^{1/2} \mathbf{v}^{(i)}$  can be computed explicitly for all  $i \in [k]$  in  $\tilde{O}(m + kz\delta^{-2})$  time by simply iterating through the changes in  $\mathbf{w}$  and  $\mathbf{v}$  and noting that each change only effects the resulting  $\tilde{O}(z\delta^{-2})$  coordinate vector in 2 coordinates. Further, this implies that  $\mathbf{d}_i \stackrel{\text{def}}{=} h^{(i)} \mathbf{SC}(\mathbf{L}_i, C)^\dagger \mathbf{B}_C^\top \mathbf{W}_i^{1/2} \mathbf{v}^{(i)}$  can be computed to high precision in  $\tilde{O}(kz\delta^{-2})$  time by using a nearly linear time Laplacian system solver to apply  $\mathbf{SC}(\mathbf{L}_i, C)^\dagger$ . Next, to compute  $\tilde{\mathbf{a}}_1 \stackrel{\text{def}}{=} \sum_{i \in [k]} \mathbf{W}_i \mathbf{B}_C \mathbf{d}_i$  note that the contribution of each row of  $\mathbf{B}_C$  for  $e \in S$  can be computed  $O(k)$  and the contribution of all the remaining rows can be computed in  $O(m)$ ; thus,  $\tilde{\mathbf{a}}_1$  can be computed from the  $\mathbf{d}_i$  in  $O(m + kz)$ . Further, given the  $\mathbf{d}_i$  by using a nearly linear time SDD solver to apply  $\mathbf{L}_{FF}^{-1}$  to a vector we see that  $\tilde{\mathbf{a}}_2$  can be computed in  $\tilde{O}(m)$ . Similarly, all the  $\mathbf{e}_i \stackrel{\text{def}}{=} \mathbf{SC}(\mathbf{L}_i, C)^\dagger \begin{bmatrix} -\mathbf{L}_{CF} \mathbf{L}_{FF}^{-1} & \mathbf{I} \end{bmatrix} \mathbf{B}_{-C}^\top \mathbf{W}^{1/2} \mathbf{v}$  can be computed in  $\tilde{O}(m + kz\delta^{-2})$  and from these  $\tilde{\mathbf{a}}_3$  can be computed in an additional  $O(m + kz)$  time (analogous to computing  $\tilde{\mathbf{a}}_1$ ). Further,  $\tilde{\mathbf{a}}_4$  can be computed  $\tilde{O}(m + kz\delta^{-2})$  since summation

can be moved to the  $\mathbf{SC}(\mathbf{L}_i, C)^\dagger$ . Putting these pieces together shows that  $\tilde{\mathbf{x}}$  can be computed in  $\tilde{O}(m + zk\delta^{-2})$ .

Next, to determine what to set  $\delta$  to. Note that

$$\begin{aligned}
& \|\mathbf{x}^* - \tilde{\mathbf{x}}\|_{\mathbf{W}_k^{-1}} \\
&= \left\| \sum_{i \in [k]} \mathbf{W}_i \mathbf{B} \begin{bmatrix} -\mathbf{L}_{FF}^{-1} \mathbf{L}_{FC} \\ \mathbf{I} \end{bmatrix} (\mathbf{SC}(\mathbf{L}_i, C)^\dagger - \widetilde{\mathbf{SC}}_i^\dagger) \begin{bmatrix} -\mathbf{L}_{FF}^{-1} \mathbf{L}_{FC} \\ \mathbf{I} \end{bmatrix}^\top \mathbf{B}^\top \mathbf{W}_i^{1/2} (h^{(i)} \mathbf{v}^{(i)}) \right\|_{\mathbf{W}_k^{-1}} \\
&\leq r \sum_{i \in [k]} \left\| \mathbf{W}_i \mathbf{B} \begin{bmatrix} -\mathbf{L}_{FF}^{-1} \mathbf{L}_{FC} \\ \mathbf{I} \end{bmatrix} (\mathbf{SC}(\mathbf{L}_i, C)^\dagger - \widetilde{\mathbf{SC}}_i^\dagger) \begin{bmatrix} -\mathbf{L}_{FF}^{-1} \mathbf{L}_{FC} \\ \mathbf{I} \end{bmatrix}^\top \mathbf{B}^\top \mathbf{W}_i^{1/2} (h^{(i)} \mathbf{v}^{(i)}) \right\|_{\mathbf{W}_i^{-1}} \\
&\leq r \sum_{i \in [k]} \left\| \mathbf{W}_i^{-1/2} \mathbf{B} \begin{bmatrix} -\mathbf{L}_{FF}^{-1} \mathbf{L}_{FC} \\ \mathbf{I} \end{bmatrix} (\mathbf{SC}(\mathbf{L}_i, C)^\dagger - \widetilde{\mathbf{SC}}_i^\dagger) \begin{bmatrix} -\mathbf{L}_{FF}^{-1} \mathbf{L}_{FC} \\ \mathbf{I} \end{bmatrix}^\top \mathbf{B}^\top \mathbf{W}_i^{1/2} \right\|_2 \\
&= r \sum_{i \in [k]} \left\| \mathbf{SC}(\mathbf{L}_i, C)^{1/2} (\mathbf{SC}(\mathbf{L}_i, C)^\dagger - \widetilde{\mathbf{SC}}_i^\dagger) \mathbf{SC}(\mathbf{L}_i, C)^{1/2} \right\|_2 = O(rk\delta)
\end{aligned}$$

where in the third line we used that assumption  $\frac{1}{r} \leq \sqrt{\langle \mathbf{w}_i \rangle_l / \langle \mathbf{w}_j \rangle_l} \leq r$ , in the fourth we used that  $\|h \mathbf{v}^{(i)}\|_2 \leq 1$ , and in the fifth we used that

$$\begin{bmatrix} -\mathbf{L}_{FF}^{-1} \mathbf{L}_{FC} \\ \mathbf{I} \end{bmatrix}^\top \mathbf{B}^\top \mathbf{W}_i \mathbf{B} \begin{bmatrix} -\mathbf{L}_{FF}^{-1} \mathbf{L}_{FC} \\ \mathbf{I} \end{bmatrix} = \mathbf{SC}(\mathbf{L}_i, C),$$

and that  $\|\mathbf{M}_1 \mathbf{M}_2 \mathbf{M}_3\|_2 = \|(\mathbf{M}_1^\top \mathbf{M}_1)^{1/2} \mathbf{M}_2 (\mathbf{M}_3 \mathbf{M}_3^\top)^{1/2}\|_2$  for matrices  $\mathbf{M}_1, \mathbf{M}_2, \mathbf{M}_3$  of appropriate dimension and that  $\widetilde{\mathbf{SC}}_i \approx_\delta \mathbf{SC}(\mathbf{L}_i, C)$ . Consequently, it suffices to set  $\delta = \Theta(\epsilon/(rk))$  and this gives the result for computing  $\tilde{\mathbf{x}}$ .

Now, we show how to find a feasible  $\tilde{\mathbf{x}}$  using  $\mathbf{x}'$ . From the first part, we can find  $\mathbf{x}'$  such that  $\|\mathbf{x}' - \mathbf{x}^*\|_{\mathbf{W}_k^{-1/2}} \leq \frac{\epsilon}{2}$  in  $\tilde{O}(m + zr^2k^3/\epsilon^2)$  time. Note that  $\mathbf{B}^\top \mathbf{x}^* = \mathbf{B}^\top \mathbf{x} = \mathbf{d}$ . However, we may not have  $\mathbf{B}^\top \mathbf{x}' = \mathbf{d}$ . To fix this, we define

$$\tilde{\mathbf{x}} \stackrel{\text{def}}{=} \mathbf{x}' + \mathbf{W}_k \mathbf{B} (\mathbf{B}^\top \mathbf{W}_k \mathbf{B})^{-1} (\mathbf{d} - \mathbf{B}^\top \mathbf{x}').$$

This can be found in an extra  $\tilde{O}(m)$  time (Theorem 3.1). Furthermore, we have

$$\begin{aligned}
\|\tilde{\mathbf{x}} - \mathbf{x}'\|_{\mathbf{W}_k^{-1}} &= \|\mathbf{W}_k^{1/2} \mathbf{B} (\mathbf{B}^\top \mathbf{W}_k \mathbf{B})^{-1} (\mathbf{B}^\top \mathbf{x}^* - \mathbf{B}^\top \mathbf{x}')\|_2 \\
&\leq \|\mathbf{W}_k^{-1/2} (\mathbf{x}^* - \mathbf{x}')\|_2 \leq \frac{\epsilon}{2}.
\end{aligned}$$

Hence, we have  $\|\tilde{\mathbf{x}} - \mathbf{x}\|_{\mathbf{W}_k^{-1}} \leq \epsilon$  and  $\mathbf{B}^\top \tilde{\mathbf{x}} = \mathbf{d}$ .

The algorithm and analysis for computing  $\mathbf{s}'$  is analogous with  $\Delta_2$  set to 0 and the signs of the exponents of some  $\mathbf{W}_i$  flipped. The main difference is that  $\mathbf{s}'$  is automatically feasible and hence we simply set  $\tilde{\mathbf{s}} = \mathbf{s}'$ . To see this, we note that the new  $\Delta_1^{(s)}$  is given by

$$\Delta_1^{(s)} = \sum_{i \in [k]} h^{(i)} \mathbf{B} \begin{bmatrix} \mathbf{I} & -\mathbf{L}_{FF}^{-1} \mathbf{L}_{FC} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{L}_{FF}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{SC}(\mathbf{L}_i, C)^\dagger \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{L}_{CF} \mathbf{L}_{FF}^{-1} & \mathbf{I} \end{bmatrix} \mathbf{B}^\top \mathbf{W}_i^{1/2} \mathbf{v}^{(i)}.$$

Note that after we replacing  $\mathbf{SC}(\mathbf{L}_i, C)^\dagger$  by its approximation, the vector above is still in the image of  $\mathbf{B}$ . Hence,  $\mathbf{s}' - \mathbf{s}^*$  is in the image of  $\mathbf{B}$ .  $\square$

## 7.5 Robust IPM Stability Bound

In this section we prove Lemma 7.8 which bounds the relative change in  $\phi$  in each iteration of the robust IPM method (See Section 7). We first provide helper Lemma 7.10 and Lemma 7.11 and then use it to prove Lemma 7.8. The first lemma is a statement about 1-dimensional log-barrier problems.

**Lemma 7.10.** *Let  $\ell, \mathbf{u} \in \mathbb{R}^m$  with  $\ell_i < \mathbf{u}_i$  for all  $i \in [m]$ ,  $c \in \mathbb{R}$ ,  $\mathbf{w}_\ell^{(0)}, \mathbf{w}_u^{(0)}, \mathbf{w}_\ell^{(1)}, \mathbf{w}_u^{(1)} \in [\frac{7}{8}, \frac{8}{7}]^m$ , and for  $j \in \{0, 1\}$  let*

$$x^{(j)} \stackrel{\text{def}}{=} \arg \min_{\max_{i \in [m]} \ell_i \leq x \leq \min_{i \in [m]} \mathbf{u}_i} c \cdot x - \sum_{i \in [m]} \mathbf{w}_{\ell,i}^{(j)} \log(x - \ell_i) - \sum_{i \in [m]} \mathbf{w}_{u,i}^{(j)} \log(\mathbf{u}_i - x)$$

Then, for  $r(a) \stackrel{\text{def}}{=} \max\{a - 3, a^{-1} - 3, 0\}$  we have

$$\sum_{i \in [m]} r\left(\frac{x^{(1)} - \ell_i}{x^{(0)} - \ell_i}\right) + \sum_{i \in [m]} r\left(\frac{\mathbf{u}_i - x^{(1)}}{\mathbf{u}_i - x^{(0)}}\right) \leq 16 \left[ \|\mathbf{w}_\ell^{(0)} - \mathbf{w}_\ell^{(1)}\|_2^2 + \|\mathbf{w}_u^{(0)} - \mathbf{w}_u^{(1)}\|_2^2 \right]$$

**Remark.** Note that  $r\left(\frac{x^{(1)} - \ell_i}{x^{(0)} - \ell_i}\right)$  large implies either  $x^{(1)}$  or  $x^{(0)}$  is much closer to  $\ell_i$  compared to the another one. The inequality above shows that if the weights do not change too much, then  $x^{(1)}$  cannot be too much closer to  $\ell_i$  compared to  $x^{(0)}$ .

*Proof.* Without loss of generality, we assume  $x^{(0)} < x^{(1)}$ . By the optimality condition of  $x^{(j)}$ , we have that

$$c - \sum_{i \in [m]} \frac{\mathbf{w}_{\ell,i}^{(j)}}{x^{(j)} - \ell_i} + \sum_{i \in [m]} \frac{\mathbf{w}_{u,i}^{(j)}}{\mathbf{u}_i - x^{(j)}} = 0.$$

Subtracting this equation for  $j = 0$  and 1, we have

$$\sum_{i \in [m]} \left( \frac{\mathbf{w}_{\ell,i}^{(0)}}{x^{(0)} - \ell_i} - \frac{\mathbf{w}_{\ell,i}^{(1)}}{x^{(1)} - \ell_i} \right) = \sum_{i \in [m]} \left( \frac{\mathbf{w}_{u,i}^{(0)}}{\mathbf{u}_i - x^{(0)}} - \frac{\mathbf{w}_{u,i}^{(1)}}{\mathbf{u}_i - x^{(1)}} \right). \quad (20)$$

We bound the left and right hand side above separately.

To lower bound the left hand side of (20), we let  $\alpha_i = \frac{x^{(1)} - x^{(0)}}{x^{(1)} - \ell_i} \in (0, 1)$  (since  $x^{(1)} > x^{(0)} > \ell_i$ ). Note that

$$\frac{\mathbf{w}_{\ell,i}^{(0)}}{x^{(0)} - \ell_i} - \frac{\mathbf{w}_{\ell,i}^{(1)}}{x^{(1)} - \ell_i} = \frac{\alpha_i}{x^{(1)} - x^{(0)}} \left( \frac{\mathbf{w}_{\ell,i}^{(0)}}{1 - \alpha_i} - \mathbf{w}_{\ell,i}^{(1)} \right).$$

If  $\alpha_i \leq 2(\mathbf{w}_{\ell,i}^{(1)} - \mathbf{w}_{\ell,i}^{(0)})$  then

$$\frac{\mathbf{w}_{\ell,i}^{(0)}}{x^{(0)} - \ell_i} - \frac{\mathbf{w}_{\ell,i}^{(1)}}{x^{(1)} - \ell_i} \geq \frac{\alpha_i(\mathbf{w}_{\ell,i}^{(0)} - \mathbf{w}_{\ell,i}^{(1)})}{x^{(1)} - x^{(0)}} \geq -\frac{2(\mathbf{w}_{\ell,i}^{(0)} - \mathbf{w}_{\ell,i}^{(1)})^2}{x^{(1)} - x^{(0)}}$$

where we used  $\mathbf{w}_{\ell,i}^{(0)} - \mathbf{w}_{\ell,i}^{(1)} \leq 0$  in the last inequality. Otherwise, we have  $\mathbf{w}_{\ell,i}^{(0)} - \mathbf{w}_{\ell,i}^{(1)} > -\alpha_i/2$  and hence

$$\frac{\mathbf{w}_{\ell,i}^{(0)}}{1 - \alpha_i} - \mathbf{w}_{\ell,i}^{(1)} = \frac{\mathbf{w}_{\ell,i}^{(0)} - \mathbf{w}_{\ell,i}^{(1)} + \alpha_i \mathbf{w}_{\ell,i}^{(1)}}{1 - \alpha_i} \geq \frac{\alpha_i(\mathbf{w}_{\ell,i}^{(1)} - 1/2)}{1 - \alpha_i} \geq \frac{1}{3} \frac{\alpha_i}{1 - \alpha_i} \geq 0$$

where we used that  $\mathbf{w}_{\ell,i}^{(1)} \geq \frac{7}{8}$  and  $\frac{7}{8} - \frac{1}{2} \geq \frac{1}{3}$ . Combining both cases, we have

$$\begin{aligned} & (x^{(1)} - x^{(0)}) \cdot \sum_{i \in [m]} \left( \frac{\mathbf{w}_{\ell,i}^{(0)}}{x^{(0)} - \ell_i} - \frac{\mathbf{w}_{\ell,i}^{(1)}}{x^{(1)} - \ell_i} \right) \\ & \geq -2 \sum_{\alpha_i \leq 2(\mathbf{w}_{\ell,i}^{(1)} - \mathbf{w}_{\ell,i}^{(0)})} (\mathbf{w}_{\ell,i}^{(0)} - \mathbf{w}_{\ell,i}^{(1)})^2 + \sum_{\alpha_i > 2(\mathbf{w}_{\ell,i}^{(1)} - \mathbf{w}_{\ell,i}^{(0)})} \frac{1}{3} \frac{\alpha_i^2}{1 - \alpha_i} \\ & \geq -2 \sum_{i \in [m]} (\mathbf{w}_{\ell,i}^{(0)} - \mathbf{w}_{\ell,i}^{(1)})^2 + \frac{1}{8} \sum_{\alpha_i \geq \frac{2}{3}} \frac{1}{1 - \alpha_i} \end{aligned} \quad (21)$$

where we used that  $\alpha_i \geq \frac{2}{3}$  implies  $\alpha_i > 2(\mathbf{w}_{\ell,i}^{(1)} - \mathbf{w}_{\ell,i}^{(0)})$  at the end.

To upper bound the right hand side of (20), we let  $\beta_i = \frac{x^{(1)} - x^{(0)}}{\mathbf{u}_i - x^{(0)}} \in (0, 1)$ . Note that

$$\frac{\mathbf{w}_{u,i}^{(0)}}{\mathbf{u}_i - x^{(0)}} - \frac{\mathbf{w}_{u,i}^{(1)}}{\mathbf{u}_i - x^{(1)}} = \frac{\beta_i}{x^{(1)} - x^{(0)}} \left( \mathbf{w}_{u,i}^{(0)} - \frac{\mathbf{w}_{u,i}^{(1)}}{1 - \beta_i} \right).$$

If  $\beta_i \leq 2(\mathbf{w}_{u,i}^{(0)} - \mathbf{w}_{u,i}^{(1)})$ , we have that

$$\frac{\mathbf{w}_{u,i}^{(0)}}{\mathbf{u}_i - x^{(0)}} - \frac{\mathbf{w}_{u,i}^{(1)}}{\mathbf{u}_i - x^{(1)}} \leq \frac{\beta_i(\mathbf{w}_{u,i}^{(0)} - \mathbf{w}_{u,i}^{(1)})}{x^{(1)} - x^{(0)}} \leq \frac{2(\mathbf{w}_{u,i}^{(0)} - \mathbf{w}_{u,i}^{(1)})^2}{x^{(1)} - x^{(0)}}$$

where we used  $\mathbf{w}_{u,i}^{(0)} - \mathbf{w}_{u,i}^{(1)} \geq 0$  at the last inequality. Otherwise, we have  $\mathbf{w}_{u,i}^{(0)} - \mathbf{w}_{u,i}^{(1)} \leq \beta_i/2$  and hence

$$\mathbf{w}_{u,i}^{(0)} - \frac{\mathbf{w}_{u,i}^{(1)}}{1 - \beta_i} = \frac{\mathbf{w}_{u,i}^{(0)} - \mathbf{w}_{u,i}^{(1)} - \beta_i \mathbf{w}_{u,i}^{(0)}}{1 - \beta_i} \leq \frac{\beta_i(\frac{1}{2} - \mathbf{w}_{u,i}^{(0)})}{1 - \beta_i} \leq -\frac{1}{3} \frac{\beta_i}{1 - \beta_i} \leq 0.$$

Combining both cases and using  $\beta_i \geq \frac{2}{3}$  implies  $\beta_i > 2(\mathbf{w}_{u,i}^{(0)} - \mathbf{w}_{u,i}^{(1)})$ , we have

$$(x^{(1)} - x^{(0)}) \cdot \sum_{i \in [m]} \left( \frac{\mathbf{w}_{u,i}^{(0)}}{\mathbf{u}_i - x^{(0)}} - \frac{\mathbf{w}_{u,i}^{(1)}}{\mathbf{u}_i - x^{(1)}} \right) \leq 2 \sum_{i \in [m]} (\mathbf{w}_{u,i}^{(0)} - \mathbf{w}_{u,i}^{(1)})^2 - \frac{1}{8} \sum_{\beta_i \geq \frac{2}{3}} \frac{1}{1 - \beta_i} \quad (22)$$

Combining (21) and (22) with (20), we have

$$-2 \sum_{i \in [m]} (\mathbf{w}_{\ell,i}^{(0)} - \mathbf{w}_{\ell,i}^{(1)})^2 + \frac{1}{8} \sum_{\alpha_i \geq \frac{2}{3}} \frac{1}{1 - \alpha_i} \leq 2 \sum_{i \in [m]} (\mathbf{w}_{u,i}^{(0)} - \mathbf{w}_{u,i}^{(1)})^2 - \frac{1}{8} \sum_{\beta_i \geq \frac{2}{3}} \frac{1}{1 - \beta_i}.$$

Using this,  $x^{(0)} \leq x^{(1)}$  and the formula of  $r$ , we have

$$\begin{aligned}
\sum_{i \in [m]} r \left( \frac{x^{(1)} - \ell_i}{x^{(0)} - \ell_i} \right) + r \left( \frac{\mathbf{u}_i - x^{(1)}}{\mathbf{u}_i - x^{(0)}} \right) &\leq \sum_{i \in [m]} \max \left\{ \frac{x^{(1)} - \ell_i}{x^{(0)} - \ell_i} - 3, 0 \right\} + \sum_{i \in [m]} \max \left\{ \frac{\mathbf{u}_i - x^{(0)}}{\mathbf{u}_i - x^{(1)}} - 3, 0 \right\} \\
&= \sum_{i \in [m]} \max \left\{ \frac{1}{1 - \alpha_i} - 3, 0 \right\} + \sum_{i \in [m]} \max \left\{ \frac{1}{1 - \beta_i} - 3, 0 \right\} \\
&\leq \sum_{\alpha_i \geq \frac{2}{3}} \frac{1}{1 - \alpha_i} + \sum_{\beta_i \geq \frac{2}{3}} \frac{1}{1 - \beta_i} \\
&\leq 16 \left[ \left\| \mathbf{w}_\ell^{(0)} - \mathbf{w}_\ell^{(1)} \right\|_2^2 + \left\| \mathbf{w}_u^{(0)} - \mathbf{w}_u^{(1)} \right\|_2^2 \right].
\end{aligned}$$

□

We leverage this lemma to generalize to higher dimensions in the following lemma.

**Lemma 7.11.** *In the setting of (16), given weights  $\mathbf{w}_\ell^{(0)}, \mathbf{w}_u^{(0)}, \mathbf{w}_\ell^{(1)}, \mathbf{w}_u^{(1)} \in [\frac{7}{8}, \frac{8}{7}]^m$  let*

$$\mathbf{x}_j \stackrel{\text{def}}{=} \arg \min_{\mathbf{x} \in \mathcal{X} \mid \mathbf{B}^\top \mathbf{x} = \mathbf{d}} \mathbf{c}^\top \mathbf{x} - \sum_{i \in [m]} \mathbf{w}_{\ell,i}^{(j)} \log(\mathbf{x} - \ell_i) - \sum_{i \in [m]} \mathbf{w}_{u,i}^{(j)} \log(\mathbf{u}_i - \mathbf{x})$$

for  $j \in \{0, 1\}$ . Then for  $r(a) \stackrel{\text{def}}{=} \max\{a - 3, a^{-1} - 3, 0\}$  we have

$$\sum_{i \in [m]} r \left( \frac{\mathbf{x}_i^{(1)} - \ell_i}{\mathbf{x}_i^{(0)} - \ell_i} \right) + \sum_{i \in [m]} r \left( \frac{\mathbf{u}_i - \mathbf{x}^{(1)}}{\mathbf{u}_i - \mathbf{x}^{(0)}} \right) \leq 16 \left[ \left\| \mathbf{w}_\ell^{(0)} - \mathbf{w}_\ell^{(1)} \right\|_2^2 + \left\| \mathbf{w}_u^{(0)} - \mathbf{w}_u^{(1)} \right\|_2^2 \right].$$

*Proof.* For all  $t \in \mathbb{R}$  let  $\mathbf{x}^{(t)} \stackrel{\text{def}}{=} \mathbf{x}^{(0)} + t\delta_x$  with  $\delta_x = \mathbf{x}^{(1)} - \mathbf{x}^{(0)}$ . By the definition of  $\mathbf{x}^{(j)}$ , we have that for  $j \in \{0, 1\}$

$$\begin{aligned}
\mathbf{x}^{(j)} &= \arg \min_{t \in \mathbb{R} \mid \mathbf{x}^{(t)} \in \mathcal{X}} \mathbf{c}^\top \mathbf{x}^{(t)} - \sum_{i \in [m]} \mathbf{w}_{\ell,i}^{(j)} \log(\mathbf{x}_i^{(t)} - \ell_i) - \sum_{i \in [m]} \mathbf{w}_{u,i}^{(j)} \log(\mathbf{u}_i - \mathbf{x}_i^{(t)}) \\
&= \arg \min_{t \in \mathbb{R} \mid \mathbf{x}^{(t)} \in \mathcal{X}} t \cdot \mathbf{c}^\top \delta_x - \sum_{i \in [m]} \mathbf{w}_{\ell,i}^{(j)} \log(t\delta_{x,i} - (\ell_i - \mathbf{x}_i^{(0)})) - \sum_{i \in [m]} \mathbf{w}_{u,i}^{(j)} \log((\mathbf{u}_i - \mathbf{x}_i^{(0)}) - t\delta_{x,i}) \\
&= \arg \min_{t \in \mathbb{R} \mid \mathbf{x}^{(t)} \in \mathcal{X}} t \cdot \tilde{\mathbf{c}} - \sum_{i \in [m] : \mathbf{x}_i^{(1)} \neq \mathbf{x}_i^{(0)}} \tilde{\mathbf{w}}_{\ell,i}^{(j)} \log(t - \tilde{\ell}_i) - \sum_{i \in [m]} \tilde{\mathbf{w}}_{u,i}^{(j)} \log(\tilde{\mathbf{u}}_i - t)
\end{aligned} \tag{23}$$

where  $\tilde{\mathbf{c}} \stackrel{\text{def}}{=} \mathbf{c}^\top \delta_x$  and

$$(\tilde{\ell}_i, \tilde{\mathbf{u}}_i, \tilde{\mathbf{w}}_{\ell,i}^{(j)}, \tilde{\mathbf{w}}_{u,i}^{(j)}) \stackrel{\text{def}}{=} \begin{cases} \left( \frac{\ell_i - \mathbf{x}_i^{(0)}}{\mathbf{x}_i^{(1)} - \mathbf{x}_i^{(0)}}, \frac{\mathbf{u}_i - \mathbf{x}_i^{(0)}}{\mathbf{x}_i^{(1)} - \mathbf{x}_i^{(0)}}, \mathbf{w}_{\ell,i}^{(j)}, \mathbf{w}_{u,i}^{(j)} \right) & \text{if } \delta_{x,i} > 0 \\ \left( \frac{\mathbf{u}_i - \mathbf{x}_i^{(0)}}{\mathbf{x}_i^{(0)} - \mathbf{x}_i^{(1)}}, \frac{\ell_i - \mathbf{x}_i^{(0)}}{\mathbf{x}_i^{(0)} - \mathbf{x}_i^{(1)}}, \mathbf{w}_{u,i}^{(j)}, \mathbf{w}_{\ell,i}^{(j)} \right) & \text{if } \delta_{x,i} < 0 \end{cases}.$$

Applying Lemma 7.10 to (23) then yields the result as for all  $i \in [m]$  with  $\delta_{x,i} > 0$ , we have that

$$\frac{1 - \tilde{\ell}_i}{0 - \tilde{\ell}_i} = \frac{1 - \frac{\ell_i - \mathbf{x}_i^{(0)}}{\mathbf{x}_i^{(1)} - \mathbf{x}_i^{(0)}}}{\frac{\mathbf{x}_i^{(0)} - \ell_i}{\mathbf{x}_i^{(1)} - \mathbf{x}_i^{(0)}}} = \frac{\mathbf{x}_i^{(1)} - \ell_i}{\mathbf{x}_i^{(0)} - \ell_i} \text{ and } \frac{1 - \tilde{\mathbf{u}}_i}{0 - \tilde{\mathbf{u}}_i} = \frac{1 - \frac{\mathbf{u}_i - \mathbf{x}_i^{(0)}}{\mathbf{x}_i^{(1)} - \mathbf{x}_i^{(0)}}}{\frac{\mathbf{x}_i^{(0)} - \mathbf{u}_i}{\mathbf{x}_i^{(1)} - \mathbf{x}_i^{(0)}}} = \frac{\mathbf{u}_i - \mathbf{x}_i^{(1)}}{\mathbf{u}_i - \mathbf{x}_i^{(0)}}$$

and similarly when  $\delta_{x,i} < 0$ , we have

$$\frac{1 - \tilde{\ell}_i}{0 - \tilde{\ell}_i} = \frac{1 - \frac{\mathbf{u}_i - \mathbf{x}_i^{(0)}}{\mathbf{x}_i^{(0)} - \mathbf{x}_i^{(1)}}}{\frac{\mathbf{x}_i^{(0)} - \mathbf{u}_i}{\mathbf{x}_i^{(0)} - \mathbf{x}_i^{(1)}}} = \frac{\mathbf{u}_i - \mathbf{x}_i^{(1)}}{\mathbf{u}_i - \mathbf{x}_i^{(0)}} \text{ and } \frac{1 - \tilde{\mathbf{u}}_i}{0 - \tilde{\mathbf{u}}_i} = \frac{1 - \frac{\ell_i - \mathbf{x}_i^{(0)}}{\mathbf{x}_i^{(0)} - \mathbf{x}_i^{(1)}}}{\frac{\mathbf{x}_i^{(0)} - \ell_i}{\mathbf{x}_i^{(0)} - \mathbf{x}_i^{(1)}}} = \frac{\mathbf{x}_i^{(1)} - \ell_i}{\mathbf{x}_i^{(0)} - \ell_i}.$$

□

Leveraging Lemma 7.11, we prove the main result of this section, Lemma 7.8.

*Proof of Lemma 7.8.* To apply Lemma 7.11 for  $j \in \{0, 1\}$  we define  $\bar{\mu} \stackrel{\text{def}}{=} \frac{1}{2}[\mu^{(0)} + \mu^{(1)}]$ ,  $\bar{\mathbf{y}}^{(j)} = \frac{\mu^{(j)}}{\bar{\mu}} \mathbf{y}^{(j)}$ ,  $\bar{\boldsymbol{\eta}}^{(j)} = \frac{\mu^{(j)}}{\bar{\mu}} \boldsymbol{\eta}^{(j)} + \bar{\mathbf{c}} - \mathbf{c}/\bar{\mu}$  and

$$\bar{\mathbf{c}}_i = \begin{cases} \mathbf{c}_i/\bar{\mu} - \frac{\mu^{(0)}}{\bar{\mu}} \boldsymbol{\eta}_i^{(0)} & \text{if } \phi_i''(\mathbf{x}^{(0)}) < \phi_i''(\mathbf{x}^{(1)}) \\ \mathbf{c}_i/\bar{\mu} - \frac{\mu^{(1)}}{\bar{\mu}} \boldsymbol{\eta}_i^{(1)} & \text{else} \end{cases}.$$

With these definitions, we note that  $\mathbf{B}\bar{\mathbf{y}}^{(j)} + \bar{\mathbf{c}} + \nabla\phi(\mathbf{x}^{(j)}) = \bar{\boldsymbol{\eta}}^{(j)}$ . Since Lemma 7.10 considers only exact minimizers of weighted log barriers, we remove the term  $\bar{\beta}^{(j)}$  using the weights as follows, we define  $\phi_{\ell,i}(\mathbf{x}) \stackrel{\text{def}}{=} -\sum_{i \in [m]} \log(\mathbf{x}_i - \ell_i)$ ,  $\phi_{u,i}(\mathbf{x}) \stackrel{\text{def}}{=} -\sum_{i \in [m]} \log(\mathbf{u}_i - \mathbf{x}_i)$ ,  $\phi_{\ell}(\mathbf{x}) = \sum_{i \in [m]} \phi_{\ell,i}(\mathbf{x})$ , and  $\phi_u(\mathbf{x}) = \sum_{i \in [m]} \phi_{u,i}(\mathbf{x})$ . Then, we define the weights

$$\alpha_i^{(j)} = 1 - \frac{\text{sign}([\nabla\phi_{\ell,i}(\mathbf{x}^{(j)})]_i)}{||[\nabla\phi_{\ell,i}(\mathbf{x}^{(j)})]_i| + |[\nabla\phi_{u,i}(\mathbf{x}^{(j)})]_i|} \bar{\eta}_i^{(j)} \text{ and } \beta_i^{(j)} = 1 - \frac{\text{sign}([\nabla\phi_{u,i}(\mathbf{x}^{(j)})]_i)}{||[\nabla\phi_{\ell,i}(\mathbf{x}^{(j)})]_i| + |[\nabla\phi_{u,i}(\mathbf{x}^{(j)})]_i|} \bar{\eta}_i^{(j)}$$

so that  $\sum_{i \in [m]} (\alpha_i^{(j)} \nabla\phi_{\ell,i}(\mathbf{x}^{(j)}) + \beta_i^{(j)} \nabla\phi_{u,i}(\mathbf{x}^{(j)})) = \nabla\phi(\mathbf{x}^{(j)}) - \bar{\boldsymbol{\eta}}$  and

$$\mathbf{B}\bar{\mathbf{y}}^{(j)} + \bar{\mathbf{c}} + \sum_{i \in [m]} (\alpha_i^{(j)} \nabla\phi_{\ell,i}(\mathbf{x}^{(j)}) + \beta_i^{(j)} \nabla\phi_{u,i}(\mathbf{x}^{(j)})) = 0.$$

Consequently,

$$\mathbf{x}^{(j)} = \arg \min_{\mathbf{x} \in \mathcal{X} \mid \mathbf{B}^\top \mathbf{x} = \mathbf{d}} \bar{\mathbf{c}}^\top \mathbf{x} + \sum_{i \in [m]} \alpha_i^{(j)} \phi_{\ell,i}(\mathbf{x}) + \sum_{i \in [m]} \beta_i^{(j)} \phi_{u,i}(\mathbf{x}).$$

Hence, we can apply Lemma 7.10 with  $\mathbf{w}_{\ell,i}^{(j)} = \alpha_i^{(j)}$  and  $\mathbf{w}_{u,i}^{(j)} = \beta_i^{(j)}$  provided  $\frac{7}{8} \leq \mathbf{w}_{\ell,i}^{(j)} \leq \frac{8}{7}$  and  $\frac{7}{8} \leq \mathbf{w}_{u,i}^{(j)} \leq \frac{8}{7}$  for all  $i \in [m]$  and  $j \in \{1, 2\}$ . To show this, note that

$$|\alpha_i^{(j)} - 1| \leq \frac{|\bar{\eta}_i^{(j)}|}{|\phi'_{\ell,i}(\mathbf{x}^{(j)})| + |\phi'_{u,i}(\mathbf{x}^{(j)})|} = \frac{|\bar{\eta}_i^{(j)}|}{\sqrt{\phi''_{\ell,i}(\mathbf{x}^{(j)})} + \sqrt{\phi''_{u,i}(\mathbf{x}^{(j)})}} \leq \frac{|\bar{\eta}_i^{(j)}|}{\sqrt{\phi_i''(\mathbf{x}^{(j)})}} \quad (24)$$

where we used the definition of  $\phi_u$  and  $\phi_\ell$  in the equality. Now, using the definition of  $\bar{\eta}_i^{(j)}$ , we have

$$|\bar{\eta}_i^{(j)}| = \begin{cases} \bar{\mu}^{-1} \left| \eta_i^{(j)} \mu^{(j)} - \eta_i^{(0)} \mu^{(0)} \right| & \text{if } \phi_i''(\mathbf{x}^{(0)}) < \phi_i''(\mathbf{x}^{(1)}) \\ \bar{\mu}^{-1} \left| \eta_i^{(j)} \mu^{(j)} - \eta_i^{(1)} \mu^{(1)} \right| & \text{otherwise} \end{cases} \\ = \begin{cases} 0 & \text{if } \phi_i''(\mathbf{x}^{(j)}) < \phi_i''(\mathbf{x}^{(1-j)}) \\ \bar{\mu}^{-1} \left| \eta_i^{(1)} \mu^{(1)} - \eta_i^{(0)} \mu^{(0)} \right| & \text{otherwise} \end{cases}.$$



Hence, we have

$$\frac{|\bar{\eta}_i^{(j)}|}{\sqrt{\phi_i''(\mathbf{x}^{(j)})}} \leq \frac{|\boldsymbol{\eta}_i^{(1)} \mu^{(1)} - \boldsymbol{\eta}_i^{(0)} \mu^{(0)}| / \bar{\mu}}{\max(\sqrt{\phi_i''(\mathbf{x}^{(0)})}, \sqrt{\phi_i''(\mathbf{x}^{(1)})})}. \quad (25)$$

Using that  $\mu^{(0)} \approx_{1/32} \mu^{(1)}$  and  $\left\| \eta^{(j)} / \sqrt{\phi''(\mathbf{x}^{(j)})} \right\|_\infty \leq \frac{1}{32}$ , we have that  $\frac{|\bar{\eta}_i^{(j)}|}{\sqrt{\phi_i''(\mathbf{x}^{(j)})}} \leq \frac{1}{8}$ . Hence, (24) shows that  $|\mathbf{w}_{\ell,i}^{(j)} - 1| \leq 1/8$  for all  $i \leq m$  and  $j$ . The same proof gives the bound of  $\beta_i^{(j)}$ .

Consequently, Lemma 7.10 shows that

$$\sum_{i \in [m]} r \left( \frac{\mathbf{x}_i^{(1)} - \ell_i}{\mathbf{x}_i^{(0)} - \ell_i} \right) + \sum_{i \in [m]} r \left( \frac{\mathbf{u}_i - \mathbf{x}_i^{(1)}}{\mathbf{u}_i - \mathbf{x}_i^{(0)}} \right) \leq 16 \left[ \left\| \mathbf{w}_\ell^{(0)} - \mathbf{w}_\ell^{(1)} \right\|_2^2 + \left\| \mathbf{w}_u^{(0)} - \mathbf{w}_u^{(1)} \right\|_2^2 \right].$$

Using (24) and (25), we have that

$$\begin{aligned} \left\| \mathbf{w}_\ell^{(0)} - \mathbf{w}_\ell^{(1)} \right\|_2^2 + \left\| \mathbf{w}_u^{(0)} - \mathbf{w}_u^{(1)} \right\|_2^2 &\leq 8 \sum_{i \in [m]} \frac{\bar{\mu}^{-2} (\boldsymbol{\eta}_i^{(1)} \mu^{(1)} - \boldsymbol{\eta}_i^{(0)} \mu^{(0)})^2}{\max(\phi_i''(\mathbf{x}^{(0)}), \phi_i''(\mathbf{x}^{(1)}))} \\ &\leq 32 \bar{\mu}^{-2} \sum_{i \in [m]} \frac{(\boldsymbol{\eta}_i^{(1)} \mu^{(1)} - \boldsymbol{\eta}_i^{(0)} \mu^{(1)})^2 + (\boldsymbol{\eta}_i^{(0)} \mu^{(1)} - \boldsymbol{\eta}_i^{(0)} \mu^{(0)})^2}{\phi_i''(\mathbf{x}^{(0)}) + \phi_i''(\mathbf{x}^{(1)})} \\ &\leq 64 \sum_{i \in [m]} \frac{(\boldsymbol{\eta}_i^{(1)} - \boldsymbol{\eta}_i^{(0)})^2}{\phi_i''(\mathbf{x}^{(0)}) + \phi_i''(\mathbf{x}^{(1)})} + \frac{1}{2} \bar{\mu}^{-2} m (\mu^{(1)} - \mu^{(0)})^2 \\ &\leq 64 \sum_{i \in [m]} \frac{(\boldsymbol{\eta}_i^{(1)} - \boldsymbol{\eta}_i^{(0)})^2}{\phi_i''(\mathbf{x}^{(0)}) + \phi_i''(\mathbf{x}^{(1)})} + m \left( \frac{\mu^{(1)} - \mu^{(0)}}{\mu^{(0)}} \right)^2. \end{aligned} \quad (26)$$

Finally, we note that  $\phi_i''(\mathbf{x}^{(0)}) \leq \max(\frac{\mathbf{x}^{(1)} - \ell_i}{\mathbf{x}^{(0)} - \ell_i}, \frac{\mathbf{u}_i - \mathbf{x}^{(1)}}{\mathbf{u}_i - \mathbf{x}^{(0)}})^2 \cdot \phi_i''(\mathbf{x}^{(1)})$ . Hence, we have

$$\begin{aligned} \sum_{\phi_i''(\mathbf{x}^{(0)})^{1/2} \geq 3 \phi_i''(\mathbf{x}^{(1)})^{1/2}} \sqrt{\frac{\phi_i''(\mathbf{x}^{(0)})}{\phi_i''(\mathbf{x}^{(1)})}} &\leq \sum_{\max(\frac{\mathbf{x}^{(1)} - \ell_i}{\mathbf{x}^{(0)} - \ell_i}, \frac{\mathbf{u}_i - \mathbf{x}^{(1)}}{\mathbf{u}_i - \mathbf{x}^{(0)}}) \geq 3} \max \left\{ \frac{\mathbf{x}^{(1)} - \ell_i}{\mathbf{x}^{(0)} - \ell_i}, \frac{\mathbf{u}_i - \mathbf{x}^{(1)}}{\mathbf{u}_i - \mathbf{x}^{(0)}} \right\} \\ &\leq \sum_{i \in [m]} r \left( \frac{\mathbf{x}^{(1)} - \ell_i}{\mathbf{x}^{(0)} - \ell_i} \right) + \sum_{i \in [m]} r \left( \frac{\mathbf{u}_i - \mathbf{x}^{(1)}}{\mathbf{u}_i - \mathbf{x}^{(0)}} \right) \\ &\leq 16 \left[ \left\| \mathbf{w}_\ell^{(0)} - \mathbf{w}_\ell^{(1)} \right\|_2^2 + \left\| \mathbf{w}_u^{(0)} - \mathbf{w}_u^{(1)} \right\|_2^2 \right]. \end{aligned} \quad (27)$$

The result then follows from (26) and (27).  $\square$

## 8 Final Runtime Bound

In this section we show Theorem 7.5 which describes how efficiently the data structures we developed in Sections 4 to 6 can implement an IPM step. Our final runtime is then achieved via Theorem 7.9. Finally, we cite previous work to explain how to get an initial point for the IPM, and how to get a mincost flow after running  $\tilde{O}(\sqrt{m})$  IPM iterations.

## 8.1 Efficient Solution Maintenance

**Theorem 7.5** (Graph Solution Maintenance). *In the special case that  $\mathbf{B}$  is the incidence matrix of a  $m$ -edge,  $n$ -node graph, if  $C_r, C_z = \tilde{O}(1)$  and  $\alpha = \tilde{\Omega}(1)$ , there is a  $(\mathcal{T}_{\text{init}}, \mathcal{T}_{\text{phase}})$ -solution maintainer (Definition 7.4) with  $\mathcal{T}_{\text{init}} = \tilde{O}(m)$  and  $\mathcal{T}_{\text{phase}} = \tilde{O}(m + m^{15/16} k^{29/8})$ .*

We first make several useful definitions. We let  $\mathbf{x}^{(j)}, \mathbf{s}^{(j)}$  be the true iterates after the  $j$ -th call to MOVE. Our algorithm will explicitly approximate iterates  $\hat{\mathbf{x}}^{(j)}, \hat{\mathbf{s}}^{(j)}$ . Using these approximate iterates, the algorithm will output  $\bar{\mathbf{x}}^{(j)}, \bar{\mathbf{s}}^{(j)}$  satisfying the desired update schedule, i.e. at most  $\tilde{O}(2^{2\ell_j} \alpha^{-2})$  coordinates are updated after the  $j$ -th call to MOVE. Additionally, call an index  $j$  corresponding to the  $j$ -th MOVE operation *special* if it occurs immediately following a STARTPHASE operation. The formal construction of  $\hat{\mathbf{x}}^{(j)}$  and  $\hat{\mathbf{s}}^{(j)}$  is given in the following definition.

**Definition 8.1** (Approximate iterates). Say there have been  $j$  MOVE operations so far. If the next operation is STARTPHASE( $\tilde{\mathbf{x}}, \tilde{\mathbf{s}}$ ), then set  $\hat{\mathbf{x}}^{(j+1)} \leftarrow \tilde{\mathbf{x}}$  and  $\hat{\mathbf{s}}^{(j+1)} \leftarrow \tilde{\mathbf{s}}$ . If the next operation is MOVE operation  $(j+1)$ , then let  $\delta_{\mathbf{x}}, \delta_{\mathbf{s}}$  satisfy

$$\left\| \delta_{\mathbf{x}} - h^{(j+1)}(\mathbf{I} - \mathbf{P}_{j+1})\mathbf{v}^{(j+1)} \right\|_{\infty} \leq \epsilon \quad \text{and} \quad \left\| \delta_{\mathbf{s}} - h^{(j+1)}\mathbf{P}_{j+1}\mathbf{v}^{(j+1)} \right\|_{\infty} \leq \epsilon, \quad (28)$$

with  $\delta_{\mathbf{x}}, \delta_{\mathbf{s}}$  supported on  $O(\epsilon^{-2})$  coordinates. If the previous operation was Move, define and let  $\hat{\mathbf{x}}^{(j+1)} \leftarrow \hat{\mathbf{x}}^{(j)} + \mathbf{W}_{j+1}^{1/2} \delta_{\mathbf{x}}$  and  $\hat{\mathbf{s}}^{(j+1)} \leftarrow \hat{\mathbf{s}}^{(j)} + \mathbf{W}_{j+1}^{-1/2} \delta_{\mathbf{s}}$ . Otherwise, if the previous operation was STARTPHASE define  $\hat{\mathbf{x}}^{(j+1)} \leftarrow \hat{\mathbf{x}}^{(j+1)} + \mathbf{W}_{j+1}^{1/2} \delta_{\mathbf{x}}$  and  $\hat{\mathbf{s}}^{(j+1)} \leftarrow \hat{\mathbf{s}}^{(j+1)} + \mathbf{W}_{j+1}^{-1/2} \delta_{\mathbf{s}}$  (so that we redefine  $\hat{\mathbf{x}}^{(j+1)}, \hat{\mathbf{s}}^{(j+1)}$ ).

We show that the  $\hat{\mathbf{x}}^{(j)}$  are slowly changing, except potentially at special indices. This is because  $\|\delta_{\mathbf{x}}\|_2 = O(1)$  as it is supported on  $\tilde{O}(\epsilon^{-2})$  nonzeros and  $h^{(j+1)}\|\mathbf{v}^{(j+1)}\|_2 \leq 1$ .

We now argue that  $\delta_{\mathbf{x}}$  and  $\delta_{\mathbf{s}}$  can be computed efficiently.

**Lemma 8.2** (Computation of  $\delta_{\mathbf{x}}, \delta_{\mathbf{s}}$ ). *In the context of Theorem 7.5, there is an operation that computes  $\delta_{\mathbf{x}}, \delta_{\mathbf{s}}$  satisfying (28) in average amortized time  $\tilde{O}(m^{15/16} \epsilon^{-7/8})$  and succeeds with high probability against adaptive adversaries.*

*Proof.* We first write  $h^{(j+1)}(\mathbf{I} - \mathbf{P}_{j+1})\mathbf{v}^{(j+1)} = h^{(j+1)}\mathbf{v}^{(j+1)} - h^{(j+1)}\mathbf{P}_{j+1}\mathbf{v}^{(j+1)}$ , and handle both parts separately up to error  $\epsilon/2$ . The first part can be trivially handled, as it can be explicitly maintained in time proportional to the number of changes in  $\mathbf{v}^{(j)}$ , and  $\|h^{(j+1)}\mathbf{v}^{(j+1)}\|_2 \leq 1$ . For the second part, we first call the dynamic LOCATOR (Theorem 5.10) to get a set  $S$  of size  $O(\epsilon^{-2})$ . Then we call the dynamic EVALUATORS (Theorem 5.9) wrapped inside Theorem 6.2 with  $\epsilon \leftarrow \epsilon/(C \log^2 n)$  on  $S$  by calling QUERY() on  $S$ . The algorithm for  $\delta_{\mathbf{s}}$  follows exactly as the second term. Also,  $\delta_{\mathbf{x}}, \delta_{\mathbf{s}}$  are supported on  $|S| = O(\epsilon^{-2})$  coordinates by Theorem 6.2.

Correctness follows directly from the guarantees of Theorems 5.9, 5.10 and 6.2. It suffices to analyze the amortized runtime. We focus on the cost of applying Theorem 5.9 inside Theorem 6.2, as the cost of Theorem 5.10 is less. Let  $\delta_i \stackrel{\text{def}}{=} 2^{-i}$  so that the  $i$ -th EVALUATOR is run with accuracy  $\epsilon_i \stackrel{\text{def}}{=} \delta_i \epsilon$  in Theorem 6.2. Let  $\beta_i$  be the terminal size parameter for the  $i$ -th EVALUATOR.

There are two possible ways to run the  $i$ -th EVALUATOR. Either it pays  $\tilde{O}(m)$  time per call to solve a Laplacian exactly (while this algorithm is randomized, we can hide randomness by adding polynomially small noise that is larger than the error we solve the Laplacian to [LS15]) or applies Theorem 5.9. Let us calculate the runtime of the latter approach. After  $\beta_i m$  edge updates or

marking, the data structure must re-initialize. Thus, after  $T$  MOVE updates, because  $C_z = \tilde{O}(1)$ , there are at most  $\tilde{O}(T^2 + T\epsilon^{-2})$  total edges we have queried or updated:  $\tilde{O}(T^2)$  from updates, and  $\tilde{O}(T\epsilon^{-2})$  from the set  $S$  returned by LOCATOR. We assume for now that the  $\tilde{O}(T^2)$  term dominates – thus the data structure must reinitialize every  $\sqrt{\beta_i m}$  iterations, where each initialization costs  $\tilde{O}(m\beta_i^{-2}\epsilon_i^{-2})$  time. Thus the amortized reinitialization time per Move is

$$\tilde{O}(m\beta_i^{-2}\epsilon_i^{-2}/\sqrt{\beta_i m}) = \tilde{O}(\sqrt{m}\beta_i^{-5/2}\epsilon^{-2}\delta_i^{-2}).$$

By Theorem 6.2, the  $i$ -th EVALUATOR is queried with probability  $O(\delta_i)$ , hence the expected query time is  $\tilde{O}(\beta_i m \epsilon_i^{-2} \delta_i) = \tilde{O}(\beta_i m \epsilon^{-2} \delta_i^{-1})$  by Theorem 5.9 QUERY(), or  $\tilde{O}(\delta_i m)$  if EVALUATOR simply solves a Laplacian every iteration. Thus, the amortized runtime for the  $i$ -th EVALUATOR is

$$\tilde{O}\left(\min\left\{\delta_i m, \beta_i m \epsilon^{-2} \delta_i^{-1} + \sqrt{m} \beta_i^{-5/2} \epsilon^{-2} \delta_i^{-2}\right\}\right).$$

For the choice  $\beta_i = m^{-1/7} \delta_i^{-2/7}$ , this becomes

$$\tilde{O}\left(\min\left\{\delta_i m, m^{6/7} \epsilon^{-2} \delta_i^{-9/7}\right\}\right).$$

This is maximized when the two expressions are equal at  $\delta_i = m^{-1/16} \epsilon^{-7/8}$ , yielding a runtime of  $\tilde{O}(m^{15/16} \epsilon^{-7/8})$  as desired. Finally, note that this means that  $\epsilon \geq m^{-1/14}$  or the previous runtime is trivial. All  $\beta_i \geq m^{-1/7}$ , so  $T\epsilon^{-2} \leq T^2$  for the choice  $T = \sqrt{\beta_i m} \geq m^{3/7} > \epsilon^{-2}$ , so the  $\tilde{O}(T^2)$  term dominated earlier, as desired.  $\square$

We now show that  $\hat{\mathbf{x}}^{(j)}$  and  $\hat{\mathbf{s}}^{(j)}$  are close to  $\mathbf{x}^{(j)}, \mathbf{s}^{(j)}$ .

**Lemma 8.3.** For  $\epsilon = \frac{\alpha}{10C_r k^3}$  and  $\hat{\mathbf{x}}^{(j)}, \hat{\mathbf{s}}^{(j)}$  defined in Definition 8.1,  $\|\mathbf{W}_j^{-1/2}(\hat{\mathbf{x}}^{(j)} - \mathbf{x}^{(j)})\|_\infty \leq \alpha/10$  and  $\|\mathbf{W}_j^{1/2}(\hat{\mathbf{s}}^{(j)} - \mathbf{s}^{(j)})\|_\infty \leq \alpha/10$ .

*Proof.* It suffices to analyze  $j$  between STARTPHASES, as  $\hat{\mathbf{x}}^{(j)}, \hat{\mathbf{s}}^{(j)}$  and  $\mathbf{x}^{(j)}, \mathbf{s}^{(j)}$  are both set to  $\tilde{\mathbf{x}}, \tilde{\mathbf{s}}$  during a STARTPHASE. Over  $L$  steps between STARTPHASES (from  $j_1$  to  $j_2 = j_1 + L$ ), we have that  $\|\mathbf{W}_{j_2}^{-1/2}(\hat{\mathbf{x}}^{(j_2)} - \mathbf{x}^{(j_2)})\|_\infty$  is at most

$$\begin{aligned} & \left\| \mathbf{W}_{j_2}^{-1/2} \sum_{j \in [j_1, j_2)} \mathbf{W}_j^{1/2} (\delta_{\mathbf{x}}^{(j)} - h^{(j+1)}(\mathbf{I} - \mathbf{P}_{j+1})\mathbf{v}^{(j+1)}) \right\|_\infty \\ & \leq \max_{j \in [j_1, j_2]} \sqrt{\|\mathbf{w}_j / \mathbf{w}_{j_2}\|_\infty} \left\| \sum_{j \in [j_1, j_2)} \delta_{\mathbf{x}}^{(j)} - h^{(j+1)}(\mathbf{I} - \mathbf{P}_{j+1})\mathbf{v}^{(j+1)} \right\|_\infty \\ & \stackrel{(i)}{\leq} C_r L^2 \epsilon L = \epsilon C_r L^3 \leq \alpha/10, \end{aligned}$$

where (i) follows from the guarantee that of Definition 7.4 that  $\sqrt{\|\mathbf{w}_{(j_2)}/\mathbf{w}_{j_1}\|_\infty} \leq rL^2$  and (28).

The bound on the error for  $\mathbf{s}^{(j)}$  follows similarly.  $\square$

*Proof of Theorem 7.5.* We show this by carefully defining  $\bar{\mathbf{x}}^{(j)}, \bar{\mathbf{s}}^{(j)}$  given  $\hat{\mathbf{x}}^{(j)}, \hat{\mathbf{s}}^{(j)}$ . We mimic the approach based on binary expansions given in previous works on robust IPMs, for example [BLSS20],

Theorem 8]. Precisely, we first calculate  $\sum_{j'=j-2^{\ell_j}}^j \delta_{\mathbf{x}}^{(j')}$ , i.e. the sum of errors in the last  $2^{\ell_j}$  steps. If this exceeds  $\frac{\alpha}{100 \log n}$ , then we set  $\bar{\mathbf{x}}^{(j)} \leftarrow \hat{\mathbf{x}}^{(j)}$ , otherwise we set  $\bar{\mathbf{x}}^{(j)} \leftarrow \bar{\mathbf{x}}^{(j-1)}$  (no change). We do the same for  $\bar{\mathbf{s}}^{(j)}$ . Now, the bounds on number of changes follows from the bounds  $\|\delta_{\mathbf{x}}\|_2 \leq O(1)$  and that  $\|\mathbf{W}_j^{-1/2}(\mathbf{x} - \bar{\mathbf{x}})\|_2 \leq 1$  in STARTPHASE. More precisely, over  $2^{\ell_j}$  steps, only  $O(2^{2\ell_j} \alpha^{-2} \log^2 n)$  could change by  $\frac{\alpha}{100 \log n}$ , because every step satisfies  $\|\delta_{\mathbf{x}}\|_2 \leq 1$  or  $\|\mathbf{W}_j^{-1/2}(\mathbf{x} - \bar{\mathbf{x}})\|_2 \leq 1$ . This completes the proof of the number of changes.

We now claim that  $\|\mathbf{W}_j^{-1/2}(\bar{\mathbf{x}}^{(j)} - \hat{\mathbf{x}}^{(j)})\|_{\infty} \leq \alpha/10$ , so  $\|\mathbf{W}_j^{-1/2}(\bar{\mathbf{x}}^{(j)} - \mathbf{x}^{(j)})\|_{\infty} \leq \alpha$  by combining with Lemma 8.3 for  $\epsilon = \frac{\alpha}{10C_r k^3} = \tilde{O}(1/k^3)$  as  $C_r = \tilde{O}(1)$ . This claim follows from the same argument as [BLSS20, Theorem 8]: each interval  $[j_1, j_2]$  can be split into  $\log n$  intervals contained in intervals  $[j - 2^{\ell_j}, j]$  for  $j \leq j_2$ . Each of these has at most  $\alpha/(100 \log n)$  error, so the total error is at most  $\alpha/(100 \log n) \cdot \log n \leq \alpha/10$ .

Finally, we must calculate the runtime of  $\mathcal{T}_{\text{phase}}$ . The first cost is  $\tilde{O}(m)$  (eg. for reading  $\tilde{\mathbf{x}}, \tilde{\mathbf{s}}$ ). The second cost is calling Lemma 8.2  $k$  times (as there are at most  $k$  MOVE operations between STARTPHASE). For our choice  $\epsilon = \tilde{O}(1/k^3)$ , the total time for this is  $\tilde{O}(m^{15/16} \epsilon^{-7/8} k) = \tilde{O}(m^{15/16} k^{29/8})$  as desired.  $\square$

## 8.2 Initial Point, Final Point, and Proof of Main Theorem

It is standard to get an initial  $\mu$ -centered feasible pair  $(\mathbf{x}, \mathbf{s})$  for path parameter  $\mu = \mu_{\text{start}} \geq (mU)^{O(1)}$ . Additionally, given a  $\mu$ -centered feasible pair  $(\mathbf{x}, \mathbf{s})$  for path parameter  $\mu = \mu_{\text{end}} \leq (mU)^{-O(1)}$  we can recover a high-accuracy mincost flow (and hence round to an exact solution).

**Lemma 8.4** ([BLL<sup>+</sup>21, Lemma 7.5, Lemma 7.8]). *Given a graph  $G = (V, E)$  and mincost flow instance with demand  $\mathbf{d} \in [-U, \dots, U]^V$ , costs  $\mathbf{c} \in [-U, \dots, U]^E$ , and capacities  $\ell, \mathbf{u} \in [-U, \dots, U]^E$ , we can build a mincost flow instance on a graph  $G'$  with at most  $O(m)$  edges with demands, costs, and capacities bounded by  $\text{poly}(mU)$ . Additionally, we can construct a  $\mu_{\text{start}}$ -centered pair  $(\mathbf{f}, \mathbf{s})$  on  $G'$  for  $\mu_{\text{start}} = \text{poly}(mU)$ . Additionally, given a  $1/\text{poly}(mU)$ -accurate mincost flow on  $G'$  we can recover an exact mincost flow on  $G$  in time  $\tilde{O}(m \log U)$ .*

*Proof of Theorem 1.1.* We apply Lemma 8.4 to get an initial point for Algorithm 9. Then, we run Algorithm 9 and round to an exact mincost flow using Lemma 8.4. This succeeds by Theorem 7.9 in time

$$\tilde{O} \left( \left( \mathcal{T}_{\text{init}} + \frac{\sqrt{m}}{k} (\mathcal{T}_{\text{phase}} + \mathcal{T}_{\text{approx}}) \right) \log \left( \frac{\mu_{\text{start}}}{\mu_{\text{end}}} \right) \right) = \tilde{O} \left( \left( \mathcal{T}_{\text{init}} + \frac{\sqrt{m}}{k} (\mathcal{T}_{\text{phase}} + \mathcal{T}_{\text{approx}}) \right) \log U \right). \quad (29)$$

It suffices to plug in the values of  $\mathcal{T}_{\text{init}}, \mathcal{T}_{\text{phase}}$  from Theorem 7.5 and  $\mathcal{T}_{\text{approx}}$  from Theorem 7.7.

We take  $k = m^{1/58}$  so  $\mathcal{T}_{\text{phase}} = \tilde{O}(m^{15/16} k^{29/8} + m) = \tilde{O}(m)$  by Theorem 7.5. Also by Theorem 7.7,  $\mathcal{T}_{\text{approx}} = \tilde{O}(m + zr^2 k^3 / \epsilon^2)$  for  $z = \tilde{O}(k^2), r = \tilde{O}(k^4), \epsilon = \tilde{\Omega}(1)$ , so  $\mathcal{T}_{\text{approx}} = \tilde{O}(m + zr^2 k^3 / \epsilon^2) = \tilde{O}(m + k^{13}) = \tilde{O}(m)$  for  $k = m^{1/58}$ . Thus, the expression in (29) evaluates to  $\tilde{O}(m^{3/2-1/58} \log U)$  as desired.  $\square$

## Acknowledgments

Aaron Sidford is supported in part by a Microsoft Research Faculty Fellowship, NSF CAREER Award CCF-1844855, NSF Grant CCF-1955039, a PayPal research award, and a Sloan Research Fellowship. Yang P. Liu is supported by the Department of Defense (DoD) through the National Defense Science and Engineering Graduate Fellowship, and NSF CAREER Award CCF-1844855 and NSF Grant CCF-1955039. Yin Tat Lee is supported in part by NSF awards CCF-1749609, DMS-1839116, DMS-2023166, CCF-2105772, a Microsoft Research Faculty Fellowship, a Sloan Research Fellowship, and a Packard Fellowship. Jan van den Brand is funded by ONR BRC grant N00014-18-1-2562 and by the Simons Institute for the Theory of Computing through a Simons-Berkeley Postdoctoral Fellowship.

Part of this work was done while Jan van den Brand, Yang P. Liu and Aaron Sidford were visiting the Discrete Optimization Trimester Program at the Hausdorff Research Institute for Mathematics.

## References

- [ACJ<sup>+</sup>21] Hilal Asi, Yair Carmon, Arun Jambulapati, Yujia Jin, and Aaron Sidford. Stochastic bias-reduced gradient methods. *Advances in Neural Information Processing Systems*, 34, 2021.
- [AKPS19] Deeksha Adil, Rasmus Kyng, Richard Peng, and Sushant Sachdeva. Iterative refinement for  $\ell_p$ -norm regression. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1405–1424. SIAM, 2019.
- [AMV20] Kyriakos Axiotis, Aleksander Madry, and Adrian Vladu. Circulation control for faster minimum cost flow in unit-capacity graphs. In *FOCS*, pages 93–104. IEEE, 2020. Available at: <https://arxiv.org/abs/2111.10368v1>.
- [AMV21] Kyriakos Axiotis, Aleksander Madry, and Adrian Vladu. Faster sparse minimum cost flow by electrical flow localization. In *FOCS*. IEEE, 2021.
- [BBG<sup>+</sup>20] Aaron Bernstein, Jan van den Brand, Maximilian Probst Gutenberg, Danupon Nanongkai, Thatchaphol Saranurak, Aaron Sidford, and He Sun. Fully-dynamic graph sparsifiers against an adaptive adversary. *CoRR*, abs/2004.08432, 2020. Available at <https://arxiv.org/abs/2004.08432>.
- [BC16] Aaron Bernstein and Shiri Chechik. Deterministic decremental single source shortest paths: beyond the  $O(mn)$  bound. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pages 389–397, 2016.
- [BC17] Aaron Bernstein and Shiri Chechik. Deterministic partially dynamic single source shortest paths for sparse graphs. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 453–469. SIAM, 2017.
- [BG15] Jose H Blanchet and Peter W Glynn. Unbiased monte carlo for optimization and functions of expectations via multi-level randomization. In *2015 Winter Simulation Conference (WSC)*, pages 3656–3667. IEEE, 2015.
- [BGS21] Aaron Bernstein, Maximilian Probst Gutenberg, and Thatchaphol Saranurak. Deterministic decremental SSSP and approximate min-cost flow in almost-linear time. *arXiv preprint arXiv:2101.07149*, 2021.
- [BLL<sup>+</sup>21] Jan van den Brand, Yin Tat Lee, Yang P. Liu, Thatchaphol Saranurak, Aaron Sidford, Zhao Song, and Di Wang. Minimum cost flows, mdps, and  $\ell_1$ -regression in nearly linear time for dense instances. In *STOC*, pages 859–869. ACM, 2021.

- [BLN<sup>+</sup>20] Jan van den Brand, Yin Tat Lee, Danupon Nanongkai, Richard Peng, Thatchaphol Saranurak, Aaron Sidford, Zhao Song, and Di Wang. Bipartite matching in nearly-linear time on moderately dense graphs. In *61st IEEE Annual Symposium on Foundations of Computer Science, FOCS 2020, Durham, NC, USA, November 16-19, 2020*, pages 919–930, 2020.
- [BLSS20] Jan van den Brand, Yin Tat Lee, Aaron Sidford, and Zhao Song. Solving tall dense linear programs in nearly linear time. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing, STOC 2020, Chicago, IL, USA, June 22-26, 2020*, pages 775–788. ACM, 2020. Available at <https://arxiv.org/abs/2002.02304>.
- [Bra20] Jan van den Brand. A deterministic linear program solver in current matrix multiplication time. In *SODA*, pages 259–278. SIAM, 2020.
- [Bra21] Jan van den Brand. Unifying matrix data structures: Simplifying and speeding up iterative algorithms. In *SOSA*, pages 1–13. SIAM, 2021.
- [CGL<sup>+</sup>20] Julia Chuzhoy, Yu Gao, Jason Li, Danupon Nanongkai, Richard Peng, and Thatchaphol Saranurak. A deterministic algorithm for balanced cut with applications to dynamic connectivity, flows, and beyond. In *61st IEEE Annual Symposium on Foundations of Computer Science, FOCS 2020, Durham, NC, USA, November 16-19, 2020*, pages 1158–1167. IEEE, 2020. Available at <https://arxiv.org/abs/1910.08025>.
- [Chu21] Julia Chuzhoy. Decremental all-pairs shortest paths in deterministic near-linear time. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 626–639, 2021.
- [CK19] Julia Chuzhoy and Sanjeev Khanna. A new algorithm for decremental single-source shortest paths with applications to vertex-capacitated flow and cut problems. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 389–400, 2019.
- [CKM<sup>+</sup>11] Paul Christiano, Jonathan A. Kelner, Aleksander Mądry, Daniel A. Spielman, and Shang-Hua Teng. Electrical flows, Laplacian systems, and faster approximation of maximum flow in undirected graphs. In *Proceedings of the 43rd ACM Symposium on Theory of Computing, STOC 2011, San Jose, CA, USA, June 6-8 2011*, pages 273–282. ACM, 2011. Available at <https://arxiv.org/abs/1010.2921>.
- [CKM<sup>+</sup>14] 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 *Proceedings of the 46th Annual ACM Symposium on Theory of Computing, STOC 2014, New York, NY, USA, June 1-3, 2014*, pages 343–352, 2014.
- [CLS19] Michael B. Cohen, Yin Tat Lee, and Zhao Song. Solving linear programs in the current matrix multiplication time. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing, STOC 2019, Phoenix, AZ, USA, June 23-26, 2019*, pages 938–942. ACM, 2019. Available at <https://arxiv.org/abs/1810.07896>.
- [CMSV17] Michael B. Cohen, Aleksander Mądry, Piotr Sankowski, and Adrian Vladu. Negative-weight shortest paths and unit capacity minimum cost flow in  $\tilde{O}(m^{10/7} \log W)$  time (extended abstract). In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2017, Barcelona, Spain, Hotel Porta Fira, January 16-19*, pages 752–771. SIAM, 2017. Available at <https://arxiv.org/abs/1605.01717>.
- [DGG<sup>+</sup>22] Sally Dong, Yu Gao, Gramoz Goranci, Yin Tat Lee, Richard Peng, Sushant Sachdeva, and Guanghao Ye. Nested dissection meets ipms : Planar min-cost flow in nearly-linear time. In *SODA*. SIAM, 2022.



- [DGGP19] David Durfee, Yu Gao, Gramoz Goranci, and Richard Peng. Fully dynamic spectral vertex sparsifiers and applications. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing, STOC 2019, Phoenix, AZ, USA, June 23-26, 2019*, pages 914–925. ACM, 2019. Available at <https://arxiv.org/abs/1906.10530>.
- [DKP<sup>+</sup>17] David Durfee, Rasmus Kyng, John Peebles, Anup B Rao, and Sushant Sachdeva. Sampling random spanning trees faster than matrix multiplication. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 730–742, 2017.
- [DLY21] Sally Dong, Yin Tat Lee, and Guanghai Ye. A nearly-linear time algorithm for linear programs with small treewidth: a multiscale representation of robust central path. In *STOC*, pages 1784–1797. ACM, 2021.
- [DR<sup>+</sup>14] Cynthia Dwork, Aaron Roth, et al. The algorithmic foundations of differential privacy. *Found. Trends Theor. Comput. Sci.*, 9(3-4):211–407, 2014.
- [DS08] Samuel I. Daitch and Daniel A. Spielman. Faster approximate lossy generalized flow via interior point algorithms. In *Proceedings of the 40th annual ACM Symposium on Theory of Computing, STOC 2008, Victoria, BC, Canada, May 17-20, 2008*, pages 451–460, New York, NY, USA, 2008. ACM. Available at <http://arxiv.org/abs/0803.0988>.
- [ET75] Shimon Even and R Endre Tarjan. Network flow and testing graph connectivity. *SIAM Journal on Computing*, 4(4):507–518, 1975.
- [Gil15] Michael B Giles. Multilevel monte carlo methods. *Acta Numerica*, 24:259–328, 2015.
- [GLP21] Yu Gao, Yang P Liu, and Richard Peng. Fully dynamic electrical flows: sparse maxflow faster than goldberg-rao. *FOCS 2021*, 2021. Available at <https://arxiv.org/abs/2101.07233>.
- [GR98] Andrew V. Goldberg and Satish Rao. Beyond the flow decomposition barrier. *Journal of the ACM*, 45(5):783–797, 1998. Announced at FOCS’97.
- [GT90] Andrew V. Goldberg and Robert E. Tarjan. Finding minimum-cost circulations by successive approximation. *Math. Oper. Res.*, 15(3):430–466, 1990.
- [GWN20] Maximilian Probst Gutenberg and Christian Wulff-Nilsen. Deterministic algorithms for decremental approximate shortest paths: Faster and simpler. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2522–2541. SIAM, 2020.
- [JS21] Arun Jambulapati and Aaron Sidford. Ultrasparse ultrasparsifiers and faster laplacian system solvers. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 540–559. SIAM, 2021.
- [JSWZ20] Shunhua Jiang, Zhao Song, Omri Weinstein, and Hengjie Zhang. Faster dynamic matrix inverse for faster lps. *CoRR*, abs/2004.07470, 2020.
- [Kar73] Alexander V Karzanov. On finding maximum flows in networks with special structure and some applications. *Matematicheskie Voprosy Upravleniya Proizvodstvom*, 5:81–94, 1973.
- [Kar84] N Karmarkar. A new polynomial-time algorithm for linear programming. *Combinatorica*, 4(4):373–395, 1984.
- [KLOS14] Jonathan A. Kelner, Yin Tat Lee, Lorenzo Orecchia, and Aaron Sidford. An almost-linear-time algorithm for approximate max flow in undirected graphs, and its multicommodity generalizations. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2014, Portland, OR, USA, January 5-7, 2014*, pages 217–226, 2014. Available at <https://arxiv.org/abs/1304.2338>.



- [KLP<sup>+</sup>16] 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 48th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2016, Cambridge, MA, USA, June 18-21, 2016*, pages 842–850, 2016. Available at <https://arxiv.org/abs/1512.01892>.
- [KLS20] Tarun Kathuria, Yang P. Liu, and Aaron Sidford. Unit capacity maxflow in almost  $O(m^{4/3})$  time. In *61st IEEE Annual Symposium on Foundations of Computer Science, FOCS 2020, Durham, NC, USA, November 16-19, 2020*, pages 119–130. IEEE, 2020.
- [KM09] Jonathan A Kelner and Aleksander Madry. Faster generation of random spanning trees. In *2009 50th Annual IEEE Symposium on Foundations of Computer Science*, pages 13–21. IEEE, 2009.
- [KMP10] Ioannis Koutis, Gary L. Miller, and Richard Peng. Approaching optimality for solving SDD linear systems. In *51th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2010, Las Vegas, NV, USA, October 23-26, 2010*, pages 235–244, 2010. Available at <https://arxiv.org/abs/1003.2958>.
- [KMP11] Ioannis Koutis, Gary L. Miller, and Richard Peng. A nearly- $m \log n$  time solver for SDD linear systems. In *52nd IEEE Annual Symposium on Foundations of Computer Science, FOCS 2011, Palm Springs, CA, USA, October 22-25, 2011*, pages 590–598, 2011. Available at <https://arxiv.org/abs/1102.4842>.
- [KNPW11] Daniel M. Kane, Jelani Nelson, Ely Porat, and David P. Woodruff. Fast moment estimation in data streams in optimal space. In *Proceedings of the 43rd ACM Symposium on Theory of Computing, STOC 2011, San Jose, CA, USA, June 6-8 2011*, pages 745–754. ACM, 2011. Available at <https://arxiv.org/abs/1007.4191>.
- [KOSA13] 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 45th Annual ACM Symposium on Theory of Computing, STOC 2013, Palo Alto, CA, USA, June 1-4, 2013*, pages 911–920, 2013. Available at <https://arxiv.org/abs/1301.6628>.
- [KPSW19] Rasmus Kyng, Richard Peng, Sushant Sachdeva, and Di Wang. Flows in almost linear time via adaptive preconditioning. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing, STOC 2019, Phoenix, AZ, USA, June 23-26, 2019*, pages 902–913. ACM, 2019. Available at <https://arxiv.org/abs/1906.10340>.
- [KS16] Rasmus Kyng and Sushant Sachdeva. Approximate gaussian elimination for Laplacians - fast, sparse, and simple. In *57th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2016, Hyatt Regency, New Brunswick, NJ, USA, October 9-11, 2016*, pages 573–582, 2016. Available at <https://arxiv.org/abs/1605.02353>.
- [LRS13] Yin Tat Lee, Satish Rao, and Nikhil Srivastava. A new approach to computing maximum flows using electrical flows. In *Proceedings of the 45th Annual ACM Symposium on Theory of Computing, STOC 2013, Palo Alto, CA, USA, June 1-4, 2013*, pages 755–764. ACM, 2013.
- [LS13] Yin Tat Lee and Aaron Sidford. Efficient accelerated coordinate descent methods and faster algorithms for solving linear systems. In *2013 IEEE 54th annual symposium on foundations of computer science*, pages 147–156. IEEE, 2013.
- [LS15] Yin Tat Lee and Aaron Sidford. Efficient inverse maintenance and faster algorithms for linear programming. In *56th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2015, Berkeley, CA, USA, October 17-20, 2015*, pages 230–249. IEEE Computer Society, 2015. Available at <https://arxiv.org/abs/1503.01752>.
- [LS18] Huan Li and Aaron Schild. Spectral subspace sparsification. In *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 385–396. IEEE, 2018.

- [LS19] Yin Tat Lee and Aaron Sidford. Solving linear programs with  $\text{sqrt}(\text{rank})$  linear system solves. *CoRR*, abs/1910.08033, 2019.
- [LS20] Yang P. Liu and Aaron Sidford. Faster energy maximization for faster maximum flow. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing, STOC 2020, Chicago, IL, USA, June 22-26, 2020*, pages 803–814. ACM, 2020. Available at <https://arxiv.org/abs/1910.14276>.
- [LSZ19] Yin Tat Lee, Zhao Song, and Qiuyi Zhang. Solving empirical risk minimization in the current matrix multiplication time. In *Conference on Learning Theory, COLT 2019, Phoenix, AZ, USA, June 25-28, 2019*, volume 99 of *Proceedings of Machine Learning Research*, pages 2140–2157. PMLR, 2019. Available at <https://arxiv.org/abs/1905.04447>.
- [Mađ13] Aleksander Mađry. Navigating central path with electrical flows: From flows to matchings, and back. In *54th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2013, Berkeley, CA, USA, October 26-29, 2013*, pages 253–262. IEEE Computer Society, 2013. Available at <https://arxiv.org/abs/1307.2205>.
- [Mađ16] Aleksander Mađry. Computing maximum flow with augmenting electrical flows. In *57th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2016, 9-11 October 2016, Hyatt Regency, New Brunswick, New Jersey, USA*, pages 593–602. IEEE Computer Society, 2016. Available at <https://arxiv.org/abs/1608.06016>.
- [MS78] Robert H. Morris Sr. Counting large numbers of events in small registers. *Commun. ACM*, 21(10):840–842, 1978.
- [MST14] Aleksander Mađry, Damian Straszak, and Jakub Tarnawski. Fast generation of random spanning trees and the effective resistance metric. In *Proceedings of the twenty-sixth annual ACM-SIAM symposium on Discrete algorithms*, pages 2019–2036. SIAM, 2014.
- [NS17] Danupon Nanongkai and Thatchaphol Saranurak. Dynamic spanning forest with worst-case update time: adaptive, Las Vegas, and  $O(n^{1/2-\epsilon})$ -time. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2017, Montreal, QC, Canada, June 19-23, 2017*, pages 1122–1129, 2017. Available at <https://arxiv.org/abs/1611.03745>.
- [NSW17] Danupon Nanongkai, Thatchaphol Saranurak, and Christian Wulff-Nilsen. Dynamic minimum spanning forest with subpolynomial worst-case update time. In *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017, Berkeley, CA, USA, October 15-17, 2017*, pages 950–961. IEEE Computer Society, 2017. Available at <https://arxiv.org/abs/1708.03962>.
- [NY20] Jelani Nelson and Huacheng Yu. Optimal bounds for approximate counting. *arXiv preprint arXiv:2010.02116*, 2020.
- [Pen16] Richard Peng. Approximate undirected maximum flows in  $O(m \text{polylog}(n))$  time. In *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2016, Arlington, VA, USA, January 10-12, 2016*, pages 1862–1867. SIAM, 2016.
- [Ren88] James Renegar. A polynomial-time algorithm, based on newton’s method, for linear programming. *Mathematical Programming*, 40(1-3):59–93, 1988.
- [Sch18] Aaron Schild. An almost-linear time algorithm for uniform random spanning tree generation. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, Los Angeles, CA, USA, June 25-29, 2018*, pages 214–227. ACM, 2018. Available at <https://arxiv.org/abs/1711.06455>.
- [She13] Jonah Sherman. Nearly maximum flows in nearly linear time. In *54th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2013, Berkeley, CA, USA, October 26-29, 2013*, pages 263–269, 2013. Available at <https://arxiv.org/abs/1304.2077>.

- [She17] Jonah Sherman. Area-convexity,  $\ell_\infty$  regularization, and undirected multicommodity flow. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 452–460, 2017.
- [ST04] Daniel A. Spielman and Shang-Hua Teng. Nearly-linear time algorithms for graph partitioning, graph sparsification, and solving linear systems. In *Proceedings of the 36th Annual ACM Symposium on Theory of Computing, STOC 2004, Chicago, IL, USA, June 13-16, 2004*, pages 81–90, 2004. Available at <https://arxiv.org/abs/0809.3232>, <https://arxiv.org/abs/0808.4134>, <https://arxiv.org/abs/cs/0607105>.
- [ST18] Aaron Sidford and Kevin Tian. Coordinate methods for accelerating  $\ell_\infty$  regression and faster approximate maximum flow. In *59th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2018, Paris, France, October 7-9, 2018*, pages 922–933, 2018. Available at <https://arxiv.org/abs/1808.01278>.
- [SW19] Thatchaphol Saranurak and Di Wang. Expander decomposition and pruning: Faster, stronger, and simpler. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2019, San Diego, California, USA, January 6-9, 2019*, pages 2616–2635. SIAM, 2019. Available at <https://arxiv.org/abs/1812.08958>.
- [Vai89] Pravin M. Vaidya. Speeding-up linear programming using fast matrix multiplication (extended abstract). In *30th IEEE Annual Symposium on Foundations of Computer Science, FOCS 1989, Research Triangle Park, NC, USA, October 30 - November 1, 1989*, pages 332–337. IEEE Computer Society, 1989.
- [Wul17] Christian Wulff-Nilsen. Fully-dynamic minimum spanning forest with improved worst-case update time. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2017, Montreal, QC, Canada, June 19-23, 2017*, pages 1130–1143, 2017. Available at <https://arxiv.org/abs/1611.02864>.