Linear-Size Singular-Value Sparsifiers for Weighted Digraphs

Research manuscript | November 2026

Abstract

We prove that every nonnegatively weighted digraph on n vertices has an ε-singular-value sparsifier with $O(n/\varepsilon^2)$ arcs. The sparsifier reweights original arcs and preserves all weighted in-degrees and out-degrees exactly. Its size is independent of the range of positive input weights. The construction builds a hierarchy of bipartite quotients. At each level, we fix a constant number of the heaviest incident arcs per vertex. Two auxiliary matrices use the fixed arcs to group the residual vertices. The quotient contracts by an arbitrarily small constant factor with at most constant distortion of the Schur-complement energies. One discrepancy walk on the original arc variables approximates the level Laplacians and enforces the exact constraints. The algorithm is deterministic in the real-arithmetic model.

Keywords. Directed graphs, singular-value approximation, sparsification, graph coarsening.

1 Introduction

Spectral sparsification asks how few edges of a graph suffice to approximate its Laplacian quadratic form simultaneously on all test vectors. Effective-resistance sampling gives $O(n \log n/\varepsilon^2)$ edges for undirected graphs [8], and the barrier method of Batson, Spielman, and Srivastava gives $O(n/\varepsilon^2)$ edges [4]. Singular-value (SV) approximation, introduced by Ahmadinejad, Peebles, Pyne, Sidford, and Vadhan [1], controls a bilinear error for directed graphs using the contraction defects of the normalized adjacency. This stronger guarantee is stable under products in the setting of [1]. It forces exact preservation of directions of zero defect energy and, in particular, of the in-degrees and out-degrees.

Ahmadinejad et al. asked whether linear-size SV sparsifiers exist [1, Section 1.5]. Lau, Wang, and Zhou gave a deterministic construction with $O(n \log^5 n/\varepsilon^2)$ arcs and obtained linear size when the bipartite lift has a constant normalized Laplacian gap [7]. Anand obtained $O(n \log(2n)/\varepsilon^2)$ arcs for unweighted inputs by combining resistance pieces with projected matrix discrepancy [3]. His construction for unweighted inputs takes nearly linear time in the real-arithmetic model. For weighted graphs, his support bound is $O(n \log(2n)/\varepsilon^2 + n \log(2U))$, where U is the ratio of the largest to the smallest positive input weight. We obtain linear size for arbitrary nonnegative weights, with a support bound independent of U.

Theorem 1 (Main theorem)**.** *Let $A \in \mathbb{R}^{n\times n}_{\geq 0}$ be the weighted adjacency matrix of a digraph and let $0 < \varepsilon < 1$. There is a nonnegative ε-SV approximation $\widetilde{A}$ in the sense of [1, Definition 3.5] such that*

$$\operatorname{supp} \widetilde{A} \subseteq \operatorname{supp} A, \qquad \widetilde{A}\mathbf{1} = A\mathbf{1}, \qquad \widetilde{A}^\top \mathbf{1} = A^\top \mathbf{1}, \qquad |\operatorname{supp} \widetilde{A}| = O(n/\varepsilon^2).$$

A deterministic construction uses polynomially many real-arithmetic operations. If the input weights are rational, the output weights can be chosen with bit length polynomial in the input encoding length and $\log(1/\varepsilon)$.

The real-arithmetic statement uses matrix operations, comparisons, and scalar square roots. The bit bound concerns the output weights. A bit-complexity bound for the entire algorithm remains open here.

1.1 Technical overview

The bipartite lift represents arcs by undirected edges between left and right copies of the vertices. The two SV defects are Schur complements of its Laplacian. Testing the ordinary spectral error on a harmonic extension controls one of the two test vectors in its Schur-complement norm. The other vector is controlled in a degree-weighted Euclidean norm. Thus ordinary spectral approximation suffices for terms in which at least one test vector varies within groups whose variance is bounded by the Schur energy.

Grouping with the help of fixed arcs. We fix the b heaviest positive entries in each row and column of a level's adjacency, where b is a large absolute constant. The residual degree in a row specifies how much within-group variance that row needs to pay. A residual row whose weights are already small relative to its degree can be used as is. If one residual arc is too heavy, the row has b fixed arcs at least as heavy. An auxiliary matrix spreads the residual degree evenly over those b fixed arcs. This leaves the row degree unchanged and keeps the auxiliary matrix entrywise below the full reference graph. We apply this construction separately on the left and right. Each auxiliary matrix has a small Frobenius norm after degree normalization, which gives a small trace budget for the filtered star covariances. A trace-sensitive PSD sparsifier selects few stars, and a resolvent argument assigns vertices to them. The resulting groups control the residual-degree variance in the full reference graph's defect energy.

Preserving degrees through a hierarchy. We contract only the residual graph. A quotient arc represents a set of original arcs, so fixing its weight costs one linear equation on the original weights. Exact original margins and the fixed weights imply exact residual and quotient margins. Every level fixes at most b arcs per vertex. The vertices decrease geometrically, so the total number of equations is $O(n)$. This count depends on the number of incidences selected at a level and is independent of their weights.

Paying for error at each scale. The SV error at one level is bounded by a constant times the sum of the ordinary relative Laplacian error at that level and the SV error of its quotient. We allow this constant to accumulate and ask for a geometrically smaller ordinary error at deeper levels. The grouping step makes the number of vertices shrink by a factor ρ that can be made small enough to offset the square of that error amplification. A shared discrepancy walk bounds each block's error in terms of its own dimension, giving support $O(n/\varepsilon^2)$.

2 Schur energies and one level of coarsening

Let B be a nonnegative rectangular matrix with row and column degree matrices D_L, D_R. Its undirected bipartite lift has Laplacian

$$L(B) = \begin{pmatrix} D_L & -B \\ -B^\top & D_R \end{pmatrix}.$$

For a PSD matrix S write $S[x] = x^\top S x$, let $\|x\|_S = S[x]^{1/2}$, and denote the Moore–Penrose pseudoinverse by $S^\dagger$. Unqualified matrix norms are operator norms. The left and right defects are

$$R_L(B) = D_L - B D_R^\dagger B^\top, \qquad R_R(B) = D_R - B^\top D_L^\dagger B.$$

In particular, if the sides have positive degrees and $P = D_L^{-1/2} B D_R^{-1/2}$, the normalized left defect is $I - PP^\top$. The energy identities

$$R_L(B)[x] = \min_y \sum_{u,v} B_{uv}(x_u - y_v)^2, \qquad R_R(B)[y] = \min_x \sum_{u,v} B_{uv}(x_u - y_v)^2 \tag{1}$$

follow by completing the square. These identities give $0 \preceq R_L(B) \preceq D_L$, and analogously on the right. The defects are monotone in the entrywise order on nonnegative B. The kernel on either side consists of potentials constant on that side of each connected component, including the singleton components from zero-degree coordinates.

We will use real vectors throughout. For an error matrix E, the real bilinear bound

$$|x^\top E y| \le t \, \|x\|_{R_L(B)} \|y\|_{R_R(B)} \tag{2}$$

implies the corresponding complex bound. Indeed, the real bound gives both kernel conditions and the operator bound on $R_L(B)^{\dagger/2} E R_R(B)^{\dagger/2}$, whose operator norm is the same over the real and complex fields. By [1, Definitions 3.1 and 3.5], $t \le \varepsilon/2$ in (2) suffices for ε-SV approximation. We refer to t as the error in product form.

A partition of each side gives a quotient $B_c = S_L^\top B S_R$, where S_L, S_R are the vertex-to-group indicator matrices. For a left potential x, write $\bar{x}$ for the vector of group means with weights given by the degrees of B. Write $\Pi_L x = S_L \bar{x}$, and use $\bar{y}, \Pi_R y$ on the right. Zero-degree vertices can be dropped throughout a grouping step.

Lemma 2 (One-level error bound)**.** *Let A be the full reference matrix and let $0 \le B \le A$ be a residual matrix, with the inequality understood entrywise. Form the group averages using the residual degrees $D_L(B), D_R(B)$. Suppose*

$$\|x - \Pi_L x\|^2_{D_L(B)} \le C_0 R_L(A)[x], \qquad \|y - \Pi_R y\|^2_{D_R(B)} \le C_0 R_R(A)[y]. \tag{3}$$

Let $\widetilde{B}$ be a reweighting on the support of B with the same margins. If its lift has ordinary relative Laplacian error at most δ and its quotient has error t in product form relative to B_c, then $(A-B)+\widetilde{B}$ has error at most $C(\delta + t)$ in product form relative to A, where $C = (1 + 2\sqrt{C_0})^2$.

Proof. Write $E = \widetilde{B} - B$ and $\Delta L = L(\widetilde{B}) - L(B)$. The diagonal blocks of ΔL vanish. The relative spectral bound gives

$$|z^\top \Delta L w| \le \delta \|z\|_{L(B)} \|w\|_{L(B)}.$$

Use $z = (x, D_R^\dagger B^\top x)$, $w = (0, y)$ to bound $|x^\top E y|$ by $\delta \|x\|_{R_L(B)} \|y\|_{D_R}$. Harmonic extension on the other side gives the bound $\delta \|x\|_{D_L} \|y\|_{R_R(B)}$.

Put $x_0 = \Pi_L x$ and $y_0 = \Pi_R y$. Split

$$x^\top E y - x_0^\top E y_0 = (x - x_0)^\top E y + x_0^\top E (y - y_0).$$

The two one-sided bounds, monotonicity $R_L(B) \preceq R_L(A)$, $R_R(B) \preceq R_R(A)$, and (3) bound the right side in absolute value by

$$(2\sqrt{C_0} + C_0)\delta \, \|x\|_{R_L(A)} \|y\|_{R_R(A)}.$$

Here we used $\|x_0\|_{R_L(B)} \le \|x\|_{R_L(B)} + \|x - x_0\|_{D_L(B)} \le (1 + \sqrt{C_0})\|x\|_{R_L(A)}$.

For the remaining term, set $h = D_R(A)^\dagger A^\top x$. The pair (x, h) is minimizing for the full reference graph. In particular, $R_R(A)[h] \le R_L(A)[x]$ and $\sum_{u,v} B_{uv}(x_u - h_v)^2 \le R_L(A)[x]$. Test the quotient's left defect at $\bar{x}$ with the right potential $\bar{h}$. Applying the triangle inequality in the residual edge-weighted norm gives

$$\begin{aligned} \|\bar{x}\|_{R_L(B_c)} &\le \left(\sum_{u,v} B_{uv} \big((\Pi_L x)_u - (\Pi_R h)_v \big)^2 \right)^{1/2} \\ &\le \|x\|_{R_L(A)} + \|x - \Pi_L x\|_{D_L(B)} + \|h - \Pi_R h\|_{D_R(B)} \\ &\le (1 + 2\sqrt{C_0}) \|x\|_{R_L(A)}. \end{aligned}$$

The right defect satisfies the same comparison. Since $x_0^\top E y_0 = \bar{x}^\top (S_L^\top E S_R) \bar{y}$, the quotient bound completes the proof. □

3 Grouping from one-sided small relative weights

Lemma 3 (One-sided grouping)**.** *Let $b \ge 2$ be an integer. Suppose a nonempty weighted bipartite graph B has positive degrees on both sides and*

$$B_{uv} \le d_L(u)/b \qquad \text{for every } u, v. \tag{4}$$

There is a partition of the left side into $O((N_L + N_R)/\sqrt{b})$ groups such that $\|x - \Pi_L x\|_{D_L}^2 \le C_0 R_L(B)[x]$ for an absolute constant C_0. The partition can be found in polynomial real arithmetic.

We use an additive matrix sparsification lemma and a rounding lemma. The former is the trace-sensitive version of the BSS barrier argument [4, 5, 6]. Its proof for arbitrary PSD summands is given in Appendix A.

Lemma 4 (Additive PSD sparsification)**.** *Let $C_i \succeq 0$ with $B_0 = \sum_i C_i \preceq I$ and let $0 < t \le 1$. Nonnegative multipliers s_i with $O(1 + \operatorname{tr}(B_0)/t^2)$ nonzero entries suffice for $\|\sum_i s_i C_i - B_0\| \le t$. They can be found in polynomial real arithmetic.*

Lemma 5 (Choosing one representative)**.** *For each index u, let v_u be a random vector with finite support and covariance $\Gamma_u = \mathbb{E} v_u v_u^\top$. Assume $\sum_u \Gamma_u = T_0 \preceq cI$ for a fixed $c > 0$. One can omit $O(\operatorname{tr} T_0/c)$ indices and choose a vector from each remaining support so that the sum of chosen outer products is at most $8cI$. The choices can be made in polynomial real arithmetic.*

Proof. Let S be the sum of selected outer products and T the sum of the unprocessed covariances. With $U = 8c$, maintain $M = UI - S - 4T \succ 0$ and keep

$$\Phi = \operatorname{tr}(M^{-1} - I/U)$$

at most its initial value. Initially $S = 0, T = T_0$, and $\Phi \le \operatorname{tr}(T_0)/(8c^2)$ by applying $4\lambda/[U(U-4\lambda)] \le \lambda/(8c^2)$ to the eigenvalues of T_0.

If an unprocessed $\Gamma = \Gamma_u \ne 0$ has $\operatorname{tr}(M^{-1}\Gamma) \le 1/2$, put $M' = M + 4\Gamma$ and $\Delta = \operatorname{tr}(M^{-1}) - \operatorname{tr}((M')^{-1})$. Convexity of the trace of the inverse along this segment gives $\Delta \ge 4 \operatorname{tr}((M')^{-2}\Gamma) > 0$. By averaging, some candidate v satisfies

$$v^\top (M')^{-1} v + \frac{v^\top (M')^{-2} v}{\Delta} \le \tfrac{1}{2} + \tfrac{1}{4}.$$

The rank-one inverse formula now shows that $M' - vv^\top \succ 0$ and that adding $vv^\top$ to S recovers at most the decrease Δ of the potential. If $\Gamma = 0$, a zero vector can be chosen directly. When no eligible index remains, their number is at most $2 \operatorname{tr}(M^{-1}T)$. Because $M \preceq UI$, $T \preceq T_0 \preceq cI$, and Φ never increased, we have

$$\operatorname{tr}(M^{-1}T) \le \operatorname{tr}(T_0)/U + c\Phi = O(\operatorname{tr}(T_0)/c).$$

Omit these indices. The invariant gives $S \preceq UI$, and all choices were among the listed candidates. □

Proof of Lemma 3. Write

$$P = D_L^{-1/2} B D_R^{-1/2}, \qquad W = PP^\top, \qquad F = I - W.$$

For each right vertex v, let $\mu_v = B_{\cdot v}/d_R(v)$. The corresponding star covariance on the left is

$$K_v = \sum_u B_{uv}(e_u - \mu_v)(e_u - \mu_v)^\top, \qquad \sum_v K_v = R_L(B).$$

Define $a_{uv} = W F^{\dagger/2} D_L^{-1/2}(e_u - \mu_v)$ and $T_v = \sum_u B_{uv} a_{uv} a_{uv}^\top$, using only u, v with $B_{uv} > 0$. Since F and W commute,

$$\sum_v T_v = W \Pi_{\operatorname{im} F} W \preceq I, \qquad \sum_v \operatorname{tr}(T_v) \le \operatorname{tr}(W) = \sum_{u,v} \frac{B_{uv}^2}{d_L(u) d_R(v)} \le \frac{N_R}{b}. \tag{5}$$

The trace bound uses (4) and sums the fractions $B_{uv}/d_R(v)$ down each column.

For coverage, put $c_v(u) = B_{uv}/d_L(u)$ and $H_v = c_v c_v^\top/(\mathbf{1}^\top c_v)$. Weighted Cauchy–Schwarz gives $H_v \preceq \operatorname{diag}(c_v)$, so

$$\sum_v H_v \preceq I, \qquad \sum_v H_v \mathbf{1} = \mathbf{1}, \qquad \operatorname{tr} H_v \le \|c_v\|_\infty \le 1/b. \tag{6}$$

Let $q = \max\{1, \sum_v \operatorname{tr} T_v\}$. Each nonempty row has at least b neighbors, so $N_R/b \ge 1$. Apply Lemma 4 at tolerance $t = b^{-1/4}$ to the PSD atoms

$$\operatorname{diag}\big(T_v, H_v, \operatorname{tr}(T_v)/q\big).$$

Their sum is bounded by I and has trace at most $2N_R/b + 1$. This yields star multipliers s_v with $O(N_R/\sqrt{b})$ nonzeros, $\sum_v s_v T_v \preceq 2I$, and $\sum_v s_v \operatorname{tr} T_v = O(N_R/b)$. The scalar block in the atoms gives the last bound. The reweighted coverage $c(u) = \sum_v s_v c_v(u)$ obeys $\|c - \mathbf{1}\|_2^2 \le t^2 N_L$. Put each u with $c(u) < 1/2$ in its own group. This costs at most $4N_L/\sqrt{b}$ singletons.

For each other u, consider choosing a center v with probability $s_v c_v(u)/c(u)$, and associate the candidate vector $\sqrt{d_L(u)}a_{uv}$ to that choice. The sum of the candidate covariances is bounded by

$$\sum_{u\,:\,c(u)\ge 1/2} \sum_v \frac{s_v c_v(u)}{c(u)} d_L(u) a_{uv} a_{uv}^\top \preceq 2 \sum_v s_v T_v.$$

It is bounded by $4I$ and has trace $O(N_R/b)$. Lemma 5 assigns centers to the rows with a constant bound on the sum of the chosen outer products, except for $O(N_R/b)$ more rows, which become singletons. The other groups are the rows assigned to the same center.

To verify their variance bound, set $x' = D_L^{-1/2} W D_L^{1/2} x$. For $B_{uv} > 0$ the vector $D_L^{-1/2}(e_u - \mu_v)$ is orthogonal to $\ker F$, since $e_u - \mu_v$ is supported on one component and has sum zero. It follows that

$$a_{uv}^\top F^{1/2} D_L^{1/2} x = x'_u - \mu_v^\top x'.$$

Using $\mu_v^\top x'$ as a candidate mean for the group assigned to v, and the chosen outer-product bound, gives

$$\|x' - \Pi_L x'\|_{D_L}^2 = O\big(\|F^{1/2} D_L^{1/2} x\|_2^2\big) = O(R_L(B)[x]).$$

Also $\|x - x'\|_{D_L}^2 = \|F D_L^{1/2} x\|_2^2 \le R_L(B)[x]$, since $0 \preceq F \preceq I$. The triangle inequality and the contraction of $I - \Pi_L$ in the D_L-norm prove (3) with an absolute constant.

The number of groups is the sum of the selected centers that received assignments and the singleton rows, bounded above by $O((N_L + N_R)/\sqrt{b})$. □

4 The hierarchy and its linear constraint count

Lemma 6 (Grouping with fixed arcs)**.** *Let A be a nonnegative bipartite adjacency matrix on a total of N vertices, and let $b \ge 2$. Fix the b heaviest positive entries of each row and each column, or all their positive entries when there are fewer than b. Break ties arbitrarily. Let B be the matrix left after removing the union of the fixed arcs. There are partitions of the active vertices of B into $O(N/\sqrt{b})$ groups satisfying (3). At most bN weights were fixed.*

Proof. For a left vertex u, put $r_u = (B\mathbf{1})_u$. Build an auxiliary adjacency $B^{(L)}$ as follows.

1. If every $B_{uv} \le r_u/b$, set the auxiliary row equal to $B_{u\cdot}$.
2. Otherwise use the b heaviest entries of row u in A, giving each auxiliary weight r_u/b and assigning zero elsewhere.

In the second case, some residual arc exceeds r_u/b. Every selected heaviest entry is at least as large as that residual arc, so $B^{(L)} \le A$ entrywise. Both cases preserve the residual left degree r_u, and every auxiliary entry is at most r_u/b. Delete zero-degree coordinates before applying Lemma 3. The active rows of $B^{(L)}$ are exactly those of B. Its active columns can include vertices that have become isolated in B, but the total size is at most N. Thus the lemma yields $O(N/\sqrt{b})$ left groups. Since $D_L(B^{(L)}) = D_L(B)$ and $R_L(B^{(L)}) \preceq R_L(A)$, these groups satisfy the first inequality of (3). Apply the same procedure to $A^\top, B^\top$ to get the right groups. The count of fixed arcs follows by summing at most b selections per vertex. □

Start with $A_0 = A$, interpreted as a bipartite graph. Fix an absolute integer b large enough that the factor $\rho = O(b^{-1/2})$ from Lemma 6 satisfies $\rho C^2 < 1$, where C is the absolute constant from Lemma 2. This choice is possible because the variance constant C_0 is independent of b. At level j, fix the arcs specified by Lemma 6 and call the residual adjacency B_j. Record an equality fixing each selected weight. If B_j is empty, stop. Otherwise, take A_{j+1} to be the quotient of B_j by the two partitions. With N_j denoting the number of active vertices of the lift of A_j, we have $N_{j+1} \le \rho N_j$ and $N_j \le 2n\rho^j$. Hence the hierarchy terminates after $O(\log(2n))$ levels.

Lemma 7 (Number of equalities)**.** *The total number of recorded equalities, together with the original degree constraints, is $O(n)$.*

Proof. At most bN_j weights are fixed at level j, so the total number of weight equalities is at most $b\sum_j N_j = O(n)$. There are at most $2n$ original degree equalities. □

To interpret the equalities, associate each reference arc with the set of original arcs that contribute to it. These sets are disjoint within a level. An original arc follows its two endpoint groups through each contraction until its reference arc is fixed. Fixing a reference weight means fixing the sum of the weights of the contributing original arcs.

Let a nonnegative weighting of the original arcs satisfy the recorded equalities and the original degree equations. Denote the corresponding level adjacencies by $\widetilde{A}_j, \widetilde{B}_j$, obtained by aggregation along the fixed reference hierarchy. Inductively, $\widetilde{A}_j$ has the margins of A_j. Subtracting the fixed weights shows that $\widetilde{B}_j$ has the margins of B_j. Summing these residual degrees within each group gives the margins at the next level.

If the residual lift at level j has ordinary relative Laplacian error at most δ_j, then Lemma 2 gives error at most $C(\delta_j + t_{j+1})$ in product form relative to A_j, where t_{j+1} bounds the quotient error. At the final level the residual graph is empty and every weight is fixed. Therefore the original error in product form is at most

$$\sum_{j\ge 0} C^{j+1}\delta_j. \tag{7}$$

5 Simultaneous sparsification of the level Laplacians

For each level j with nonempty residual graph, write $L_j = L(B_j)$. For an original arc e of weight w_e whose image in level j is a residual edge (u, v), define

$$X_{ej} = w_e L_j^{\dagger/2}(e_u - e_v)(e_u - e_v)^\top L_j^{\dagger/2}$$

using incidence coordinates on the lift of B_j. Set $X_{ej} = 0$ when the arc has been fixed at or before level j. Then $\sum_e X_{ej} = \Pi_{\operatorname{im} L_j} \preceq I$. Let s_e be the multiplier of original arc e, so its new weight is $s_e w_e$. We have

$$\sum_e (s_e - 1)X_{ej} = L_j^{\dagger/2}\big(L(\widetilde{B}_j) - L_j\big)L_j^{\dagger/2}.$$

Since residual support is contained in the reference support, its Laplacian annihilates the kernel of L_j. An operator norm bound on this block thus gives the ordinary relative Laplacian error. We may use $d_j \le N_j$ for the dimension of block j.

The following result gives different error bounds for different blocks while imposing one shared set of exact constraints. Its proof in Appendix B uses a weighted sum of the deterministic discrepancy potentials of Lau, Wang, and Zhou [7], which build on the regularized spectral optimization framework of Allen-Zhu, Liao, and Orecchia [2].

Lemma 8 (Simultaneous PSD sparsification)**.** *Suppose $X_{ej} \succeq 0$ for $e \in [m]$ and $j = 0, \dots, \ell - 1$, where each block has dimension $d_j \ge 1$ and $\sum_e X_{ej} \preceq I$. Given K exact linear constraints on the coefficient changes and tolerances $0 < \delta_j \le 1$, there exist nonnegative coefficients s_e satisfying the constraints on $s - \mathbf{1}$ and*

$$\left\| \sum_e (s_e - 1)X_{ej} \right\| \le \delta_j \qquad (0 \le j < \ell)$$

with support size

$$O\left(K + 1 + \max_{0 \le j < \ell} \frac{(j+1)^8 d_j}{\delta_j^2} \right). \tag{8}$$

The coefficients can be found deterministically in polynomial real arithmetic.

Proof of Theorem 1. Use the hierarchy above, and take $\delta_j = c_1 \varepsilon C^{-j-1}/(j+1)^2$ with a sufficiently small absolute constant $c_1 > 0$. Equation (7) then gives error at most $\varepsilon/2$ in product form. Lemma 7 gives $K = O(n)$ and $d_j \le 2n\rho^j$. Since $\rho C^2 < 1$,

$$\max_j \frac{(j+1)^8 d_j}{\delta_j^2} \le O(n/\varepsilon^2) \sup_{j \ge 0} (j+1)^{12} (\rho C^2)^j = O(n/\varepsilon^2).$$

Therefore Lemma 8 on the original arc variables gives the claimed support. If the initial graph is empty, take $\widetilde{A} = A$. If no residual block is needed, every original arc was removed at level zero, so Lemma 7 gives $O(n)$ input arcs and we return A. The bounds on all construction steps give polynomial real-arithmetic time.

For the arithmetic implementation, the displayed symmetric pseudoinverse square roots are convenient notation. Each can be replaced by a Cholesky whitening on the complement of its kernel. In detail, take an orthonormal basis U of the image of the PSD matrix H to be whitened, and factor $U^\top HU = GG^\top$. The map $G^{-1}U^\top$ agrees with $H^{\dagger/2}$ up to an isometry on its image, so it gives the same norm, trace, and covariance bounds. The kernels of the lift Laplacians and of the normalized Schur defects are known from the connected components and the degree matrices. Thus these operations use only basis construction and Cholesky factorization. Appendix B treats the scalar optimizers in the discrepancy walk.

For the output precision assertion, suppose the input is rational and multiply all weights by an integer D that clears their denominators. The bit length of D can be bounded by the total input encoding length. Work at accuracy $\varepsilon/2$ on this scaled graph. Let a_e be the output weights and Q a positive integer. On the output support, Qa_e is feasible for the transportation problem with degree margins multiplied by Q and capacities $[\lfloor Qa_e \rfloor, \lceil Qa_e \rceil]$. The margins of the scaled graph are integral. Integral bipartite flow supplies an integral solution. Dividing by Q preserves the margins and changes each scaled weight by at most $1/Q$.

The change kills both original defect kernels, by the exact margins and component-wise support. In the scaled graph, all original positive weights are at least one, so each nontrivial component's Laplacian dominates its unweighted Laplacian. A path bound gives smallest positive Laplacian eigenvalue $\Omega(n^{-2})$. The positive Schur-complement eigenvalues inherit the same lower bound: in a component, if a side potential x is orthogonal to constants, then the squared norm of the full vector (x, h) after centering is at least $\|x\|_2^2$, even when h is the minimizing harmonic extension. The adjacency perturbation has operator norm at most n/Q, and thus its defect-normalized norm is $O(n^3/Q)$. Choose $Q = \lceil c_2 n^3/\varepsilon \rceil$ for a sufficiently large absolute c_2, and divide the rounded weights by D to undo the input scaling. Each numerator is bounded by Q times the largest scaled degree, and the denominator divides QD. These bounds establish the bit-length claim. □

Tightness. The complete bipartite digraph with k vertices per side and all arcs oriented left to right gives a matching worst-case order in the unweighted theorem. Its normalized biadjacency is J_k/k. If the output block Q has s nonzero entries and exact margins, its total mass is k, so $\|Q - J_k/k\|_F^2 = \|Q\|_F^2 - 1 \geq k^2/s - 1$. Both normalized defects are the identity on the subspace orthogonal to $\mathbf{1}$. Thus the SV guarantee gives $\|Q - J_k/k\|_F^2 \leq (k-1)\varepsilon^2/4$, using the operator bound and rank at most $k-1$. Hence $s = \Omega(\min\{k^2, k/\varepsilon^2\})$. This agrees with the lower bound in [3].

A Trace-sensitive additive sparsification

Proof of Lemma 4. Write $B_0 = \sum_i C_i$, $\tau = \operatorname{tr} B_0$, and assume $\tau > 0$. The case $\tau = 0$ has all atoms zero. Maintain a nonnegative combination Z between lower and upper walls $L \prec Z \prec U$ with potentials

$$\operatorname{tr}\left(B_0(U - Z)^{-1}\right) \leq 1, \qquad \operatorname{tr}\left(B_0(Z - L)^{-1}\right) \leq 1.$$

Initialize $Z = 0$, $L = -\tau I$, $U = \tau I$. Put $a = t/16$, $b' = (1 + t/2)a$. At each step, shift L by aB_0 and U by $b'B_0$. Denote the shifted gaps by $M_l = Z - L - aB_0$ and $M_u = U + b'B_0 - Z$. The old lower

gap dominates B_0, because the trace of B_0 whitened by that gap is at most one. Thus $M_l \succ 0$ and $\operatorname{tr}(B_0 M_l^{-1}) \le (1-a)^{-1}$. We also have $\operatorname{tr}(B_0 M_u^{-1}) \le 1$.

Let $\Delta_l > 0$ be the increase in the lower potential and $\Delta_u > 0$ the decrease in the upper potential due to the shifts. For an atom C_i and either gap M, set

$$r_i(M) = \operatorname{tr}(M^{-1} C_i), \qquad h_i(M) = \operatorname{tr}(B_0 M^{-1} C_i M^{-1}).$$

Adding wC_i to Z preserves both potentials if

$$\frac{h_i(M_u)}{\Delta_u} + r_i(M_u) \;\le\; \frac{1}{w} \;\le\; \frac{h_i(M_l)}{\Delta_l} - r_i(M_l). \tag{9}$$

Indeed, set $H = C_i^{1/2} M^{-1} C_i^{1/2}$ and $V = C_i^{1/2} M^{-1} B_0 M^{-1} C_i^{1/2}$. The Woodbury identity expresses the upper potential increase as $w \operatorname{tr}(V(I - wH)^{-1})$, which is at most $w h_i(M)/(1 - w r_i(M))$ when $w r_i(M) < 1$. It expresses the lower potential decrease as $w \operatorname{tr}(V(I + wH)^{-1})$, which is at least $w h_i(M)/(1 + w r_i(M))$. These bounds give (9). For any nonzero C_i, $h_i(M) > 0$ since $B_0 \succeq C_i$, so the upper update has positive gap.

The derivative of the potential along either shift is the trace of a product with two inverse gaps. Convexity of the trace inverse, or the resolvent identity, bounds

$$\Delta_u \ge b' \operatorname{tr}(B_0 M_u^{-1} B_0 M_u^{-1}), \qquad \Delta_l \le a \operatorname{tr}(B_0 M_l^{-1} B_0 M_l^{-1}).$$

Summing the left thresholds in (9) gives at most $1/b' + 1$; summing the right thresholds gives at least $1/a - 1/(1-a)$. The latter is strictly larger for the chosen a, b', since $1/a - 1/b' = 8/(1 + t/2)$. Hence some nonzero atom admits an update of positive weight.

After $T = \lceil 32\tau/t^2 \rceil$ steps, the walls imply

$$B_0 - \frac{\tau}{Ta} I \;\preceq\; \frac{Z}{Ta} \;\preceq\; (1 + t/2) B_0 + \frac{\tau}{Ta} I.$$

Since $\tau/(Ta) \le t/2$ and $B_0 \preceq I$, the additive error is at most t. At most one atom is added per step. The thresholds require only matrix inversion and trace evaluations. □

B A discrepancy walk with different block tolerances

Proof of Lemma 8. Start with $s = \mathbf{1}$. At each phase, restrict to the m' live coordinates, of which we require m' to be larger than a sufficiently large constant times (8). Assume the accumulated block error is at most its target tolerance, so $\sum_e s_e X_{ej} \preceq 2I$ for every j. Define signed, doubled atoms $A_{ej} = \operatorname{diag}(s_e X_{ej}/2, -s_e X_{ej}/2)$ on the live coordinates, and write $A_j(x) = \sum_e x_e A_{ej}$, $D'_j = 2d_j$. Then $\sum_e |A_{ej}| \preceq I$, and $\lambda_{\max}(A_j(x)) = \|A_j(x)\|$. The homogeneous constraint subspace for a phase update $s_e \leftarrow s_e(1 + x_e)$ has codimension at most K.

We use the potential of [7, Lemma 3.1] in each block, with the common parameter $\eta = \sqrt{m'}$:

$$\Phi_j(x) = \max_{\substack{M \succeq 0 \\ \operatorname{tr} M = 1}} \left\{ \operatorname{tr}(M A_j(x)) + \frac{2}{\eta} \operatorname{tr}(M^{1/2}) \right\}, \qquad \Psi(x) = \sum_j \frac{\Phi_j(x)}{\sqrt{D'_j}(j+1)^4}.$$

In particular $\Psi(0) = O(1/\sqrt{m'})$ and $\Phi_j(x) \geq \|A_j(x)\| \geq 0$. Let M_j be the optimizing density matrix at the current point and write $S_j = M_j^{1/2}$. For a unit direction y and a step $0 \leq h \leq 1/(2\eta)$, the potential increase bound of [7, Lemma 3.1] gives

$$\Phi_j(x + hy) - \Phi_j(x) \leq h \operatorname{tr}(M_j A_j(y)) + 2\eta h^2 \operatorname{tr}\left(S_j A_j(y) S_j A_j(y) S_j\right). \tag{10}$$

The step cap suffices because $\|A_j(y)\| \leq 1$ and $\|S_j\| \leq 1$.

Begin each phase at $x = 0$. Until at least $m'/4$ box coordinates have reached an endpoint, choose directions supported on the fractional coordinates. For each j, disallow coordinates with

$$\operatorname{tr}(S_j|A_{ej}|) > \frac{b_j}{m'} \operatorname{tr} S_j, \qquad b_j = c_3(j+1)^2.$$

The union of disallowed coordinates has size at most $\sum_j m'/b_j \leq m'/16$ for sufficiently large absolute c_3. On the remaining coordinates the second-order term of block j is a PSD quadratic form $y^\top G_j y$ with trace at most $b_j\sqrt{D'_j}/m'$. To see this, the diagonal entry for e is at most $\operatorname{tr}(M_j|A_{ej}|)\operatorname{tr}(S_j|A_{ej}|)$ by the trace inequality of [7, Lemma 2.3]. The sum of the first factors is at most one, and $\operatorname{tr} S_j \leq \sqrt{D'_j}$. Consequently the PSD matrix

$$G = \sum_j G_j / \left(\sqrt{D'_j}(j+1)^4\right)$$

has trace $O(1/m')$.

In addition to support and the exact phase constraints, require $y \perp x$ and $y \perp \nabla\Psi(x)$. There is still a subspace of dimension $\Omega(m')$, by the phase threshold and the number of fractional coordinates. An orthonormal basis of this subspace has a unit vector with $y^\top G y = O(1/(m')^2)$, by averaging. Take the largest step in this direction up to $1/(2\eta)$ that stays in $[-1,1]^{m'}$. Since $y \perp x$, the sum of squared steps in the phase is at most m'. Summing (10) for Ψ bounds its final value by $O(1/\sqrt{m'})$. Each of its summands is nonnegative, so the final block discrepancy is at most $O((j+1)^4\sqrt{d_j/m'})$, including the rescaling from A_{ej} to $s_e X_{ej}$.

Choose between x and $-x$ so at least $m'/8$ coordinates equal -1. The multiplicative update preserves nonnegativity and the exact constraints, and deletes those coordinates. The block errors sum geometrically over phases because their live support sizes decrease by a fixed factor. In particular, before reaching the cutoff, each accumulated error is at most $O((j+1)^4\sqrt{d_j/M_*})$, where M_* is the cutoff support size. Choosing the absolute constant in $M_* = \Theta(1 + K + \max_j (j+1)^8 d_j/\delta_j^2)$ sufficiently large both closes the induction and gives the target errors at termination.

We give details of the real-arithmetic implementation of the potential step. For each block $M_j = (u_j I - \eta A_j(x))^{-2}$, where $u_j > \eta\lambda_{\max}(A_j(x))$ is the unique scalar with $\operatorname{tr} M_j = 1$ [7, Lemma 3.1]. Its gap to $\eta\lambda_{\max}(A_j(x))$ is at least one. It can be approximated from above by bisection, testing positive definiteness and the inverse-square trace, starting with the interval $[-\eta, \eta + \sqrt{D'_j} + 1]$. Use absolute error at most $\xi = (m')^{-4}/(1 + \sum_j D'_j)$. For the resulting resolvent $\widehat{S}_j$, we have $\widehat{S}_j \preceq S_j \preceq (1+\xi)\widehat{S}_j$, and $\|\widehat{S}_j^2 - M_j\| \leq 2\xi$. The trace budget and the bound on the second-order

form apply to the approximate matrices with $\operatorname{tr}(\widehat{S}_j^2) \le 1$. In the common eigenbasis of $S_j, \widehat{S}_j$, the true curvatures are at most $(1+\xi)^3$ times the approximate ones. Orthogonality to the approximate summed gradient leaves a first-order directional error of at most $2\xi \sum_j D'_j$. Indeed $\|A_j(y)\| \le 1$, and every gradient weight is at most one. The walk takes $O((m')^2)$ steps, since each step either hits an endpoint or increases $\|x\|_2^2$ by $1/(4\eta^2)$. Thus the sum of the first-order errors after multiplication by the step lengths is $O((m')^{-5/2})$. It is absorbed by the discrepancy bounds. Inverting the shifted matrices, finding the constrained subspace, and choosing a basis direction all use polynomially many real-arithmetic operations. □

References

[1] AmirMahdi Ahmadinejad, John Peebles, Edward Pyne, Aaron Sidford, and Salil Vadhan. Singular value approximation and sparsifying random walks on directed graphs. In *64th IEEE Symposium on Foundations of Computer Science (FOCS)*, pp. 846–854, 2023. arXiv:2301.13541.

[2] Zeyuan Allen-Zhu, Zhenyu Liao, and Lorenzo Orecchia. Spectral sparsification and regret minimization beyond matrix multiplicative updates. In *47th ACM Symposium on Theory of Computing (STOC)*, 2015. arXiv:1506.04838.

[3] Emile Anand. Directed singular-value sparsifiers with $O(n \log n/\varepsilon^2)$ arcs. Preprint, 2026. arXiv:2611.14342.

[4] Joshua Batson, Daniel A. Spielman, and Nikhil Srivastava. Twice-Ramanujan sparsifiers. *SIAM Journal on Computing*, 41(6):1704–1721, 2012. arXiv:0808.0163.

[5] Michael B. Cohen, Jelani Nelson, and David P. Woodruff. Optimal approximate matrix product in terms of stable rank. In *43rd International Colloquium on Automata, Languages, and Programming (ICALP)*, 2016. arXiv:1507.02268.

[6] Marcel K. de Carli Silva, Nicholas J. A. Harvey, and Cristiane M. Sato. Sparse sums of positive semidefinite matrices. *ACM Transactions on Algorithms*, 12(1), 2015. arXiv:1107.0088.

[7] Lap Chi Lau, Robert Wang, and Hong Zhou. Spectral sparsification by deterministic discrepancy walk. In *Symposium on Simplicity in Algorithms (SOSA)*, pp. 315–340, 2025. arXiv:2408.06146.

[8] Daniel A. Spielman and Nikhil Srivastava. Graph sparsification by effective resistances. *SIAM Journal on Computing*, 40(6):1913–1926, 2011. arXiv:0803.0929.