



# A scaling-invariant algorithm for linear programming whose running time depends only on the constraint matrix

Daniel Dadush<sup>1</sup> · Sophie Huiberts<sup>2</sup> · Bento Natura<sup>3</sup> · László A. Végh<sup>4</sup>

Received: 17 June 2021 / Accepted: 7 March 2023  
© The Author(s) 2023

## Abstract

Following the breakthrough work of Tardos (Oper Res 34:250–256, 1986) in the bit-complexity model, Vavasis and Ye (Math Program 74(1):79–120, 1996) gave the first exact algorithm for linear programming in the real model of computation with running time depending only on the constraint matrix. For solving a linear program (LP)  $\max c^\top x$ ,  $Ax = b$ ,  $x \geq 0$ ,  $A \in \mathbb{R}^{m \times n}$ , Vavasis and Ye developed a primal-dual interior point method using a ‘layered least squares’ (LLS) step, and showed that  $O(n^{3.5} \log(\bar{\chi}_A + n))$  iterations suffice to solve (LP) exactly, where  $\bar{\chi}_A$  is a condition measure controlling the size of solutions to linear systems related to  $A$ . Monteiro and Tsuchiya (SIAM J Optim 13(4):1054–1079, 2003), noting that the central path is invariant under rescalings of the columns of  $A$  and  $c$ , asked whether there exists an LP algorithm depending instead on the measure  $\bar{\chi}_A^*$ , defined as the minimum  $\bar{\chi}_{AD}$  value achievable by a column rescaling  $AD$  of  $A$ , and gave strong evidence that this should be the case. We resolve this open question affirmatively. Our first main contribution is an  $O(m^2n^2 + n^3)$  time algorithm which works on the linear matroid of  $A$  to compute a nearly optimal diagonal rescaling  $D$  satisfying  $\bar{\chi}_{AD} \leq n(\bar{\chi}_A^*)^3$ . This algorithm also allows us to approximate the value of  $\bar{\chi}_A$  up to a factor  $n(\bar{\chi}_A^*)^2$ . This result is in surprising contrast to that of Tunçel (Math Program 86(1):219–223, 1999), who showed NP-hardness for approximating  $\bar{\chi}_A$  to within  $2^{\text{poly}(\text{rank}(A))}$ . The key

---

This work was done while SH was at Centrum Wiskunde & Informatica, and BN was at the London School of Economics and Political Science. This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme: DD and SH from Grant Agreement No. 805241-QIP, BN and LAV from Grant Agreement No. 757481-ScaleOpt. A preliminary version of this paper has appeared in the proceedings of the 52nd Annual ACM Symposium on Theory of Computing (STOC) [8].

---

✉ László A. Végh  
L.Vegh@lse.ac.uk

- 1 Centrum Wiskunde & Informatica, Amsterdam, The Netherlands
- 2 Columbia University, New York, USA
- 3 Georgia Institute of Technology, Atlanta, USA
- 4 London School of Economics and Political Science, London, UK

insight for our algorithm is to work with ratios  $g_i/g_j$  of circuits of  $A$ —i.e., minimal linear dependencies  $Ag = 0$ —which allow us to approximate the value of  $\bar{\chi}_A^*$  by a maximum geometric mean cycle computation in what we call the ‘*circuit ratio digraph*’ of  $A$ . While this resolves Monteiro and Tsuchiya’s question by appropriate preprocessing, it falls short of providing either a truly scaling invariant algorithm or an improvement upon the base LLS analysis. In this vein, as our second main contribution we develop a *scaling invariant* LLS algorithm, which uses and dynamically maintains improving estimates of the circuit ratio digraph, together with a refined potential function based analysis for LLS algorithms in general. With this analysis, we derive an improved  $O(n^{2.5} \log(n) \log(\bar{\chi}_A^* + n))$  iteration bound for optimally solving (LP) using our algorithm. The same argument also yields a factor  $n/\log n$  improvement on the iteration complexity bound of the original Vavasis–Ye algorithm.

**Keywords** Linear programming · Interior point methods · Layered least squares methods · Circuit imbalances

**Mathematics Subject Classification** 90C05 (Linear programming) · 90C51 (Interior point methods)

## 1 Introduction

The linear programming (LP) problem in primal-dual form is to solve

$$\begin{array}{ll} \min c^\top x & \max y^\top b \\ Ax = b & A^\top y + s = c \\ x \geq 0, & s \geq 0, \end{array} \quad (\text{LP})$$

where  $A \in \mathbb{R}^{m \times n}$ ,  $\text{rank}(A) = m$ ,  $b \in \mathbb{R}^m$ ,  $c \in \mathbb{R}^n$  are given in the input, and  $x, s \in \mathbb{R}^n$ ,  $y \in \mathbb{R}^m$  are the variables. The program in  $x$  will be referred to as the primal problem and the program in  $(y, s)$  as the dual problem.

Khachiyan [23] used the ellipsoid method to give the first polynomial time LP algorithm in the bit-complexity model, that is, polynomial in the bit description length of  $(A, b, c)$ . An outstanding open question is the existence of a *strongly polynomial* algorithm for LP, listed by Smale as one of the most prominent mathematical challenges for the 21st century [46]. Such an algorithm amounts to solving LP using  $\text{poly}(n, m)$  basic arithmetic operations in the real model of computation.<sup>1</sup> Known strongly polynomially solvable LP problems classes include: feasibility for two variable per inequality systems [33], the minimum-cost circulation problem [50], the maximum generalized flow problem [41, 61], and discounted Markov decision problems [65, 67].

Towards this goal, the principal line of attack has been to develop LP algorithms whose running time is bounded in terms of natural *condition measures*. Such condition measures attempt to measure the ‘‘intrinsic complexity’’ of LPs. An important line of work in this area has been to parametrize LPs by the ‘‘niceness’’ of their solutions

<sup>1</sup> In the bit-complexity model, a further requirement is that the algorithm must be in PSPACE.

(e.g. the depth of the most interior point), where relevant examples include the Goffin measure [19] for conic systems and Renegar’s distance to ill-posedness for general LPs [43, 44], and bounded ratios between the nonzero entries in basic feasible solutions [6, 24].

**Parametrizing by the constraint matrix** A second line of research, and the main focus of this work, focuses on the complexity of the constraint matrix  $A$ . The first breakthrough in this area was given by Tardos [51], who showed that if  $A$  has integer entries and all square submatrices of  $A$  have determinant at most  $\Delta$  in absolute value, then (LP) can be solved in  $\text{poly}(n, m, \log \Delta)$  arithmetic operations, independent of the encoding length of the vectors  $b$  and  $c$ . This is achieved by finding the exact solutions to  $O(nm)$  rounded LPs derived from the original LP, with the right hand side vector and cost function being integers of absolute value bounded in terms of  $n$  and  $\Delta$ . From  $m$  such rounded problem instances, one can infer, via proximity results, that  $x_i = 0$  must hold for every optimal solution for some index  $i$ . The process continues by induction until the optimal primal face is identified.

**Path-following methods and the Vavasis–Ye algorithm** In a seminal work, Vavasis and Ye [63] introduced a new type of interior-point method that optimally solves (LP) within  $O(n^{3.5} \log(\bar{\chi}_A + n))$  iterations, where the condition number  $\bar{\chi}_A$  controls the size of solutions to certain linear systems related to the kernel of  $A$  (see Sect. 2 for the formal definition).

Before detailing the Vavasis–Ye (henceforth VY) algorithm, we recall the basics of path following interior-point methods. If both the primal and dual problems in (LP) are strictly feasible, the *central path* for (LP) is the curve  $((x(\mu), y(\mu), s(\mu)) : \mu > 0)$  defined by

$$\begin{aligned} x(\mu)_i s(\mu)_i &= \mu, \quad \forall i \in [n] \\ Ax(\mu) &= b, \quad x(\mu) > 0, \\ A^\top y(\mu) + s(\mu) &= c, \quad s(\mu) > 0, \end{aligned} \tag{CP}$$

which converges to complementary optimal primal and dual solutions  $(x^*, y^*, s^*)$  as  $\mu \rightarrow 0$ , recalling that the duality gap at time  $\mu$  is exactly  $x(\mu)^\top s(\mu) = n\mu$ . We thus refer to  $\mu$  as the normalized duality gap. Methods that “follow the path” generate iterates that stay in a certain neighborhood around it while trying to achieve rapid multiplicative progress w.r.t. to  $\mu$ , where given  $(x, y, s)$  ‘close’ to the path, we define the normalized duality gap as  $\mu(x, y, s) = \sum_{i=1}^n x_i s_i / n$ . Given a target parameter  $\mu'$  and starting point close to the path at parameter  $\mu$ , standard path following methods [20] can compute a point at parameter below  $\mu'$  in at most  $O(\sqrt{n} \log(\mu/\mu'))$  iterations, and hence the quantity  $\log(\mu/\mu')$  can be usefully interpreted as the length of the corresponding segment of the central path.

**Crossover events and layered least squares steps** At a very high level, Vavasis and Ye show that the central path can be decomposed into at most  $\binom{n}{2}$  short but curved segments, possibly joined by long (apriori unbounded) but very straight segments. At the end of each curved segment, they show that a new ordering relation  $x_i(\mu) > x_j(\mu)$ —called a ‘crossover event’—is implicitly learned. This inequality did not hold

at the start of the segment, but is guaranteed to hold at every point from the end of the segment onwards. These  $\binom{n}{2}$  relations give a combinatorial way to measure progress along the central path. In contrast to Tardos's algorithm, where the main progress is setting variables to zero explicitly, the variables participating in crossover events cannot be identified; the analysis only shows their existence.

At a technical level, the VY-algorithm is a variant of the Mizuno–Todd–Ye [39] predictor–corrector method (MTY P-C). In predictor–corrector methods, corrector steps bring an iterate closer to the path, i.e., improve centrality, and predictor steps “shoot down” the path, i.e., reduce  $\mu$  without losing too much centrality. Vavasis and Ye's main algorithmic innovation was the introduction of a new predictor step, called the ‘*layered least squares*’ (LLS) step, which crucially allowed them to cross each aforementioned “straight” segment of the central path in a *single step*, recalling that these straight segments may be arbitrarily long. To traverse the short and curved segments of the path, the standard predictor step, known as *affine scaling* (AS), in fact suffices.

To compute the LLS direction, the variables are decomposed into ‘layers’  $J_1 \cup J_2 \cup \dots \cup J_p = [n]$ . The goal of such a decomposition is to eventually learn a refinement of the optimal partition of the variables  $B^* \cup N^* = [n]$ , where  $B^* := \{i \in [n] : x_i^* > 0\}$  and  $N^* := \{i \in [n] : s_i^* > 0\}$  for the limit optimal solution  $(x^*, y^*, s^*)$ .

The primal affine scaling direction can be equivalently described by solving a weighted least squares problem in  $\text{Ker}(A)$ , with respect to a weighting defined according to the current iterate. The primal LLS direction is obtained by solving a series of weighted least squares problems, starting with focusing only on the final layer  $J_p$ . This solution is gradually extended to the higher layers (i.e., layers with lower indices). The dual directions have analogous interpretations, with the solutions on the layers obtained in the opposite direction, starting with  $J_1$ . If we use the two-level layering  $J_1 = B^*$ ,  $J_2 = N^*$ , and are sufficiently close to the limit  $(x^*, y^*, s^*)$  of the central path, then the LLS step reaches an exact optimal solution in a single step. We note that standard AS steps generically never find an exact optimal solution, and thus some form of “LLS rounding” in the final iteration is always necessary to achieve finite termination with an exact optimal solution.

Of course, guessing  $B^*$  and  $N^*$  correctly is just as hard as solving (LP). Still, if we work with a “good” layerings, these will reveal new information about the “optimal order” of the variables, where  $B^*$  is placed on higher layers than  $N^*$ . The crossover events correspond to swapping two wrongly ordered variables into the correct ordering. Namely, a variable  $i \in B^*$  and  $j \in N^*$  are currently ordered on the same layer, or  $j$  is in a higher layer than  $i$ . After the crossover event,  $i$  will always be placed on a higher layer than  $j$ .

**Computing good layerings and the  $\bar{\chi}_A$  condition measure** Given the above discussion, the obvious question is how to come up with “good” layerings? The philosophy behind LLS can be stated as saying that if modifying a set of variables  $x_I$  barely affects the variables in  $x_{[n] \setminus I}$  (recalling that movement is constrained to  $\Delta x \in \text{Ker}(A)$ ), then one should optimize over  $x_I$  without regard to the effect on  $x_{[n] \setminus I}$ ; hence  $x_I$  should be placed on lower layers.

VY’s strategy for computing such layerings was to directly use the size of the coordinates of the current iterate  $x$  (where  $(x, y, s)$  is a point near the central path). In particular, assuming  $x_1 \geq x_2 \geq \dots \geq x_n$ , the layering  $J_1 \cup J_2 \cup \dots \cup J_p = [n]$  corresponds to consecutive intervals constructed in decreasing order of  $x_i$  values. The break between  $J_i$  and  $J_{i+1}$  occurs if the gap  $x_r/x_{r+1} > g$ , where  $r$  is the rightmost element of  $J_i$  and  $g > 0$  is a threshold parameter. Thus, the expectation is that if  $x_i > gx_j$ , then a small multiplicative change to  $x_j$ , subject to moving in  $\text{Ker}(A)$ , should induce a small multiplicative change to  $x_i$ . By proximity to the central path, the dual ordering is reversed as mentioned above.

The threshold  $g$  for which this was justified in the VY-algorithm is a function of the  $\bar{\chi}_A$  condition measure. We now provide a convenient definition that immediately yields this justification (see Proposition 2.4). Letting  $W = \text{Ker}(A)$  and  $\pi_I(W) = \{x_I : x \in W\}$ , we define  $\bar{\chi}_A := \bar{\chi}_W$  as the minimum number  $M \geq 1$  such that for any  $\emptyset \neq I \subseteq [n]$  and  $z \in \pi_I(W)$ , there exists  $y \in W$  with  $y_I = z$  and  $\|y\| \leq M\|z\|$ . Thus, a change of norm  $\epsilon$  in the variables in  $I$  can be lifted to a change of norm at most  $\bar{\chi}_A \epsilon$  in the variables in  $[n] \setminus I$ . Crucially,  $\bar{\chi}$  is a “self-dual” quantity. That is,  $\bar{\chi}_W = \bar{\chi}_{W^\perp}$ , where  $W^\perp = \text{range}(A^\top)$  is the movement subspace for the dual problem, justifying the reversed layering for the dual (see Sects. 2 for more details).

**The question of scale invariance and  $\bar{\chi}_A^*$**  While the VY layering procedure is powerful, its properties are somewhat mismatched with those of the central path. In particular, variable ordering information has *no intrinsic meaning* on the central path, as the path itself is *scaling invariant*. Namely, the central path point  $(x(\mu), y(\mu), s(\mu))$  w.r.t. the problem instance  $(A, b, c)$  is in bijective correspondence with the central path point  $(D^{-1}x(\mu), Dy(\mu), Ds(\mu))$  w.r.t. the problem instance  $(AD, Dc, b)$  for any positive diagonal matrix  $D$ . The standard path following algorithms are also scaling invariant in this sense.

This lead Monteiro and Tsuchiya [36] to ask whether a scaling invariant LLS algorithm exists. They noted that any such algorithm would then depend on the potentially much smaller parameter

$$\bar{\chi}_A^* := \inf_D \bar{\chi}_{AD}, \tag{1}$$

where the infimum is taken over the set of  $n \times n$  positive diagonal matrices. Thus, Monteiro and Tsuchiya’s question can be rephrased as to whether there exists an exact LP algorithm with running time  $\text{poly}(n, m, \log \bar{\chi}_A^*)$ .

Substantial progress on this question was made in the followup works [28, 37]. The paper [37] showed that the number of iterations of the MTY predictor–corrector algorithm [39] can get from  $\mu_0 > 0$  to  $\eta > 0$  on the central path in

$$O\left(n^{3.5} \log \bar{\chi}_A^* + \min\{n^2 \log \log(\mu^0/\eta), \log(\mu^0/\eta)\}\right)$$

iterations. This is attained by showing that the standard AS steps are reasonably close to the LLS steps. This proximity can be used to show that the AS steps can traverse the “curved” parts of the central path in the same iteration complexity bound as the VY

algorithm. Moreover, on the “straight” parts of the path, the rate of progress amplifies geometrically, thus attaining a log log convergence on these parts. Subsequently, [28] developed an affine invariant *trust region* step, which traverses the full path in  $O(n^{3.5} \log(\bar{\chi}_A^* + n))$  iterations. However, the running time of each iteration is weakly polynomial in  $b$  and  $c$ . The question of developing an LP algorithm with complexity bound  $\text{poly}(n, m, \log \bar{\chi}_A^*)$  thus remained open.

A related open problem to the above is whether it is possible to compute a near-optimal rescaling  $D$  for program (1)? This would give an alternate pathway to the desired LP algorithm by simply preprocessing the matrix  $A$ . The related question of approximating  $\bar{\chi}_A$  was already studied by Tunçel [54], who showed NP-hardness for approximating  $\bar{\chi}_A$  to within a  $2^{\text{poly}(\text{rank}(A))}$  factor. Taken at face value, this may seem to suggest that approximating the rescaling  $D$  should be hard.

A further open question is whether Vavasis and Ye’s cross-over analysis can be improved. Ye showed in [66] that the iteration complexity can be reduced to  $O(n^{2.5} \log(\bar{\chi}_A + n))$  for feasibility problems and further to  $O(n^{1.5} \log(\bar{\chi}_A + n))$  for homogeneous systems, though the  $O(n^{3.5} \log(\bar{\chi}_A + n))$  bound for optimization has not been improved since [63].

## 1.1 Our contributions

In this work, we resolve all of the above questions in the affirmative. We detail our contributions below.

*1. Finding an approximately optimal rescaling.* As our first contribution, we give an  $O(m^2n^2 + n^3)$  time algorithm that works on the linear matroid of  $A$  to compute a diagonal rescaling matrix  $D$  which achieves  $\bar{\chi}_{AD} \leq n(\bar{\chi}_A^*)^3$ , given any  $m \times n$  matrix  $A$ . Furthermore, this same algorithm allows us to approximate  $\bar{\chi}_A$  to within a factor  $n(\bar{\chi}_A^*)^2$ . The algorithm bypasses Tunçel’s hardness result by allowing the approximation factor to depend on  $A$  itself, namely on  $\bar{\chi}_A^*$ . This gives a simple first answer to Monteiro and Tsuchiya’s question: by applying the Vavasis–Ye algorithm directly on the preprocessed  $A$  matrix, we may solve any LP with constraint matrix  $A$  using  $O(n^{3.5} \log(\bar{\chi}_A^* + n))$  iterations. Note that the approximation factor  $n(\bar{\chi}_A^*)^2$  increases the runtime only by a constant factor.

To achieve this result, we work with the circuits of  $A$ , where a circuit  $C \subseteq [n]$  corresponds to an inclusion-wise minimal set of linearly dependent columns. With each circuit, we can associate a vector  $g^C \in \text{Ker}(A)$  with  $\text{supp}(g^C) = C$  that is unique up to scaling. By the ‘circuit ratio’  $\kappa_{ij}$  associated with the pair of nodes  $(i, j)$ , we mean the largest ratio  $|g_j^C/g_i^C|$  taken over every circuit  $C$  of  $A$  such that  $i, j \in C$ . As our first observation, we show that the maximum of all circuit ratios, which we call the ‘circuit imbalance measure’, in fact characterizes  $\bar{\chi}_A$  up to a factor  $n$ . This measure was first studied by Vavasis [56], who showed that it lower bounds  $\bar{\chi}_A$ , though, as far as we are aware, our upper bound is new. The circuit ratios of each pair  $(i, j)$  induce a weighted directed graph we call the ‘circuit ratio digraph’ of  $A$ . From here, our main result is that  $\bar{\chi}_A^*$  is up to a factor  $n$  equal to the maximum geometric mean cycle in the circuit ratio digraph. Our algorithm populates the circuit ratio digraph with approximations of the  $\kappa_{ij}$  ratios for each  $i, j \in [n]$  using standard techniques from

matroid theory, and then computes a rescaling by solving the dual of the maximum geometric mean ratio cycle on the ‘*approximate circuit ratio digraph*’.

2. *Scaling invariant LLS algorithm.* While the above yields an LP algorithm with  $\text{poly}(n, m, \log \bar{\chi}_A^*)$  running time, it does not satisfactorily address Monteiro and Tsuchiya’s question on a scaling invariant algorithm. As our second contribution, we use the circuit ratio digraph directly to give a natural scaling invariant LLS layering algorithm together with a scaling invariant crossover analysis.

At a conceptual level, we show that the circuit ratios give a scale invariant way to measure whether ‘ $x_i > x_j$ ’ and enable a natural layering algorithm. Assume for now that the circuit imbalance value  $\kappa_{ij}$  is known for every pair  $(i, j)$ . Given the circuit ratio graph induced by the  $\kappa_{ij}$ ’s and given a primal point  $x$  near the path, our layering algorithm can be described as follows. We first rescale the variables so that  $x$  becomes the all ones vector, which rescales  $\kappa_{ij}$  to  $\kappa_{ij}x_i/x_j$ . We then restrict the graph to its edges of length  $\kappa_{ij}x_i/x_j \geq 1/\text{poly}(n)$ —the *long edges* of the (rescaled) circuit ratio graph—and let the layering  $J_1 \cup J_2 \cup \dots \cup J_p$  be a topological ordering of its strongly connected components (SCC) with edges going from left to right. Intuitively, variables that “affect each other” should be in the same layer, which motivates the SCC definition.

We note that our layering algorithm does not have access to the true circuit ratios  $\kappa_{ij}$ ; these are in fact NP-hard to compute. Getting a good enough initial estimate for our purposes however is easy: we let  $\hat{\kappa}_{ij}$  be the ratio corresponding to an *arbitrary* circuit containing  $i$  and  $j$ . This already turns out to be within a factor  $(\bar{\chi}_A^*)^2$  from the true value  $\kappa_{ij}$ —recall this is the maximum over all such circuits. Our layering algorithm learns better circuit ratio estimates if the ‘lifting costs of our SCC layering, i.e., how much it costs to lift changes from lower layer variables to higher layers (as in the definition of  $\bar{\chi}_A$ ), are larger than we expected them to be based on the previous estimates.

We develop a scaling-invariant analogue of cross-over events as follows. Before the crossover event,  $\text{poly}(n)(\bar{\chi}_A^*)^n > \kappa_{ij}x_i/x_j$ , and after the crossover event,  $\text{poly}(n)(\bar{\chi}_A^*)^n < \kappa_{ij}x_i/x_j$  for all further central path points. Our analysis relies on  $\bar{\chi}_A^*$  in only a minimalistic way, and does not require an estimate on the value of  $\bar{\chi}_A^*$ . Namely, it is only used to show that if  $i, j \in J_q$ , for a layer  $q \in [p]$ , then the rescaled circuit ratio  $\kappa_{ij}x_i/x_j$  is in the range  $(\text{poly}(n)\bar{\chi}_A^*)^{\pm O(|J_q|)}$ . The argument to show this crucially utilizes the maximum geometric mean cycle characterization. Furthermore, unlike prior analyses [36, 63], our definition of a “good” layering (i.e., ‘balanced’ layerings, see Sect. 3.5), is completely independent of  $\bar{\chi}_A^*$ .

3. *Improved potential analysis.* As our third contribution, we improve the Vavasis–Ye crossover analysis using a new and simple potential function based approach. When applied to our new LLS algorithm, we derive an  $O(n^{2.5} \log n \log(\bar{\chi}_A^* + n))$  iteration bound for path following, improving the polynomial term by an  $\Omega(n/\log n)$  factor compared to the VY analysis.

Our potential function can be seen as a fine-grained version of the crossover events as described above. In case of such a crossover event, it is guaranteed that in every subsequent iteration,  $i$  is in a layer before  $j$ . We analyze less radical changes instead: an “event” parametrized by  $\tau$  means that  $i$  and  $j$  are currently together on a layer of

size  $\leq \tau$ , and after the event,  $i$  is on a layer before  $j$ , or if they are together on the same layer, then this layer must have size  $\geq 2\tau$ . For every LLS step, we can find a parameter  $\tau$  such that an event of this type happens concurrently for at least  $\tau - 1$  pairs within the next  $O(\sqrt{n}\tau \log(\bar{\chi}_A^* + n))$  iterations,

Our improved analysis is also applicable to the original VY-algorithm. Let us now comment on the relation between the VY-algorithm and our new algorithm. The VY-algorithm starts a new layer once  $x_{\pi(i)} > g x_{\pi(i+1)}$  between two consecutive variables where the permutation  $\pi$  is a non-increasing order of the  $x_i$  variables, and  $g = \text{poly}(n)\bar{\chi}_A$ . Setting the initial ‘estimates’  $\hat{\kappa}_{ij} = \bar{\chi}_A$  for a suitable polynomial, our algorithm runs the same way as the VY algorithm. Using these estimates, the layering procedure becomes much simpler: there is no need to verify ‘balancedness’ as in our algorithm.

However, using estimates  $\hat{\kappa}_{ij} = \bar{\chi}_A$  has drawbacks. Most importantly, it does not give a lower bound on the true circuit ratio  $\kappa_{ij}$ —to the contrary,  $g$  will be an upper bound. In effect, this causes VY’s layers to be “much larger” than ours, and for this reason, the connection to  $\bar{\chi}_A^*$  is lost. Nevertheless, our potential function analysis can still be adapted to the VY-algorithm to obtain the same  $\Omega(n/\log n)$  improvement on the iteration complexity bound; see Sect. 4.1 for more details.

## 1.2 Related work

Since the seminal works of Karmarkar [22] and Renegar [42], there has been a tremendous amount of work on speeding up and improving interior-point methods. In contrast to the present work, the focus of these works has mostly been to improve complexity of *approximately solving* LPs. Progress has taken many forms, such as the development of novel barrier methods, such as Vaidya’s volumetric barrier [55] and the recent entropic barrier of Bubeck and Eldan [5] and the weighted log-barrier of Lee and Sidford [29, 31], together with new path following techniques, such as the predictor–corrector framework [34, 39], as well as advances in fast linear system solving [30, 48]. For this last line, there has been substantial progress in improving IPM by amortizing the cost of the iterative updates, and working with approximate computations, see e.g. [42, 55] for classical results. Recently, Cohen, Lee and Song [7] developed a new inverse maintenance scheme to get a randomized  $\tilde{O}(n^\omega \log(1/\varepsilon))$ -time algorithm for  $\varepsilon$ -approximate LP, which was derandomized by van den Brand [57]; here  $\omega \approx 2.37$  is the matrix multiplication exponent. A very recent result by van den Brand et al. [60] obtained a randomized  $\tilde{O}(nm + m^3)$  algorithm. For special classes of LP such as network flow and matching problems, even faster algorithms have been obtained using, among other techniques, fast Laplacian solvers, see e.g. [15, 32, 58, 59]. Given the progress above, we believe it to be an interesting problem to understand to what extent these new numerical techniques can be applied to speed up LLS computations, though we expect that such computations will require very high precision. We note that no attempt has been made in the present work to optimize the complexity of the linear algebra.

Subsequent to the conference version of this paper [8], some of the authors extended Tardos’s framework to the real model of computation [14], showing



that  $\text{poly}(n, m, \log \bar{\chi}_A)$  running time can be achieved using approximate solvers in a black box manner. Combined with [57], one obtains a deterministic  $O(mn^{\omega+1} \log^{O(1)}(n) \log(\bar{\chi}_A))$  LP algorithm; using the initial rescaling subroutine from this paper, the dependence can be improved to  $\bar{\chi}_A^*$  resulting in a running time of  $O(mn^{\omega+1} \log^{O(1)}(n) \log(\bar{\chi}_A^* + n))$ . A weaker extension of Tardos's framework to the real model of computation was previously given by Ho and Tunçel [21].

With regard to LLS algorithms, the original VY-algorithm required explicit knowledge of  $\bar{\chi}_A$  to implement their layering algorithm. The paper [35] showed that this could be avoided by computing all LLS steps associated with  $n$  candidate partitions and picking the best one. In particular, they showed that all such LLS steps can be computed in  $O(m^2n)$  time. In [36], an alternate approach was presented to compute an LLS partition directly from the coefficients of the AS step. We note that these methods crucially rely on the variable ordering, and hence are not scaling invariant. Kitahara and Tsuchiya [27], gave a 2-layer LLS step which achieves a running time depending only on  $\bar{\chi}_A^*$  and right-hand side  $b$ , but with no dependence on the objective, assuming the primal feasible region is bounded.

A series of papers have studied the central path from a differential geometry perspective. Monteiro and Tsuchiya [38] showed that a curvature integral of the central path, first introduced by Sonnevend, Stoer, and Zhao [47], is in fact upper bounded by  $O(n^{3.5} \log(\bar{\chi}_A^* + n))$ . This has been extended to SDP and symmetric cone programming [26], and also studied in the context of information geometry [25].

Circuits have appeared in several papers on linear and integer optimization (see [13] and references within). The idea of using circuits within the context of LP algorithms also appears in [12]. They develop a circuit augmentation framework for LP (as well ILP) and show that simplex-like algorithms that take steps according to the “best circuit” direction achieves linear convergence, though these steps are hard to compute. Recently, [11] used circuit imbalance measures to obtain a circuit augmentation algorithm for LP with  $\text{poly}(n, \log(\bar{\chi}_A))$  iterations. We refer to [16] for an overview on circuit imbalances and their applications.

Our algorithm makes progress towards strongly polynomial solvability of LP, by improving the dependence  $\text{poly}(n, m, \log \bar{\chi}_A)$  to  $\text{poly}(n, m, \log \bar{\chi}_A^*)$ . However, in a remarkable recent paper, Allamigeon, Benchimol, Gaubert, and Joswig [1] have shown, using tools from tropical geometry, that path-following methods for the standard logarithmic barrier *cannot* be strongly polynomial. In particular, they give a parametrized family of instances, where, for sufficiently large parameter values, any sequence of iterations following the central path must be of exponential length—thus,  $\bar{\chi}_A^*$  will be doubly exponential. We note that very recently, Allamigeon, Gaubert, and Vandame [3] strengthened this result, showing that no interior point method using a self-concordant barrier function may be strongly polynomial.

As a further recent development, Allamigeon, Dadush, Loho, Natura, and Végh [2] complement these negative results by giving a weakly polynomial interior point method that always terminates in at most  $O(2^n n^{1.5} \log n)$  iterations—even when  $\log \bar{\chi}_A^*$  is unbounded. Moreover, their interior point method is ‘universal’: it matches the number of iterations of any interior point method that uses a self-concordant barrier function up to a factor  $O(n^{1.5} \log n)$ . The ‘subspace LLS’ step used in the paper is a

generalization of the LLS step, using restricted movements in general subspaces, not only coordinate subspaces.

### 1.3 Organization

The rest of the paper is organized as follows. We conclude this section by introducing some notation. Section 2 discusses our results on the circuit imbalance measure. It starts with Sect. 2.1 on the necessary background on the condition measures  $\bar{\chi}_A$  and  $\bar{\chi}_A^*$ . Section 2.2 introduces the circuit imbalance measure, and formulates and explains all main results of Sect. 2. The proofs are given in the rest of the sections: basic properties in Sect. 2.3, the min-max characterization in Sect. 2.4, the circuit finding algorithm in Sect. 2.5, the algorithms for approximating  $\bar{\chi}_A^*$  and  $\bar{\chi}_A$  in Sect. 2.6.

In Sect. 3, we develop our scaling invariant interior-point method. Interior-point preliminaries are given in Sect. 3.1. Section 3.2 introduces the affine scaling and layered-least-squares directions, and proves some basic properties. Section 3.3 provides a detailed overview of the high level ideas and a roadmap to the analysis. Section 3.4 further develops the theory of LLS directions and introduces partition lifting scores. Section 3.5 gives our scaling invariant layering procedure, and our overall algorithm can be found in Sect. 3.6.

In Sect. 4, we give the potential function proof for the improved iteration bound, relying on technical lemmas. The full proof of these lemmas is deferred to Sect. 6; however, Sect. 4 provides the high-level ideas to each proof. Section 4.1 shows that our argument also leads to a factor  $\Omega(n/\log n)$  improvement in the iteration complexity bound of the VY-algorithm.

In Sect. 5, we prove the technical properties of our LLS step, including its proximity to AS and step length estimates. Finally, in Sect. 7, we discuss the initialization of the interior-point method.

Besides reading the paper linearly, we suggest two other possible ways of navigating the paper. Readers mainly interested in the circuit imbalance measure and its approximation may focus only on Sect. 2; this part can be understood without any familiarity with interior point methods. Other readers, who wish to mainly focus on our interior point algorithm may read Sect. 2 only up to Sect. 2.2; this includes all concepts and statements necessary for the algorithm.

### 1.4 Notation

Our notation will largely follow [36, 37]. We let  $\mathbb{R}_{++}$  denote the set of positive reals, and  $\mathbb{R}_+$  the set of nonnegative reals. For  $n \in \mathbb{N}$ , we let  $[n] = \{1, 2, \dots, n\}$ . Let  $e^i \in \mathbb{R}^n$  denote the  $i$ th unit vector, and  $e \in \mathbb{R}^n$  the all 1s vector. For a vector  $x \in \mathbb{R}^n$ , we let  $\text{Diag}(x) \in \mathbb{R}^{n \times n}$  denote the diagonal matrix with  $x$  on the diagonal. We let  $\mathbf{D}$  denote the set of all positive  $n \times n$  diagonal matrices and  $\mathbf{I}_k$  denote the  $k \times k$  identity matrix. For  $x, y \in \mathbb{R}^n$ , we use the notation  $xy \in \mathbb{R}^n$  to denote  $xy = \text{Diag}(x)y = (x_i y_i)_{i \in [n]}$ . The inner product of the two vectors is denoted as  $x^\top y$ . For  $p \in \mathbb{Q}$ , we also use the notation  $x^p$  to denote the vector  $(x_i^p)_{i \in [n]}$ . Similarly, for  $x, y \in \mathbb{R}^n$ , we let  $x/y$  denote the vector  $(x_i/y_i)_{i \in [n]}$ . We denote the support of a vector  $x \in \mathbb{R}^n$  by  $\text{supp}(x) = \{i \in [n] : x_i \neq 0\}$ .

For an index subset  $I \subseteq [n]$ , we use  $\pi_I : \mathbb{R}^n \rightarrow \mathbb{R}^I$  for the coordinate projection. That is,  $\pi_I(x) = x_I$ , and for a subset  $S \subseteq \mathbb{R}^n$ ,  $\pi_I(S) = \{x_I : x \in S\}$ . We let  $\mathbb{R}_J^n = \{x \in \mathbb{R}^n : x_{[n] \setminus J} = 0\}$ .

For a matrix  $B \in \mathbb{R}^{n \times k}$ ,  $I \subset [n]$  and  $J \subset [k]$  we let  $B_{I,J}$  denote the submatrix of  $B$  restricted to the set of rows in  $I$  and columns in  $J$ . We also use  $B_{I,\bullet} = B_{I,[k]}$  and  $B_J = B_{\bullet,J} = B_{[n],J}$ . We let  $B^\dagger \in \mathbb{R}^{k \times n}$  denote the pseudo-inverse of  $B$ .

We let  $\text{Ker}(A)$  denote the kernel of the matrix  $A \subseteq \mathbb{R}^{m \times n}$ . Throughout, we assume that the matrix  $A$  in (LP) has full row rank, and that  $n \geq 3$ .

We use the real model of computation, allowing basic arithmetic operations  $+$ ,  $-$ ,  $\times$ ,  $/$ , comparisons, and square root computations. We keep (exact) square root computations for simplicity but we note that these could be avoided.

**Subspace formulation** Throughout the paper, we let  $W = \text{Ker}(A) \subseteq \mathbb{R}^n$  denote the kernel of the matrix  $A$ . Using this notation, (LP) can be written in the form

$$\begin{aligned} \min \quad & c^\top x & \max \quad & d^\top (c - s) \\ & x \in W + d & & s \in W^\perp + c \\ & x \geq 0, & & s \geq 0, \end{aligned} \tag{2}$$

where  $d \in \mathbb{R}^n$  satisfies  $Ad = b$ . One can e.g., choose  $d$  as the minimum norm solution  $d = \arg \min\{\|x\| : Ax = b\} = A^\top(AA^\top)^{-1}b$ . Note that  $s \in W^\perp + c$  is equivalent to  $\exists y \in \mathbb{R}^m$  such that  $A^\top y + c = s$ . Hence, the original variable  $y$  is implicit in this formulation.

## 2 Finding an approximately optimal rescaling

### 2.1 The condition number $\bar{\chi}$

The condition number  $\bar{\chi}_A$  is defined as

$$\begin{aligned} \bar{\chi}_A &= \sup \left\{ \|A^\top (ADA^\top)^{-1} AD\| : D \in \mathbf{D} \right\} \\ &= \sup \left\{ \frac{\|A^\top y\|}{\|p\|} : y \text{ minimizes } \|D^{1/2}(A^\top y - p)\| \text{ for some } 0 \neq p \in \mathbb{R}^n \text{ and } D \in \mathbf{D} \right\}. \end{aligned} \tag{3}$$

This condition number was first studied by Dikin [9, 10], Stewart [49], and Todd [52], among others, and plays a key role in the analysis of the Vavasis–Ye interior point method [63]. There is an extensive literature on the properties and applications of  $\bar{\chi}_A$ , as well as its relations to other condition numbers. We refer the reader to the papers [21, 36, 63] for further results and references.

It is important to note that  $\bar{\chi}_A$  only depends on the subspace  $W = \text{Ker}(A)$ . Hence, we can also write  $\bar{\chi}_W$  for a subspace  $W \subseteq \mathbb{R}^n$ , defined to be equal to  $\bar{\chi}_A$  for some matrix  $A \in \mathbb{R}^{k \times n}$  with  $W = \text{Ker}(A)$ . We will use the notations  $\bar{\chi}_A$  and  $\bar{\chi}_W$  interchangeably.

The next lemma summarizes some important known properties of  $\bar{\chi}_A$ .

**Proposition 2.1** Let  $A \in \mathbb{R}^{m \times n}$  with full row rank and  $W = \text{Ker}(A)$ .

- (i) If the entries of  $A$  are all integers, then  $\bar{\chi}_A$  is bounded by  $2^{O(L_A)}$ , where  $L_A$  is the input bit length of  $A$ .
- (ii)  $\bar{\chi}_A = \max\{\|B^{-1}A\| : B \text{ non-singular } m \times m \text{-submatrix of } A\}$ .
- (iii) Let the columns of  $B \in \mathbb{R}^{n \times (n-m)}$  form an orthonormal basis of  $W$ . Then

$$\bar{\chi}_W = \max \left\{ \|BB_{I,\cdot}^\dagger\| : \emptyset \neq I \subset [n] \right\}.$$

(iv)  $\bar{\chi}_W = \bar{\chi}_{W^\perp}$ .

**Proof** Part (i) was proved in [63, Lemma 24]. For part (ii), see [53, Theorem 1] and [63, Lemma 3]. In part (iii), the direction  $\geq$  was proved in [49], and the direction  $\leq$  in [40]. The duality statement (iv) was shown in [18]. □

In Proposition 3.8, we will also give another proof of (iv). We now define the lifting map, a key operation in this paper, and explain its connection to  $\bar{\chi}_A$ .

**Definition 2.2** Let us define the *lifting map*  $L_I^W : \pi_I(W) \rightarrow W$  by

$$L_I^W(p) = \arg \min \{ \|z\| : z_I = p, z \in W \}.$$

Note that  $L_I^W$  is the unique linear map from  $\pi_I(W)$  to  $W$  such that  $(L_I^W(p))_I = p$  and  $L_I^W(p)$  is orthogonal to  $W \cap \mathbb{R}_{[n] \setminus I}^n$ .

**Lemma 2.3** Let  $W \subseteq \mathbb{R}^n$  be an  $(n - m)$ -dimensional linear subspace. Let the columns of  $B \in \mathbb{R}^{n \times (n-m)}$  denote an orthonormal basis of  $W$ . Then, viewing  $L_I^W$  as a matrix in  $\mathbb{R}^{n \times |I|}$ ,

$$L_I^W = BB_{I,\cdot}^\dagger.$$

**Proof** If  $p \in \pi_I(W)$ , then  $p = B_{I,\cdot}y$  for some  $y \in \mathbb{R}^{n-m}$ . By the well-known property of the pseudo-inverse we get  $B_{I,\cdot}^\dagger p = \arg \min_{p=B_{I,\cdot}y} \|y\|$ . This solution satisfies  $\pi_I(BB_{I,\cdot}^\dagger p) = p$  and  $BB_{I,\cdot}^\dagger p \in W$ . Since the columns of  $B$  form an orthonormal basis of  $W$ , we have  $\|BB_{I,\cdot}^\dagger p\| = \|B_{I,\cdot}^\dagger p\|$ . Consequently,  $BB_{I,\cdot}^\dagger p$  is the minimum-norm point with the above properties. □

The above lemma and Proposition 2.1(iii) yield the following characterization. This will be the most suitable characterization of  $\bar{\chi}_W$  for our purposes.

**Proposition 2.4** For a linear subspace  $W \subseteq \mathbb{R}^n$ ,

$$\bar{\chi}_W = \max \left\{ \|L_I^W\| : I \subseteq [n], I \neq \emptyset \right\}.$$

The following notation will be convenient for our algorithm. For a subspace  $W \subseteq \mathbb{R}^n$  and an index set  $I \subseteq [n]$ , if  $\pi_I(W) \neq \{0\}$  then we define the *lifting score*

$$\ell^W(I) := \sqrt{\|L_I^W\|^2 - 1}. \tag{4}$$

Otherwise, we define  $\ell^W(I) = 0$ . This means that for any  $z \in \pi_I(W)$  and  $x = L_I^W(z)$ ,  $\|x_{[n] \setminus I}\| \leq \ell^W(I) \|z\|$ .

**The condition number  $\bar{\chi}_A^*$**  For every  $D \in \mathbf{D}$ , we can consider the condition number  $\bar{\chi}_{DW} = \bar{\chi}_{AD^{-1}}$ . We let

$$\bar{\chi}_W^* = \bar{\chi}_A^* = \inf\{\bar{\chi}_{DW} : D \in \mathbf{D}\}$$

denote the best possible value of  $\bar{\chi}$  that can be attained by rescaling the coordinates of  $W$ . The main result of this section is the following theorem.

**Theorem 2.5** (Proof in Sect. 2.6) *There is an  $O(n^2m^2 + n^3)$  time algorithm that for any matrix  $A \in \mathbb{R}^{m \times n}$  computes an estimate  $\xi$  of  $\bar{\chi}_W$  such that*

$$\xi \leq \bar{\chi}_W \leq n(\bar{\chi}_W^*)^2 \xi$$

and a  $D \in \mathbf{D}$  such that

$$\bar{\chi}_W^* \leq \bar{\chi}_{DW} \leq n(\bar{\chi}_W^*)^3.$$

## 2.2 The circuit imbalance measure

The key tool in proving Theorem 2.5 is to study a more combinatorial condition number, the *circuit imbalance measure* which turns out to give a good proxy to  $\bar{\chi}_A$ .

**Definition 2.6** For a linear subspace  $W \subseteq \mathbb{R}^n$  and a matrix  $A$  such that  $W = \text{Ker}(A)$ , a circuit is an inclusion-wise minimal dependent set of columns of  $A$ . Equivalently, a circuit is a set  $C \subseteq [n]$  such that  $W \cap \mathbb{R}_C^n$  is one-dimensional and that no strict subset of  $C$  has this property. The set of circuits of  $W$  is denoted by  $\mathcal{C}_W$ .

Note that circuits defined above are the same as the circuits in the linear matroid associated with  $A$ . Every circuit  $C \in \mathcal{C}_W$  can be associated with a vector  $g^C \in W$  such that  $\text{supp}(g^C) = C$ ; this vector is unique up to scalar multiplication.

**Definition 2.7** For a circuit  $C \in \mathcal{C}_W$  and  $i, j \in C$ , we let

$$\kappa_{ij}^W(C) = \frac{|g_j^C|}{|g_i^C|}. \tag{5}$$

Note that since  $g^C$  is unique up to scalar multiplication, this is independent of the choice of  $g^C$ . For any  $i, j \in [n]$ , we define the *circuit ratio* as the maximum of  $\kappa_{ij}^W(C)$  over all choices of the circuit  $C$ :

$$\kappa_{ij}^W = \max \left\{ \kappa_{ij}^W(C) : C \in \mathcal{C}_W, i, j \in C \right\}. \tag{6}$$

By convention we set  $\kappa_{ij}^W = 0$  if there is no circuit supporting  $i$  and  $j$ . Further, we define the *circuit imbalance measure* as

$$\kappa_W = \max \left\{ \kappa_{ij}^W : i, j \in [n] \right\}.$$

Minimizing over all coordinate rescalings, we define

$$\kappa_W^* = \min \{ \kappa_{DW} : D \in \mathbf{D} \}.$$

We omit the index  $W$  whenever it is clear from context. Further, for a vector  $d \in \mathbb{R}_{++}^n$ , we write  $\kappa_{ij}^d = \kappa_{ij}^{\text{Diag}(d)W}$  and  $\kappa^d = \kappa_W^d = \kappa_{\text{Diag}(d)W}$ .

We want to remark that a priori it is not clear that  $\kappa_W^*$  is well-defined. Theorem 2.12 will show that the minimum of  $\{ \kappa_{DW} : D \in \mathbf{D} \}$  is indeed attained.

We next formulate the main statements on the circuit imbalance measure; proofs will be given in the subsequent subsections. Crucially, we show that the circuit imbalance  $\kappa_W$  is a good proxy to the condition number  $\bar{\chi}_W$ . The lower bound was already proven in [56], and the upper bound is from [14]. A slightly weaker upper bound  $\sqrt{1 + (n\kappa_W)^2}$  was previously given in the conference version of this paper [8].

**Theorem 2.8** (Proof in Sect. 2.3) *For a linear subspace  $W \subseteq \mathbb{R}^n$ ,*

$$\sqrt{1 + (\kappa_W)^2} \leq \bar{\chi}_W \leq n\kappa_W.$$

We now overview some basic properties of  $\kappa_W$ . Proposition 2.4 asserts that  $\bar{\chi}_W$  is the maximum  $\ell_2 \rightarrow \ell_2$  operator norm of the mappings  $L_I^W$  over  $I \subseteq [n]$ . In [14], it was shown that  $\kappa_W$  is in contrast the maximum  $\ell_1 \rightarrow \ell_\infty$  operator norm of the same mappings; this easily implies the upper bound  $\bar{\chi}_W \leq n\kappa_W$ .

**Proposition 2.9** [14] *For a linear subspace  $W \subseteq \mathbb{R}^n$ ,*

$$\kappa_W = \max \left\{ \frac{\|L_I^W(p)\|_\infty}{\|p\|_1} : I \subseteq [n], I \neq \emptyset, p \in \pi_I(W) \setminus \{0\} \right\}.$$

Similarly to  $\bar{\chi}_W$ ,  $\kappa_W$  is self-dual; this holds for all individual  $\kappa_{ij}^W$  values as well.

**Lemma 2.10** (Proof in Sect. 2.3) *For any subspace  $W \subseteq \mathbb{R}^n$  and  $i, j \in [n]$ ,  $\kappa_{ij}^W = \kappa_{ji}^{W^\perp}$ .*

The next lemma provides a subroutine that efficiently yields upper bounds on  $\ell^W(I)$  or lower bounds on some circuit imbalance values. Recall the definition of the lifting score  $\ell^W(I)$  from (4).

**Lemma 2.11** (Proof in Sect. 2.3) *There exists a subroutine `VERIFY-LIFT( $W, I, \theta$ )` that, given a linear subspace  $W \subseteq \mathbb{R}^n$ , an index set  $I \subseteq [n]$ , and a threshold  $\theta \in \mathbb{R}_{++}$ , either returns the answer ‘pass’, verifying  $\ell^W(I) \leq \theta$ , or returns the answer ‘fail’, and a pair  $i \in I, j \in [n] \setminus I$  such that  $\theta/n \leq \kappa_{ij}^W$ . The running time can be bounded as  $O(n(n-m)^2)$ .*

The proofs of the above statements are given in Sect. 2.3.

**A min-max theorem** We next provide a combinatorial min-max characterization of  $\kappa_W^*$ . Consider the *circuit ratio digraph*  $G = ([n], E)$  on the node set  $[n]$  where  $(i, j) \in E$  if  $\kappa_{ij} > 0$ , that is, there exists a circuit  $C \in \mathcal{C}$  with  $i, j \in C$ . We will refer to  $\kappa_{ij} = \kappa_{ij}^W$  as the weight of the edge  $(i, j)$ . (Note that  $(i, j) \in E$  if and only if  $(j, i) \in E$ , but the weight of these two edges can be different.)

Let  $H$  be a *cycle* in  $G$ , that is, a sequence of indices  $i_1, i_2, \dots, i_k, i_{k+1} = i_1$ . We use  $|H| = k$  to denote the length of the cycle. (In our terminology, ‘cycles’ always refer to objects in  $G$ , whereas ‘circuits’ refer to the minimum supports in  $\text{Ker}(A)$ .)

We use the notation  $\kappa(H) = \kappa_W(H) = \prod_{j=1}^k \kappa_{i_j i_{j+1}}^W$ . For a vector  $d \in \mathbb{R}_{++}^n$ , we denote  $\kappa_W^d(H) = \kappa_{\text{Diag}(d)W}(H)$ . A simple but important observation is that such a rescaling does not change the value associated with the cycle, that is,

$$\kappa_W^d(H) = \kappa_W(H) \quad \forall d \in \mathbb{R}_{++}^n \quad \text{for any cycle } H \text{ in } G. \tag{7}$$

**Theorem 2.12** (Proof in Sect. 2.4) *For a subspace  $W \subset \mathbb{R}^n$ , we have*

$$\kappa_W^* = \min_{d>0} \kappa_W^d = \max \left\{ \kappa_W(H)^{1/|H|} : H \text{ is a cycle in } G \right\}.$$

The proof relies on the following formulation:

$$\begin{aligned} \kappa_W^* = \quad & \min t \\ & \kappa_{ij} d_j / d_i \leq t \quad \forall (i, j) \in E \\ & d > 0. \end{aligned}$$

Taking logarithms, we can rewrite this problem as

$$\begin{aligned} \quad & \min s \\ \log \kappa_{ij} + z_j - z_i & \leq s \quad \forall (i, j) \in E \\ z & \in \mathbb{R}^n. \end{aligned}$$

This is the dual of the minimum-mean cycle problem with weights  $\log \kappa_{ij}$ , and can be solved in polynomial time (see e.g. [4, Theorem 5.8]).

Whereas this formulation verifies Theorem 2.12, it does not give a polynomial-time algorithm to compute  $\kappa_W^*$ . The caveat is that the values  $\kappa_{ij}^W$  are typically not available; in fact, approximating them up to a factor  $2^{O(m)}$  is NP-hard, as follows from the work of Tunçel [54].

Nevertheless, the following corollary of Theorem 2.12 shows that any arbitrary circuit containing  $i$  and  $j$  yields a  $(\kappa^*)^2$  approximation to  $\kappa_{ij}$ .

**Corollary 2.13** (Proof in Sect. 2.4) *Let us be given a linear subspace  $W \subseteq \mathbb{R}^n$  and  $i, j \in [n], i \neq j$ , and a circuit  $C \in \mathcal{C}_W$  with  $i, j \in C$ . Let  $g \in W$  be the corresponding*

vector with  $\text{supp}(g) = C$ . Then,

$$\frac{\kappa_{ij}^W}{(\kappa_W^*)^2} \leq \frac{|g_j|}{|g_i|} \leq \kappa_{ij}^W.$$

The above statements are shown in Sect. 2.4. In Sect. 2.5, we use techniques from matroid theory and linear algebra to efficiently identify a circuit for any pair of variables that are contained in the same circuit. A matroid is non-separable if the circuit hypergraph is connected; precise definitions and background will be described in Sect. 2.5.

**Theorem 2.14** (Proof in Sect. 2.5) *Given  $A \in \mathbb{R}^{m \times n}$ , there exists an  $O(n^2m^2)$  time algorithm  $\text{FIND-CIRCUITS}(A)$  that obtains a decomposition of  $\mathcal{M}(A)$  to a direct sum of non-separable linear matroids, and returns a family  $\hat{C}$  of circuits such that if  $i$  and  $j$  are in the same non-separable component, then there exists a circuit in  $\hat{C}$  containing both  $i$  and  $j$ . Further, for each  $i \neq j$  in the same component, the algorithm returns a value  $\hat{\kappa}_{ij}$  as the the maximum of  $|g_j/g_i|$  such that  $g \in W$ ,  $\text{supp}(g) = C$  for some  $C \in \hat{C}$  containing  $i$  and  $j$ . For these values,  $\hat{\kappa}_{ij} \leq \kappa_{ij} \leq (\kappa^*)^2 \hat{\kappa}_{ij}$ .*

Finally, in Sect. 2.6, we combine the above results to prove Theorem 2.5 on approximating  $\bar{\chi}_W^*$  and  $\kappa_W^*$ .

Section 2.5 contains an interesting additional statement, namely that the logarithms of the circuit ratios satisfy the triangle inequality. This will also be useful in the analysis of the LLS algorithm. The proof uses similar arguments as the proof of Theorem 2.14. A simpler proof of this statement was subsequently given in [16].

**Lemma 2.15** (Proof in Sect. 2.5)

- (i) *For any distinct  $i, j, k$  in the same connected component of  $\mathcal{C}_W$ , and any  $g^C$  with  $i, j \in C, C \in \mathcal{C}_W$ , there exist circuits  $C_1, C_2 \in \mathcal{C}_W, i, k \in C_1, j, k \in C_2$  such that  $|g_j^C/g_i^C| = |g_j^{C_2}/g_k^{C_2}| \cdot |g_k^{C_1}/g_i^{C_1}|$ .*
- (ii) *For any distinct  $i, j, k$  in the same connected component of  $\mathcal{C}_W, \kappa_{ij} \leq \kappa_{ik} \cdot \kappa_{kj}$ .*

### 2.3 Basic properties of $\kappa_W$

**Theorem 2.8** (Restatement). *For a linear subspace  $W \subseteq \mathbb{R}^n$ ,*

$$\sqrt{1 + (\kappa_W)^2} \leq \bar{\chi}_W \leq n\kappa_W.$$

**Proof** For the first inequality, let  $C \in \mathcal{C}_W$  be the circuit and  $i \neq j \in C$  such that  $|g_j/g_i| = \kappa_W$  for the corresponding solution  $g = g^C$ . Let us use the characterization of  $\bar{\chi}_W$  in Proposition 2.4. Let  $I = ([n] \setminus C) \cup \{i\}$ , and  $p = g_i e^i$ , that is, the vector with  $p_i = g_i$  and  $p_k = 0$  for  $k \neq i$ . Then, the unique vector  $z \in W$  such that  $z_I = p$  is  $z = g$ . Therefore,

$$\bar{\chi}_W \geq \min_{z \in W, z_I = p} \frac{\|z\|}{\|p\|} = \frac{\|g\|}{|g_i|} \geq \frac{\sqrt{|g_i|^2 + |g_j|^2}}{|g_i|} = \sqrt{1 + \kappa_W^2}.$$



The second inequality is immediate from Propositions 2.4 and 2.9, and the inequalities between  $\ell_1$ ,  $\ell_2$ , and  $\ell_\infty$  norms. The proof of the slightly weaker  $\bar{\chi}_W \leq \sqrt{1 + (n\kappa_W)^2}$  follows from Lemma 2.11.  $\square$

The next lemma will be needed to prove Lemma 2.11 and also to analyze the LLS algorithm. Let us say that the vector  $y \in \mathbb{R}^n$  conforms to  $x \in \mathbb{R}^n$  if  $x_i y_i > 0$  whenever  $y_i \neq 0$ .

**Lemma 2.16** For  $i \in I \subset [n]$  with  $e^i \in \pi_I(W)$ , let  $z = L_I^W(e^i)$ . Then for any  $j \in \text{supp}(z)$  we have  $\kappa_{ij}^W \geq |z_j|$ .

**Proof** We consider the cone  $F \subset W$  of vectors that conform to  $z$ . The faces of  $F$  are bounded by inequalities of the form  $z_k y_k \geq 0$  or  $y_k = 0$ . The edges (rays) of  $F$  are of the form  $\{\alpha g : \alpha \geq 0\}$  with  $\text{supp}(g) \in \mathcal{C}_W$ . It is easy to see from the Minkowski–Weyl theorem that  $z$  can be written as

$$z = \sum_{k=1}^h g^k,$$

where  $h \leq n$ ,  $C_1, C_2, \dots, C_h \in \mathcal{C}_W$  are circuits, and the vectors  $g^1, g^2, \dots, g^h \in W$  conform to  $z$  and  $\text{supp}(g^k) = C_k$  for all  $k \in [h]$ . Note that  $i \in C_k$  for all  $k \in [h]$ , as otherwise,  $z' = z - g^k$  would also satisfy  $z'_i = e^i$ , but  $\|z'\| < \|z\|$  due to  $g^k$  being conformal to  $z$ , a contradiction to the definition of  $z$ .

At least one  $k \in [h]$  contributes at least as much to  $|z_j| = \frac{\sum_{k=1}^h |g_j^k|}{\sum_{k=1}^h g_i^k}$  as the average.

Hence we find  $\kappa_{ij}^W \geq |g_j^k/g_i^k| \geq |z_j|$ .  $\square$

**Lemma 2.11 (Restatement).** There exists a subroutine VERIFY-LIFT( $W, I, \theta$ ) that, given a linear subspace  $W \subseteq \mathbb{R}^n$ , an index set  $I \subseteq [n]$ , and a threshold  $\theta \in \mathbb{R}_{++}$ , either returns the answer ‘pass’, verifying  $\ell^W(I) \leq \theta$ , or returns the answer ‘fail’, and a pair  $i \in I, j \in [n] \setminus I$  such that  $\theta/n \leq \kappa_{ij}^W$ . The running time can be bounded as  $O(n(n - m)^2)$ .

**Proof** Take any minimal  $I' \subset I$  such that  $\dim(\pi_{I'}(W)) = \dim(\pi_I(W))$ . Then we know that  $\pi_{I'}(W) = \mathbb{R}^{I'}$  and for  $p \in \pi_I(W)$  we can compute  $L_{I'}^W(p) = L_{I'}^W(p_{I'})$ . Let  $B \in \mathbb{R}^{([n] \setminus I) \times I'}$  be the matrix sending any  $q \in \pi_{I'}(W)$  to the corresponding vector  $(L_{I'}^W(q))_{[n] \setminus I}$ . The column  $B_i$  can be computed as  $(L_{I'}^W(e^i))_{[n] \setminus I}$  for  $e^i \in \mathbb{R}^{I'}$ . We have  $\|L_{I'}^W(p)\|^2 = \|p\|^2 + \|(L_{I'}^W(p_{I'}))_{[n] \setminus I}\|^2 \leq \|p\|^2 + \|B\|^2 \|p_{I'}\|^2$  for any  $p \in \pi_I(W)$ , and so  $\ell^W(I) = \sqrt{\|L_{I'}^W\|^2 - 1} \leq \|B\|$ . We upper bound the operator norm by the Frobenius norm as  $\|B\| \leq \|B\|_F = \sqrt{\sum_{ji} B_{ji}^2} \leq n \max_{ji} |B_{ji}|$ . By Lemma 2.16 it follows that  $|B_{ji}| = |(L_{I'}^W(e^i))_j| \leq \kappa_{ij}^W$ . The algorithm returns the answer ‘pass’ if  $n \max_{ji} |B_{ji}| \leq \theta$  and ‘fail’ otherwise.

To implement the algorithm, we first need to select a minimal  $I' \subset I$  such that  $\dim(\pi_{I'}(W)) = \dim(\pi_I(W))$ . This can be found by computing a matrix

$M \in \mathbb{R}^{n \times (n-m)}$  such that  $\text{range}(M) = W$ , and selecting a maximal number of linearly independent columns of  $M_{I, \cdot}$ . Then, we compute the matrix  $B \in \mathbb{R}^{([n] \setminus I) \times I'}$  that implements the transformation  $[L_{I'}^W]_{[n] \setminus I} : \pi_{I'}(W) \rightarrow \pi_{[n] \setminus I}(W)$ . The algorithm returns the pair  $(i, j)$  corresponding to the entry maximizing  $|B_{ji}|$ . The running time analysis will be given in the proof of Lemma 3.15, together with an amortized analysis of a sequence of calls to the subroutine.  $\square$

**Remark 2.17** We note that the algorithm VERIFY-LIFT does not need to compute the circuit as in Lemma 2.16. The following observation will be important in the analysis: the algorithm returns the answer ‘fail’ even if  $\ell^W(I) \leq \theta < n|B_{ji}|$ .

We now prove the duality property of the circuit imbalances.

**Lemma 2.10** (Restatement). *For any subspace  $W \subseteq \mathbb{R}^n$  and  $i, j \in [n]$ ,  $\kappa_{ij}^W = \kappa_{ji}^{W^\perp}$ .*

**Proof** Choose a circuit  $C \in \mathcal{C}_W$  and corresponding circuit solution  $g := g^C \in W \cap \mathbb{R}_C^n$  such that  $\kappa_{ij} = \kappa_{ij}(C) = |g_j/g_i|$ . We will construct a circuit solution in  $W^\perp$  that certifies  $\kappa_{ji}^{W^\perp} \geq \kappa_{ij}^W$ .

Define  $h \in \mathbb{R}^C$  by  $h_i = g_j, h_j = -g_i$  and  $h_k = 0$  for all  $k \in C \setminus \{i, j\}$ . Then,  $h$  is orthogonal to  $g_C$  by construction, and hence  $h \in (\pi_C(W \cap \mathbb{R}_C^n))^\perp = \pi_C(W^\perp)$ . Furthermore, we have  $\text{supp}(h) \in \mathcal{C}_{\pi_C(W^\perp)}$  since  $h \in \mathbb{R}^C$  is a support minimal vector orthogonal to  $g^C$ .

Take any vector  $\bar{h} \in W^\perp$  satisfying  $\bar{h}_C = h$  that is support minimal subject to these constraints. We claim that  $\text{supp}(\bar{h}) \in \mathcal{C}_{W^\perp}$ . Assume not, then there exists a non-zero  $v \in W^\perp$  with  $\text{supp}(v) \subset \text{supp}(\bar{h})$ . Since  $\text{supp}(\pi_C(v)) \subseteq \text{supp}(\pi_C(\bar{h})) = \text{supp}(h)$ , we must have either  $v_C = 0$  or  $v_C = sh$  for  $s \neq 0$ . If  $v_C = 0$ , then  $\bar{h} - \alpha v$  is also in  $W^\perp$  satisfying  $\pi_C(\bar{h}_C - \alpha v) = h$  for all  $\alpha \in \mathbb{R}$ , and since  $v \neq 0$  we can choose  $\alpha$  such that  $\bar{h} - \alpha v$  has smaller support than  $\bar{h}$ , a contradiction. If  $s \neq 0$  then  $v/s \in W^\perp$  satisfies  $\pi_C(v/s) = h$  and has smaller support than  $\bar{h}$ , again a contradiction.

By the above construction, we have

$$\kappa_{ji}^{W^\perp} \geq \left| \frac{\bar{h}_i}{\bar{h}_j} \right| = \left| \frac{h_i}{h_j} \right| = \left| \frac{g_j}{g_i} \right| = \kappa_{ij}^W.$$

By swapping the role of  $W$  and  $W^\perp$  and  $i$  and  $j$ , we obtain  $\kappa_{ij}^W \geq \kappa_{ji}^{W^\perp}$ . The statement follows.  $\square$

### 2.4 A min–max theorem on $\kappa_W^*$

The proof of the characterization of  $\kappa_W^*$  follows.

**Theorem 2.12** (Restatement). *For a subspace  $W \subset \mathbb{R}^n$ , we have*

$$\kappa_W^* = \min_{d>0} \kappa_W^d = \max \left\{ \kappa_W(H)^{1/|H|} : H \text{ is a cycle in } G \right\}.$$

**Proof** For the direction  $\kappa_W(H)^{1/|H|} \leq \kappa_W^*$  we use (7). Let  $d > 0$  be a scaling and  $H$  a cycle. We have  $\kappa_{ij}^d \leq \kappa_W^d$  for every  $i, j \in [n]$ , and hence  $\kappa_W(H) = \kappa_W^d(H) \leq (\kappa_W^d)^{|H|}$ . Since this inequality holds for every  $d > 0$ , it follows that  $\kappa_W(H) \leq (\kappa_W^*)^{|H|}$ .

For the reverse direction, consider the following optimization problem.

$$\begin{aligned} \min t \\ \kappa_{ij}d_j/d_i \leq t \quad \forall(i, j) \in E \\ d > 0. \end{aligned} \tag{8}$$

For any feasible solution  $(d, t)$  and  $\lambda > 0$ , we get another feasible solution  $(\lambda d, t)$  with the same objective value. As such, we can strengthen the condition  $d > 0$  to  $d \geq 1$  without changing the objective value. This makes it clear that the optimum value is achieved by a feasible solution.

Any rescaling  $d > 0$  provides a feasible solution with objective value  $\kappa^d$ , which means that the optimal value  $t^*$  of (8) is  $t^* = \kappa^*$ . Moreover, with the variable substitution  $z_i = \log d_i, s = \log t$ , (8) can be written as a linear program:

$$\begin{aligned} \min s \\ \log \kappa_{ij} + z_j - z_i \leq s \quad \forall(i, j) \in E \\ z \in \mathbb{R}^n. \end{aligned} \tag{9}$$

This is the dual of a minimum-mean cycle problem with respect to the cost function  $\log(\kappa_{ij})$ . Therefore, an optimal solution corresponds to the cycle maximizing  $\sum_{i,j \in H} \log \kappa_{ij}/|H|$ , or in other words, maximizing  $\kappa(H)^{1/|H|}$ .  $\square$

The following example shows that  $\kappa^* \leq \bar{\chi}^*$  can be arbitrarily big.

**Example 2.18** Take  $W = \text{span}((0, 1, 1, M)^\top, (1, 0, M, 1)^\top)$ , where  $M > 0$ . Then  $\{2, 3, 4\}$  and  $\{1, 3, 4\}$  are circuits with  $\kappa_{34}^W(\{2, 3, 4\}) = M$  and  $\kappa_{43}^W(\{1, 3, 4\}) = M$ . Hence, by Theorem 2.12, we see that  $\kappa^* \geq M$ .

**Corollary 2.13** (Restatement). *Let us be given a linear subspace  $W \subseteq \mathbb{R}^n$  and  $i, j \in [n], i \neq j$ , and a circuit  $C \in \mathcal{C}_W$  with  $i, j \in C$ . Let  $g \in W$  be the corresponding vector with  $\text{supp}(g) = C$ . Then,*

$$\frac{\kappa_{ij}^W}{(\kappa_W^*)^2} \leq \frac{|g_j|}{|g_i|} \leq \kappa_{ij}^W.$$

**Proof** The second inequality follows by definition. For the first inequality, note that the same circuit  $C$  yields  $|g_i/g_j| \leq \kappa_{ji}^W(C) \leq \kappa_{ji}^W$ . Therefore,  $|g_j/g_i| \geq 1/\kappa_{ji}^W$ .

From Theorem 2.12 we see that  $\kappa_{ij}^W \kappa_{ji}^W \leq (\kappa_W^*)^2$ , giving  $1/\kappa_{ji}^W \geq \kappa_{ij}^W/(\kappa_W^*)^2$ , completing the proof.  $\square$

## 2.5 Finding circuits: a detour in matroid theory

We next prove Theorem 2.14, showing how to efficiently obtain a family  $\hat{\mathcal{C}} \subseteq \mathcal{C}_W$  such that for any  $i, j \in [n]$ ,  $\hat{\mathcal{C}}$  includes a circuit containing both  $i$  and  $j$ , provided there exists such a circuit.

We need some simple concepts and results from matroid theory. We refer the reader to [45, Chapter 39] or [17, Chapter 5] for definitions and background. Let  $\mathcal{M} = ([n], \mathcal{I})$  be a matroid on ground set  $[n]$  with independent sets  $\mathcal{I} \subseteq 2^{[n]}$ . The rank  $\text{rk}(S)$  of a set  $S \subseteq [n]$  is the maximum size of an independent set contained in  $S$ . The maximal independent sets are called *bases*. All bases have the same cardinality  $\text{rk}([n])$ .

For the matrix  $A \in \mathbb{R}^{m \times n}$ , we will work with the linear matroid  $\mathcal{M}(A) = ([n], \mathcal{I}(A))$ , where a subset  $I \subseteq [n]$  is independent if the columns  $\{A_i : i \in I\}$  are linearly independent. Note that  $\text{rk}([n]) = m$  under the assumption that  $A$  has full row rank.

The *circuits* of the matroid are the inclusion-wise minimal non-independent sets. Let  $I \in \mathcal{I}$  be an independent set, and  $i \in [n] \setminus I$  such that  $I \cup \{i\} \notin \mathcal{I}$ . Then, there exists a unique circuit  $C(I, i) \subseteq I \cup \{i\}$  that is called the *fundamental circuit* of  $i$  with respect to  $I$ . Note that  $i \in C(I, i)$ .

The matroid  $\mathcal{M}$  is *separable*, if the ground set  $[n]$  can be partitioned to two nonempty subsets  $[n] = S \cup T$  such that  $I \in \mathcal{I}$  if and only if  $I \cap S, I \cap T \in \mathcal{I}$ . In this case, the matroid is the direct sum of its restrictions to  $S$  and  $T$ . In particular, every circuit is fully contained in  $S$  or in  $T$ .

For the linear matroid  $\mathcal{M}(A)$ , separability means that  $\text{Ker}(A) = \text{Ker}(A_S) \times \text{Ker}(A_T)$ . In this case, solving (LP) can be decomposed into two subproblems, restricted to the columns in  $A_S$  and in  $A_T$ , and  $\kappa_A = \max\{\kappa_{A_S}, \kappa_{A_T}\}$ .

Hence, we can focus on *non-separable* matroids. The following characterization is well-known, see e.g. [17, Theorems 5.2.5, 5.2.7 – 5.2.9]. For a hypergraph  $H = ([n], \mathcal{E})$ , we define the underlying graph  $H_G = ([n], E)$  such that  $(i, j) \in E$  if there is a hyperedge  $S \in \mathcal{E}$  with  $i, j \in S$ . That is, we add a clique corresponding to each hyperedge. The hypergraph is called *connected* if the underlying graph  $G = ([n], E)$  is connected.

**Proposition 2.19** *For a matroid  $\mathcal{M} = ([n], \mathcal{I})$ , the following are equivalent:*

- (i)  $\mathcal{M}$  is non-separable.
- (ii) The hypergraph of the circuits is connected.
- (iii) For any base  $B$  of  $\mathcal{M}$ , the hypergraph formed by the fundamental circuits  $\mathcal{C}^B = \{C(B, i) : i \in [n] \setminus B\}$  is connected.
- (iv) For any  $i, j \in [n]$ , there exists a circuit containing  $i$  and  $j$ .

**Proof** The implications (i)  $\Leftrightarrow$  (ii), (iii)  $\Rightarrow$  (ii), and (iv)  $\Rightarrow$  (ii) are immediate from the definitions.

For the implication (ii)  $\Rightarrow$  (iii), assume for a contradiction that the hypergraph of the fundamental circuits with respect to  $B$  is not connected. This means that we can partition  $[n] = S \cup T$  such that for each  $i \in S$ ,  $C(B, i) \subseteq S$ , and for each  $i \in T$ ,  $C(B, i) \subseteq T$ . Consequently,  $\text{rk}(S) = |B \cap S|$ ,  $\text{rk}(T) = |B \cap T|$ , and therefore  $\text{rk}([n]) = \text{rk}(S) + \text{rk}(T)$ . It is easy to see that this property is equivalent to separability to  $S$  and  $T$ ; see e.g. [17, Theorem 5.2.7] for a proof.

Finally, for the implication (ii)  $\Rightarrow$  (iv), consider the undirected graph  $([n], E)$  where  $(i, j) \in E$  if there is a circuit containing both  $i$  and  $j$ . This graph is transitive according to [17, Theorem 5.2.5]: if  $(i, j), (j, k) \in E$ , then also  $(i, k) \in E$ . Consequently, whenever  $([n], E)$  is connected, it must be a complete directed graph.  $\square$

We give a different proof of (iii)  $\Rightarrow$  (iv) in Lemma 2.21 that will be convenient for our algorithmic purposes. First, we need a simple lemma that is commonly used in matroid optimization, see e.g. [17, Lemma 13.1.11] or [45, Theorem 39.13].

**Lemma 2.20** *Let  $I$  be an independent set of a matroid  $\mathcal{M} = ([n], \mathcal{I})$ , and  $U = \{u_1, u_2, \dots, u_\ell\} \subseteq I, V = \{v_1, v_2, \dots, v_\ell\} \subseteq [n] \setminus I$  such that  $I \cup \{v_i\}$  is dependent for each  $i \in [\ell]$ . Further, assume that for each  $t \in [\ell], u_t \in C(I, v_t)$  and  $u_t \notin C(I, v_h)$  for all  $h < t$ . Then,  $(I \setminus U) \cup V \in \mathcal{I}$ .*

We give a sketch of the proof. First, we note that for each  $t \in [\ell], u_t \in C(I, v_t)$  means that exchanging  $v_t$  for  $u_t$  maintains independence. The statement follows by induction on  $\ell$ : we consider the independent set  $I' = (I \setminus \{u_\ell\}) \cup \{v_\ell\}$ . We can apply induction for  $I', U' = \{u_1, u_2, \dots, u_{\ell-1}\}$ , and  $V' = \{v_1, v_2, \dots, v_{\ell-1}\}$ , noting that the assumption guarantees that  $C(I', v_t) = C(I, v_t)$  for all  $t \in [\ell - 1]$ . Based on this lemma, we show the following exchange property.

**Lemma 2.21** *Let  $B$  be a basis of the matroid  $\mathcal{M} = ([n], \mathcal{I})$ , and let  $U = \{u_1, u_2, \dots, u_\ell\} \subseteq B$ , and  $V = \{v_1, v_2, \dots, v_\ell, v_{\ell+1}\} \subseteq [n] \setminus B$ . Assume  $C(B, v_1) \cap U = \{u_1\}, C(B, v_{\ell+1}) \cap U = \{u_\ell\}$ , and for each  $2 \leq t \leq \ell, C(B, v_t) \cap U = \{u_{t-1}, u_t\}$ . Then  $(B \setminus U) \cup V$  contains a unique circuit  $C$ , and  $V \subseteq C$ .*

The situation described here corresponds to a minimal path in the hypergraph  $\mathcal{C}^B$  of the fundamental circuits with respect to a basis  $B$ . The hyperedges  $C(B, v_i)$  form a path from  $v_1$  to  $v_{\ell+1}$  such that no shortcut is possible (note that this is weaker than requiring a shortest path).

**Proof of Lemma 2.21** Note that  $S = (B \setminus U) \cup V \notin \mathcal{I}$  since  $|S| > |B|$  and  $B$  is a basis. For any  $i \in [\ell + 1]$ , we can use Lemma 2.20 to show that  $S \setminus \{v_i\} = (B \setminus U) \cup (V \setminus \{v_i\}) \in \mathcal{I}$  (and thus, is a basis). To see this, we apply Lemma 2.20 for the ordered sets  $V' = \{v_1, \dots, v_{i-1}, v_{i+1}, v_\ell, \dots, v_{\ell+1}\}$  and  $U' = \{u_1, \dots, u_{i-1}, u_i, u_{i+1}, \dots, u_\ell\}$ .

Consequently, every circuit in  $S$  must contain the entire set  $V$ . The uniqueness of the circuit in  $S$  follows by the well-known circuit axiom asserting that if  $C, C' \in \mathcal{C}, C \neq C'$  and  $v \in C \cap C'$ , then there exists a circuit  $C'' \in \mathcal{C}$  such that  $C'' \subseteq (C \cup C') \setminus \{v\}$ , contradicting the claim that every circuit in  $S$  contains the entire set  $V$ .  $\square$

We are ready to describe the algorithm that will be used to obtain lower bounds on all  $\kappa_{ij}$  values.

**Theorem 2.14** (Restatement). *Given  $A \in \mathbb{R}^{m \times n}$ , there exists an  $O(n^2 m^2)$  time algorithm FIND-CIRCUITS( $A$ ) that obtains a decomposition of  $\mathcal{M}(A)$  to a direct sum of non-separable linear matroids, and returns a family  $\hat{\mathcal{C}}$  of circuits such that if  $i$  and  $j$  are in the same non-separable component, then there exists a circuit in  $\hat{\mathcal{C}}$  containing both  $i$  and  $j$ . Further, for each  $i \neq j$  in the same component, the algorithm returns a value  $\hat{\kappa}_{ij}$  as the maximum of  $|g_j/g_i|$  such that  $g \in W, \text{supp}(g) = C$  for some  $C \in \hat{\mathcal{C}}$  containing  $i$  and  $j$ . For these values,  $\hat{\kappa}_{ij} \leq \kappa_{ij} \leq (\kappa^*)^2 \hat{\kappa}_{ij}$ .*

**Proof** Once we have found the set of circuits  $\hat{\mathcal{C}}$ , and computed  $\hat{\kappa}_{ij}$  as in the statement, the inequalities  $\hat{\kappa}_{ij} \leq \kappa_{ij} \leq (\kappa^*)^2 \hat{\kappa}_{ij}$  follow easily. The first inequality is by the definition of  $\kappa_{ij}$ , and the second inequality is from Corollary 2.13.

We now turn to the computation of  $\hat{\mathcal{C}}$ . We first obtain a basis  $B \subseteq [n]$  of  $\text{Ker}(A)$  via Gauss-Jordan elimination in time  $O(nm^2)$ . Recall the assumption that  $A$  has full row-rank. Let us assume that  $B = [m]$  is the set of first  $m$  indices. The elimination transforms it to the form  $A = (\mathbf{I}_m | H)$ , where  $H \in \mathbb{R}^{m \times (n-m)}$  corresponds to the non-basis elements. In this form, the fundamental circuit  $C(B, i)$  is the support of the  $i$ th column of  $A$  together with  $i$  for every  $m + 1 \leq i \leq n$ . We let  $\mathcal{C}^B$  denote the set of all these fundamental circuits.

We construct an undirected graph  $G = (B, E)$  as follows. For each  $i \in [n] \setminus B$ , we add a clique between the nodes in  $C(B, i) \setminus \{i\}$ . This graph can be constructed in  $O(nm^2)$  time.

The connected components of  $G$  correspond to the connected components of  $\mathcal{C}^B$  restricted to  $B$ . Thus, due to the equivalence shown in Proposition 2.19 we can obtain the decomposition by identifying the connected components of  $G$ . For the rest of the proof, we assume that the entire hypergraph is connected; connectivity can be checked in  $O(m^2)$  time.

We initialize  $\hat{\mathcal{C}}$  as  $\mathcal{C}^B$ . We will then check all pairs  $i, j \in [n], i \neq j$ . If no circuit  $C \in \hat{\mathcal{C}}$  exists with  $i, j \in C$ , then we will add such a circuit to  $\hat{\mathcal{C}}$  as follows.

Assume first  $i, j \in [n] \setminus B$ . We can find a shortest path in  $G$  between the sets  $C(B, i) \setminus \{i\}$  and  $C(B, j) \setminus \{j\}$  in time  $O(m^2)$ . This can be represented by the sequences of points  $V = \{v_1, v_2, \dots, v_{\ell+1}\} \subseteq [n] \setminus B$ ,  $v_1 = i$ ,  $v_{\ell+1} = j$ , and  $U = \{u_1, u_2, \dots, u_{\ell}\} \subseteq B$  as in Lemma 2.21. According to the lemma,  $S = (B \setminus U) \cup V$  contains a unique circuit  $C$  that contains all  $v_t$ 's, including  $i$  and  $j$ .

We now show how this circuit can be identified in  $O(m)$  time, along with the vector  $g^C$ . Let  $A_S$  be the submatrix corresponding to the columns in  $S$ . Since  $g = g^C$  is unique up to scaling, we can set  $g_{v_1} = 1$ . Note that for each  $t \in [\ell]$ , the row of  $A_S$  corresponding to  $u_t$  contains only two nonzero entries:  $A_{u_t, v_t}$  and  $A_{u_t, v_{t+1}}$ . Thus, the value  $g_{v_1} = 1$  can be propagated to assigning unique values to  $g_{v_2}, g_{v_3}, \dots, g_{v_{\ell+1}}$ . Once these values are set, there is a unique extension of  $g$  to the indices  $t \in B \cap S$  in the basis. Thus, we have identified  $g$  as the unique element of  $\text{Ker}(A_S)$  up to scaling. The circuit  $C$  is obtained as  $\text{supp}(g)$ . Clearly, the above procedure can be implemented in  $O(m)$  time.

The argument easily extends to finding circuits for the case  $\{i, j\} \cap B \neq \emptyset$ . If  $i \in B$ , then for any choice of  $V = \{v_1, v_2, \dots, v_{\ell+1}\}$  and  $U = \{u_1, u_2, \dots, u_{\ell}\}$  as in Lemma 2.21 such that  $i \in C(B, v_1)$  and  $i \notin C(B, v_t)$  for  $t > 1$ , the unique circuit in  $(B \setminus U) \cup V$  also contains  $i$ . This follows from Lemma 2.20 by taking  $V' = \{v_{\ell+1}, v_{\ell}, \dots, v_1\}$  and  $U' = \{u_{\ell}, \dots, u_1, i\}$ , which proves that  $S \setminus \{i\} = (B \setminus U') \cup V' \in \mathcal{I}$ . Similarly, if  $j \in B$  with  $j \in C(B, v_{\ell+1})$  and  $j \notin C(B, v_t)$  for  $t < \ell + 1$ , taking  $V'' = V$  and  $U'' = \{u_1, \dots, u_{\ell}, j\}$  gives  $S \setminus \{j\} \in \mathcal{I}$ .

The bottleneck for the running time is finding the shortest paths for the  $n(n-1)$  pairs, in time  $O(m^2)$  each.  $\square$

**The triangle inequality** An interesting additional fact about the circuit ratio graph is that the logarithm of the weights satisfy the triangle inequality. The proof uses similar arguments as the proof of Theorem 2.14 above.

**Lemma 2.15** (Restatement).

- (i) For any distinct  $i, j, k$  in the same connected component of  $\mathcal{C}_W$ , and any  $g^C$  with  $i, j \in C, C \in \mathcal{C}_W$ , there exist circuits  $C_1, C_2 \in \mathcal{C}_W, i, k \in C_1, j, k \in C_2$  such that  $|g_j^C/g_i^C| = |g_j^{C_2}/g_k^{C_2}| \cdot |g_k^{C_1}/g_i^{C_1}|$ .
- (ii) For any distinct  $i, j, k$  in the same connected component of  $\mathcal{C}_W, \kappa_{ij} \leq \kappa_{ik} \cdot \kappa_{kj}$ .

**Proof** Note that part (ii) immediately follows from part (i) when taking  $C \in \mathcal{C}_W$  such that  $\kappa_{ij}(C) = \kappa_{ij}$ . We now prove part (i).

Let  $A \in \mathbb{R}^{m \times n}$  be a full-rank matrix with  $W = \text{Ker}(A)$ . If  $C = \{i, j\}$ , then the columns  $A_i, A_j$  are linearly dependent. Writing  $A_i = \lambda A_j$ , we have  $\lambda = -g_j^C/g_i^C$ . Let  $h$  be any circuit solution with  $i, k \in \text{supp}(h)$ , and hence  $j \notin \text{supp}(h)$ . By assumption, the vector  $h' = h - h_i e_i + \lambda h_i e_j$  will satisfy  $Ah' = 0$  and have  $i \notin \text{supp}(h'), j, k \in \text{supp}(h')$ . We know that  $h'$  is a circuit solution, because any circuit  $C' \subset \text{supp}(h')$  could, by the above process in reverse, be used to produce a kernel solution with strictly smaller support than  $h$ , contradicting the assumption that  $h$  is a circuit solution. Now we have  $|h'_j/h'_k| \cdot |h_k/h_i| = |h'_j/h_i| = |\lambda|$  by construction. Thus,  $h$  and  $h'$  are the circuit solutions we are looking for.

Now assume  $C \neq \{i, j\}$ . If  $k \in C$ , the statement is trivially true with  $C = C_1 = C_2$ , so assume  $k \notin C$ . Pick  $l \in C, l \notin \{i, j\}$  and set  $B = C \setminus \{l\}$ . Assume without loss of generality that  $B \subseteq [m]$  and apply row operations to  $A$  such that  $A_{B,B} = \mathbf{I}_{B \times B}$  is an identity submatrix and  $A_{[m] \setminus B, B} = 0$ . Then the column  $A_l$  has support given by  $B$ , for otherwise  $g^C$  could not be in the kernel. The given circuit solution satisfies  $g_t^C = -A_{t,l} g_l^C$  for all  $t \in B$ , and in particular  $g_j^C/g_i^C = A_{j,l}/A_{i,l}$ .

Take any circuit solution  $h \in \text{Ker}(A)$  such that  $l, k \in \text{supp}(h)$  and such that  $C \cup \text{supp}(h)$  is inclusion-wise minimal. Such a vector exists by Proposition 2.19(iv). Now let  $J = \text{supp}(h) \setminus C$ . Because  $A_{[m] \setminus B, C} = 0$  and  $Ah = 0$ , we must have  $0 \neq h_J \in \text{Ker}(A_{[m] \setminus B, J})$ . We show that we can uniquely lift any vector  $x \in \text{Ker}(A_{B, C \cup \{k\}})$  to a vector  $x' \in \text{Ker}(A_{C \cup J})$  with  $x'_{C \cup k} = x$ . Since this lift will send circuit solutions to circuit solutions by uniqueness, it suffices to find our desired circuits as solutions to the smaller linear system.

We first prove that  $\dim(\text{Ker}(A_{[m] \setminus B, J})) = 1$ . For suppose that  $\dim(\text{Ker}(A_{[m] \setminus B, J})) \geq 2$ , then  $|J| \geq 2$  and there would exist some vector  $y \in \text{Ker}(A_{[m] \setminus B, J})$  linearly independent from  $h_J$  with  $k \in \text{supp}(y)$ . This vector could be uniquely lifted to a vector  $\bar{y} \in \text{Ker}(A)$ , and we could then find a linear combination  $h + \alpha \bar{y}$  such that  $\text{supp}(h + \alpha \bar{y}) \subsetneq C \cup J$  but  $l, k \in \text{supp}(h + \alpha \bar{y})$ . The existence of such a vector contradicts the minimality of  $C \cup \text{supp}(h)$ . As such, we know that  $\dim(\text{Ker}(A_{[m] \setminus B, J})) = 1$ .

This clear linear relation between any two entries in  $J$  for any vector in  $\text{Ker}(A_{[m] \setminus B, J})$  implies that we can apply row operations to  $A$  such that  $A_{B, J}$  has non-zero entries only in the column  $A_{B, \{k\}}$ . Note that these row operations leave  $A_C$  unchanged because  $A_{[m] \setminus B, C} = 0$ . From this, we can see that any element in  $\text{Ker}(A_{B, C \cup \{k\}})$  can be uniquely lifted to an element in  $\text{Ker}(A_{C \cup J})$ . Hence we can focus on  $\text{Ker}(A_{B, C \cup \{k\}})$ .

If  $A_{i,k} = A_{j,k} = 0$ , then any  $x \in \text{Ker}(A_{B, C \cup \{k\}})$  satisfies  $x_i + A_{i,l}x_l = x_j + A_{j,l}x_l = 0$  and, in particular, any circuit  $l, k \in \bar{C} \subset C \cup \{k\}$  contains  $\{i, j\} \subset \bar{C}$  and fulfills  $|g_j^C/g_i^C| = |A_{j,l}/A_{i,l}| = |g_j^{\bar{C}}/g_i^{\bar{C}}| = |g_j^{\bar{C}}/g_k^{\bar{C}}||g_k^{\bar{C}}/g_i^{\bar{C}}|$ . Choosing  $C_1 = C_2 = \bar{C}$  concludes the case.

Otherwise we know that  $A_{i,k} \neq 0$  or  $A_{j,k} \neq 0$ , meaning that  $\text{Ker}(A_{\{i,j\}, \{i,j,l,k\}})$  contains at least one circuit solution with  $k$  in its support. Observe that any circuit in  $\text{Ker}(A_{\{i,j\}, \{i,j,l,k\}})$  can be lifted uniquely to an element in  $\text{Ker}(A_{B, C \cup \{k\}})$  since  $A_{B,B}$  is an identity matrix and we can set the entries of  $B \setminus \{i, j\}$  individually to satisfy the equalities. Note that this lifted vector is a circuit as well, again by uniqueness of the lift. Hence we may restrict our attention to the matrix  $A_{\{i,j\}, \{i,j,l,k\}}$ . If the columns  $A_{\{i,j\},k}, A_{\{i,j\},l}$  are linearly dependent, then any circuit solution to  $A_{\{i,j\}, \{i,j,l\}}x = 0, x_l \neq 0$ , such as  $g_{\{i,j,l\}}^C$ , is easily transformed into a circuit solution to  $A_{\{i,j\}, \{i,j,k\}}x = 0, x_k \neq 0$  and we are done.

If  $A_{\{i,j\},k}, A_{\{i,j\},l}$  are independent, we can write  $A_{\{i,j\}, \{i,j,l,k\}} = \begin{pmatrix} 1 & 0 & a & c \\ 0 & 1 & b & d \end{pmatrix}$ , where  $g_j^C/g_i^C = b/a$ . For  $\alpha = ad - bc$ , which is non-zero since  $\alpha = \det\left(\begin{pmatrix} a & c \\ b & d \end{pmatrix}\right) \neq 0$  by the independence assumption, we can check that  $(\alpha, 0, -d, b)^\top$  and  $(0, \alpha, c, -a)^\top$  are the circuits we are looking for.  $\square$

### 2.6 Approximating $\bar{\chi}$ and $\bar{\chi}^*$

Equipped with Theorems 2.12 and 2.14, we are ready to prove Theorem 2.5. Recall that we defined  $\kappa_{ij}^d := \kappa_{ij}^{\text{Diag}(d)W} = \kappa_{ij}d_j/d_i$  when  $d > 0$ . We can similarly define  $\hat{\kappa}_{ij}^d := \hat{\kappa}_{ij}d_j/d_i$ , and  $\hat{\kappa}_{ij}^d$  approximates  $\kappa_{ij}^d$  just as in Theorem 2.14.

**Theorem 2.5 (Restatement).** *There is an  $O(n^2m^2 + n^3)$  time algorithm that for any matrix  $A \in \mathbb{R}^{m \times n}$  computes an estimate  $\xi$  of  $\bar{\chi}_W$  such that*

$$\xi \leq \bar{\chi}_W \leq n(\bar{\chi}_W^*)^2\xi$$

and a  $D \in \mathbf{D}$  such that

$$\bar{\chi}_W^* \leq \bar{\chi}_{DW} \leq n(\bar{\chi}_W^*)^3.$$

**Proof** Let us run the algorithm FINDING-CIRCUITS( $A$ ) described in Theorem 2.14 to obtain the values  $\hat{\kappa}_{ij}$  such that  $\hat{\kappa}_{ij} \leq \kappa_{ij} \leq (\kappa_W^*)^2\hat{\kappa}_{ij}$ . We let  $G = ([n], E)$  be the circuit ratio digraph, that is,  $(i, j) \in E$  if  $\kappa_{ij} > 0$ .

To show the first statement on approximating  $\bar{\chi}$ , we simply set  $\xi = \max_{(i,j) \in E} \hat{\kappa}_{ij}$ . Then,

$$\xi \leq \kappa_W \leq \bar{\chi}_W \leq n\kappa_W \leq n(\kappa_W^*)^2\xi \leq n(\bar{\chi}_W^*)^2\xi$$

follows by Theorem 2.8.

For the second statement on finding a nearly optimal rescaling for  $\bar{\chi}_W^*$ , we consider the following optimization problem, which is an approximate version of (8) from



**Theorem 2.12.**

$$\begin{aligned} \min t \\ \hat{\kappa}_{ij}d_j/d_i \leq t \quad \forall (i, j) \in E \\ d > 0. \end{aligned} \tag{10}$$

Let  $\hat{d}$  be an optimal solution to (10) with value  $\hat{t}$ . We will prove that  $\kappa^{\hat{d}} \leq (\kappa_W^*)^3$ .

First, observe that  $\kappa_{ij}^{\hat{d}} = \kappa_{ij}\hat{d}_j/\hat{d}_i \leq (\kappa_W^*)^2\hat{\kappa}_{ij}\hat{d}_j/\hat{d}_i \leq (\kappa_W^*)^2\hat{t}$  for any  $(i, j) \in E$ . Now, let  $d^* > 0$  be such that  $\kappa^{d^*} = \kappa_W^*$ . The vector  $d^*$  is a feasible solution to (10), and so  $\hat{t} \leq \max_{i \neq j} \hat{\kappa}_{ij}d_j^*/d_i^* \leq \max_{i \neq j} \kappa_{ij}d_j^*/d_i^* = \kappa^{d^*}$ . Hence we find that  $\hat{d}$  gives a rescaling with

$$\bar{\chi}_W\hat{D} \leq n\kappa^{\hat{d}} \leq n(\kappa_W^*)^3 \leq n(\bar{\chi}_W)^3,$$

where we again used Theorem 2.8.

We can obtain the optimal value  $\hat{t}$  of (10) by solving the corresponding maximum-mean cycle problem (see Theorem 2.12). It is easy to develop a multiplicative version of the standard dynamic programming algorithm of the classical minimum-mean cycle problem (see e.g. [4, Theorem 5.8]) that allows finding the optimum to (10) directly, in the same  $O(n^3)$  time.

It is left to find the labels  $d_i > 0, i \in [n]$  such that  $\hat{\kappa}_{ij}d_j/d_i \leq \hat{t}$  for all  $(i, j) \in E$ . We define the following weighted directed graph. We associate the weight  $w_{ij} = \log \hat{t} - \log \hat{\kappa}_{ij}$  with every  $(i, j) \in E$ , and add an extra source vertex  $r$  with edges  $(r, i)$  of weight  $w_{ri} = 0$  for all  $i \in [n]$ .

By the choice of  $\hat{t}$ , this graph does not contain any negative weight directed cycles. We can compute the shortest paths from  $r$  to all nodes in  $O(n^3)$  using the Bellman-Ford algorithm; let  $\sigma_i$  be the shortest path label for  $i$ . We then set  $d_i = \exp(\sigma_i)$ . One can avoid computing logarithms by using a multiplicative variant of the Bellman-Ford algorithm instead.

The running time of the whole algorithm will be bounded by  $O(n^2 m^2 + n^3)$ . The running time is dominated by the  $O(n^2 m^2)$  complexity of FINDING-CIRCUITS( $A$ ) and the  $O(n^3)$  complexity of solving the minimum-mean cycle problem and shortest path computation. □

### 3 A scaling-invariant layered least squares interior-point algorithm

#### 3.1 Preliminaries on interior-point methods

In this section, we introduce the standard definitions, concepts and results from the interior-point literature that will be required for our algorithm. We consider an LP problem in the form (LP), or equivalently, in the subspace form (2) for  $W = \text{Ker}(A)$ . We let

$$\mathcal{P}^{++} = \{x \in \mathbb{R}^n : Ax = b, x > 0\},$$

$$\mathcal{D}^{++} = \{(y, s) \in \mathbb{R}^{m+n} : A^\top y + s = c, s > 0\}.$$

Recall the *central path* defined in (CP), with  $w(\mu) = (x(\mu), y(\mu), s(\mu))$  denoting the central path point corresponding to  $\mu > 0$ . We let  $w^* = (x^*, y^*, s^*)$  denote the primal and dual optimal solutions to (LP) that correspond to the limit of the central path for  $\mu \rightarrow 0$ .

For a point  $w = (x, y, s) \in \mathcal{P}^{++} \times \mathcal{D}^{++}$ , the *normalized duality gap* is  $\mu(w) = x^\top s/n$ .

The  $\ell_2$ -neighborhood of the central path with opening  $\beta > 0$  is the set

$$\mathcal{N}(\beta) = \left\{ w \in \mathcal{P}^{++} \times \mathcal{D}^{++} : \left\| \frac{xs}{\mu(w)} - e \right\| \leq \beta \right\}.$$

Furthermore, we let  $\overline{\mathcal{N}}(\beta) := \text{cl}(\mathcal{N}(\beta))$  denote the closure of  $\mathcal{N}(\beta)$ . Throughout the paper, we will assume  $\beta$  is chosen from  $(0, 1/4]$ ; in Algorithm 2 we use the value  $\beta = 1/8$ . The following proposition gives a bound on the distance between  $w$  and  $w(\mu)$  if  $w \in \mathcal{N}(\beta)$ . See e.g., [20, Lemma 5.4], [36, Proposition 2.1].

**Proposition 3.1** *Let  $w = (x, y, s) \in \mathcal{N}(\beta)$  for  $\beta \in (0, 1/4]$  and  $\mu = \mu(w)$ , and consider the central path point  $w(\mu) = (x(\mu), y(\mu), s(\mu))$ . For each  $i \in [n]$ ,*

$$\begin{aligned} \frac{x_i}{1+2\beta} &\leq \frac{1-2\beta}{1-\beta} \cdot x_i \leq x_i(\mu) \leq \frac{x_i}{1-\beta}, \quad \text{and} \\ \frac{s_i}{1+2\beta} &\leq \frac{1-2\beta}{1-\beta} \cdot s_i \leq s_i(\mu) \leq \frac{s_i}{1-\beta}. \end{aligned}$$

We will often use the following proposition which is immediate from definition of  $\beta$ .

**Proposition 3.2** *Let  $w = (x, y, s) \in \mathcal{N}(\beta)$  for  $\beta \in (0, 1/4]$ , and  $\mu = \mu(w)$ . Then for each  $i \in [n]$*

$$(1-\beta)\sqrt{\mu} \leq \sqrt{s_i x_i} \leq (1+\beta)\sqrt{\mu}.$$

**Proof** By definition of  $\mathcal{N}(\odot)$  we have for all  $i \in [n]$  that  $|\frac{x_i s_i}{\mu} - 1| \leq \|\frac{xs}{\mu} - e\| \leq \beta$  and so  $(1-\beta)\mu \leq x_i s_i \leq (1+\beta)\mu$ . Taking roots gives the results.  $\square$

A key property of the central path is “near monotonicity”, formulated in the following lemma, see [63, Lemma 16].

**Lemma 3.3** *Let  $w = (x, y, s)$  be a central path point for  $\mu$  and  $w' = (x', y', s')$  be a central path point for  $\mu' \leq \mu$ . Then  $\|x'/x + s'/s\|_\infty \leq n$ . Further, for the optimal solution  $w^* = (x^*, y^*, s^*)$  corresponding to the central path limit  $\mu \rightarrow 0$ , we have  $\|x^*/x\|_1 + \|s^*/s\|_1 = n$ .*

**Proof** We show that  $\|x'/x\|_1 + \|s'/s\|_1 \leq 2n$  for any feasible primal  $x'$  and dual  $(y', s')$  such that  $(x')^\top s' \leq x^\top s = n\mu$ ; this implies the first statement with the weaker bound  $2n$ . For the stronger bound  $\|x'/x + s'/s\|_\infty \leq n$ , see the proof of [63, Lemma

16]. Since  $x - x' \in W$  and  $s - s' \in W^\perp$ , we have  $(x - x')^\top (s - s') = 0$ . This can be rewritten as  $x^\top s' + (x')^\top s = x^\top s + (x')^\top s'$ . By our assumption on  $x'$  and  $s'$ , the right hand side is bounded by  $2n\mu$ . Dividing by  $\mu$ , and noting that  $x_i s_i = \mu$  for all  $i \in [n]$ , we obtain

$$\left\| \frac{x'}{x} \right\|_1 + \left\| \frac{s'}{s} \right\|_1 = \sum_{i=1}^n \frac{x'_i}{x_i} + \frac{s'_i}{s_i} \leq 2n.$$

The second statement follows by using this to central path points  $(x', y', s')$  with parameter  $\mu'$ , and taking the limit  $\mu' \rightarrow 0$ . □

### 3.2 The affine scaling and layered-least-squares steps

Given  $w = (x, y, s) \in \mathcal{P}^{++} \times \mathcal{D}^{++}$ , the search directions commonly used in interior-point methods are obtained as the solution  $(\Delta x, \Delta y, \Delta s)$  to the following linear system for some  $\sigma \in [0, 1]$ .

$$A\Delta x = 0 \tag{11}$$

$$A^\top \Delta y + \Delta s = 0 \tag{12}$$

$$s\Delta x + x\Delta s = \sigma\mu e - xs \tag{13}$$

Predictor–corrector methods, such as the Mizuno–Todd–Ye Predictor–Corrector (MTY P-C) algorithm [39], alternate between two types of steps. In *predictor steps*, we use  $\sigma = 0$ . This direction is also called the *affine scaling direction*, and will be denoted as  $\Delta w^a = (\Delta x^a, \Delta y^a, \Delta s^a)$  throughout. In *corrector steps*, we use  $\sigma = 1$ . This gives the *centrality direction*, denoted as  $\Delta w^c = (\Delta x^c, \Delta y^c, \Delta s^c)$ .

In the predictor steps, we make progress along the central path. Given the search direction on the current iterate  $w = (x, y, s) \in \mathcal{N}(\beta)$ , the step-length is chosen such that the line segment between the current and next steps remain in  $\overline{\mathcal{N}}(2\beta)$ , i.e.,

$$\alpha^a \leq \sup\{\alpha \in [0, 1] : \forall \alpha' \in [0, \alpha] : w + \alpha' \Delta w^a \in \mathcal{N}(2\beta)\}.$$

Thus, we obtain a point  $w^+ = w + \alpha^a \Delta w^a \in \overline{\mathcal{N}}(2\beta)$ . The corrector step finds a next iterate  $w^c = w^+ + \Delta w^c$ , where  $\Delta w^c$  is the centrality direction computed at  $w^+$ . The next proposition summarizes well-known properties, see e.g. [64, Section 4.5.1].

**Proposition 3.4** *Let  $w = (x, y, s) \in \mathcal{N}(\beta)$  for  $\beta \in (0, 1/4]$ .*

- (i) *For the affine scaling step, we have  $\mu(w^+) = (1 - \alpha)\mu(w)$ .*
- (ii) *The affine scaling step-length can be chosen as*

$$\alpha^a \geq \max \left\{ \frac{\beta}{\sqrt{n}}, 1 - \frac{\|\Delta x^a \Delta s^a\|}{\beta\mu(w)} \right\}.$$

- (iii) *For  $w^+ \in \overline{\mathcal{N}}(2\beta)$  with  $\mu(w^+) > 0$ , let  $\Delta w^c$  be the centrality direction at  $w^+$ . Then for  $w^c = w^+ + \Delta w^c$ , we have  $\mu(w^c) = \mu(w^+)$  and  $w^c \in \mathcal{N}(\beta)$ .*

(iv) After a sequence of  $O(\sqrt{nt})$  predictor and corrector steps, we obtain an iterate  $w' = (x', y', s') \in \mathcal{N}(\beta)$  such that  $\mu(w') \leq \mu(w)/2^t$ .

**Minimum norm viewpoint and residuals** For any point  $w = (x, y, s) \in \mathcal{P}^{++} \times \mathcal{D}^{++}$  we define

$$\delta = \delta(w) = s^{1/2}x^{-1/2} \in \mathbb{R}^n. \tag{14}$$

With this notation, we can write (13) for  $\sigma = 0$  in the form

$$\delta \Delta x + \delta^{-1} \Delta s = -s^{1/2}x^{1/2}. \tag{15}$$

Note that for a point  $w(\mu) = (x(\mu), y(\mu), s(\mu))$  on the central path, we have  $\delta_i(w(\mu)) = s_i(\mu)/\sqrt{\mu} = \sqrt{\mu}/x_i(\mu)$  for all  $i \in [n]$ . From Proposition 3.1, we see that if  $w \in \mathcal{N}(\beta)$ , and  $\mu = \mu(w)$ , then for each  $i \in [n]$ ,

$$\sqrt{1 - 2\beta} \cdot \delta_i(w(\mu)) \leq \delta_i(w) \leq \frac{1}{\sqrt{1 - 2\beta}} \cdot \delta_i(w(\mu)). \tag{16}$$

The matrix  $\text{Diag}(\delta(w))$  will be often used for rescaling in the algorithm. That is, for the current iterate  $w = (x, y, s)$  in the interior-point method, we will perform projections in the space  $\text{Diag}(\delta(w))W$ . To simplify notation, for  $\delta = \delta(w)$ , we use  $L_I^\delta$  and  $\kappa_{ij}^\delta$  as shorthands for  $L_I^{\text{Diag}(\delta)W}$  and  $\kappa_{ij}^{\text{Diag}(\delta)W}$ . The subspace  $W = \text{Ker}(A)$  will be fixed throughout.

It is easy to see from the optimality conditions that the components of the affine scaling direction  $\Delta w^a = (\Delta x^a, \Delta y^a, \Delta s^a)$  are the optimal solutions of the following minimum-norm problems.

$$\begin{aligned} \Delta x^a &= \arg \min_{\Delta x \in \mathbb{R}^n} \{ \|\delta(x + \Delta x)\|^2 : A \Delta x = 0 \} \\ (\Delta y^a, \Delta s^a) &= \arg \min_{(\Delta y, \Delta s) \in \mathbb{R}^m \times \mathbb{R}^n} \{ \|\delta^{-1}(s + \Delta s)\|^2 : A^\top \Delta y + \Delta s = 0 \} \end{aligned} \tag{17}$$

Following [37], for a search direction  $\Delta w = (\Delta x, \Delta y, \Delta s)$ , we define the *residuals* as

$$R_x := \frac{\delta(x + \Delta x)}{\sqrt{\mu}}, \quad R_s := \frac{\delta^{-1}(s + \Delta s)}{\sqrt{\mu}}. \tag{18}$$

We let  $R_x^a$  and  $R_s^a$  denote the residuals for the affine scaling direction  $\Delta w^a$ . Hence, the primal affine scaling direction  $\Delta x^a$  is the one that minimizes the  $\ell_2$ -norm of the primal residual  $R_x^a$ , and the dual affine scaling direction  $(\Delta y^a, \Delta s^a)$  minimizes the  $\ell_2$ -norm of the dual residual  $R_s^a$ . The next lemma summarizes simple properties of the residuals, see [37].

**Lemma 3.5** For  $\beta \in (0, 1/4]$  such that  $w = (x, y, s) \in \mathcal{N}(\beta)$  and the affine scaling direction  $\Delta w = (\Delta x^a, \Delta y^a, \Delta s^a)$ , we have

(i)

$$Rx^aRs^a = \frac{\Delta x^a \Delta s^a}{\mu}, \quad Rx^a + Rs^a = \frac{x^{1/2}s^{1/2}}{\sqrt{\mu}}, \quad (19)$$

(ii)

$$\|Rx^a\|^2 + \|Rs^a\|^2 = n,$$

(iii) We have  $\|Rx^a\|, \|Rs^a\| \leq \sqrt{n}$ , and for each  $i \in [n]$ ,  $\max\{Rx_i^a, Rs_i^a\} \geq \frac{1}{2}(1 - \beta)$ .

(iv)

$$Rx^a = -\frac{1}{\sqrt{\mu}}\delta^{-1}\Delta s^a, \quad Rs^a = -\frac{1}{\sqrt{\mu}}\delta\Delta x^a.$$

**Proof** Parts (i) and (iv) are immediate from the definitions and from (11)-(13) and (15). In part (ii), we use part (i) and  $(Rx^a)^\top Rs^a = 0$ . In part, (iii), the first statement follows by part (ii), and the second statement follows from (i) and Proposition 3.2.  $\square$

For a subset  $I \subset [n]$ , we define

$$\epsilon_I^a(w) := \max_{i \in I} \min\{|Rx_i^a|, |Rs_i^a|\}, \quad \text{and} \quad \epsilon^a(w) := \epsilon_{[n]}^a(w). \quad (20)$$

The next claim shows that for the affine scaling direction, a small  $\epsilon(w)$  yields a long step; see [37, Lemma 2.5].

**Lemma 3.6** *Let  $w = (x, y, s) \in \mathcal{N}(\beta)$  for  $\beta \in (0, 1/4]$ . Then the affine scaling step can be chosen such that*

$$\frac{\mu(w + \alpha^a \Delta w^a)}{\mu(w)} \leq \min \left\{ 1 - \frac{\beta}{\sqrt{n}}, \frac{2\sqrt{n}\epsilon^a(w)}{\beta} \right\}.$$

**Proof** Let  $\epsilon := \epsilon^a(w)$ . From Lemma 3.5(i), we get  $\|\Delta x^a \Delta s^a\|/\mu = \|Rx^aRs^a\|$ . We can bound  $\|Rx^aRs^a\| \leq \epsilon(\|Rx^a\| + \|Rs^a\|) \leq 2\epsilon\sqrt{n}$ , where the latter inequality follows by Lemma 3.5(iii). From Proposition 3.4(ii), we get  $\alpha^a \geq \max\{\beta/\sqrt{n}, 1 - 2\sqrt{n}\epsilon/\beta\}$ . The claim follows by part (i) of the same proposition.  $\square$

### 3.2.1 The layered-least-squares direction

Let  $\mathcal{J} = (J_1, J_2, \dots, J_p)$  be an ordered partition of  $[n]$ .<sup>2</sup> For  $k \in [p]$ , we use the notations  $J_{<k} := J_1 \cup \dots \cup J_{k-1}$ ,  $J_{>k} := J_{k+1} \cup \dots \cup J_p$ , and similarly  $J_{\leq k}$  and  $J_{\geq k}$ . We will also refer to the sets  $J_k$  as layers, and  $\mathcal{J}$  as a layering. Layers with lower indices will be referred to as ‘higher’ layers.

<sup>2</sup> In contrast to how ordered partitions were defined in [37], we use the term *ordered* only to the  $p$ -tuple  $(J_1, \dots, J_p)$ , which is to be viewed independently of  $\delta$ .

Given  $w = (x, y, s) \in \mathcal{P}^{++} \times \mathcal{D}^{++}$ , and the layering  $\mathcal{J}$ , the *layered-least-squares (LLS) direction* is defined as follows. For the primal direction, we proceed backwards, with  $k = p, p - 1, \dots, 1$ . Assume the components on the lower layers  $\Delta x_{J_{>k}}^{\text{ll}}$  have already been determined. We define the components in  $J_k$  as the coordinate projection  $\Delta x_{J_k}^{\text{ll}} = \pi_{J_k}(X_k)$ , where the affine subspace  $X_k$  is defined as the set of minimizers

$$X_k := \arg \min_{\Delta x \in \mathbb{R}^n} \left\{ \left\| \delta_{J_k}(x_{J_k} + \Delta x_{J_k}) \right\|^2 : A \Delta x = 0, \Delta x_{J_{>k}} = \Delta x_{J_{>k}}^{\text{ll}} \right\}. \tag{21}$$

The dual direction  $\Delta s^{\text{ll}}$  is determined in the forward order of the layers  $k = 1, 2, \dots, p$ . Assume we already fixed the components  $\Delta s_{J_{<k}}^{\text{ll}}$  on the higher layers. Then,  $\Delta s_{J_k}^{\text{ll}} = \pi_{J_k}(S_k)$  for

$$S_k = \arg \min_{\Delta s \in \mathbb{R}^n} \left\{ \left\| \delta_{J_k}^{-1}(s_{J_k} + \Delta s_{J_k}) \right\|^2 : \exists y \in \mathbb{R}^m, A^\top \Delta y + \Delta s = 0, \Delta s_{J_{<k}} = \Delta s_{J_{<k}}^{\text{ll}} \right\}. \tag{22}$$

The component  $\Delta y^{\text{ll}}$  is obtained as the optimal  $\Delta y$  for the final layer  $k = p$ . We use the notation  $Rx^{\text{ll}}$  and  $\varepsilon^{\text{ll}}(w)$  analogously to the affine scaling direction. This search direction was first introduced in [63].

The affine scaling direction is a special case for the single element partition. In this case, the definitions (21) and (22) coincide with those in (17).

### 3.3 Overview of ideas and techniques

A key technique in the analysis of layered least-squares algorithms [28, 36, 63] is to argue about variables that have ‘converged’. According to Proposition 3.1 and Lemma 3.3, for any iterate  $w = (x, y, s) \in \mathcal{N}(\beta)$  and the limit optimal solution  $w^* = (x^*, y^*, s^*)$ , the bounds  $x_i^* \leq O(n)x_i$  and  $s_i^* \leq O(n)s_i$  hold. We informally say that  $x_i$  (or  $s_i$ ) has *converged*, if  $x_i \leq O(n)x_i^*$  ( $s_i \leq O(n)s_i^*$ ) hold for the current iterate. Thus, the value of  $x_i$  (or  $s_i$ ) remains within a multiplicative factor  $O(n^2)$  for the rest of the algorithm. Note that if  $\mu > \mu'$  and  $x_i$  has converged at  $\mu$ , then  $\frac{s_i(\mu')/s_i(\mu)}{\mu'/\mu} \in \left[ \frac{1}{O(n^2)}, O(n^2) \right]$ ; thus,  $s_i$  keeps “shooting down” with the central path parameter.

**Converged variables in the affine scaling algorithm** Let us start by showing that at any point of the algorithm, at least one primal or dual variable has converged.

Suppose for simplicity that our current iterate is exactly on the central path, i.e., that  $xs = \mu e$ . This assumption will be maintained throughout this overview. In this case, the residuals can be simply written as  $Rx^a = (x + \Delta x^a)/x$ ,  $Rs^a = (s + \Delta s^a)/s$ . Recall from (17) that the affine scaling direction corresponds to minimizing the residuals  $Rx^a$  and  $Rs^a$ . From this choice, we see that

$$\left\| \frac{x^*}{x} \right\| \geq \left\| \frac{x + \Delta x^a}{x} \right\|, \quad \left\| \frac{s^*}{s} \right\| \geq \left\| \frac{s + \Delta s^a}{s} \right\|. \tag{23}$$

We have  $\|Rx^a\|^2 + \|Rs^a\|^2 = n$  by Lemma 3.5(ii). Let us assume  $\|Rx^a\|^2 \geq n/2$ ; thus, there exists a  $i \in [n]$  such that  $x_i^* \geq x_i/\sqrt{2}$ . In other words, just by looking at the residuals, we get the guarantee that a primal or a dual variable has already converged. Based on the value of the residuals, we can guarantee this to be a primal or a dual variable, but cannot identify which particular  $x_i$  or  $s_i$  this might be.

For  $\|Rx^a\|^2 \geq n/2$ , a primal variable has already converged *before* performing the predictor and corrector steps. We now show that even if  $\|Rx^a\|$  is small, a primal variable will have converged *after* a single iteration. From (23), we see that there is an index  $i$  with  $x_i^*/x_i \geq \|Rx^a\|/\sqrt{n}$ .

Furthermore, Proposition 3.4(ii) and Lemma 3.5 imply that  $1 - \alpha \leq \|Rx^a\| \cdot \|Rs^a\|/\beta \leq \sqrt{n}\|Rx^a\|/\beta$ , since  $\|Rs^a\| \leq \sqrt{n}$ . The predictor step moves to  $x^+ := x + \alpha \Delta x^a = (1 - \alpha)x + \alpha(x + \Delta x^a)$ . Hence,  $x^+ \leq \left(\frac{\sqrt{n}\|Rx^a\|}{\beta} + \|Rx^a\|\right)x$ . Putting the two inequalities together, we learn that  $x_i^+ \leq O(n)x_i^*$  for some  $i \in [n]$ . Since  $w^+ = (x^+, y^+, s^+) \in \bar{\mathcal{N}}(2\beta)$ , Proposition 3.1 implies that  $x_i$  will have converged after this iteration. An analogous argument proves that some  $s_j$  will also have converged after the iteration. We again emphasize that the argument only shows the existence of converged variables, but we cannot identify them in general.

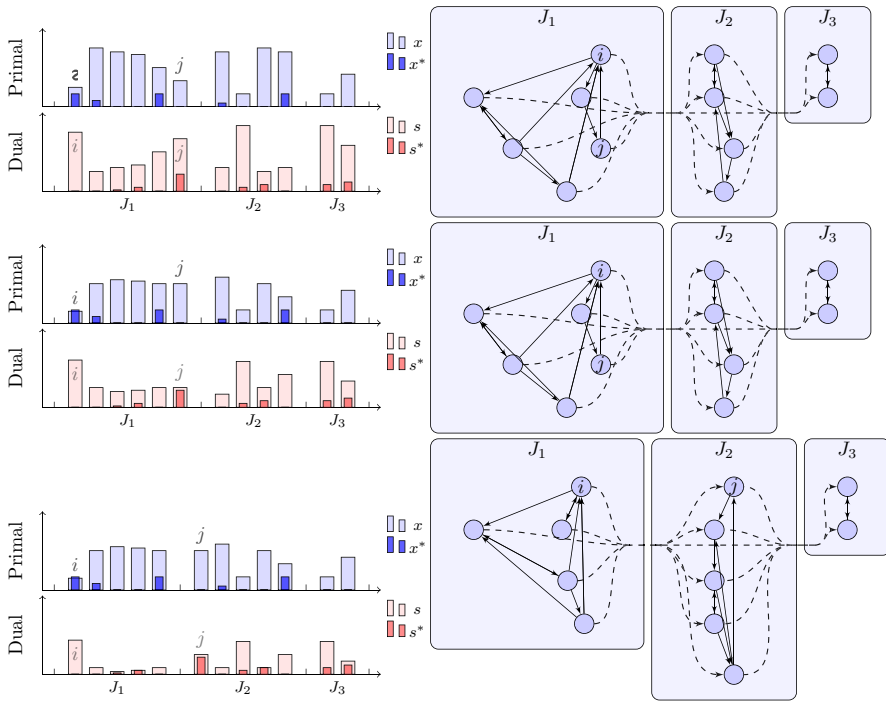
**Measuring combinatorial progress** Tying the above together, we find that after a single affine scaling step, at least one primal variable  $x_i$  and at least one dual variable  $s_j$  has converged. This means that for any  $\mu' < \mu$ ,  $\frac{x_i(\mu')/x_j(\mu')}{x_i(\mu)/x_j(\mu)} \in \left[\frac{\mu}{O(n^4)\mu'}, \frac{O(n^4)\mu}{\mu'}\right]$ ; thus, the ratio of these variables keeps asymptotically increasing. The  $x_i/x_j$  ratios serve as the main progress measure in the Vavasis–Ye algorithm. If  $x_i/x_j$  is between  $1/(\text{poly}(n)\bar{\chi})$  and  $\text{poly}(n)\bar{\chi}$  before the affine scaling step for the pair of converged variables  $x_i$  and  $s_j$ , then after  $\text{poly}(n) \log \bar{\chi}$  iterations, the  $x_i/x_j$  ratio must leave this interval and never return. Thus, we obtain a ‘crossover-event’ that cannot again occur for the same pair of variables. In the affine scaling algorithm, there is no guarantee that  $x_i/x_j$  falls in such a bounded interval for the converging variables  $x_i$  and  $s_j$ ; in particular, we may obtain the same pairs of converged variables after each step.

The main purpose of layered-least-squares methods is to proactively force that in every certain number of iterations, some ‘bounded’  $x_i/x_j$  ratios become ‘large’ and remain so for the rest of the algorithm.

In our approach, the first main insight is to focus on the scaling invariant quantities  $\kappa_{ij}^W x_i/x_j$  instead. For simplicity’s sake, we first present the algorithm with the assumption that all values  $\kappa_{ij}^W$  are known. We will then explain how this assumption can be removed by using gradually improving estimates on the values.

The combinatorial progress will be observed in the ‘long edge graph’. For a primal-dual feasible point  $w = (x, y, s)$  and  $\sigma = 1/O(n^6)$ , this is defined as  $G_{w,\sigma} = ([n], E_{w,\sigma})$  with edges  $(i, j)$  such that  $\kappa_{ij}^W x_i/x_j \geq \sigma$ . Observe that for any  $i, j \in [n]$ , at least one of  $(i, j)$  and  $(j, i)$  are long edges: this follows since for any circuit  $C$  with  $i, j \in C$ , we get lower bounds  $|g_j^C/g_i^C| \leq \kappa_{ij}^W$  and  $|g_i^C/g_j^C| \leq \kappa_{ji}^W$ .

Intuitively, our algorithm will enforce the following two types of events. The analysis in Sect. 4 is based on a potential function analysis capturing roughly the same progress.



**Fig. 1** Top-down we have a chart of primal/dual variables and the *estimated* subgraph of the circuit ratio digraph (Definition 3.11) for three different iterations: 1) All variables except  $x_i$  are far away from their optimal values. 2) On  $J_1$  there is a primal variable ( $i$ ) and dual variable ( $j$ ) that have converged, i.e.  $x_i$  is close to  $x_i^*$  and  $s_i$  is close to  $s_i^*$ . 3)  $j$  moves to layer  $J_2$  due to a change in the underlying subgraph of the circuit ratio digraph

- For an iterate  $w$  and a value  $\mu > 0$ , we have  $i, j \in [n]$  in a strongly connected component in  $G_{w,\sigma}$  of size  $\leq \tau$ , and for any iterate  $w'$  with  $\mu(w') > \mu$ , if  $i, j$  are in a strongly connected component of  $G_{w',\sigma}$  then this component has size  $\geq 2\tau$ .
- For an iterate  $w$  and a value  $\mu > 0$ , we have  $(i, j) \notin E_{w,\sigma}$ , and for any iterate  $w'$  with  $\mu(w') > \mu$  we have  $(i, j) \in E_{w',\sigma}$ .

At most  $O(n^2 \log n)$  such events can happen overall, so if we can prove that on average an event will happen every  $O(\sqrt{n} \log(\bar{\chi}_A^* + n))$  iterations or the algorithm terminates, then we have the desired convergence bound of  $O(n^{2.5} \log(n) \log(\bar{\chi}_A^* + n))$  iterations.

**Converged variables cause combinatorial progress** We now show that combinatorial progress as above must happen in the affine scaling step in the case when the graph  $G_{w,\sigma}$  is strongly connected. As noted above, for the pair of converged variables  $x_i$  and  $s_j$  after the affine scaling step,  $x_i/x_j$ , and thus  $\kappa_{ij}^W x_i/x_j$ , will asymptotically increase by a factor 2 in every  $O(\sqrt{n})$  iterations.

By the strong connectivity assumption, there is a directed path in the long edge graph from  $i$  to  $j$  of length at most  $n - 1$ . Each edge has length at least  $\sigma$ , and by the cycle characterization (Theorem 2.12) we know that  $(\kappa_{ji}^W x_j/x_i) \cdot \sigma^{n-1} \leq (\kappa_W^*)^n$ .



As such,  $\kappa_{ji}^W x_j/x_i \leq (\kappa_W^*/\sigma)^n/\sigma^{n-1}$ . Since  $\kappa_{ij}^W \kappa_{ji}^W \geq 1$ , we obtain the lower bound  $\kappa_{ij}^W x_i/x_j \geq \sigma^{n-1}/(\kappa_W^*)^n$ .

This means that after  $O(\sqrt{n} \log((\kappa_W^*/\sigma)^n)) = O(n^{1.5} \log(\kappa_W^* + n))$  affine scaling steps, the weight of the edge  $(i, j)$  will be more than  $(\kappa_W^*/\sigma)^{4n}$ . There can never again be a length  $n$  or shorter path from  $j$  to  $i$  in the long edge graph, for otherwise the resulting cycle would violate Theorem 2.12. Moreover, by the triangle inequality (Lemma 2.15), any other  $k \neq i, j$  will have either  $(i, k)$  or  $(k, j)$  of length at least  $(\kappa_W^*/\sigma)^{2n}$ , similarly causing a pair of variables to never again be in the same connected component. As such, we took  $O(n^{1.5} \log(\kappa_W^* + n))$  affine scaling steps and in that time at least  $n - 1$  combinatorial progress events have occurred.

**The layered least squares step** Similarly to the Vavasis–Ye algorithm [63] and subsequent literature, our algorithm is a predictor–corrector method using *layered least squares (LLS)* steps as in Sect. 3.2.1 for certain predictor iterations. Our algorithm (Algorithm 2) uses LLS steps only sometimes, and most steps are the simpler affine scaling steps; but for simplicity of this overview, we can assume every predictor iteration uses an LLS step.

We define the ordered partition  $\mathcal{J} = (J_1, J_2, \dots, J_p)$  corresponding to the strongly connected components in topological ordering. Recalling that either  $(i, j)$  or  $(j, i)$  is a long edge for every pair  $i, j \in [n]$ , this order is unique and such that there is a complete directed graph of long edges from every  $J_k$  to  $J_{k'}$  for  $1 \leq k < k' \leq p$ .

The first important property of the LLS step is that it is very close to the affine scaling step. In Sect. 3.4.1, we introduce the partition lifting cost  $\ell^W(\mathcal{J}) = \max_{2 \leq k \leq p} \ell^W(J_{\geq k})$  as the cost of lifting from lower to higher layers; we let  $\ell^{1/x}(\mathcal{J})$  be a shorthand for  $\ell^{\text{Diag}(1/x)W}(\mathcal{J})$ . Note that this same rescaling is used for the affine scaling step in (17), since  $\delta = \sqrt{\mu}/x$  if  $w$  is on the central path. In Lemma 3.10(ii), we show that for a small partition lifting cost, the LLS residuals will remain near the affine scaling residuals. Namely,

$$\|R_x^{\text{ll}} - R_x^{\text{a}}\|, \|R_s^{\text{ll}} - R_s^{\text{a}}\| \leq 6n^{3/2} \ell^{1/x}(\mathcal{J}).$$

Recall that the LLS residuals can be written as  $R_x^{\text{ll}} = (x + \Delta x^{\text{ll}})/x$ ,  $R_s^{\text{ll}} = (s + \Delta s^{\text{ll}})/s$  for a point on the central path. For  $\mathcal{J}$  defined as above, Lemma 2.11 yields  $\ell^{1/x}(\mathcal{J}) \leq n \max_{i \in J_{>k}, j \in J_{\leq k}, k \in [p]} \kappa_{ij}^W x_i/x_j$ . This will be sufficiently small as this maximum is taken over ‘short’ edges (not in  $E_{w,\sigma}$ ).

A second, crucial property of the LLS step is that it “splits” our LP into  $p$  separate LPs that have “negligible” interaction. Namely, the direction  $(\Delta x_{J_k}^{\text{ll}}, \Delta s_{J_k}^{\text{ll}})$  will be very close to the affine scaling step obtained in the problem restricted to the subspace  $W_{\mathcal{J},k} = \{x_{J_k} : x \in W, x_{J_{>k}} = 0\}$  (Lemma 3.10(i))

Since each component  $J_k$  is strongly connected in the long edge graph  $G_{w,\sigma}$ , if there is at least one primal  $x_i$  and dual  $s_j$  in  $J_k$  that have converged after the LLS step, we can use the above argument to show combinatorial progress regarding the  $\kappa_{ij}^W x_i/x_j$  value (Lemma 4.3).

Exploiting the proximity between the LLS and affine scaling steps, Lemma 3.10(iv) gives a lower bound on the step size  $\alpha \geq 1 - \frac{3\sqrt{n}}{\beta} \max_{i \in [n]} \min\{|R_x^{\text{ll}}|, |R_s^{\text{ll}}|\}$ . Let  $J_k$

be the component where  $\min\{\|Rx_{J_k}^{\parallel}\|, \|Rs_{J_k}^{\parallel}\|\}$  is the largest. Hence, the step size  $\alpha$  can be lower bounded in terms of  $\min\{\|Rx_{J_k}^{\parallel}\|, \|Rs_{J_k}^{\parallel}\|\}$ .

The analysis now distinguishes two cases. Let  $w^+ = w + \alpha \Delta s^{\parallel}$  be the point obtained by the predictor LLS step. If the corresponding partition lifting cost  $\ell^{1/x^+}(\mathcal{J})$  is still small, then a similar argument that has shown the convergence of primal and dual variables in the affine scaling step will imply that after the LLS step, at least one  $x_i$  and one  $s_j$  will have converged for  $i, j \in J_k$ . Thus, in this case we obtain the combinatorial progress (Lemma 4.4).

The remaining case is when  $\ell^{1/x^+}(\mathcal{J})$  becomes large. In Lemma 4.5, we show that in this case a new edge will enter the long edge graph, corresponding to the second combinatorial event listed previously. Intuitively, in this case one layer “crashes” into another.

**Refined estimates on circuit imbalances** In the above overview, we assumed the circuit imbalance values  $\kappa_{ij}^W$  are given, and thus the graph  $G_{w,\sigma}$  is available. Whereas these quantities are difficult to compute, we can naturally work with lower estimates. For each  $i, j \in [n]$  that are contained in a circuit together, we start with the lower bound  $\hat{\kappa}_{ij}^W = |g_j^C/g_i^C|$  obtained for an arbitrary circuit  $C$  with  $i, j \in C$ . We use the graph  $\hat{G}_{w,\sigma} = ([n], \hat{E}_{w,\sigma})$  corresponding to these estimates. Clearly,  $\hat{E}_{w,\sigma} \subseteq E_{w,\sigma}$ , but some long edges may be missing. We determine the partition  $\mathcal{J}$  of the strongly connected components of  $\hat{G}_{w,\sigma}$  and estimate the partition lifting cost  $\ell^{1/x}(\mathcal{J})$ . If this is below the desired bound, the argument works correctly. Otherwise, we can identify a pair  $i, j$  responsible for this failure. Namely, we find a circuit  $C$  with  $i, j \in C$  such that  $\hat{\kappa}_{ij}^W < |g_j^C/g_i^C|$ . In this case, we update our estimate, and recompute the partition; this is described in Algorithm 1. At each LLS step, the number of updates is bounded by  $n$ , since every update leads to a decrease in the number of partition classes. This finishes the overview of the algorithm.

### 3.4 A linear system viewpoint of layered least squares

We now continue with the detailed exposition of our algorithm. We present an equivalent definition of the LLS step introduced in Sect. 3.2.1, generalizing the linear system (12)–(13). We use the subspace notation. With this notation, (12)–(13) for the affine scaling direction can be written as

$$s \Delta x^a + x \Delta s^a = -xs, \quad \Delta x^a \in W, \quad \text{and} \quad \Delta s^a \in W^\perp, \tag{24}$$

which is further equivalent to  $\delta \Delta x^a + \delta^{-1} \Delta s^a = -x^{1/2} s^{1/2}$ .

Given the layering  $\mathcal{J}$  and  $w = (x, y, s)$ , for each  $k \in [p]$  we define the subspaces

$$W_{\mathcal{J},k} := \{x_{J_k} : x \in W, x_{J_{>k}} = 0\} \quad \text{and} \quad W_{\mathcal{J},k}^\perp := \{x_{J_k} : x \in W^\perp, x_{J_{<k}} = 0\}.$$

We emphasize that  $W_{\mathcal{J},k}$  and  $W_{\mathcal{J},k}^\perp$  live on the variables in layer  $k$ . That is,  $W_{\mathcal{J},k}, W_{\mathcal{J},k}^\perp \subseteq \mathbb{R}^{J_k}$ . It is easy to see that these two subspaces are orthogonal complements. Our next goal is to show that, analogously to (24), the primal LLS step  $\Delta x^{\parallel}$  is

obtained as the unique solution to the linear system

$$\delta \Delta x^{\text{ll}} + \delta^{-1} \Delta s = -x^{1/2} s^{1/2}, \quad \Delta x^{\text{ll}} \in W, \quad \text{and} \quad \Delta s \in W_{\mathcal{J},1}^{\perp} \times \cdots \times W_{\mathcal{J},p}^{\perp}, \tag{25}$$

and the dual LLS step  $\Delta s^{\text{ll}}$  is the unique solution to

$$\delta \Delta x + \delta^{-1} \Delta s^{\text{ll}} = -x^{1/2} s^{1/2}, \quad \Delta x \in W_{\mathcal{J},1} \times \cdots \times W_{\mathcal{J},p}, \quad \text{and} \quad \Delta s^{\text{ll}} \in W^{\perp}. \tag{26}$$

It is important to note that  $\Delta s$  in (25) may be different from  $\Delta s^{\text{ll}}$ , and  $\Delta x$  in (26) may be different from  $\Delta x^{\text{ll}}$ . In fact,  $\Delta s^{\text{ll}} = \Delta s$  and  $\Delta x^{\text{ll}} = \Delta x$  can only be the case for the affine scaling step.

The following lemma proves that the above linear systems are indeed uniquely solved by the LLS step.

**Lemma 3.7** For  $t \in \mathbb{R}^n$ ,  $W \subseteq \mathbb{R}^n$ ,  $\delta \in \mathbb{R}_{++}^n$ , and  $\mathcal{J} = (J_1, J_2, \dots, J_p)$ , let  $w = \text{LLS}_{\mathcal{J}}^{W,\delta}(t)$  be defined by

$$\delta w + \delta^{-1} v = \delta t, \quad w \in W, \quad v \in W_{\mathcal{J},1}^{\perp} \times \cdots \times W_{\mathcal{J},p}^{\perp}.$$

Then  $\text{LLS}_{\mathcal{J}}^{W,\delta}(t)$  is well-defined and

$$\|\delta_{J_k}(t_{J_k} - w_{J_k})\| = \min \{ \|\delta_{J_k}(t_{J_k} - z_{J_k})\| : z \in W, z_{J_{>k}} = w_{J_{>k}} \}$$

for every  $k \in [p]$ .

In the notation of the above lemma we have, for ordered partitions  $\mathcal{J} = (J_1, J_2, \dots, J_p)$ ,  $\tilde{\mathcal{J}} = (J_p, J_{p-1}, \dots, J_1)$ , and  $(x, y, s) \in \mathcal{P}^{++} \times \mathcal{D}^{++}$  with  $\delta = s^{1/2} x^{-1/2}$ , that  $\Delta x^{\text{ll}} = \text{LLS}_{\mathcal{J}}^{W,\delta}(-x)$  and  $\Delta s^{\text{ll}} = \text{LLS}_{\tilde{\mathcal{J}}}^{W^{\perp},\delta^{-1}}(-s)$ .

**Proof of Lemma 3.7** We first prove the equality  $W \cap (W_{\mathcal{J},1}^{\perp} \times \cdots \times W_{\mathcal{J},p}^{\perp}) = \{0\}$ , and by a similar argument we have  $W^{\perp} \cap (W_{\mathcal{J},1} \times \cdots \times W_{\mathcal{J},p}) = \{0\}$ . By duality, this last equality tells us that

$$(W^{\perp} \cap (W_{\mathcal{J},1} \times \cdots \times W_{\mathcal{J},p}))^{\perp} = W + (W_{\mathcal{J},1}^{\perp} \times \cdots \times W_{\mathcal{J},p}^{\perp}) = \mathbb{R}^n.$$

Thus, the linear decomposition defining  $\text{LLS}_{\mathcal{J}}^{W,\delta}(t)$  has a solution and its solution is unique.

Suppose  $y \in W \cap (W_{\mathcal{J},1}^{\perp} \times \cdots \times W_{\mathcal{J},p}^{\perp})$ . We prove  $y_{J_k} = 0$  by induction on  $k$ , starting at  $k = p$ . The induction hypothesis is that  $y_{J_{>k}} = 0$ , which is an empty requirement when  $k = p$ . The hypothesis  $y_{J_{>k}} = 0$  together with the assumption  $y \in W$  is equivalent to  $y \in W \cap \mathbb{R}_{J_{\leq k}}^n$ , and implies  $y_{J_k} \in \pi_{J_k}(W \cap \mathbb{R}_{J_{\leq k}}^n) := W_{\mathcal{J},k}$ . Since we also have  $y_{J_k} \in W_{\mathcal{J},k}^{\perp}$  by assumption, which is the orthogonal complement

of  $W_{\mathcal{J},k}$ , we must have  $y_{J_k} = 0$ . Hence, by induction  $y = 0$ . This finishes the proof that  $\text{LLS}_{\mathcal{J}}^{W,\delta}(t)$  is well-defined.

Next we prove that  $w$  is a minimizer of  $\min \{ \|\delta_{J_k}(t_{J_k} - z_{J_k})\| : z \in W, z_{J_{>k}} = w_{J_{>k}} \}$ . The optimality condition is for  $\delta_{J_k}(t_{J_k} - z_{J_k})$  to be orthogonal to  $\delta_{J_k}u$  for any  $u \in W_{\mathcal{J},k}$ . By the LLS equation, we have  $\delta_{J_k}(t_{J_k} - w_{J_k}) = \delta_{J_k}^{-1}v_{J_k}$ , where  $v_{J_k} \in W_{\mathcal{J},k}^\perp$ . Noting then that  $\langle \delta_{J_k}u, \delta_{J_k}^{-1}v \rangle = \langle u_{J_k}, v_{J_k} \rangle = 0$  for  $u \in W_{\mathcal{J},k}$ , the optimality condition follows immediately.  $\square$

With these tools, we can prove that the lifting costs are self-dual. This explains the reverse order in the dual vs primal LLS step and justifies our attention on the lifting cost in a self-dual algorithm. The next proposition generalizes the result of [18].

**Proposition 3.8** (Proof in Sect. 5) *For a linear subspace  $W \subseteq \mathbb{R}^n$  and index set  $I \subseteq [n]$  with  $J = [n] \setminus I$ ,*

$$\|L_I^W\| \leq \max\{1, \|L_J^{W^\perp}\|\}.$$

*In particular,  $\ell^W(I) = \ell^{W^\perp}(J)$ .*

We defer the proof to Sect. 5. Note that this proposition also implies Proposition 2.1(iv).

### 3.4.1 Partition lifting scores

A key insight is that if the layering  $\mathcal{J}$  is “well-separated”, then we indeed have  $x \Delta s^{\text{ll}} + s \Delta x^{\text{ll}} \approx -xs$ , that is, the LLS direction is close to the affine scaling direction. This will be shown in Lemma 3.10. The notion of “well-separatedness” can be formalized as follows. Recall the definition of the lifting score (4). The lifting score of the layering  $\mathcal{J} = (J_1, J_2, \dots, J_p)$  of  $[n]$  with respect to  $W$  is defined as

$$\ell^W(\mathcal{J}) := \max_{2 \leq k \leq p} \ell^W(J_{\geq k}).$$

For  $\delta \in \mathbb{R}_{++}^n$ , we use  $\ell^{W,\delta}(I) := \ell^{\text{Diag}(\delta)W}(I)$  and  $\ell^{W,\delta}(\mathcal{J}) := \ell^{\text{Diag}(\delta)W}(\mathcal{J})$ . When the context is clear, we omit  $W$  and write  $\ell^\delta(I) := \ell^{W,\delta}(I)$  and  $\ell^\delta(\mathcal{J}) := \ell^{W,\delta}(\mathcal{J})$ .

The following important duality claim asserts that the lifting score of a layering equals the lifting score of the reverse layering in the orthogonal complement subspace. It is an immediate consequence of Proposition 3.8.

**Lemma 3.9** *Let  $W \subseteq \mathbb{R}^n$  be a linear subspace,  $\delta \in \mathbb{R}_{++}^n$ . For an ordered partition  $\mathcal{J} = (J_1, J_2, \dots, J_p)$ , let  $\bar{\mathcal{J}} = (J_p, J_{p-1}, \dots, J_1)$  denote the reverse ordered partition. Then, we have*

$$\ell^{W,\delta}(\mathcal{J}) = \ell^{W^\perp, \delta^{-1}}(\bar{\mathcal{J}}).$$

**Proof** Let  $U = \text{Diag}(\delta)W$ . Note that  $U^\perp = \text{Diag}(\delta^{-1})W^\perp$ . Then by Proposition 3.8, for  $2 \leq k \leq p$ , we have that

$$\ell^{W,\delta}(J_{\geq k}) = \ell^U(J_{\geq k}) = \ell^{U^\perp}(J_{\leq k-1}) = \ell^{U^\perp}(\bar{J}_{\geq p-k+2}) = \ell^{W^\perp, \delta^{-1}}(\bar{J}_{\geq p-k+2}).$$

In particular,  $\ell^{W,\delta}(\mathcal{J}) = \ell^{W^\perp,\delta^{-1}}(\bar{\mathcal{J}})$ , as needed. □

The next lemma summarizes key properties of the LLS steps, assuming the partition has a small lifting score. We show that if  $\ell^\delta(\mathcal{J})$  is sufficiently small, then on the one hand, the LLS step will be very close to the affine scaling step, and on the other hand, on each layer  $k \in [p]$ , it will be very close to the affine scaling step restricted to this layer for the subspace  $W_{\mathcal{J},k}$ . The proof is deferred to Sect. 5.

**Lemma 3.10** (Proof on p. 46) *Let  $w = (x, y, s) \in \mathcal{N}(\beta)$  for  $\beta \in (0, 1/4]$ , let  $\mu = \mu(w)$  and  $\delta = \delta(w)$ . Let  $\mathcal{J} = (J_1, \dots, J_p)$  be a layering with  $\ell^\delta(\mathcal{J}) \leq \beta/(32n^2)$ , and let  $\Delta w^{\text{ll}} = (\Delta x^{\text{ll}}, \Delta y^{\text{ll}}, \Delta s^{\text{ll}})$  denote the LLS direction for the layering  $\mathcal{J}$ . Let furthermore  $\epsilon^{\text{ll}}(w) = \max_{i \in [n]} \min\{|Rx_i^{\text{ll}}|, |Rs_i^{\text{ll}}|\}$ , and define the maximal step length as*

$$\alpha^* := \sup\{\alpha' \in [0, 1] : \forall \bar{\alpha} \in [0, \alpha'] : w + \bar{\alpha} \Delta w^{\text{ll}} \in \mathcal{N}(2\beta)\}.$$

Then the following properties hold.

(i) We have

$$\|\delta_{J_k} \Delta x_{J_k}^{\text{ll}} + \delta_{J_k}^{-1} \Delta s_{J_k}^{\text{ll}} + x_{J_k}^{1/2} s_{J_k}^{1/2}\| \leq 6n \ell^\delta(\mathcal{J}) \sqrt{\mu}, \quad \forall k \in [p], \text{ and} \quad (27)$$

$$\|\delta \Delta x^{\text{ll}} + \delta^{-1} \Delta s^{\text{ll}} + x^{1/2} s^{1/2}\| \leq 6n^{3/2} \ell^\delta(\mathcal{J}) \sqrt{\mu}. \quad (28)$$

(ii) For the affine scaling direction  $\Delta w^{\text{a}} = (\Delta x^{\text{a}}, \Delta y^{\text{a}}, \Delta s^{\text{a}})$ ,

$$\|Rx^{\text{ll}} - Rx^{\text{a}}\|, \|Rs^{\text{ll}} - Rs^{\text{a}}\| \leq 6n^{3/2} \ell^\delta(\mathcal{J}).$$

(iii) For the residuals of the LLS steps we have  $\|Rx^{\text{ll}}\|, \|Rs^{\text{ll}}\| \leq \sqrt{2n}$ . For each  $i \in [n]$ ,  $\max\{|Rx_i^{\text{ll}}|, |Rs_i^{\text{ll}}|\} \geq \frac{1}{2} - \frac{3}{4}\beta$ .

(iv) We have

$$\alpha^* \geq 1 - \frac{3\sqrt{n}\epsilon^{\text{ll}}(w)}{\beta}, \quad (29)$$

and for any  $\alpha \in [0, 1]$

$$\mu(w + \alpha \Delta w^{\text{ll}}) = (1 - \alpha)\mu,$$

(v) We have  $\epsilon^{\text{ll}}(w) = 0$  if and only if  $\alpha^* = 1$ . These are further equivalent to  $w + \Delta w^{\text{ll}} = (x + \Delta x^{\text{ll}}, y + \Delta y^{\text{ll}}, s + \Delta s^{\text{ll}})$  being an optimal solution to (LP).

### 3.5 The layering procedure

Our algorithm performs LLS steps on a layering with a low lifting score. A further requirement is that within each layer, the circuit imbalances  $\kappa_{ij}^\delta$  defined in (6) are suitably bounded. The rescaling here is with respect to  $\delta = \delta(w)$  for the current iterate

$w = (x, y, s)$ . To define the precise requirement on the layering, we first introduce an auxiliary graph. Throughout we use the parameter

$$\gamma := \frac{\beta}{2^{10}n^5}. \quad (30)$$

**The auxiliary graph** For a vector  $\delta \in \mathbb{R}_{++}^n$  and  $\sigma > 0$ , we define the directed graph  $G_{\delta, \sigma} = ([n], E_{\delta, \sigma})$  such that  $(i, j) \in E_{\delta, \sigma}$  if  $\kappa_{ij}^\delta \geq \sigma$ . This is a subgraph of the *circuit ratio digraph* studied in Sect. 2, including only the edges where the circuit ratio is at least the threshold  $\sigma$ . Note that we do not have direct access to this graph, as we cannot efficiently compute the values  $\kappa_{ij}^\delta$ .

At the beginning of the entire algorithm, we run the subroutine `FIND-CIRCUITS(A)` as in Theorem 2.14, where  $W = \text{Ker}(A)$ . We assume the matroid  $\mathcal{M}(A)$  is non-separable. For a separable matroid, we can solve the subproblems of our LP on the components separately. Thus, for each  $i \neq j$ ,  $i, j \in [n]$ , we obtain an estimate  $\hat{\kappa}_{ij} \leq \kappa_{ij}$ . These estimates will be gradually improved throughout the algorithm.

Note that  $\kappa_{ij}^\delta = \kappa_{ij}\delta_j/\delta_i$  and  $\hat{\kappa}_{ij}^\delta = \hat{\kappa}_{ij}\delta_j/\delta_i$ . If  $\hat{\kappa}_{ij}^\delta \geq \sigma$ , then we are guaranteed  $(i, j) \in E_{\delta, \sigma}$ .

**Definition 3.11** Define  $\hat{G}_{\delta, \sigma} = ([n], \hat{E}_{\delta, \sigma})$  to be the directed graph with edges  $(i, j)$  such that  $\hat{\kappa}_{ij}^\delta \geq \sigma$ ; clearly,  $\hat{G}_{\delta, \sigma}$  is a subgraph of  $G_{\delta, \sigma}$ .

**Lemma 3.12** Let  $\delta \in \mathbb{R}_{++}^n$ . For every  $i \neq j$ ,  $i, j \in [n]$ ,  $\hat{\kappa}_{ij}^\delta \cdot \hat{\kappa}_{ji}^\delta \geq 1$ . Consequently, for any  $0 < \sigma \leq 1$ , at least one of  $(i, j) \in \hat{E}_{\delta, \sigma}$  or  $(j, i) \in \hat{E}_{\delta, \sigma}$ .

**Proof** We show that this property holds at the initialization. Since the estimates can only increase, it remains true throughout the algorithm. Recall the definition of  $\hat{\kappa}_{ij}$  from Theorem 2.14. This is defined as the maximum of  $|g_j/g_i|$  such that  $g \in W$ ,  $\text{supp}(g) = C$  for some  $C \in \hat{\mathcal{C}}$  containing  $i$  and  $j$ . For the same vector  $g$ , we get  $\hat{\kappa}_{ji} \geq |g_i/g_j|$ . Consequently,  $\hat{\kappa}_{ij} \cdot \hat{\kappa}_{ji} \geq 1$ , and also  $\hat{\kappa}_{ij}^\delta \cdot \hat{\kappa}_{ji}^\delta \geq 1$ . The second claim follows by the assumption  $\sigma \leq 1$ .  $\square$

**Balanced layerings** We are ready to define the requirements on the layering in the algorithm. In the algorithm,  $\delta = \delta(w)$  will correspond to the scaling of the current iterate  $w = (x, y, s)$ .

**Definition 3.13** Let  $\delta \in \mathbb{R}_{++}^n$ . The layering  $\mathcal{J} = (J_1, J_2, \dots, J_p)$  of  $[n]$  is  $\delta$ -balanced if

- (i)  $\ell^\delta(\mathcal{J}) \leq \gamma$ , and
- (ii)  $J_k$  is strongly connected in  $G_{\delta, \gamma/n}$  for all  $k \in [p]$ .

The following lemma shows that within each layer, the  $\kappa_{ij}^\delta$  values are within a bounded range. This will play an important role in our potential analysis.

**Lemma 3.14** Let  $0 < \sigma < 1$  and  $t > 0$ , and  $i, j \in [n]$ ,  $i \neq j$ .

(i) If the graph  $G_{\delta,\sigma}$  contains a directed path of at most  $t - 1$  edges from  $j$  to  $i$ , then

$$\kappa_{ij}^\delta < \left(\frac{\kappa^*}{\sigma}\right)^t.$$

(ii) If  $G_{\delta,\sigma}$  contains a directed path of at most  $t - 1$  edges from  $i$  to  $j$ , then

$$\kappa_{ij}^\delta > \left(\frac{\sigma}{\kappa^*}\right)^t.$$

**Proof** For part (i), let  $j = i_1, i_2, \dots, i_h = i$  be a path in  $G_{\delta,\sigma}$  in  $J$  from  $j$  to  $i$  with  $h \leq t$ . That is,  $\kappa_{i_\ell i_{\ell+1}}^\delta \geq \sigma$  for each  $\ell \in [h]$ . Theorem 2.12 yields

$$(\bar{\kappa}^*)^t \geq \kappa_{ij}^\delta \cdot \sigma^{h-1} > \kappa_{ij}^\delta \cdot \sigma^t,$$

since  $h \leq t$  and  $\sigma < 1$ . Part (ii) follows using part (i) for  $j$  and  $i$ , and that  $\kappa_{ij}^\delta \cdot \kappa_{ji}^\delta \geq 1$  according to Lemma 3.12.  $\square$

**Description of the layering subroutine** Consider an iterate  $w = (x, y, s) \in \mathcal{N}(\beta)$  of the algorithm with  $\delta = \delta(w)$ . The subroutine  $\text{LAYERING}(\delta, \hat{\kappa})$ , described in Algorithm 1, constructs a  $\delta$ -balanced layering. We recall that the approximated auxilliary graph  $\hat{G}_{\delta,\gamma/n}$  with respect to  $\hat{\kappa}$  is as in Definition 3.11

---

**Algorithm 1:**  $\text{LAYERING}(\delta, \hat{\kappa})$

---

**Input** :  $\delta \in \mathbb{R}_{++}^n$  and  $\hat{\kappa} \in \mathbb{R}_{++}^E$ .  
**Output:**  $\delta$ -balanced layering  $\mathcal{J} = (J_1, \dots, J_p)$  and updated values  $\hat{\kappa} \in \mathbb{R}_{++}^E$ .

- 1 Compute the strongly connected components  $C_1, C_2, \dots, C_\ell$  of  $\hat{G}_{\delta,\gamma/n}$ , listed in the ordering imposed by  $\hat{G}_{\delta,\gamma/n}$ ;
- 2  $\bar{E} \leftarrow \hat{E}_{\delta,\gamma/n}$ ;
- 3 **for**  $k = 2, \dots, \ell$  **do**
- 4     Call  $\text{VERIFY-LIFT}(\text{Diag}(\delta)W, C_{\geq k}, \gamma)$  that answers ‘pass’ or ‘fail’;
- 5     **if** the answer is ‘fail’ **then**
- 6         Let  $i \in C_{\geq k}, j \in C_{<k}$ , and  $t$  be the output of  $\text{VERIFY-LIFT}$  such that  $\gamma/n \leq t \leq \kappa_{ij}^\delta$ ;
- 7          $\hat{\kappa}_{ij} \leftarrow t\delta_i/\delta_j$ ;
- 8          $\bar{E} \leftarrow \bar{E} \cup \{(i, j)\}$ ;
- 9 Compute strongly connected components  $J_1, J_2, \dots, J_p$  of  $([n], \bar{E})$ , listed in the ordering imposed by  $\hat{G}_{\delta,\gamma/n}$ ;
- 10 **return**  $\mathcal{J} = (J_1, J_2, \dots, J_p), \hat{\kappa}$ .

---

We now give an overview of the subroutine  $\text{LAYERING}(\delta, \hat{\kappa})$ . We start by computing the strongly connected components (SCCs) of the directed graph  $\hat{G}_{\delta,\gamma/n}$ . The edges of this graph are obtained using the current estimates  $\hat{\kappa}_{ij}^\delta$ . According to Lemma 3.12, we have  $(i, j) \in \hat{E}_{\delta,\gamma/n}$  or  $(j, i) \in \hat{E}_{\delta,\gamma/n}$  for every  $i, j \in [n], i \neq j$ . Hence, there is a

linear ordering of the components  $C_1, C_2, \dots, C_\ell$  such that  $(u, v) \in \hat{E}_{\delta, \gamma/n}$  whenever  $u \in C_i, v \in C_j$ , and  $i < j$ . We call this the ordering imposed by  $\hat{G}_{\delta, \gamma/n}$ .

Next, for each  $k = 2, \dots, \ell$ , we use the subroutine VERIFY-LIFT(Diag( $\delta$ ) $W, C_{\geq k}, \gamma$ ) described in Lemma 2.11. If the subroutine returns ‘pass’, then we conclude  $\ell^\delta(C_{\geq k}) \leq \gamma$ , and proceed to the next layer. If the answer is ‘fail’, then the subroutine returns as certificates  $i \in C_{\geq k}, j \in C_{<k}$ , and  $t$  such that  $\gamma/n \leq t \leq \kappa_{ij}^\delta$ . In this case, we update  $\hat{\kappa}_{ij}^\delta$  to the higher value  $t$ . We add  $(i, j)$  to an edge set  $\bar{E}$ ; this edge set was initialized to contain  $\hat{E}_{\delta, \gamma/n}$ . After adding  $(i, j)$ , all components  $C_\ell$  between those containing  $i$  and  $j$  will be merged into a single strongly connected component. To see this, recall that if  $i' \in C_\ell$  and  $j' \in C_{\ell'}$  for  $\ell < \ell'$ , then  $(i', j') \in \hat{E}_{\delta, \gamma/n}$  according to Lemma 3.12.

Finally, we compute the strongly connected components of  $([n], \bar{E})$ . We let  $J_1, J_2, \dots, J_p$  denote their unique acyclic order, and return these layers.

**Lemma 3.15** *The subroutine LAYERING( $\delta, \hat{\kappa}$ ) returns a  $\delta$ -balanced layering in  $O(nm^2 + n^2)$  time.*

The difficult part of the proof is showing the running time bound. We note that the weaker bound  $O(n^2m^2)$  can be obtained by a simpler argument.

**Proof** We first verify that the output layering is indeed  $\delta$ -balanced. For property (i) of Definition 3.13, note that each  $J_q$  component is the union of some of the  $C_k$ ’s. In particular, for every  $q \in [p]$ , the set  $J_{\geq q} = C_{\geq k}$  for some  $k \in [\ell]$ . Assume now  $\ell^\delta(C_{\geq k}) > \gamma$ . At step  $k$  of the main cycle, the subroutine VERIFY-LIFT returned the answer ‘fail’, and a new edge  $(i, j) \in E$  was added with  $i \in C_{\geq k}, j \in C_{<k}$ . Note that we already had  $(j, i) \in \hat{E}_{\delta, \gamma/n}$ , since  $j \in C_r$  for some  $r < k$ , and  $i \in C_{r'}$  for  $r' \geq k$ . This contradicts the choice of  $J_{\geq q}$  as a maximal strongly connected component in  $([n], E)$ .

Property (ii) follows since all new edges added to  $E$  have  $\kappa_{ij} \geq \gamma/n$ . Therefore,  $([n], E)$  is a subgraph of  $G_{\delta, \gamma/n}$ .

Let us now turn to the computational cost. The initial strongly-connected components can be obtained in time  $O(n^2)$ , and the same bound holds for the computation of the final components. (The latter can be also done in linear time, exploiting the special structure that the components  $C_i$  have a complete linear ordering.)

The second computational bottleneck is the subroutine VERIFY-LIFT. We assume a matrix  $M \in \mathbb{R}^{n \times (n-m)}$  is computed at the very beginning such that  $\text{range}(M) = W$ . We first explain how to implement one call to VERIFY-LIFT in  $O(n(n-m)^2)$  time. We then sketch how to amortize the work across the different calls to VERIFY-LIFT, using the nested structure of the layering, to implement the whole procedure in  $O(n(n-m)^2)$  time. To turn this into  $O(nm^2)$ , we recall that the layering procedure is the same for  $W$  and  $W^\perp$  due to duality (Proposition 3.8). Since  $\dim(W^\perp) = m$ , applying this subroutine on  $W^\perp$  instead of  $W$  achieves the same result but in time  $O(nm^2)$ .

We now explain the implementation of VERIFY-LIFT, where we are given as input  $C \subseteq [n]$  and the basis matrix  $M \in \mathbb{R}^{n \times (n-m)}$  as above with  $\text{range}(M) = W$ . Clearly, the running time is dominated by the computation of the set  $I \subseteq C$  and the matrix  $B \in \mathbb{R}^{([n] \setminus C) \times |I|}$  satisfying  $L_C^W(x)_{[n] \setminus C} = Bx_I$ , for  $x \in \pi_C(W)$ . We explain how to compute  $I$  and  $B$  from  $M$  using column operations (note that these preserve the



range). The valid choices for  $I \subseteq C$  are in correspondence with maximal sets of linear independent rows of  $M_{C,\bullet}$ , noting then that  $|I| = r$  where  $r := \text{rk}(M_{C,\bullet})$ . Let  $D_1 = [n - m - r]$  and  $D_2 = [n - m] \setminus [n - m - r]$ . By applying column operations to  $M$ , we can compute  $I \subseteq C$  such that  $M_{I,D_2} = \mathbf{I}_r$  ( $r \times r$  identity) and  $M_{C,D_1} = 0$ . This requires  $O(n(n - m)|C|)$  time using Gaussian elimination. At this point, note that  $\pi_C(W) = \text{range}(M_{C,D_2})$ ,  $\pi_I(W) = \mathbb{R}^I$  and  $\text{range}(M_{\bullet,D_1}) = W \cap \mathbb{R}^n_{[n] \setminus C}$ . To compute  $B$ , we must transform the columns of  $M_{\bullet,D_2}$  into minimum norm lifts of  $e_i \in \pi_I(W)$  into  $W$ , for all  $i \in I$ . For this purpose, it suffices to make the columns of  $M_{[n] \setminus C, D_2}$  orthogonal to the range of  $M_{[n] \setminus C, D_1}$ . Applying Gram-Schmidt orthogonalization, this requires  $O((n - |C|)(n - m)(n - m - r))$  time. From here, the desired matrix  $B = M_{[n] \setminus C, D_2}$ . Thus, the total running time of VERIFY- LIFT is  $O(n(n - m)|C| + (n - |C|)(n - m)(n - m - r)) = O(n(n - m)^2)$ .

We now sketch how to amortize the work of all the calls of VERIFY- LIFT during the layering algorithm, to achieve a total  $O(n(n - m)^2)$  running time. Let  $C_1, \dots, C_\ell$  denote the candidate SCC layering. Our task is to compute the matrices  $B_k, 2 \leq k \leq \ell$ , needed in the calls to VERIFY- LIFT on  $W, C_{\geq k}, 2 \leq k \leq \ell$ , in total  $O(n(n - m)^2)$  time. We achieve this in three steps working with the basis matrix  $M$  as above. Firstly, by applying column operations to  $M$ , we compute sets  $I_k \subseteq C_k$  and  $D_k = [I_{\leq k}] \setminus [I_{< k}]$ ,  $k \in [\ell]$ , such that  $M_{I_k, D_k} = \mathbf{I}_{r_k}$ , where  $r_k = |I_k|$ , and  $M_{C_{\geq k}, D_{< k}} = 0, 2 \leq k \leq \ell$ . Note that this enforces  $\sum_{k=1}^\ell r_k = (n - m)$ . This computation requires  $O(n(n - m)^2)$  time using Gaussian elimination. This computation achieves  $\text{range}(M_{C_k, D_k}) = \pi_{C_k}(W \cap \mathbb{R}^n_{C_{< k}})$ ,  $\text{range}(M_{C_{\geq k}, D_{\geq k}}) = \pi_{C_{\geq k}}(W)$  and  $\text{range}(M_{\bullet, D_{\leq k}}) = W \cap \mathbb{R}^n_{C_{\leq k}}$ , for all  $k \in [\ell]$ .

From here, we block orthogonalize  $M$ , such that the columns of  $M_{\bullet, D_k}$  are orthogonal to the range of  $M_{\bullet, D_{< k}}, 2 \leq k \leq \ell$ . Applying an appropriately adapted Gram-Schmidt orthogonalization, this requires  $O(n(n - m)^2)$  time. Note that this operation maintains  $M_{I_k, D_k} = \mathbf{I}_{r_k}, k \in [\ell]$ , since  $M_{C_{\geq k}, D_{< k}} = 0$ . At this point, for  $k \in [\ell]$  the columns of  $M_{\bullet, D_k}$  are in correspondence with minimum norm lifts of  $e_i \in \pi_{D_{\geq k}}(W)$  into  $W$ , for all  $i \in I_k$ . Note that to compute the matrix  $B_k$  we need the lifts of  $e_i \in \pi_{D_{\geq k}}(W)$ , for all  $i \in I_{\geq k}$  instead of just  $i \in I_k$ .

We now compute the matrices  $B_\ell, \dots, B_2$  in this order via the following iterative procedure. Let  $k$  denote the iteration counter, which decrements from  $\ell$  to 2. For  $k = \ell$  (first iteration), we let  $B_\ell = M_{C_{< \ell}, D_\ell}$  and decrement  $k$ . For  $k < \ell$ , we eliminate the entries of  $M_{I_k, D_{> k}}$  by using the columns of  $M_{\bullet, D_k}$ . We then let  $B_k = M_{C_{< k}, D_{\geq k}}$  and decrement  $k$ . To justify correctness, one only has to notice that at the end of iteration  $k$ , we maintain the orthogonality of  $M_{\bullet, D_{\geq k}}$  to the range of  $M_{\bullet, D_{< k}}$  and that  $M_{I_{\geq k}, D_{\geq k}} = \mathbf{I}_{|I_{\geq k}|}$  is the appropriate identity. The cost of this procedure is the same as a full run of Gaussian elimination and thus is bounded by  $O(n(n - m)^2)$ . The calls to VERIFY- LIFT during the layering procedure can thus be executed in  $O(n(n - m)^2)$  amortized time as claimed.  $\square$

### 3.6 The overall algorithm

Algorithm 2 presents the overall algorithm LP- SOLVE( $A, b, c, w^0$ ). We assume that an initial feasible solution  $w^0 = (x^0, y^0, s^0) \in \mathcal{N}(\beta)$  is given. We address this in Sect. 7, by adapting the extended system used in [63]. We note that this subroutine

**Algorithm 2:** LP- SOLVE( $A, b, c, w^0$ )

---

**Input** :  $A \in \mathbb{R}^{m \times n}$ ,  $b \in \mathbb{R}^m$ ,  $c \in \mathbb{R}^n$ , and an initial feasible solution  $w^0 = (x^0, y^0, s^0) \in \mathcal{N}(1/8)$  to (LP).

**Output**: Optimal solution  $w^* = (x^*, y^*, s^*)$  to (LP).

- 1 Call FIND- CIRCUITS( $A$ ) to obtain the lower bounds  $\hat{\kappa}_{ij}$  for each  $i, j \in [n]$ ,  $i \neq j$ ;
- 2  $k \leftarrow 0$ ,  $\alpha \leftarrow 0$ ;
- 3 **repeat**
- 4   /\* Predictor step \*/
- 5   Compute affine scaling direction  $\Delta w^a = (\Delta x^a, \Delta y^a, \Delta s^a)$  for  $w$ ;
- 6   **if**  $\epsilon^a(w) < 10n^{3/2}\gamma$  **then** // Recall  $\epsilon^a(w)$  defined in (20)
- 7      $\delta \leftarrow (s^k)^{1/2}(x^k)^{-1/2}$ ;
- 8      $(\mathcal{J}, \hat{\kappa}) \leftarrow \text{LAYERING}(\delta, \hat{\kappa})$ ;
- 9     Compute Layered Least Squares direction  $\Delta w^{\text{ll}} = (\Delta x^{\text{ll}}, \Delta y^{\text{ll}}, \Delta s^{\text{ll}})$  for the layering  $\mathcal{J}$  and  $w$ ;
- 10     $\Delta w \leftarrow \Delta w^{\text{ll}}$ ;
- 11     $\alpha \leftarrow 1 - 24\sqrt{n}\epsilon^{\text{ll}}(w)$ ; // As in Lemma 3.10(ii)
- 12   **else**
- 13      $\Delta w \leftarrow \Delta w^a$ ;
- 14      $\alpha \leftarrow \min\{1/(8\sqrt{n}), 1 - 8\|\Delta x^a \Delta s^a\|/\mu(w)\}$ ; // As in Proposition 3.4(ii)
- 15     $w' \leftarrow w^k + \alpha \Delta w$ ;
- 16    /\* Corrector step \*/
- 17    Compute centrality direction  $\Delta w^c = (\Delta x^c, \Delta y^c, \Delta s^c)$  for  $w'$ ;
- 18     $w^{k+1} \leftarrow w' + \Delta w^c$ ;
- 19     $k \leftarrow k + 1$ ;
- 20 **until**  $\mu(w^k) = 0$ ;
- 21 **return**  $w^k = (x^k, y^k, s^k)$ .

---

requires an upper bound on  $\bar{\chi}^*$ . Since computing  $\bar{\chi}^*$  is hard, we can implement it by a doubling search on  $\log \bar{\chi}^*$ , as explained in Sect. 7. Other than for initialization, the algorithm does not require an estimate on  $\bar{\chi}^*$ .

The algorithm starts with the subroutine FIND- CIRCUITS( $A$ ) as in Theorem 2.14. The iterations are similar to the MTY Predictor–Corrector algorithm [39]. The main difference is that certain affine scaling steps are replaced by LLS steps. In every predictor step, we compute the affine scaling direction, and consider the quantity  $\epsilon^a(w) = \max_{i \in [n]} \min\{|Rx_i^a|, |Rs_i^a|\}$ . If this is above the threshold  $10n^{3/2}\gamma$ , then we perform the affine scaling step. However, in case  $\epsilon^a(w) < 10n^{3/2}\gamma$ , we use the LLS direction instead. In each such iteration, we call the subroutine LAYERING( $\delta, \hat{\kappa}$ ) (Algorithm 1) to compute the layers, and we compute the LLS step for this layering.

Another important difference is that the algorithm does not require a final rounding step. It terminates with the exact optimal solution  $w^*$  once a predictor step is able to perform a full step with  $\alpha = 1$ .

**Theorem 3.16** *For given  $A \in \mathbb{R}^{m \times n}$ ,  $b \in \mathbb{R}^m$ ,  $c \in \mathbb{R}^n$ , and an initial feasible solution  $w^0 = (x^0, y^0, s^0) \in \mathcal{N}(1/8)$ , Algorithm 2 finds an optimal solution to (LP) in  $O(n^{2.5} \log n \log(\bar{\chi}_A^* + n))$  iterations.*

**Remark 3.17** Whereas using LLS steps enables us to give a strong bound on the total number of iterations, finding LLS directions has a significant computational overhead

as compared to finding affine scaling directions. The layering  $\mathcal{J}$  can be computed in time  $O(nm^2)$  (Lemma 3.15), and the LLS steps also require  $O(nm^2)$  time, see [35, 63]. This is in contrast to the computational cost  $O(n^\omega)$  of an affine scaling direction. Here  $\omega < 2.373$  is the matrix multiplication constant [62].

We now sketch a possible approach to amortize the computational cost of the LLS steps over the sequence of affine scaling steps. It was shown in [37] that for the MTY P-C algorithm, the “bad” scenario between two crossover events amounts to a series of affine scaling steps where the progress in  $\mu$  increases exponentially from every iteration to the next. This corresponds to the term  $O(\min\{n^2 \log \log(\mu_0/\eta), \log(\mu_0/\eta)\})$  in their running time analysis. Roughly speaking, such a sequence of affine scaling steps indicates that an LLS step is necessary.

Hence, we could observe these accelerating sequences of affine scaling steps, and perform an LLS step after we see a sequence of length  $O(\log n)$ . The progress made by these affine scaling steps offsets the cost of computing the LLS direction.

#### 4 The potential function and the overall analysis

Let  $\mu > 0$  and  $\delta(\mu) = s(\mu)^{1/2}x(\mu)^{-1/2} = \sqrt{\mu}/x(\mu) = s(\mu)/\sqrt{\mu}$  correspond to the point on the central path and recall the definition of  $\gamma$  in (30). For  $i, j \in [n], i \neq j$ , we define

$$\rho^\mu(i, j) := \frac{\log \kappa_{ij}^{\delta(\mu)}}{\log(4n\kappa_W^*/\gamma)},$$

and the main potentials in the algorithm as

$$\Psi^\mu(i, j) := \max \left\{ 1, \min \left\{ 2n, \inf_{0 < \mu' < \mu} \rho^{\mu'}(i, j) \right\} \right\} \quad \text{and}$$

$$\Psi(\mu) := \sum_{i, j \in [n], i \neq j} \log \Psi^\mu(i, j).$$

The motivation for  $\rho^\mu(i, j)$  and  $\Psi^\mu(i, j)$  comes from Lemma 3.14, using  $\sigma = \gamma/(4n)$ . Thus,  $\log \kappa_{ij}^{\delta(\mu)} / \log(4n\kappa_W^*/\gamma)$  can be seen as a lower bound on the length of the shortest  $j-i$  path. Recall that the layers are defined as strongly connected components of  $\hat{G}_{\delta, \gamma/n}$ , which is a subgraph of  $G_{\delta(\mu), \gamma/(4n)}$  (using the bound (16)). Consequently, whenever  $\rho^\mu(i, j) \geq n$ , the nodes  $i$  and  $j$  cannot be in the same strongly connected component for the normalized duality gap  $\mu$ . Thus, our potentials  $\Psi^\mu(i, j)$  can be seen as fine-grained analogues of the crossover events analyzed in [36, 37, 63]: the definition of  $\Psi^\mu(i, j)$  contains a minimization over  $0 < \mu' < \mu$ ; therefore,  $\Psi^\mu(i, j) > n$  implies that  $i$  and  $j$  may never appear on the same layer for any  $\mu' \leq \mu$ . On the other hand, these potentials are more fine-grained: even for  $t < n$ , if  $\Psi^\mu(i, j) \geq t$  then whenever a layer contains both  $i$  and  $j$  for  $\mu' \leq \mu$ , this layer must have size  $\geq t$ .

By definition, for all pairs  $(i, j) \in [n] \times [n]$  we have  $\Psi^{\mu'}(i, j) \geq \Psi^\mu(i, j)$  for  $0 < \mu' \leq \mu$ ; and we enforce  $\Psi^\mu(i, j) \in [1, 2n]$ . The upper bound can be imposed

since values  $\Psi^{\mu'}(i, j) \geq n$  do not yield any new information on the layering. Hence, the overall potential  $\Psi(\mu)$  is between 0 and  $O(n^2 \log n)$ . The overall analysis in the proof of Theorem 3.16 divides the iterations into phases. In each phase, we can identify a set  $J \subseteq [n]$ ,  $|J| > 1$  arising as a layer or as the union of two layers in the LLS step at the beginning of the phase. We show that  $\Psi^{\mu}(i, j)$  doubles for at least  $|J| - 1$  pairs  $(i, j) \in J \times J$  during the subsequent  $O(\sqrt{n}|J| \log(\bar{\chi}^* + n))$  iterations; consequently,  $\Psi(\mu)$  increases by at least  $|J| - 1$  during these iterations. This leads to the overall iteration bound  $O(n^{2.5} \log(n) \log(\bar{\chi}^* + n))$ . In comparison, the crossover analysis would correspond to showing that within  $O(n^{1.5} \log(\bar{\chi}^* + n))$  iterations, one of the  $\Psi^{\mu}(i, j)$  values previously  $< n$  becomes larger than  $n$ . The following statement formalizes the above mentioned properties of  $\Psi^{\mu}(i, j)$ .

**Lemma 4.1** *Let  $w = (x, y, s) \in \mathcal{N}(\beta)$  for  $\beta \in (0, 1/4]$ . Let  $i, j \in [n]$ ,  $i \neq j$ , and let  $\mu = \mu(w)$ .*

1. *If  $\hat{G}_{\delta, \gamma/n}$  contains a path from  $j$  to  $i$  of at most  $t - 1$  edges, then  $\rho^{\mu}(i, j) < t$ .*
2. *If  $\hat{G}_{\delta, \gamma/n}$  contains a path from  $i$  to  $j$  of at most  $t - 1$  edges, then  $\rho^{\mu}(i, j) > -t$ .*
3. *If  $\Psi^{\mu}(i, j) \geq t$ , then in any  $\delta(w')$ -balanced layering, where  $w' = (x', y', s') \in \mathcal{N}(\beta)$  with  $\mu(w') \leq \mu$ ,*
  - *$i$  and  $j$  cannot be together on a layer of size at most  $t$ , and*
  - *$j$  cannot be on a layer preceding the layer containing  $i$ .*

**Proof** From (16), we see that for any  $i, j$ ,

$$\hat{\kappa}_{ij}^{\delta} \leq \kappa_{ij}^{\delta} \leq (1 - 2\beta)^{-1} \kappa_{ij}^{\delta(\mu)} \leq 4\kappa_{ij}^{\delta(\mu)}.$$

Consequently,  $\hat{G}_{\delta, \gamma/n}$  is a subgraph of  $G_{\delta(\mu), \gamma/(4n)}$ . The statement now follows from Lemma 3.14 with  $\sigma = \gamma/(4n)$ . □

In what follows, we formulate four important lemmas crucial for the proof of Theorem 3.16. For the lemmas, we only highlight some key ideas here, and defer the full proofs to Sect. 6.

For a triple  $w \in \mathcal{N}(\beta)$ ,  $\Delta w^{\parallel}$  refers to the LLS direction found in the algorithm, and  $Rx^{\parallel}$  and  $Rs^{\parallel}$  denote the residuals as in (18). For a subset  $I \subset [n]$  recall the definition

$$\epsilon_I^{\parallel}(w) := \max_{i \in I} \min\{|Rx_i^{\parallel}|, |Rs_i^{\parallel}|\}.$$

We introduce another important quantity  $\xi$  for the analysis:

$$\xi_I^{\parallel}(w) := \min\{\|Rx_I^{\parallel}\|, \|Rs_I^{\parallel}\|\} \tag{31}$$

for a subset  $I \subset [n]$ . For a layering  $\mathcal{J} = (J_1, J_2, \dots, J_p)$ , we let

$$\xi_{\mathcal{J}}^{\parallel}(w) = \max_{k \in [p]} \xi_{J_k}^{\parallel}(w).$$

The key idea of the analysis is to extract information about the optimal solution  $w^* = (x^*, y^*, s^*)$  from the LLS direction. The first main lemma shows that if  $\|Rx_{J_q}^{\text{ll}}\|$  is large on some layer  $J_q$ , then for at least one index  $i \in J_q$ ,  $x_i^*/x_i \geq 1/\text{poly}(n)$ , i.e., the variable  $x_i$  has “converged”. The analogous statement holds on the dual side for  $\|Rs_{J_q}^{\text{ll}}\|$  and an index  $j \in J_q$ .

**Lemma 4.2** (Proof in Sect. 6) *Let  $w = (x, y, s) \in \mathcal{N}(\beta)$  for  $\beta \in (0, 1/8]$  and let  $w^* = (x^*, y^*, s^*)$  be the optimal solution corresponding to  $\mu^* = 0$  on the central path. Let further  $\mathcal{J} = (J_1, \dots, J_p)$  be a  $\delta(w)$ -balanced layering (Definition 3.13), and let  $\Delta w^{\text{ll}} = (\Delta x^{\text{ll}}, \Delta y^{\text{ll}}, \Delta s^{\text{ll}})$  be the corresponding LLS direction. Then the following statement holds for every  $q \in [p]$ :*

(i) *There exists  $i \in J_q$  such that*

$$x_i^* \geq \frac{2x_i}{3\sqrt{n}} \cdot (\|Rx_{J_q}^{\text{ll}}\| - 2\gamma n). \tag{32}$$

(ii) *There exists  $j \in J_q$  such that*

$$s_j^* \geq \frac{2s_j}{3\sqrt{n}} \cdot (\|Rs_{J_q}^{\text{ll}}\| - 2\gamma n). \tag{33}$$

We outline the main idea of the proof of part (i); part (ii) follows analogously using the duality of the lifting scores (Lemma 3.9). On layer  $q$ , the LLS step minimizes  $\|\delta_{J_q}(x_{J_q} + \Delta x_{J_q})\|$ , subject to  $\Delta x_{J_{>q}} = \Delta x_{J_{>q}}^{\text{ll}}$  and subject to existence of  $\Delta x_{J_{<q}}$  such that  $\Delta x \in W$ . By making use of  $\ell^{\delta(w)}(J_{>q}) \leq \gamma$  due to  $\delta(w)$ -balancedness, we can show the existence of a point  $z \in W + x^*$  such that  $\|\delta_{J_q}(z_{J_q} - x_{J_q}^*)\|$  is small, and  $z_{J_{>q}} = x_{J_{>q}} + \Delta x_{J_{>q}}^{\text{ll}}$ . By the choice of  $\Delta x_{J_q}^{\text{ll}}$ , we have  $\|\delta_{J_q} z_{J_q}\| \geq \|\delta_{J_q}(x_{J_q} + \Delta x_{J_q}^{\text{ll}})\| = \sqrt{\mu} \|Rx_{J_q}^{\text{ll}}\|$ . Therefore,  $\|\delta_{J_q} x_{J_q}^*/\sqrt{\mu}\|$  cannot be much smaller than  $\|Rx_{J_q}^{\text{ll}}\|$ . Noting that  $\delta_{J_q} x_{J_q}^*/\sqrt{\mu} \approx x_{J_q}^*/x_{J_q}$ , we obtain a lower bound on  $x_i^*/x_i$  for some  $i \in J_q$ .

We emphasize that the lemma only shows the existence of such indices  $i$  and  $j$ , but does not provide an efficient algorithm to identify them. It is also useful to note that for any  $i \in [n]$ ,  $\max\{|Rx_i^{\text{ll}}|, |Rs_i^{\text{ll}}|\} \geq \frac{1}{2} - \frac{3}{4}\beta$  according to Lemma 3.10(iii). Thus, for each  $q \in [p]$ , we obtain a strong and positive lower bound in either case (i) on  $x_i/x_i^*$  or case (ii) on  $s_i/s_i^*$  for some  $i \in J_q$ .

The next lemma allows us to argue that the potential function  $\Psi(\cdot, \cdot)$  increases for multiple pairs of variables, if we have strong lower bounds on both  $x_i^*$  and  $s_j^*$  for some  $i, j \in [n]$ , along with a lower and upper bound on  $\rho^\mu(i, j)$ .

**Lemma 4.3** (Proof in Sect. 6) *Let  $w = (x, y, s) \in \mathcal{N}(2\beta)$  for  $\beta \in (0, 1/8]$ , let  $\mu = \mu(w)$  and  $\delta = \delta(w)$ . Let  $i, j \in [n]$  and  $2 \leq \tau \leq n$  such that for the optimal solution  $w^* = (x^*, y^*, s^*)$ , we have  $x_i^* \geq \beta x_i/(2^{10}n^{5.5})$  and  $s_j^* \geq \beta s_j/(2^{10}n^{5.5})$ , and assume  $\rho^\mu(i, j) \geq -\tau$ . After  $O(\beta^{-1}\sqrt{n}\tau \log(\bar{\chi}^* + n))$  further iterations the duality gap  $\mu'$  fulfills  $\Psi^{\mu'}(i, j) \geq 2\tau$ , and for every  $\ell \in [n] \setminus \{i, j\}$ , either  $\Psi^{\mu'}(i, \ell) \geq 2\tau$ , or  $\Psi^{\mu'}(\ell, j) \geq 2\tau$ .*

We note that  $i$  and  $j$  as in the lemma are necessarily different, since  $i = j$  would imply  $0 = x_i^* s_i^* \geq \beta^2 \mu / (2^{20} n^{11}) > 0$ .

Let us illustrate the idea of the proof of  $\Psi^{\mu'}(i, j) \geq 2\tau$ . For  $i$  and  $j$  as in the lemma, and for a central path element  $w' = w(\mu')$  for  $\mu' < \mu$ , we have  $x'_i \geq x_i^*/n \geq \beta x_i / (2^{10} n^{6.5})$  and  $s'_j \geq s_j^*/n \geq \beta s_j / (2^{10} n^{6.5})$  by the near-monotonicity of the central path (Lemma 3.3). Note that

$$\kappa_{ij}^{\delta'} = \kappa_{ij} \cdot \frac{\delta'_j}{\delta'_i} = \kappa_{ij} \cdot \frac{x'_i s'_j}{\mu'} \geq \kappa_{ij} \cdot \frac{\beta^2 x_i s_j}{2^{20} n^{13} \mu'} \geq \frac{\beta^2 (1 - \beta)^2}{2^{20} n^{13}} \cdot \kappa_{ij}^{\delta} \cdot \frac{\mu}{\mu'},$$

where the last inequality uses Proposition 3.2. Consequently, as  $\mu'$  sufficiently decreases,  $\kappa_{ij}^{\delta'}$  will become much larger than  $\kappa_{ij}^{\delta}$ . The claim on  $\ell \in [n] \setminus \{i, j\}$  can be shown by using the triangle inequality  $\kappa_{ik} \cdot \kappa_{kj} \geq \kappa_{ij}$  shown in Lemma 2.15.

Assume now  $\xi_{J_q}^{\text{ll}}(w) \geq 4\gamma n$  for some  $q \in [p]$  in the LLS step. Then, Lemma 4.2 guarantees the existence of  $i, j \in J_q$  such that  $x_i^*/x_i, s_j^*/s_j \geq \frac{4}{3\sqrt{n}}\gamma n > \beta / (2^{10} n^{5.5})$ . Further, Lemma 4.1 gives  $\rho^\mu(i, j) \geq -|J_q|$ . Hence, Lemma 4.3 is applicable for  $i$  and  $j$  with  $\tau = |J_q|$ .

The overall potential argument in the proof of Theorem 3.16 uses Lemma 4.3 in three cases:  $\xi_{\mathcal{J}}^{\text{ll}}(w) \geq 4\gamma n$  (Lemma 4.2 is applicable as above);  $\xi_{\mathcal{J}}^{\text{ll}}(w) < 4\gamma n$  and  $\ell^{\delta^+}(\mathcal{J}) \leq 4\gamma n$  (Lemma 4.4); and  $\xi_{\mathcal{J}}^{\text{ll}}(w) < 4\gamma n$  and  $\ell^{\delta^+}(\mathcal{J}) > 4\gamma n$  (Lemma 4.5). Here,  $\delta^+$  refers to the value of  $\delta$  after the LLS step. Note that  $\delta^+ > 0$  is well-defined, unless the algorithm terminated with an optimal solution.

To prove these lemmas, we need to study how the layers “move” during the LLS step. We let  $\mathbf{B} = \{t \in [n] : |R_s^{\text{ll}}| < 4\gamma n\}$  and  $\mathbf{N} = \{t \in [n] : |R_x^{\text{ll}}| < 4\gamma n\}$ . The assumption  $\xi_{\mathcal{J}}^{\text{ll}}(w) < 4\gamma n$  means that for each layer  $J_k$ , either  $J_k \subseteq \mathbf{B}$  or  $J_k \subseteq \mathbf{N}$ ; we accordingly refer to  $\mathbf{B}$ -layers and  $\mathbf{N}$ -layers.

**Lemma 4.4** (Proof in Sect. 6) *Let  $w = (x, y, s) \in \mathcal{N}(\beta)$  for  $\beta \in (0, 1/8]$ , and let  $\mathcal{J} = (J_1, \dots, J_p)$  be a  $\delta(w)$ -balanced partition. Assume that  $\xi_{\mathcal{J}}^{\text{ll}}(w) < 4\gamma n$ , and let  $w^+ = (x^+, y^+, s^+) \in \mathcal{N}(2\beta)$  be the next iterate obtained by the LLS step with  $\mu^+ = \mu(w^+)$  and assume  $\mu^+ > 0$ . Let  $q \in [p]$  such that  $\xi_{\mathcal{J}}^{\text{ll}}(w) = \xi_{J_q}^{\text{ll}}(w)$ . If  $\ell^{\delta^+}(\mathcal{J}) \leq 4\gamma n$ , then there exist  $i, j \in J_q$  such that  $x_i^* \geq \beta x_i^+ / (16n^{3/2})$  and  $s_j^* \geq \beta s_j^+ / (16n^{3/2})$ . Further, for any  $\ell, \ell' \in J_q$ , we have  $\rho^{\mu^+}(\ell, \ell') \geq -|J_q|$ .*

For the proof sketch, without loss of generality, let  $\xi_{\mathcal{J}}^{\text{ll}} = \xi_{J_q}^{\text{ll}} = \|R_x^{\text{ll}}\|$ , that is,  $J_q$  is an  $\mathbf{N}$ -layer. The case  $\xi_{J_q}^{\text{ll}} = \|R_s^{\text{ll}}\|$  can be treated analogously. Since the residuals  $\|R_x^{\text{ll}}\|$  and  $\|R_s^{\text{ll}}\|$  cannot be both small, Lemma 4.2 readily provides a  $j \in J_q$  such that  $s_j^*/s_j \geq 1/(6\sqrt{n})$ . Using Lemma 3.3 and Proposition 3.1,  $s_j^*/s_j^+ = s_j^*/s_j \cdot s_j/s_j^+ > (1 - \beta)/(6(1 + 4\beta)n^{3/2}) > \beta/(16n^{3/2})$ .

The key ideas of showing the existence of an  $i \in J_q$  such that  $x_i^* \geq x_i^+ / (16n^{3/2})$  are the following. With  $\approx, \lesssim$  and  $\gtrsim$ , we write equalities and inequalities that hold up to small polynomial factors. First, we show that (i)  $\|\delta_{J_q} x_{J_q}^+\| \lesssim \mu^+ / \sqrt{\mu}$ , and then, that (ii)  $\|\delta_{J_q} x_{J_q}^*\| \gtrsim \mu^+ / \sqrt{\mu}$ .

If we can show (i) and (ii) as above, we obtain that  $\|\delta_{J_q} x_{J_q}^*\| \gtrsim \|\delta_{J_q} x_{J_q}^+\|$ , and thus,  $x_i^* \gtrsim x_i^+$  for some  $i \in J_q$ .

Let us now sketch the first step. By the assumption  $J_q \subset N$ , one can show  $x_{J_q}^+ / x_{J_q} \approx \mu^+ / \mu$ , and therefore

$$\|\delta_{J_q} x_{J_q}^+\| \approx \frac{\mu^+}{\mu} \|\delta_{J_q} x_{J_q}\| \approx \frac{\mu^+}{\mu} \sqrt{\mu} = \frac{\mu^+}{\sqrt{\mu}}.$$

The second part of the proof, namely, lower bounding  $\|\delta_{J_q} x_{J_q}^*\|$ , is more difficult. Here, we only sketch it for the special case when  $J_q = [n]$ . That is, we have a single layer only; in particular, the LLS step is the same as the affine scaling step  $\Delta x^{\text{ll}} = \Delta x^a$ . The general case of multiple layers follows by making use of Lemma 3.10, i.e. exploiting that for a sufficiently small  $\ell^\delta(\mathcal{J})$ , the LLS step is close to the affine scaling step.

Hence, assume that  $\Delta x^{\text{ll}} = \Delta x^a$ . Using the equivalent definition of the affine scaling step (17) as a minimum-norm point, we have  $\|\delta x^*\| \geq \|\delta(x + \Delta x^{\text{ll}})\| = \sqrt{\mu} \|R x^{\text{ll}}\| = \sqrt{\mu} \xi_{\mathcal{J}}^{\text{ll}}$ . From Lemma 3.6,  $\mu^+ / \mu \leq 2\sqrt{n} \epsilon^a(w) / \beta \leq 2\sqrt{n} \xi_{\mathcal{J}}^{\text{ll}} / \beta$ . Thus, we see that  $\|\delta x^*\| \geq \beta \mu^+ / (2\sqrt{n\mu})$ .

The final statement on lower bounding  $\rho^{\mu^+}(\ell, \ell') \geq -|J_q|$  for any  $\ell, \ell' \in J_q$  follows by showing that  $\delta_\ell^+ / \delta_{\ell'}^+$  remains close to  $\delta_\ell / \delta_{\ell'}$ , and hence the values of  $\kappa^{\mu^+}(\ell, \ell')$  and  $\kappa^\mu(\ell, \ell')$  are sufficiently close for indices on the same layer (Lemma 6.1).

**Lemma 4.5** (Proof in Sect. 6) *Let  $w = (x, y, s) \in \mathcal{N}(\beta)$  for  $\beta \in (0, 1/8]$ , and let  $\mathcal{J} = (J_1, \dots, J_p)$  be a  $\delta(w)$ -balanced partition. Assume that  $\xi_{\mathcal{J}}^{\text{ll}}(w) < 4\gamma n$ , and let  $w^+ = (x^+, y^+, s^+) \in \overline{\mathcal{N}}(2\beta)$  be the next iterate obtained by the LLS step with  $\mu^+ = \mu(w^+)$  and assume  $\mu^+ > 0$ . If  $\ell^{\delta^+}(\mathcal{J}) > 4\gamma n$ , then there exist two layers  $J_q$  and  $J_r$  and  $i \in J_q$  and  $j \in J_r$  such that  $x_i^* \geq x_i^+ / (8n^{3/2})$ , and  $s_j^* \geq s_j^+ / (8n^{3/2})$ . Further,  $\rho^{\mu^+}(i, j) \geq -|J_q \cup J_r|$ , and for all  $\ell, \ell' \in J_q \cup J_r$ ,  $\ell \neq \ell'$  we have  $\Psi^\mu(\ell, \ell') \leq |J_q \cup J_r|$ .*

Consider now any  $\ell \in J_k \subseteq \mathbf{B}$ . Then, since  $R x_\ell^{\text{ll}}$  is multiplicatively close to 1,  $x_\ell^+ \approx x_\ell$ ; on the other hand  $s_\ell^+$  will “shoot down” close to the small value  $R s_\ell^{\text{ll}} \cdot s_\ell$ . Conversely, for  $\ell \in J_k \subseteq \mathbf{N}$ ,  $s_\ell^+ \approx s_\ell$ , and  $x_\ell^+$  will “shoot down” to a small value.

The key step of the analysis is showing that the increase in  $\ell^{\delta^+}(\mathcal{J})$  can be attributed to an  $N$ -layer  $J_r$  “crashing into” a  $\mathbf{B}$ -layer  $J_q$ . That is, we show the existence of an edge  $(i', j') \in E_{\delta^+, \gamma/(4n)}$  for  $i' \in J_q$  and  $j' \in J_r$ , where  $r < q$  and  $J_q \subseteq \mathbf{B}$ ,  $J_r \subseteq \mathbf{N}$ . This can be achieved by analyzing the matrix  $B$  used in the subroutine VERIFY-LIFT.

For the layers  $J_q$  and  $J_r$ , we can use Lemma 4.2 to show that there exists an  $i \in J_q$  where  $x_i^* / x_i$  is lower bounded, and there exists a  $j \in J_r$  where  $s_j^* / s_j$  is lower bounded. The lower bound on  $\rho^{\mu^+}(i, j)$  and the upper bounds on the  $\Psi^\mu(\ell, \ell')$  values can be shown by tracking the changes between the  $\kappa^\delta(\ell, \ell')$  and  $\kappa^{\delta^+}(\ell, \ell')$  values, and applying Lemma 4.1 both at  $w$  and at  $w^+$ .

**Proof of Theorem 3.16** We analyze the overall potential function  $\Psi(\mu)$ . By the iteration at  $\mu$  we mean the iteration where the normalized duality gap of the current iterate is  $\mu$ .

By Proposition 3.4(ii) and Lemma 3.10(ii), the predictor step gives  $w' \in \overline{\mathcal{N}}(1/4)$  in every iteration, and thus by Proposition 3.4(iii), if  $\mu(w') > 0$ , the iterate  $w^c$  after a corrector step fulfills  $w^c \in \mathcal{N}(1/8)$ . If  $\mu^+ = 0$  at the end of an iteration, the algorithm terminates with an optimal solution. Recall from Lemma 3.10(v) that this happens if and only if  $\epsilon^{\text{ll}}(w) = 0$  at a certain iteration.

From now on, assume that  $\mu^+ > 0$ . We distinguish three cases at each iteration. These cases are well-defined even at iterations where affine scaling steps are used. At such iterations,  $\xi_{\mathcal{J}}^{\text{ll}}(w)$  still refers to the LLS residuals, even if these have not been computed by the algorithm. (*Case I*)  $\xi_{\mathcal{J}}^{\text{ll}}(w) \geq 4\gamma n$ ; (*Case II*)  $\xi_{\mathcal{J}}^{\text{ll}}(w) < 4\gamma n$  and  $\ell^{\delta^+}(\mathcal{J}) \leq 4\gamma n$ ; and (*Case III*)  $\xi_{\mathcal{J}}^{\text{ll}}(w) < 4\gamma n$  and  $\ell^{\delta^+}(\mathcal{J}) > 4\gamma n$ .

Recall that the algorithm uses an LLS direction instead of the affine scaling direction whenever  $\epsilon^a(w) < 10n^{3/2}\gamma$ . Consider now the case when an affine scaling direction is used, that is,  $\epsilon^a(w) \geq 10n^{3/2}\gamma$ . According to Lemma 3.10(ii),  $\|Rx^{\text{ll}} - Rx^a\|$ ,  $\|Rs^{\text{ll}} - Rs^a\| \leq 6n^{3/2}\gamma$ . This implies that  $\xi_{\mathcal{J}}^{\text{ll}}(w) \geq 4n^{3/2}\gamma \geq 4n\gamma$ . Therefore, in cases II and III, an LLS step will be performed.

Starting with any given iteration, in each case we will identify a set  $J \subseteq [n]$  of indices with  $|J| > 1$ , and start a *phase* of  $O(\sqrt{n}|J| \log(\bar{\chi}^* + n))$  iterations (that can be either affine scaling or LLS steps). In each phase, we will guarantee that  $\Psi$  increases by at least  $|J| - 1$ . By definition,  $0 \leq \Psi(\mu) \leq n(n-1)(\log_2 n + 1)$ , and if  $\mu' < \mu$  then  $\Psi(\mu') \geq \Psi(\mu)$ . As we can partition the union of all iterations into disjoint phases, this yields the bound  $O(n^{2.5} \log n \log(\bar{\chi}^* + n))$  on the total number of iterations.

We now consider each of the cases. We always let  $\mu$  denote the normalized duality gap at the current iteration, and we let  $q \in [p]$  be the layer such that  $\xi_{\mathcal{J}}^{\text{ll}}(w) = \xi_{J_q}^{\text{ll}}(w)$ .

**Case I:**  $\xi_{\mathcal{J}}^{\text{ll}}(w) \geq 4\gamma n$ . Lemma 4.2 guarantees the existence of  $x_i, s_j \in J_q$  such that  $x_i^*/x_i, s_j^*/s_j \geq 4\gamma n/(3\sqrt{n}) > 1/(2^{10}n^{5.5})$ . Further, according to Lemma 4.1,  $\rho^\mu(i, j) \geq -|J_q|$ . Thus, Lemma 4.3 is applicable for  $J = J_q$ . The phase starting at  $\mu$  comprises  $O(\sqrt{n}|J_q| \log(\bar{\chi}^* + n))$  iterations, after which we get a normalized duality gap  $\mu'$  such that  $\Psi^{\mu'}(i, j) \geq 2|J_q|$ , and for each  $\ell \in [n] \setminus \{i, j\}$ , either  $\Psi^{\mu'}(i, \ell) \geq 2|J_q|$ , or  $\Psi^{\mu'}(\ell, j) \geq 2|J_q|$ .

We can take advantage of these bounds for indices  $\ell \in J_q$ . Again by Lemma 4.1, for any  $\ell, \ell' \in J_q$ , we have  $\Psi^\mu(\ell, \ell') \leq \rho^\mu(\ell, \ell') \leq |J_q|$ . Thus, there are at least  $|J_q| - 1$  pairs of indices  $(\ell, \ell')$  for which  $\Psi^\mu(\ell, \ell')$  increases by at least a factor 2 between iterations at  $\mu$  and  $\mu'$ . The increase in the contribution of these terms to  $\Psi(\mu)$  is at least  $|J_q| - 1$  during these iterations.

We note that this analysis works regardless whether an LLS step or an affine scaling step was performed in the iteration at  $\mu$ .

**Case II:**  $\xi_{\mathcal{J}}^{\text{ll}}(w) < 4\gamma n$  and  $\ell^{\delta^+}(\mathcal{J}) \leq 4\gamma n$ . As explained above, in this case we perform an LLS step in the iteration at  $\mu$ , and we let  $w^+$  denote the iterate obtained by the LLS step. For  $J = J_q$ , Lemma 4.4 guarantees the existence of  $i, j \in J_q$  such that  $x_i^*/x_i^+, s_j^*/s_j^+ > \beta/(16n^{3/2})$ , and further,  $\rho^{\mu^+}(i, j) > -|J_q|$ . We can therefore apply Lemma 4.3. The phase starting at  $\mu$  includes the LLS step leading to  $\mu^+$  (and the subsequent centering step), and the additional  $O(\sqrt{n}|J_q| \log(\bar{\chi}^* + n))$  iterations



( $\beta$  is a fixed constant in Algorithm 2) as in Lemma 4.3. As in Case I, we get the desired potential increase compared to the potentials at  $\mu$  in layer  $J_q$ .

**Case III:**  $\xi_{\mathcal{J}}^{\text{ll}}(w) < 4\gamma n$  and  $\ell^{\delta^+}(\mathcal{J}) > 4\gamma n$ . Again, the iteration at  $\mu$  will use an LLS step. We apply Lemma 4.5, and set  $J = J_q \cup J_r$  as in the lemma. The argument is the same as in Case II, using that Lemma 4.5 explicitly states that  $\Psi^\mu(\ell, \ell') \leq |J|$  for any  $\ell, \ell' \in J, \ell \neq \ell'$ . □

### 4.1 The iteration complexity bound for the Vavasis–Ye algorithm

We now show that the potential analysis described above also gives an improved bound  $O(n^{2.5} \log n \log(\bar{\chi}_A + n))$  for the original VY algorithm [63].

We recall the VY layering step. Order the variables via  $\pi$  such that  $\delta_{\pi(1)} \leq \delta_{\pi(2)} \leq \dots \leq \delta_{\pi(n)}$ . The layers will be consecutive sets in the ordering; a new layer starts with  $\pi(i + 1)$  each time  $\delta_{\pi(i+1)} > g\delta_{\pi(i)}$ , for a parameter  $g = \text{poly}(n)\bar{\chi}_A$ .

As outlined in the Introduction, the VY algorithm can be seen as a special implementation of our algorithm by setting  $\hat{\kappa}_{ij} = g\gamma/n$ . With these edge weights, we have that  $\hat{\kappa}_{ij}^\delta \geq \gamma/n$  precisely if  $g\delta_j \geq \delta_i$ .<sup>3</sup>

With these edge weights, it is easy to see that our LAYERING( $\delta, \hat{\kappa}$ ) subroutine finds the exact same components as VY. Moreover, the layers will be the initial strongly connected components  $C_i$  of  $G_{\delta, \gamma/n}$ : due to the choice of  $g$ , this partition is automatically  $\delta$ -balanced. There is no need to call VERIFY-LIFT.

The essential difference compared to our algorithm is that the values  $\hat{\kappa}_{ij} = g\gamma/n$  are not lower bounds on  $\kappa_{ij}$  as we require, but upper bounds instead. This is convenient to simplify the construction of the layering. On the negative side, the strongly connected components of  $\hat{G}_{\delta, \gamma/n}$  may not anymore be strongly connected in  $G_{\delta, \gamma/n}$ . Hence, we cannot use Lemma 4.1, and consequently, Lemma 4.3 does not hold.

Still, the  $\hat{\kappa}_{ij}$  bounds are overestimating  $\kappa_{ij}$  by at most a factor  $\text{poly}(n)\bar{\chi}_A$ . Therefore, the strongly connected components of  $\hat{G}_{\delta, n/\gamma}$  are strongly connected in  $G_{\delta, \sigma}$  for some  $\sigma = 1/(\text{poly}(n)\bar{\chi}_A)$ .

Hence, the entire argument described in this section is applicable to the VY algorithm, with a different potential function defined with  $\bar{\chi}_A$  instead of  $\bar{\chi}_A^*$ . This is the reason why the iteration bound in Lemma 4.3, and therefore in Theorem 3.16, also changes to  $\bar{\chi}_A$  dependency.

It is worth noting that due to the overestimation of the  $\kappa_{ij}$  values, the VY algorithm uses a coarser layering than our algorithm. Our algorithm splits up the VY layers into smaller parts so that  $\ell^\delta(\mathcal{J})$  remains small, but within each part, the gaps between the variables are bounded as a function of  $\bar{\chi}_A^*$  instead of  $\bar{\chi}_A$ .

## 5 Properties of the layered least square step

This section is dedicated to the proofs of Proposition 3.8 on the duality of lifting scores and Lemma 3.10 on properties of LLS steps.

<sup>3</sup> For simplicity, in the Introduction we used  $gx_i \geq x_j$  instead, which is almost the same in the proximity in the central path.

**Proposition 3.8** (Restatement). *For a linear subspace  $W \subseteq \mathbb{R}^n$  and index set  $I \subseteq [n]$  with  $J = [n] \setminus I$ ,*

$$\|L_I^W\| \leq \max\{1, \|L_J^{W^\perp}\|\}.$$

*In particular,  $\ell^W(I) = \ell^{W^\perp}(J)$*

**Proof** We first treat the case where  $\pi_I(W) = \{0\}$  or  $\pi_J(W^\perp) = \{0\}$ . If  $\pi_I(W) = \{0\}$  then  $\|L_I^W\| = \ell^W(I) = 0$ . Furthermore, in this case  $\mathbb{R}^I = \pi_I(W)^\perp = \pi_I(W^\perp \cap \mathbb{R}^n)$ , and thus  $\{(0, w_j) : w \in W^\perp\} \subseteq W^\perp$ . In particular,  $\|L_J^{W^\perp}\| \leq 1$  and  $\ell^{W^\perp}(J) = 0$ . Symmetrically, if  $\pi_J(W^\perp) = \{0\}$  then  $\|L_J^{W^\perp}\| = \ell^{W^\perp}(J) = 0$ ,  $\|L_I^W\| \leq 1$  and  $\ell^W(I) = 0$ .

We now restrict our attention to the case where both  $\pi_I(W), \pi_J(W^\perp) \neq \{0\}$ . Under this assumption, we show that  $\|L_I^W\| = \|L_J^{W^\perp}\|$  and thus that  $\ell^W(I) = \ell^{W^\perp}(J)$ . Note that by non-emptiness, we clearly have that  $\|L_I^W\|, \|L_J^{W^\perp}\| \geq 1$ .

We formulate a more general claim. Let  $\{0\} \neq U, V \subset \mathbb{R}^n$  be linear subspaces such that  $U + V = \mathbb{R}^n$  and  $U \cap V = \{0\}$ . Note that for the orthogonal complements in  $\mathbb{R}^n$ , we also have  $\{0\} \neq U^\perp, V^\perp, U^\perp + V^\perp = \mathbb{R}^n$  and  $U^\perp \cap V^\perp = \{0\}$ .

**Claim 5.1** *Let  $\{0\} \neq U, V \subset \mathbb{R}^n$  be linear subspaces such that  $U + V = \mathbb{R}^n$  and  $U \cap V = \{0\}$ . Thus, for  $z \in \mathbb{R}^n$ , there are unique decompositions  $z = u + v$  with  $u \in U, v \in V$  and  $z = u' + v'$  with  $u' \in U^\perp$  and  $v' \in V^\perp$ . Let  $T : \mathbb{R}^n \rightarrow V$  be the map sending  $Tz = v$ . Let  $T' : \mathbb{R}^n \rightarrow V^\perp$  be the map sending  $T'z = v'$ . Then,  $\|T\| = \|T'\|$ .*

**Proof** To prove the statement, we claim that it suffices to show that if  $\|T\| > 1$  then  $\|T'\| \geq \|T\|$ . To prove sufficiency, note that by symmetry, we also get that if  $\|T'\| > 1$  then  $\|T\| \geq \|T'\|$ . Note that  $V, V^\perp \neq \{0\}$  by assumption, and  $Tz = z$  for  $z \in V, T'z = z$  for  $z \in V^\perp$ . Thus, we always have  $\|T\|, \|T'\| \geq 1$ , and therefore the equality  $\|T\| = \|T'\|$  must hold in all cases. We now assume  $\|T\| > 1$  and show  $\|T'\| \geq \|T\|$ .

Representing  $T$  as an  $n \times n$  matrix, we write  $T = \sum_{i=1}^k \sigma_i v_i u_i^\top$  using a singular value decomposition with  $\sigma_1 \geq \dots \geq \sigma_k > 0$ . As such,  $v_1, \dots, v_k$  is an orthonormal basis of  $V$ , since the range( $T$ ) =  $V$ , and  $u_1, \dots, u_k$  is an orthonormal basis of  $U^\perp$ , since Ker( $T$ ) =  $U$ , noting that we have restricted to the singular vectors associated with positive singular values. By assumption, we have that  $\|T\| = \|Tu_1\| = \sigma_1 > 1$ .

The proof is complete by showing that

$$\|T'(v_1 - u_1/\sigma_1)\| \geq \sigma_1 \|v_1 - u_1/\sigma_1\|, \tag{34}$$

and that  $\|v_1 - u_1/\sigma_1\| > 0$ , since then the vector  $v_1 - u_1/\sigma_1$  will certify that  $\|T'\| \geq \sigma_1$ .

The map  $T$  is a linear projection with  $T^2 = T$ . Hence  $\langle u_i, v_i \rangle = \sigma_i^{-1}$  and  $\langle u_i, v_j \rangle = 0$  for all  $i \neq j$ .

We show that  $v_1 - \sigma_1^{-1}u_1$  can be decomposed as  $v_1 - \sigma_1 u_1 + (\sigma_1 - \sigma_1^{-1})u_1$  such that  $v_1 - \sigma_1 u_1 \in V^\perp$  and  $(\sigma_1 - \sigma_1^{-1})u_1 \in U^\perp$ . Therefore,  $T'(v_1 - \sigma_1^{-1}u_1) = v_1 - \sigma_1 u_1$ .

The containment  $(\sigma_1 - \sigma_1^{-1})u_1 \in U^\perp$  is immediate. To show  $v_1 - \sigma_1 u_1 \in V^\perp$ , we need that  $\langle v_1 - \sigma_1 u_1, v_i \rangle = 0$  for all  $i \in [k]$ . For  $i \geq 2$ , this is true since  $\langle u_i, v_j \rangle = 0$

and  $\langle v_i, v_j \rangle = 0$ . For  $i = 1$ , we have  $\langle v_1 - \sigma_1 u_1, v_1 \rangle = 0$  since  $\|v_1\| = 1$  and  $\langle u_1, v_1 \rangle = \sigma_1^{-1}$ . Consequently,  $T'(v_1 - \sigma_1^{-1} u_1) = v_1 - \sigma_1 u_1$ .

We compute  $\|v_1 - \sigma_1^{-1} u_1\| = \sqrt{1 - \sigma_1^{-2}} > 0$ , since  $\sigma_1 > 1$ , and  $\|v_1 - \sigma_1 u_1\| = \sqrt{\sigma_1^2 - 1}$ . This verifies (34), and thus  $\|T'\| \geq \sigma_1 = \|T\|$ .  $\square$

To prove the lemma, we define  $\mathcal{J} = (J, I)$ ,  $U = W_{\mathcal{J},1}^\perp \times W_{\mathcal{J},2}^\perp$  and  $V = W$  and let  $T : \mathbb{R}^n \rightarrow V$  and  $T' : \mathbb{R}^n \rightarrow V^\perp$  be as in Claim 5.1. By assumption,  $\{0\} \neq \pi_I(W) \Rightarrow \{0\} \neq V$  and  $\{0\} \neq \pi_J(W^\perp) = W_{\mathcal{J},1}^\perp \Rightarrow \{0\} \neq U$ . Applying Lemma 3.7,  $U, V$  satisfy the conditions of Claim 5.1 and  $T = \text{LLS}_{\mathcal{J}}^{W,1}$ . In particular,  $\|T'\| = \|T\|$ . Using the fact that  $U^\perp = W_{\mathcal{J},1} \times W_{\mathcal{J},2}$  and  $V^\perp = W^\perp$ , we similarly get that  $T' = \text{LLS}_{\tilde{\mathcal{J}}}^{W^\perp,1}$ , where  $\tilde{\mathcal{J}} = (I, J)$ . By (21) we have, for any  $t \in \pi_{\mathbb{R}^n}(W)$ , that  $Tt = \text{LLS}_{\mathcal{J}}^{W,1}(t) = L_I^W(t)$ . Thus,  $\|T\| \geq \|L_I^W\| \geq 1$ .

To finish the proof of the lemma from the claim, we show that  $\|T\| \leq \|L_I^W\|$ . By a symmetric argument we get  $\|T'\| = \|L_J^{W^\perp}\|$ .

If  $x \in \mathbb{R}_J^n$ , then  $Tx \in W \cap \mathbb{R}_J^n$  because any  $s \in W_{\mathcal{J},2}^\perp, t \in \pi_I(W)$  with  $s + t = 0$  must have  $s = t = 0$  since  $W_{\mathcal{J},2}^\perp$  is orthogonal to  $\pi_I(W)$ . But  $W \cap \mathbb{R}_J^n$  and  $W_{\mathcal{J},1}^\perp$  are orthogonal, so  $\|Tx\| \leq \|x\|$  because  $x = Tx + (x - Tx)$  is an orthogonal decomposition.

If  $y \in \mathbb{R}_I^n$ , then  $y_J = 0$  and hence  $(Ty)_J = (Ty - y)_J$ . Since  $(Ty - y)_J \in W_{\mathcal{J},1}^\perp = \pi_J(W \cap \mathbb{R}_J^n)^\perp$ , we see that  $Ty \in (W \cap \mathbb{R}_J^n)^\perp$ . As such, for any  $x \in \mathbb{R}_J^n, y \in \mathbb{R}_I^n$ , we see that  $x \perp y$  and  $Tx \perp Ty$ . For  $x, y \neq 0$ , we thus have that

$$\begin{aligned} \frac{\|T(x + y)\|^2}{\|x + y\|^2} &= \frac{\|T(x)\|^2 + \|T(y)\|^2}{\|x\|^2 + \|y\|^2} \leq \max \left\{ \frac{\|T(x)\|^2}{\|x\|^2}, \frac{\|T(y)\|^2}{\|y\|^2} \right\} \\ &\leq \max \left\{ 1, \frac{\|T(y)\|^2}{\|y\|^2} \right\}. \end{aligned}$$

Since  $\|L_I^W\| \geq 1$ , we must have that  $\|Tt\|/\|t\|$  is maximized by some  $t \in \mathbb{R}_I^n$ . From  $\text{Ker}(T) = U$  it is clear that  $\|Tt\|/\|t\|$  is maximized by some  $t \in U^\perp$ . Now,  $U^\perp \cap \mathbb{R}_I^n = \pi_{\mathbb{R}^n}(W)$ , so any  $t$  maximizing  $\|Tt\|/\|t\|$  satisfies  $Tt = L_I^W(t)$ . Therefore,  $\|L_I^W\| \geq \|T\|$ .  $\square$

Our next goal is to show Lemma 3.10: for a layering with small enough  $\ell^\delta(\mathcal{J})$ , the LLS step approximately satisfies (13), that is,  $\delta \Delta x^\parallel + \delta^{-1} \Delta s^\parallel \approx -x^{1/2} s^{1/2}$ . This also enables us to derive bounds on the norm of the residuals and on the step-length. We start by proving a few auxiliary technical claims. The next simple lemma allows us to take advantage of low lifting scores in the layering.

**Lemma 5.2** *Let  $u, v \in \mathbb{R}^n$  be two vectors such that  $u - v \in W$ . Let  $I \subseteq [n]$ , and  $\delta \in \mathbb{R}_{++}^n$ . Then there exists a vector  $u' \in W + u$  satisfying  $u'_I = v_I$  and*

$$\|\delta_{[n] \setminus I}(u'_{[n] \setminus I} - u_{[n] \setminus I})\| \leq \ell^\delta(I) \|\delta_I(u_I - v_I)\|.$$

**Proof** We let

$$u' := u + \delta^{-1} L_I^\delta (\delta_I (v_I - u_I)).$$

The claim follows by the definition of the lifting score  $\ell^\delta(I)$ . □

The next lemma will be the key tool to prove Lemma 3.10. It is helpful to recall the characterization of the LLS step in Sect. 3.4.

**Lemma 5.3** *Let  $w = (x, y, s) \in \mathcal{N}(\beta)$  for  $\beta \in (0, 1/4]$ , let  $\mu = \mu(w)$  and  $\delta = \delta(w)$ . Let  $\mathcal{J} = (J_1, \dots, J_p)$  be a  $\delta(w)$ -balanced layering, and let  $\Delta w^{\text{ll}} = (\Delta x^{\text{ll}}, \Delta y^{\text{ll}}, \Delta s^{\text{ll}})$  denote the corresponding LLS direction. Let  $\Delta x \in \times_{k=1}^p W_{\mathcal{J},k}$  and  $\Delta s \in \times_{k=1}^p W_{\mathcal{J},k}^\perp$  as in (25) and (26), that is*

$$\delta \Delta x^{\text{ll}} + \delta^{-1} \Delta s + x^{1/2} s^{1/2} = 0, \tag{35}$$

$$\delta \Delta x + \delta^{-1} \Delta s^{\text{ll}} + x^{1/2} s^{1/2} = 0. \tag{36}$$

Then, there exist vectors  $\Delta \bar{x} \in \times_{k=1}^p W_{\mathcal{J},k}$  and  $\Delta \bar{s} \in \times_{k=1}^p W_{\mathcal{J},k}^\perp$  such that

$$\|\delta_{J_k} (\Delta \bar{x}_{J_k} - \Delta x_{J_k}^{\text{ll}})\| \leq 2n \ell^\delta(\mathcal{J}) \sqrt{\mu} \quad \forall k \in [p] \quad \text{and} \tag{37}$$

$$\|\delta_{J_k}^{-1} (\Delta \bar{s}_{J_k} - \Delta s_{J_k}^{\text{ll}})\| \leq 2n \ell^\delta(\mathcal{J}) \sqrt{\mu} \quad \forall k \in [p]. \tag{38}$$

**Proof** Throughout, we use the shorthand notation  $\lambda = \ell^\delta(\mathcal{J})$ . We construct  $\Delta \bar{x}$ ; one can obtain  $\Delta \bar{s}$ , using that the reverse layering has lifting score  $\lambda$  in  $W^\perp \text{Diag}(\delta^{-1})$  according to Lemma 3.9.

We proceed by induction, constructing  $\Delta \bar{x}_{J_k} \in W_{\mathcal{J},k}$  for  $k = p, p-1, \dots, 1$ . This will be given as  $\Delta \bar{x}_{J_k} = \Delta x_{J_k}^{(k)}$  for a vector  $\Delta x^{(k)} \in W$  such that  $\Delta x_{J_{>k}}^{(k)} = 0$ . We prove the inductive hypothesis

$$\left\| \delta_{J_{\leq k}} \left( \Delta x_{J_{\leq k}}^{(k)} - \Delta x_{J_{\leq k}}^{\text{ll}} \right) \right\| \leq 2\lambda \sqrt{\mu} \sum_{q=k+1}^p \sqrt{|J_q|}. \tag{39}$$

Note that (37) follows by restricting the norm on the LHS to  $J_k$  and since the sum on the RHS is  $\leq n$ .

For  $k = p$ , the RHS is 0. We simply set  $\Delta x^{(p)} = \Delta x^{\text{ll}}$ , that is,  $\Delta \bar{x}_{J_p} = \Delta x_{J_p}^{\text{ll}}$ , trivially satisfying the hypothesis. Consider now  $k < p$ , and assume that we have a  $\Delta \bar{x}_{J_{k+1}} = \Delta x_{J_{k+1}}^{(k+1)}$  satisfying (39) for  $k + 1$ . From (35) and the induction hypothesis, we get that

$$\begin{aligned} \|\delta_{J_{k+1}} \Delta \bar{x}_{J_{k+1}} + \delta_{J_{k+1}}^{-1} \Delta s_{J_{k+1}}\| &\leq \|x_{J_{k+1}}^{1/2} s_{J_{k+1}}^{1/2}\| + \|\delta_{J_{k+1}} (\Delta \bar{x}_{J_{k+1}} - \Delta x_{J_{k+1}}^{\text{ll}})\| \\ &\leq \|x_{J_{k+1}}^{1/2} s_{J_{k+1}}^{1/2}\| + 2\lambda \sqrt{\mu} \sum_{q=k+2}^p \sqrt{|J_q|} \leq \sqrt{1 + \beta} \sqrt{\mu |J_{k+1}|} + 2n\lambda \sqrt{\mu} < 2\sqrt{\mu |J_{k+1}|}, \end{aligned}$$

using also that  $w \in \mathcal{N}(\beta)$ , Proposition 3.2, and the assumptions  $\beta \leq 1/4$ ,  $\lambda \leq \beta/(32n^2)$ . Note that  $\Delta \bar{x}_{J_{k+1}} \in W_{\mathcal{J},k}$  and  $\Delta s_{J_{k+1}} \in W_{\mathcal{J},k}^\perp$  are orthogonal vectors. The above inequality therefore implies

$$\|\delta_{J_{k+1}} \Delta \bar{x}_{J_{k+1}}\| \leq 2\sqrt{\mu|J_{k+1}|}.$$

Let us now use Lemma 5.2 to obtain  $\Delta x^{(k)}$  for  $u = \Delta x^{(k+1)}$ ,  $v = 0$ , and  $I = J_{>k}$ . That is, we get  $\Delta x_{J_{>k}}^{(k)} = 0$ ,  $\Delta x^{(k)} \in W$ , and

$$\begin{aligned} \|\delta_{J_{\leq k}}(\Delta x_{J_{\leq k}}^{(k)} - \Delta x_{J_{\leq k}}^{(k+1)})\| &\leq \lambda \|\delta_{J_{>k}} \Delta x_{J_{>k}}^{(k+1)}\| \\ &= \lambda \|\delta_{J_{k+1}} \Delta \bar{x}_{J_{k+1}}\| \leq 2\lambda \sqrt{\mu|J_{k+1}|}. \end{aligned}$$

By the triangle inequality and the induction hypothesis (39) for  $k + 1$ ,

$$\begin{aligned} \|\delta_{J_{\leq k}}(\Delta x_{J_{\leq k}}^{(k)} - \Delta x_{J_{\leq k}}^{\parallel})\| &\leq \|\delta_{J_{\leq k}}(\Delta x_{J_{\leq k}}^{(k)} - \Delta x_{J_{\leq k}}^{(k+1)})\| + \|\delta_{J_{\leq k}}(\Delta x_{J_{\leq k}}^{(k+1)} - \Delta x_{J_{\leq k}}^{\parallel})\| \\ &\leq 2\lambda \sqrt{\mu|J_{k+1}|} + 2\lambda \sum_{q=k+2}^p \sqrt{\mu|J_q|}, \end{aligned}$$

yielding the induction hypothesis for  $k$ . □

**Lemma 3.10 (Restatement).** *Let  $w = (x, y, s) \in \mathcal{N}(\beta)$  for  $\beta \in (0, 1/4]$ , let  $\mu = \mu(w)$  and  $\delta = \delta(w)$ . Let  $\mathcal{J} = (J_1, \dots, J_p)$  be a layering with  $\ell^\delta(\mathcal{J}) \leq \beta/(32n^2)$ , and let  $\Delta w^\parallel = (\Delta x^\parallel, \Delta y^\parallel, \Delta s^\parallel)$  denote the LLS direction for the layering  $\mathcal{J}$ . Let furthermore  $\epsilon^\parallel(w) = \max_{i \in [n]} \min\{|R x_i^\parallel|, |R s_i^\parallel|\}$ , and define the maximal step length as*

$$\alpha^* := \sup\{\alpha' \in [0, 1] : \forall \bar{\alpha} \in [0, \alpha'] : w + \bar{\alpha} \Delta w^\parallel \in \mathcal{N}(2\beta)\}.$$

Then the following properties hold.

(i) We have

$$\|\delta_{J_k} \Delta x_{J_k}^\parallel + \delta_{J_k}^{-1} \Delta s_{J_k}^\parallel + x_{J_k}^{1/2} s_{J_k}^{1/2}\| \leq 6n \ell^\delta(\mathcal{J}) \sqrt{\mu}, \quad \forall k \in [p], \text{ and} \quad (27)$$

$$\|\delta \Delta x^\parallel + \delta^{-1} \Delta s^\parallel + x^{1/2} s^{1/2}\| \leq 6n^{3/2} \ell^\delta(\mathcal{J}) \sqrt{\mu}. \quad (28)$$

(ii) For the affine scaling direction  $\Delta w^a = (\Delta x^a, \Delta y^a, \Delta s^a)$ ,

$$\|R x^\parallel - R x^a\|, \|R s^\parallel - R s^a\| \leq 6n^{3/2} \ell^\delta(\mathcal{J}).$$

(iii) For the residuals of the LLS steps we have  $\|R x^\parallel\|, \|R s^\parallel\| \leq \sqrt{2n}$ . For each  $i \in [n]$ ,  $\max\{|R x_i^\parallel|, |R s_i^\parallel|\} \geq \frac{1}{2} - \frac{3}{4}\beta$ .

(iv) We have

$$\alpha^* \geq 1 - \frac{3\sqrt{n}\epsilon^\parallel(w)}{\beta}, \quad (29)$$

and for any  $\alpha \in [0, 1]$

$$\mu(w + \alpha \Delta w^{\parallel}) = (1 - \alpha)\mu,$$

(v) We have  $\epsilon^{\parallel}(w) = 0$  if and only if  $\alpha^* = 1$ . These are further equivalent to  $w + \Delta w^{\parallel} = (x + \Delta x^{\parallel}, y + \Delta y^{\parallel}, s + \Delta s^{\parallel})$  being an optimal solution to (LP).

**Proof** Again, we use  $\lambda = \ell^{\delta}(\mathcal{J})$ .

**Part (i).** Clearly, (27) implies (28). To show (27), we use Lemma 5.3 to obtain  $\Delta \bar{x}$  and  $\Delta \bar{s}$  as in (37) and (38). We will also use  $\Delta x \in \times_{k=1}^p W_{\mathcal{J},k}$  and  $\Delta s \in \times_{k=1}^p W_{\mathcal{J},k}^{\perp}$  as in (35) and (36).

Select any layer  $k \in [p]$ . From (35), we get that

$$\|\delta_{J_k} \Delta \bar{x}_{J_k} + \delta_{J_k}^{-1} \Delta s_{J_k} + x_{J_k}^{1/2} s_{J_k}^{1/2}\| = \|\delta_{J_k} (\Delta \bar{x}_{J_k} - \Delta x_{J_k}^{\parallel})\| \leq 2n\lambda\sqrt{\mu}. \quad (40)$$

Similarly, from (36), we see that

$$\|\delta_{J_k}^{-1} \Delta \bar{s}_{J_k} + \delta_{J_k} \Delta x_{J_k} + x_{J_k}^{1/2} s_{J_k}^{1/2}\| = \|\delta_{J_k}^{-1} (\Delta \bar{s}_{J_k} - \Delta s_{J_k}^{\parallel})\| \leq 2n\lambda\sqrt{\mu}.$$

From the above inequalities, we see that

$$\|\delta_{J_k} (\Delta \bar{x}_{J_k} - \Delta x_{J_k}) + \delta_{J_k}^{-1} (\Delta s_{J_k} - \Delta \bar{s}_{J_k})\| \leq 4n\lambda\sqrt{\mu}.$$

Since  $\delta_{J_k} (\Delta \bar{x}_{J_k} - \Delta x_{J_k})$  and  $\delta_{J_k}^{-1} (\Delta s_{J_k} - \Delta \bar{s}_{J_k})$  are orthogonal vectors, we have

$$\|\delta_{J_k} (\Delta \bar{x}_{J_k} - \Delta x_{J_k})\|, \|\delta_{J_k}^{-1} (\Delta s_{J_k} - \Delta \bar{s}_{J_k})\| \leq 4n\lambda\sqrt{\mu}.$$

Together with (37), this yields  $\|\delta_{J_k} (\Delta x_{J_k}^{\parallel} - \Delta x_{J_k})\| \leq 6n\lambda\sqrt{\mu}$ . Combined with (26), we get

$$\|\delta_{J_k} \Delta x_{J_k}^{\parallel} + \delta_{J_k}^{-1} \Delta s_{J_k}^{\parallel} + x_{J_k}^{1/2} s_{J_k}^{1/2}\| = \|\delta_{J_k} (\Delta x_{J_k}^{\parallel} - \Delta x_{J_k})\| \leq 6n\lambda\sqrt{\mu},$$

thus, (27) follows.

**Part (ii).** Recall from Lemma 3.5(i) that  $\sqrt{\mu}Rx^a + \sqrt{\mu}Rs^a = x^{1/2}s^{1/2}$ . From part (i), we can similarly see that

$$\|\sqrt{\mu}Rx^{\parallel} + \sqrt{\mu}Rs^{\parallel} - x^{1/2}s^{1/2}\| \leq 6n^{3/2}\lambda\sqrt{\mu}.$$

From these, we get

$$\|(Rx^{\parallel} - Rx^a) + (Rs^{\parallel} - Rs^a)\| \leq 6n^{3/2}\lambda.$$

The claim follows since  $Rx^{\parallel} - Rx^a \in \text{Diag}(\delta)W$  and  $Rs^{\parallel} - Rs^a \in \text{Diag}(\delta^{-1})W^{\perp}$  are orthogonal vectors.

**Part (iii).** Both bounds follow from the previous part and Lemma 3.5(iii), using the assumption  $\ell^\delta(\mathcal{J}) \leq \beta/(32n^2)$ .

**Part (iv).** Let  $w^+ = w + \alpha \Delta w^{\text{ll}}$ . We need to find the largest value  $\alpha > 0$  such that  $w^+ \in \mathcal{N}(2\beta)$ . To begin, we first show that the normalized duality gap  $\mu(w^+)$  fulfills  $\mu(w^+) = (1 - \alpha)\mu$  for any  $\alpha \in \mathbb{R}$ . For this purpose, we use the decomposition:

$$(x + \alpha \Delta x^{\text{ll}})(s + \alpha \Delta s^{\text{ll}}) = (1 - \alpha)xs + \alpha(x + \Delta x^{\text{ll}})(s + \Delta s^{\text{ll}}) - \alpha(1 - \alpha)\Delta x^{\text{ll}}\Delta s^{\text{ll}}. \tag{41}$$

Recall from Part (i) that there exists  $\Delta x \in \times_{k=1}^p W_{\mathcal{J},k}$  and  $\Delta s \in \times_{k=1}^p W_{\mathcal{J},k}^\perp$  as in (35) and (36) such that  $\delta \Delta x^{\text{ll}} + \delta^{-1} \Delta s = -\delta x$  and  $\delta \Delta x + \delta^{-1} \Delta s^{\text{ll}} = -\delta^{-1} s$ . In particular,  $x + \Delta x^{\text{ll}} = -\delta^{-2} \Delta s$  and  $s + \Delta s^{\text{ll}} = -\delta^2 \Delta x$ . Noting that  $\Delta x^{\text{ll}} \perp \Delta s^{\text{ll}}$  and  $\Delta x \perp \Delta s$ , taking the average of the coordinates on both sides of (41), we get that

$$\begin{aligned} \mu(w + \alpha \Delta w^{\text{ll}}) &= (1 - \alpha)\mu(w) + \alpha \langle x + \Delta x^{\text{ll}}, s + \Delta s^{\text{ll}} \rangle / n - \alpha(1 - \alpha) \langle \Delta x^{\text{ll}}, \Delta s^{\text{ll}} \rangle / n \\ &= (1 - \alpha)\mu(w) + \alpha \langle \delta^{-2} \Delta s, \delta^2 \Delta x \rangle / n \\ &= (1 - \alpha)\mu(w), \end{aligned} \tag{42}$$

as needed.

Let  $\epsilon := \epsilon^{\text{ll}}(w)$ . To obtain the desired lower bound on the step-length, given (42) it suffices to show that for all  $0 \leq \alpha < 1 - \frac{3\sqrt{n}\epsilon}{\beta}$  that

$$\left\| \frac{(x + \alpha \Delta x^{\text{ll}})(s + \alpha \Delta s^{\text{ll}})}{(1 - \alpha)\mu} - e \right\| \leq 2\beta. \tag{43}$$

We will need a bound on the product of the LLS residuals:

$$\begin{aligned} \left\| Rx^{\text{ll}}Rs^{\text{ll}} - \frac{1}{\mu} \Delta x^{\text{ll}}\Delta s^{\text{ll}} \right\| &= \left\| \frac{x^{1/2}s^{1/2}}{\sqrt{\mu}} \cdot \frac{\delta \Delta x^{\text{ll}} + \delta^{-1} \Delta s^{\text{ll}} + x^{1/2}s^{1/2}}{\sqrt{\mu}} \right\| \\ &\leq 6(1 + 2\beta)n^{3/2}\lambda \leq \frac{\beta}{4}, \end{aligned} \tag{44}$$

using Proposition 3.1, part (i), and the assumptions  $\lambda \leq \beta/(32n^2)$ ,  $\beta \leq 1/4$ . Another useful bound will be

$$\begin{aligned} \|Rx^{\text{ll}}Rs^{\text{ll}}\|^2 &= \sum_{i \in [n]} |Rx_i^{\text{ll}}|^2 |Rs_i^{\text{ll}}|^2 \leq \epsilon^2 \sum_{i \in [n]} \max \{ |Rx_i^{\text{ll}}|^2, |Rs_i^{\text{ll}}|^2 \} \\ &\leq \epsilon^2 (\|Rx^{\text{ll}}\|^2 + \|Rs^{\text{ll}}\|^2) \leq 2n\epsilon^2. \end{aligned} \tag{45}$$

The last inequality uses part (iii). With (41) we are ready to get the bound in (43), as

$$\begin{aligned} \left\| \frac{(x + \alpha \Delta x^{\parallel})(s + \alpha \Delta s^{\parallel})}{(1 - \alpha)\mu} - e \right\| &\leq \beta + \left\| \frac{\alpha}{(1 - \alpha)\mu} (x + \Delta x^{\parallel})(s + \Delta s^{\parallel}) - \frac{\alpha}{\mu} \Delta x^{\parallel} \Delta s^{\parallel} \right\| \\ &= \beta + \left\| \left( \frac{\alpha}{1 - \alpha} - \alpha \right) R x^{\parallel} R s^{\parallel} + \alpha \left( R x^{\parallel} R s^{\parallel} - \frac{1}{\mu} \Delta x^{\parallel} \Delta s^{\parallel} \right) \right\| \\ &\leq \beta + \frac{\alpha^2}{1 - \alpha} \|R x^{\parallel} R s^{\parallel}\| + \alpha \left\| R x^{\parallel} R s^{\parallel} - \frac{1}{\mu} \Delta x^{\parallel} \Delta s^{\parallel} \right\| \\ &\leq \beta + \frac{\sqrt{2n}\epsilon}{1 - \alpha} + \frac{\beta}{4} \leq \frac{5}{4}\beta + \frac{\sqrt{2n}\epsilon}{1 - \alpha}. \end{aligned}$$

This value is  $\leq 2\beta$  whenever  $2\sqrt{n}\epsilon/(1 - \alpha) \leq (3/4)\beta \iff \alpha < 1 - \frac{3\sqrt{n}\epsilon}{\beta}$ , as needed.

**Part(v).** From part (iv), it is immediate that  $\epsilon^{\parallel}(w) = 0$  implies  $\alpha = 1$ . If  $\alpha = 1$ , we have that  $w + \Delta w^{\parallel}$  is the limit of (strictly) feasible solutions to (LP) and thus is also a feasible solution. Optimality of  $w + \Delta w^{\parallel}$  now follows from Part (iv), since  $\alpha = 1$  implies  $\mu(w + \Delta w^{\parallel}) = 0$ . The remaining implication is that if  $w + \Delta w^{\parallel}$  is optimal, then  $\epsilon^{\parallel}(w) = 0$ . Recall that  $Rx_i^{\parallel} = \delta_i(x_i + \Delta x_i^{\parallel})/\sqrt{\mu}$  and  $Rs_i^{\parallel} = \delta_i^{-1}(s_i + \Delta s_i^{\parallel})/\sqrt{\mu}$ . The optimality of  $w + \Delta w^{\parallel}$  means that for each  $i \in [n]$ , either  $x_i + \Delta x_i^{\parallel} = 0$  or  $s_i + \Delta s_i^{\parallel} = 0$ . Therefore,  $\epsilon^{\parallel}(w) = 0$ .  $\square$

### 6 Proofs of the main lemmas for the potential analysis

**Lemma 4.2** *Let  $w = (x, y, s) \in \mathcal{N}(\beta)$  for  $\beta \in (0, 1/8]$  and let  $w^* = (x^*, y^*, s^*)$  be the optimal solution corresponding to  $\mu^* = 0$  on the central path. Let further  $\mathcal{J} = (J_1, \dots, J_p)$  be a  $\delta(w)$ -balanced layering (Definition 3.13), and let  $\Delta w^{\parallel} = (\Delta x^{\parallel}, \Delta y^{\parallel}, \Delta s^{\parallel})$  be the corresponding LLS direction. Then the following statement holds for every  $q \in [p]$ :*

(i) *There exists  $i \in J_q$  such that*

$$x_i^* \geq \frac{2x_i}{3\sqrt{n}} \cdot (\|R x_{J_q}^{\parallel}\| - 2\gamma n). \tag{32}$$

(ii) *There exists  $j \in J_q$  such that*

$$s_j^* \geq \frac{2s_j}{3\sqrt{n}} \cdot (\|R s_{J_q}^{\parallel}\| - 2\gamma n). \tag{33}$$

**Proof of Lemma 4.2** We prove part (i); part (ii) follows analogously using Lemma 3.9. Let  $z$  be a vector fulfilling the statement of Lemma 5.2 for  $u = x^*$ ,  $v = x + \Delta x^{\parallel}$ , and  $I = J_{>q}$ . Then  $z \in W + x$ ,  $z_{J_{>q}} = x_{J_{>q}} + \Delta x_{J_{>q}}^{\parallel}$  and by  $\ell^\delta(\mathcal{J}) \leq \circ$

$$\left\| \delta_{J_{\leq q}}(x_{J_{\leq q}}^* - z_{J_{\leq q}}) \right\| \leq \gamma \left\| \delta_{J_{>q}}(x_{J_{>q}}^* - (x_{J_{>q}} + \Delta x_{J_{>q}}^{\parallel})) \right\|.$$



Restricting to the components in  $J_q$ , and dividing by  $\sqrt{\mu}$ , we get

$$\begin{aligned} \left\| \frac{\delta_{J_q}(x_{J_q}^* - z_{J_q})}{\sqrt{\mu}} \right\| &\leq \gamma \left\| \frac{\delta_{J_{>q}}(x_{J_{>q}}^* - (x_{J_{>q}} + \Delta x_{J_{>q}}^{\parallel}))}{\sqrt{\mu}} \right\| \\ &\leq \gamma \left\| \frac{\delta_{J_{>q}} x_{J_{>q}}^*}{\sqrt{\mu}} \right\| + \gamma \|R x_{J_{>q}}^{\parallel}\|. \end{aligned} \tag{46}$$

Since  $w \in \mathcal{N}(\beta)$ , from Proposition 3.1 and (16) we see that for  $i \in [n]$

$$\frac{\delta_i}{\sqrt{\mu}} \leq \frac{1}{\sqrt{1-2\beta}} \cdot \frac{\delta_i(w(\mu))}{\sqrt{\mu}} = \frac{1}{\sqrt{1-2\beta}} \cdot \frac{1}{x_i(\mu)},$$

and therefore

$$\left\| \frac{\delta_{J_{>q}} x_{J_{>q}}^*}{\sqrt{\mu}} \right\| \leq \frac{1}{\sqrt{1-2\beta}} \|x(\mu)_{J_{>q}}^{-1} x_{J_{>q}}^*\| \leq \frac{1}{\sqrt{1-2\beta}} \cdot \|x(\mu)_{J_{>q}}^{-1} x_{J_{>q}}^*\|_1 \leq \frac{n}{\sqrt{1-2\beta}},$$

where the last inequality follows by Lemma 3.3.

Using the above bounds with (46), along with  $\|R x_{J_{>q}}^{\parallel}\| \leq \|R x^{\parallel}\| \leq \sqrt{2n}$  from Lemma 3.10(iii), we get

$$\left\| \frac{\delta_{J_q} z_{J_q}}{\sqrt{\mu}} \right\| \leq \left\| \frac{\delta_{J_q} x_{J_q}^*}{\sqrt{\mu}} \right\| + \frac{\gamma n}{\sqrt{1-2\beta}} + \gamma \sqrt{2n} \leq \left\| \frac{\delta_{J_q} x_{J_q}^*}{\sqrt{\mu}} \right\| + 2\gamma n,$$

using that  $\beta \leq 1/8$  and  $n \geq 3$ . Note that  $z$  is a feasible solution to the least-squares problem which is optimally solved by  $x_{J_q}^{\parallel}$  for layer  $J_q$  and so

$$\|R x_{J_q}^{\parallel}\| \leq \left\| \frac{\delta_{J_q} z_{J_q}}{\sqrt{\mu}} \right\|.$$

It follows that

$$\left\| \frac{\delta_{J_q} x_{J_q}^*}{\sqrt{\mu}} \right\| \geq \|R x_{J_q}^{\parallel}\| - 2\gamma n.$$

Let us pick  $i = \arg \max_{t \in J_q} |\delta_t x_t^*|$ . Using Proposition 3.2,

$$\frac{x_i^*}{x_i} \geq \frac{1}{1+\beta} \cdot \frac{\delta_i x_i^*}{\sqrt{\mu}} \geq \frac{\|R x_{J_q}^{\parallel}\| - 2\gamma n}{(1+\beta)\sqrt{n}} \geq \frac{2}{3\sqrt{n}} \cdot (\|R x_{J_q}^{\parallel}\| - 2\gamma n),$$

completing the proof. □

**Lemma 4.3** (Restatement). *Let  $w = (x, y, s) \in \mathcal{N}(2\beta)$  for  $\beta \in (0, 1/8]$ , let  $\mu = \mu(w)$  and  $\delta = \delta(w)$ . Let  $i, j \in [n]$  and  $2 \leq \tau \leq n$  such that for the optimal solution  $w^* = (x^*, y^*, s^*)$ , we have  $x_i^* \geq \beta x_i / (2^{10} n^{5.5})$  and  $s_j^* \geq \beta s_j / (2^{10} n^{5.5})$ , and assume  $\rho^\mu(i, j) \geq -\tau$ . After  $O(\beta^{-1} \sqrt{n} \tau \log(\bar{\chi}^* + n))$  further iterations the duality gap  $\mu'$  fulfills  $\Psi^{\mu'}(i, j) \geq 2\tau$ , and for every  $\ell \in [n] \setminus \{i, j\}$ , either  $\Psi^{\mu'}(i, \ell) \geq 2\tau$ , or  $\Psi^{\mu'}(\ell, j) \geq 2\tau$ .*

**Proof of Lemma 4.3** Let us select a value  $\mu'$  such that

$$\log \mu - \log \mu' \geq 5\tau \log \left( \frac{4n\kappa^*}{\gamma} \right) + 31 \log n + 44 - 4 \log \beta.$$

The normalized duality gap decreases to such value within  $O(\beta^{-1} \sqrt{n} \tau \cdot \log(\bar{\chi}^* + n))$  iterations, recalling that  $\log(\bar{\chi}^* + n) = \Theta(\log(\kappa^* + n))$ . The step-lengths for the affine scaling and LLS steps are stated in Proposition 3.4 and Lemma 3.10(iv). Whenever the algorithm chooses an LLS step,  $\epsilon^a(w) < 10n^{3/2}\gamma$ . Thus, the progress in  $\mu$  will be at least as much (in fact, much better) than the  $1 - \beta/\sqrt{n}$  guarantee for the affine scaling step in Proposition 3.4.

Let  $w' = (x', y', s')$  be the central path element corresponding to  $\mu'$ , and let  $\delta' = \delta(w')$ . From now on we use the shorthand notation

$$\Gamma := \log \left( \frac{4n\kappa^*}{\gamma} \right).$$

We first show that

$$\Gamma \rho^{\mu'}(i, j) \geq 4\Gamma\tau + 18 \log n + 22 \log 2 - 2 \log \beta \tag{47}$$

for  $\mu'$ , and therefore,  $\Gamma \Psi^{\mu'}(i, j) \geq \min(2\Gamma n, 4\Gamma\tau + 18 \log n + 22 \log 2 - 2 \log \beta) \geq 2\Gamma\tau$  as  $\tau \leq n$ . Recalling the definition  $\kappa_{ij}^\delta = \kappa_{ij} \delta_j / \delta_i$ , we see that according to Proposition 3.2,

$$\kappa_{ij}^\delta \leq \frac{\kappa_{ij}}{(1 - \beta)^2} \cdot \frac{x_i s_j}{\mu}, \quad \text{and} \quad \kappa_{ij}^{\delta'} = \kappa_{ij} \cdot \frac{x'_i s'_j}{\mu'}.$$

Thus,

$$\begin{aligned} \Gamma \rho^{\mu'}(i, j) &\geq \Gamma \rho^\mu(i, j) + \log \mu - \log \mu' + 2 \log(1 - \beta) + \log x'_i - \log x_i + \log s'_j - \log s_j \\ &\geq \Gamma \rho^\mu(i, j) + 5\Gamma\tau + 31 \log n + 44 - 4 \log \beta + 2 \log(1 - \beta) + \log x'_i - \log x_i \\ &\quad + \log s'_j - \log s_j. \end{aligned}$$

Using the near-monotonicity of the central path (Lemma 3.3), we have  $x'_i \geq x_i^*/n$  and  $s'_j \geq s_j^*/n$ . Together with our assumptions  $x_i^* \geq \beta x_i / (2^{10} n^{5.5})$  and  $s_j^* \geq$

$\beta s_i / (2^{10} n^{5.5})$ , we see that

$$\log x'_i - \log x_i + \log s'_j - \log s_j \geq -13 \log n - 20 \log 2 + 2 \log \beta.$$

Using the assumption  $\rho^\mu(i, j) > -\tau$  of the lemma, we can establish (47) as  $\beta < 1/8$ .

Next, consider any  $\ell \in [n] \setminus \{i, j\}$ . From the triangle inequality Lemma 2.15(ii) it follows that  $\kappa_{ij}^{\delta'} \leq \kappa_{i\ell}^{\delta'} \cdot \kappa_{\ell j}^{\delta'}$ , which gives  $\rho^{\mu'}(i, \ell) + \rho^{\mu'}(\ell, j) \geq \rho^{\mu'}(i, j)$ . We therefore get

$$\max\{\Gamma\rho^{\mu'}(i, \ell), \Gamma\rho^{\mu'}(\ell, j)\} \geq \frac{1}{2}\Gamma\rho^{\mu'}(i, j) \stackrel{(47)}{\geq} 2\Gamma\tau + 9 \log n + 11 \log 2 - \log \beta.$$

We next show that if  $\Gamma\rho^{\mu'}(i, \ell) \geq 2\Gamma\tau + 9 \log n + 11 \log 2 - \log \beta$ , then  $\Psi^{\mu'}(i, \ell) \geq 2\tau$ . The case  $\Gamma\rho^{\mu'}(\ell, j) \geq 2\Gamma\tau + 9 \log n + 11 \log 2 - \log \beta$  follows analogously.

Consider any  $0 < \bar{\mu} < \mu'$  with the corresponding central path point  $\bar{w} = (\bar{x}, \bar{y}, \bar{s})$ . The proof is complete by showing  $\Gamma\rho^{\bar{\mu}}(i, \ell) \geq \Gamma\rho^{\mu'}(i, \ell) - 9 \log n - 11 \log 2 + \log \beta$ . Recall that for central path elements, we have  $\kappa_{ij}^{\delta'} = \kappa_{ij} x'_i / x'_j$ , and  $\kappa_{ij}^{\bar{\delta}} = \kappa_{ij} \bar{x}_i / \bar{x}_j$ . Therefore

$$\Gamma\rho^{\bar{\mu}}(i, j) = \Gamma\rho^{\mu'}(i, j) + \log \bar{x}_i - \log x'_i - \log \bar{x}_j + \log x'_j.$$

Using Proposition 3.1, Lemma 3.3 and the assumption  $x_i^* \geq \beta x_i / (2^{10} n^{5.5})$ , we have  $\bar{x}_j \leq n x'_j$  and

$$\bar{x}_i \geq \frac{x_i^*}{n} \geq \frac{\beta x_i}{2^{10} n^{6.5}} \geq \frac{\beta(1 - \beta)x'_i}{2^{10} n^{7.5}} \geq \frac{\beta x'_i}{2^{11} n^{7.5}}.$$

Using these bounds, we get

$$\Gamma\rho^{\bar{\mu}}(i, j) \geq \Gamma\rho^{\mu'}(i, j) - 9 \log n - 11 \log 2 + \log \beta,$$

completing the proof. □

It remains to prove Lemma 4.4 and Lemma 4.5, addressing the more difficult case  $\xi_{\mathcal{J}}^{\parallel} < 4\gamma n$ . It is useful to decompose the variables into two sets. We let

$$\mathbf{B} := \{t \in [n] : |Rs_t^{\parallel}| < 4\gamma n\}, \quad \text{and} \quad \mathbf{N} := \{t \in [n] : |Rx_t^{\parallel}| < 4\gamma n\}. \quad (48)$$

The assumption  $\xi_{\mathcal{J}}^{\parallel} < 4\gamma n$  implies that for every layer  $J_k$ , either  $J_k \subseteq \mathbf{B}$  or  $J_k \subseteq \mathbf{N}$ . The next two lemmas describe the relations between  $\delta$  and  $\delta^+$ .

**Lemma 6.1** *Let  $w \in \mathcal{N}(\beta)$  for  $\beta \in (0, 1/8]$ , and assume  $\ell^\delta(\mathcal{J}) \leq \gamma$  and  $\epsilon^{\parallel}(w) < 4\gamma n$ . For the next iterate  $w^+ = (x^+, y^+, s^+) \in \bar{\mathcal{N}}(2\beta)$ , we have*

(i) For  $i \in \mathbf{B}$ ,

$$\frac{1}{2} \cdot \sqrt{\frac{\mu^+}{\mu}} \leq \frac{\delta_i^+}{\delta_i} \leq 2 \cdot \sqrt{\frac{\mu^+}{\mu}} \quad \text{and} \quad \delta_i^{-1} s_i^+ \leq \frac{3\mu^+}{\sqrt{\mu}}.$$

(ii) For  $i \in \mathbf{N}$ ,

$$\frac{1}{2} \cdot \sqrt{\frac{\mu}{\mu^+}} \leq \frac{\delta_i^+}{\delta_i} \leq 2 \cdot \sqrt{\frac{\mu}{\mu^+}} \quad \text{and} \quad \delta_i x_i^+ \leq \frac{3\mu^+}{\sqrt{\mu}}.$$

(iii) If  $i, j \in \mathbf{B}$  or  $i, j \in \mathbf{N}$ , then

$$\frac{1}{4} \leq \frac{\kappa_{ij}^\delta}{\kappa_{ij}^{\delta^+}} = \frac{\delta_i^+ \delta_j}{\delta_i \delta_j^+} \leq 4.$$

(iv) If  $i \in \mathbf{N}$  and  $j \in \mathbf{B}$ , then

$$\frac{\kappa_{ij}^\delta}{\kappa_{ij}^{\delta^+}} \geq 4n^{3.5}.$$

**Proof Part (i).** By Lemma 3.10(i), we see that

$$\begin{aligned} \|\delta_B \Delta x_B^{\parallel}\|_\infty &\leq \|\delta_B \Delta x_B^{\parallel} + \delta_B^{-1} \Delta s_B^{\parallel} + x_B^{1/2} s_B^{1/2}\|_\infty + \|\delta_B^{-1} (\Delta s_B^{\parallel} + s_B)\|_\infty \\ &= \|\delta_B \Delta x_B^{\parallel} + \delta_B^{-1} \Delta s_B^{\parallel} + x_B^{1/2} s_B^{1/2}\|_\infty + \sqrt{\mu} \|R s_B^{\parallel}\|_\infty \\ &\leq \sqrt{\mu} (6n\ell^\delta(\mathcal{J}) + 4n\gamma) \leq 10n\gamma\sqrt{\mu} \leq \sqrt{\mu}/64, \end{aligned}$$

by the assumption on  $\ell^\delta(\mathcal{J})$  and the definition of  $\mathbf{B}$ .

By construction of the LLS step,  $|x_i^+ - x_i| = \alpha^+ |\Delta x_i^{\parallel}| \leq |\Delta x_i^{\parallel}|$ , recalling that  $0 \leq \alpha^+ \leq 1$ . Using the bound derived above, for  $i \in \mathbf{B}$  we get

$$\left| \frac{x_i^+}{x_i} - 1 \right| \leq \left| \frac{\Delta x_i^{\parallel}}{x_i} \right| = \frac{|\delta_i \Delta x_i^{\parallel}|}{\delta_i x_i} \leq \frac{\sqrt{\mu}}{64\delta_i x_i} \leq \frac{1}{32},$$

where the last inequality follows from Proposition 3.2. As

$$\frac{\delta_i^+}{\delta_i} = \sqrt{\frac{x_i^+ s_i^+}{x_i s_i}} \cdot \frac{x_i}{x_i^+} \quad \text{and} \quad \frac{1 - 2\beta}{1 + \beta} \frac{\sqrt{\mu^+}}{\sqrt{\mu}} \leq \sqrt{\frac{x_i^+ s_i^+}{x_i s_i}} \leq \frac{1 + 2\beta}{1 - \beta} \frac{\sqrt{\mu^+}}{\sqrt{\mu}}$$

by Proposition 3.2 the claimed bounds follow with  $\beta \leq 1/8$ .

To get the upper bound on  $\delta_i^{-1} s_i^+$ , again with Proposition 3.2

$$\delta_i^{-1} s_i^+ = \frac{\delta_i^+}{\delta_i \delta_i^+} s_i^+ = \frac{\delta_i^+}{\delta_i} \cdot \sqrt{x_i^+ s_i^+} \leq 2 \sqrt{\frac{\mu^+}{\mu}} \cdot (1 + 2\beta) \sqrt{\mu^+} \leq \frac{3\mu^+}{\sqrt{\mu}}.$$

**Part (ii).** Analogously to (i).

**Part (iii).** Immediate from parts (i) and (ii).

**Part (iv).** Follows by parts (i) and (ii), and by the lower bound on  $\sqrt{\mu/\mu^+}$  obtained from Lemma 3.10(iv) as follows

$$\frac{\kappa_{ij}^\delta}{\kappa_{ij}^{\delta^+}} = \frac{\delta_i^+ \delta_j}{\delta_i \delta_j^+} \geq \frac{\mu}{4\mu^+} = \frac{1}{4(1 - \alpha^+)} \geq \frac{\beta}{12\sqrt{n}\epsilon^{\text{ll}}(w)} \geq 4n^{3.5}.$$

□

**Lemma 4.4 (Restatement).** Let  $w = (x, y, s) \in \mathcal{N}(\beta)$  for  $\beta \in (0, 1/8]$ , and let  $\mathcal{J} = (J_1, \dots, J_p)$  be a  $\delta(w)$ -balanced partition. Assume that  $\xi_{\mathcal{J}}^{\text{ll}}(w) < 4\gamma n$ , and let  $w^+ = (x^+, y^+, s^+) \in \overline{\mathcal{N}}(2\beta)$  be the next iterate obtained by the LLS step with  $\mu^+ = \mu(w^+)$  and assume  $\mu^+ > 0$ . Let  $q \in [p]$  such that  $\xi_{\mathcal{J}}^{\text{ll}}(w) = \xi_{J_q}^{\text{ll}}(w)$ . If  $\ell^{\delta^+}(\mathcal{J}) \leq 4\gamma n$ , then there exist  $i, j \in J_q$  such that  $x_i^* \geq \beta x_i^+ / (16n^{3/2})$  and  $s_j^* \geq \beta s_j^+ / (16n^{3/2})$ . Further, for any  $\ell, \ell' \in J_q$ , we have  $\rho^{\mu^+}(\ell, \ell') \geq -|J_q|$ .

**Proof of Lemma 4.4** Without loss of generality, let  $\xi_{\mathcal{J}}^{\text{ll}} = \xi_{J_q}^{\text{ll}} = \|Rx_{J_q}^{\text{ll}}\|$  for a layer  $q$  with  $J_q \subseteq N$ . The case  $\xi_{J_q}^{\text{ll}} = \|Rs_{J_q}^{\text{ll}}\|$  and  $J_q \subseteq \mathbf{B}$  can be treated analogously.

By Lemma 3.10(iii),  $\|Rs_{J_q}^{\text{ll}}\| \geq \frac{1}{2} - \frac{3}{4}\beta > \frac{1}{4} + 2n\gamma$ , and therefore Lemma 4.2 provides a  $j \in J_q$  such that  $s_j^*/s_j \geq 1/(6\sqrt{n})$ . Using Lemmas 3.3 and 3.1 we find that  $s_j^+/s_j \leq 2n$  and so  $s_j^*/s_j^+ = s_j^*/s_j \cdot s_j/s_j^+ \geq 1/(12n^{3/2}) > 1/(16n^{3/2})$ .

The final statement  $\rho^{\mu^+}(\ell, \ell') \geq -|J_q|$  for any  $\ell, \ell' \in J_q$  is also straightforward. From Lemma 6.1(iii) and the strong connectivity of  $J_q$  in  $G_{\delta, \gamma/n}$ , we obtain that  $J_q$  is strongly connected in  $G_{\delta^+, \gamma/(4n)}$ . Hence,  $\rho^{\mu^+}(\ell, \ell') \geq -|J_q|$  follows by Lemma 4.1.

The rest of the proof is dedicated to showing the existence of an  $i \in J_q$  such that  $x_i^* \geq \beta x_i^+ / (16n^{3/2})$ . For this purpose, we will prove following claim.

**Claim 1**  $\|\delta_{J_q} x_{J_q}^*\| \geq \frac{\beta \mu^+}{8\sqrt{n\mu}}$ .

In order to prove Claim 1, we define

$$z := (\delta^+)^{-1} L_{J_{>q}}^{\delta^+} \left( \delta_{J_{>q}}^+ (x_{J_{>q}}^* - x_{J_{>q}}^+) \right) \quad \text{and } w := x^* - x^+ - z,$$

as in Lemma 5.2. By construction,  $w \in W$  and  $w_{J_{>q}} = 0$ . Thus,  $w_{J_q} \in W_{\mathcal{J}, q}$  as defined in Sect. 3.4.

Using the triangle inequality, we get

$$\|\delta_{J_q} x_{J_q}^*\| \geq \|\delta_{J_q} (x_{J_q}^+ + w_{J_q})\| - \|\delta_{J_q} z_{J_q}\|. \tag{49}$$

We bound the two terms separately, starting with an upper bound on  $\|\delta_{J_q} z_{J_q}\|$ . Since  $\ell^{\delta^+}(\mathcal{J}) \leq 4\gamma n$ , we have with Lemma 5.2 that

$$\begin{aligned} \|\delta_{J_q}^+ z_{J_q}\| &\leq \ell^{\delta^+}(\mathcal{J}) \left\| \delta_{J_{>q}}^+ \left( x_{J_{>q}}^* - x_{J_{>q}}^+ \right) \right\| \\ &\leq 4n\gamma \left\| \delta_{J_{>q}}^+ \left( x_{J_{>q}}^* - x_{J_{>q}}^+ \right) \right\| \\ &= 4n\gamma \left\| \delta_{J_{>q}}^+ x_{J_{>q}}^+ \left( \frac{x_{J_{>q}}^*}{x_{J_{>q}}^+} - e \right) \right\| \\ &\leq 4n\gamma \left( \|\delta^+ x^+\|_\infty \cdot \left\| \frac{x^*}{x^+} \right\|_1 + \sqrt{n\mu^+} \right) \\ &\leq 4n\gamma \left( \frac{3}{2}\sqrt{\mu^+} \cdot \frac{4}{3}n + \sqrt{n\mu^+} \right) \\ &\leq 16n^2\sqrt{\mu^+}\gamma, \end{aligned} \tag{50}$$

where the penultimate inequality follows by Proposition 3.2 and Lemma 3.3. We can use this and Lemma 6.1(ii) to obtain

$$\|\delta_{J_q} z_{J_q}\| \leq \|\delta_{J_q} / \delta_{J_q}^+\|_\infty \cdot \|\delta_{J_q}^+ z_{J_q}\| \leq \frac{32n^2\gamma\mu^+}{\sqrt{\mu}} \leq \frac{\beta\mu^+}{32n^3\sqrt{\mu}}, \tag{51}$$

using the definition of  $\gamma$ .

The first RHS term in (49) will be bounded as follows.

**Claim 2**  $\|\delta_{J_q}(x_{J_q}^+ + w_{J_q})\| \geq \frac{1}{2}\sqrt{\mu}\xi_{\mathcal{J}}^{\text{ll}}$ .

**Proof of Claim 2** We recall the characterization (25) of the LLS step  $\Delta x^{\text{ll}} \in W$ . Namely, there exists  $\Delta s \in W_{\mathcal{J},1}^\perp \times \dots \times W_{\mathcal{J},q}^\perp$  that is the unique solution to  $\delta^{-1}\Delta s + \delta\Delta x^{\text{ll}} = -\delta x$ . From the above, note that

$$\|\delta_{J_q}^{-1}\Delta s_{J_q}\| = \|\delta_{J_q}(x_{J_q} + \Delta x_{J_q}^{\text{ll}})\| = \sqrt{\mu}\|R x_{J_q}^{\text{ll}}\| = \sqrt{\mu}\xi_{\mathcal{J}}^{\text{ll}}.$$

From the Cauchy-Schwarz inequality,

$$\begin{aligned} \|\delta_{J_q}^{-1}\Delta s_{J_q}\| \cdot \|\delta_{J_q}(x_{J_q}^+ + w_{J_q})\| &\geq \left| \left\langle \delta_{J_q}^{-1}\Delta s_{J_q}, \delta_{J_q}(x_{J_q}^+ + w_{J_q}) \right\rangle \right| \\ &= \left| \left\langle \delta_{J_q}^{-1}\Delta s_{J_q}, \delta_{J_q} x_{J_q}^+ \right\rangle \right|. \end{aligned} \tag{52}$$

Here, we used that  $\Delta s_{J_q} \in W_{\mathcal{J},q}^\perp$  and  $w_{J_q} \in W_{\mathcal{J},q}$ . Note that

$$x^+ = x + \alpha\Delta x^{\text{ll}} = x + \Delta x^{\text{ll}} - (1 - \alpha)\Delta x^{\text{ll}} = -\delta^{-2}\Delta s - (1 - \alpha)\Delta x^{\text{ll}}.$$

Therefore,

$$\begin{aligned} \left| \left\langle \delta_{J_q}^{-1} \Delta s_{J_q}, \delta_{J_q} x_{J_q}^+ \right\rangle \right| &= \left| \left\langle \delta_{J_q}^{-1} \Delta s_{J_q}, -\delta_{J_q}^{-1} \Delta s_{J_q} - (1 - \alpha) \delta_{J_q} \Delta x_{J_q}^{\parallel} \right\rangle \right| \\ &\geq \|\delta_{J_q}^{-1} \Delta s_{J_q}\|^2 - (1 - \alpha) \left| \left\langle \delta_{J_q}^{-1} \Delta s_{J_q}, \delta_{J_q} \Delta x_{J_q}^{\parallel} \right\rangle \right|. \end{aligned}$$

By Lemma 5.3, there exists  $\Delta \bar{x} \in W_{\mathcal{J},1} \times \dots \times W_{\mathcal{J},p}$  such that  $\|\delta_{J_q}(\Delta x_{J_q}^{\parallel} - \Delta \bar{x}_{J_q})\| \leq 2n\ell^\delta(\mathcal{J})\sqrt{\mu}$ . Therefore, using the orthogonality of  $\Delta s_{J_q}$  and  $\Delta \bar{x}_{J_q}$ , we get that

$$\left| \left\langle \delta_{J_q}^{-1} \Delta s_{J_q}, \delta_{J_q} \Delta x_{J_q}^{\parallel} \right\rangle \right| = \left| \left\langle \delta_{J_q}^{-1} \Delta s_{J_q}, \delta_{J_q}(\Delta x_{J_q}^{\parallel} - \Delta \bar{x}_{J_q}^{\parallel}) \right\rangle \right| \leq 2n\ell^\delta(\mathcal{J})\sqrt{\mu} \cdot \|\delta_{J_q}^{-1} \Delta s_{J_q}\|.$$

From the above inequalities, we see that

$$\|\delta_{J_q}(x_{J_q}^+ + w_{J_q})\| \geq \|\delta_{J_q}^{-1} \Delta s_{J_q}\| - 2(1 - \alpha)n\ell^\delta(\mathcal{J})\sqrt{\mu} = \sqrt{\mu}\xi_{\mathcal{J}}^{\parallel} - 2(1 - \alpha)n\ell^\delta(\mathcal{J})\sqrt{\mu}.$$

It remains to show  $(1 - \alpha)n\ell^\delta(\mathcal{J}) \leq \xi_{\mathcal{J}}^{\parallel}/4$ . From Lemma 3.10(iv), we obtain

$$(1 - \alpha)n\ell^\delta(\mathcal{J}) \leq 3n^{3/2}\ell^\delta(\mathcal{J})\xi_{\mathcal{J}}^{\parallel}\beta^{-1},$$

using  $\xi_{\mathcal{J}}^{\parallel} \geq \varepsilon^{\parallel}$ . The claim now follows by the assumption  $\ell^\delta(\mathcal{J}) \leq \gamma$ , and the choice of  $\gamma$ . □

**Proof of Claim 1** Using Lemma 3.10(iv),

$$\mu^+ \leq \frac{3\sqrt{n}\xi_{\mathcal{J}}^{\parallel}\mu}{\beta},$$

implying  $\|\delta_{J_q}(x_{J_q}^+ + w_{J_q})\| \geq \beta\mu^+/(6\sqrt{n\mu})$  by Claim 2. Now the claim follows using (49) and (51). □

By Lemma 6.1(ii), we see that

$$\|\delta_{J_q} x_{J_q}^+\| \leq \sqrt{n}\|\delta_{J_q} x_{J_q}^+\|_\infty \leq \frac{3\sqrt{n}\mu^+}{\sqrt{\mu}}.$$

Thus, the lemma follows immediately from Claim 1: for at least one  $i \in J_q$ , we must have

$$\frac{x_i^*}{x_i} \geq \frac{\|\delta_{J_q} x_{J_q}^*\|}{\|\delta_{J_q} x_{J_q}^+\|} \geq \frac{\beta}{24n} \geq \frac{\beta}{16n^{3/2}}.$$

□

**Lemma 4.5** (Restatement). *Let  $w = (x, y, s) \in \mathcal{N}(\beta)$  for  $\beta \in (0, 1/8]$ , and let  $\mathcal{J} = (J_1, \dots, J_p)$  be a  $\delta(w)$ -balanced partition. Assume that  $\xi_{\mathcal{J}}^{\text{ll}}(w) < 4\gamma n$ , and let  $w^+ = (x^+, y^+, s^+) \in \overline{\mathcal{N}}(2\beta)$  be the next iterate obtained by the LLS step with  $\mu^+ = \mu(w^+)$  and assume  $\mu^+ > 0$ . If  $\ell^{\delta^+}(\mathcal{J}) > 4\gamma n$ , then there exist two layers  $J_q$  and  $J_r$  and  $i \in J_q$  and  $j \in J_r$  such that  $x_i^* \geq x_i^+/(8n^{3/2})$ , and  $s_j^* \geq s_j^+/(8n^{3/2})$ . Further,  $\rho^{\mu^+}(i, j) \geq -|J_q \cup J_r|$ , and for all  $\ell, \ell' \in J_q \cup J_r$ ,  $\ell \neq \ell'$  we have  $\Psi^\mu(\ell, \ell') \leq |J_q \cup J_r|$ .*

**Proof of Lemma 4.5** Recall the sets  $\mathbf{B}$  and  $N$  defined in (48). The key is to show the existence of an edge

$$(i', j') \in E_{\delta^+, \gamma/(4n)} \quad \text{such that} \quad i' \in J_q \subseteq \mathbf{B}, \quad j' \in J_r \subseteq N, \quad r < q. \quad (53)$$

Before proving the existence of such  $i'$  and  $j'$ , we show how the rest of the statements follow. Note that  $x^+ \leq (1 - \beta)^{-1}(1 + 2 \cdot 2\beta)nx \leq \frac{7}{4}nx$  by Lemma 3.3 and Proposition 3.1. Further, we have  $\|R_{x_{J_q}}^{\text{ll}}\| - 2\gamma n \geq \frac{1}{2} - \frac{3}{4}\beta - 2\gamma n \geq \frac{2}{5}$  by Lemma 3.10 (iii). The existence of  $i \in J_q$  such that  $x_i^* \geq x_i^+/(8n^{3/2})$  now follows immediately from Lemma 4.2, as there is an  $i \in J_q$  such that

$$x_i^* \geq \frac{2x_i}{3\sqrt{n}} \cdot (\|R_{x_{J_q}}^{\text{ll}}\| - 2\gamma n) \geq \frac{2}{3\sqrt{n}} \frac{4x_i^+}{7n} \frac{2}{5} \geq \frac{x_i^+}{8n^{3/2}}. \quad (54)$$

With analogous argumentation it can be shown that there exists  $j \in J_r$  such that  $s_j^* \geq s_j^+/(8n^{3/2})$ . The other statements are that  $\rho^{\mu^+}(i, j) \geq -|J_q \cup J_r|$ , and for each  $\ell, \ell' \in J_q \cup J_r$ ,  $\ell \neq \ell'$ ,  $\Psi^\mu(\ell, \ell') \leq |J_q \cup J_r|$ . According to Lemma 4.1, the latter is true (even with the stronger bound  $\max\{|J_q|, |J_r|\}$ ) whenever  $\ell, \ell' \in J_q$ , or  $\ell, \ell' \in J_r$ , or if  $\ell \in J_q$  and  $\ell' \in J_r$ . It is left to show the lower bound on  $\rho^{\mu^+}(i, j)$  and  $\Psi^\mu(\ell, \ell') \leq |J_q \cup J_r|$  for  $\ell' \in J_q$  and  $\ell \in J_r$ .

From Lemma 6.1(iii), we have that if  $\ell, \ell' \in J_q \subseteq \mathbf{B}$  or  $\ell, \ell' \in J_r \subseteq N$ , then  $\kappa_{\ell\ell'}^\delta/4 \leq \kappa_{\ell\ell'}^{\delta^+}$ . Hence, the strong connectivity of  $J_r$  and  $J_q$  in  $G_{\delta, \gamma}$  implies the strong connectivity of these sets in  $G_{\delta^+, \gamma/(4n)}$ . Together with the edge  $(i', j')$ , we see that every  $\ell' \in J_q$  can reach every  $\ell \in J_r$  on a directed path of length  $\leq |J_q \cup J_r| - 1$  in  $G_{\delta^+, \gamma/(4n)}$ . Applying Lemma 4.1 for this setting, we obtain  $\Psi^\mu(\ell, \ell') \leq \rho^{\mu^+}(\ell, \ell') \leq |J_q \cup J_r|$  for all such pairs, and also  $\rho^{\mu^+}(i, j) \geq -|J_q \cup J_r|$ .

The rest of the proof is dedicated to showing the existence of  $i'$  and  $j'$  as in (53). We let  $k \in [p]$  such that  $\ell^{\delta^+}(J_{\geq k}) = \ell^{\delta^+}(\mathcal{J}) > 4n\gamma$ . To simplify the notation, we let  $I = J_{\geq k}$ .

When constructing  $\mathcal{J}$  in  $\text{LAYERING}(\delta, \hat{\kappa})$ , the subroutine  $\text{VERIFY-LIFT}(\text{Diag}(\delta)W, I, \gamma)$  was called for the set  $I = J_{\geq k}$ , with the answer ‘pass’. Besides  $\ell^\delta(I) \leq \gamma$ , this guaranteed the stronger property that  $\max_{ji} |B_{ji}| \leq \gamma$  for the matrix  $B$  implementing the lift (see Remark 2.17).

Let us recall how this matrix  $B$  was obtained. The subroutine starts by finding a minimal  $I' \subset I$  such that  $\dim(\pi_{I'}(W)) = \dim(\pi_I(W))$ . Recall that  $\pi_{I'}(W) = \mathbb{R}^{I'}$  and  $L_i^\delta(p) = L_{i'}^\delta(p_{I'})$  for every  $p \in \pi_I(\text{Diag}(\delta)W)$ .



Consider the optimal lifting  $L_I^\delta : \pi_I(\text{Diag}(\delta)W) \rightarrow \text{Diag}(\delta)W$ . We defined  $B \in \mathbb{R}^{([n]\setminus I) \times I'}$  as the matrix sending any  $q \in \pi_{I'}(\text{Diag}(\delta)W)$  to the corresponding vector  $[L_{I'}^\delta(q)]_{[n]\setminus I}$ . The column  $B_i$  can be computed as  $[L_{I'}^\delta(e^i)]_{[n]\setminus I}$  for  $e^i \in \mathbb{R}^{I'}$ .

We consider the transformation

$$\bar{B} := \text{Diag}(\delta^+ \delta^{-1}) B \text{Diag}((\delta_{I'}^+)^{-1} \delta_{I'}).$$

This maps  $\pi_{I'}(\text{Diag}(\delta^+)W) \rightarrow \pi_{[n]\setminus I}(\text{Diag}(\delta^+)W)$ .

Let  $z \in \pi_{I'}(\text{Diag}(\delta^+)W)$  be the singular vector corresponding to the maximum singular value of  $L_{I'}^{\delta^+}$ , namely,  $\|[L_{I'}^{\delta^+}(z)]_{[n]\setminus I}\| > 4n\gamma\|z\|$ . Let us normalize  $z$  such that  $\|z_{I'}\| = 1$ . Thus,

$$\|[L_{I'}^{\delta^+}(z_{I'})]_{[n]\setminus I}\| > 4n\gamma.$$

Let us now apply  $\bar{B}$  to  $z_{I'} \in \pi_{I'}(\text{Diag}(\delta^+)W)$ . Since  $L_{I'}^{\delta^+}$  is the minimum-norm lift operator, we see that

$$\|\bar{B}z_{I'}\| \geq \|[L_{I'}^{\delta^+}(z_{I'})]_{[n]\setminus I}\| > 4n\gamma.$$

We can upper bound the operator norm by the Frobenius norm  $\|\bar{B}\| \leq \|\bar{B}\|_F = \sqrt{\sum_{ji} \bar{B}_{ji}^2} \leq n \max_{ji} |\bar{B}_{ji}|$ , and therefore

$$\max_{ji} |\bar{B}_{ji}| > 4\gamma.$$

Let us fix  $i' \in I'$  and  $j' \in [n]\setminus I$  as the indices giving the maximum value of  $\bar{B}$ . Note that  $\bar{B}_{j'i'} = B_{j'i'} \delta_{j'}^+ \delta_{i'} / (\delta_{i'}^+ \delta_{j'})$ .

Let us now use Lemma 2.16 for the pair  $i', j'$ , the matrix  $B$  and the subspace  $\text{Diag}(\delta)W$ . Noting that  $B_{j'i'} = [L_{I'}^\delta(e^{i'})]_{j'}$ , we obtain  $\kappa_{i'j'}^\delta \geq |B_{j'i'}|$ . Now,

$$\kappa_{i'j'}^{\delta^+} = \kappa_{i'j'}^\delta \cdot \frac{\delta_{j'}^+ \delta_{i'}}{\delta_{i'}^+ \delta_{j'}} \geq |B_{j'i'}| \cdot \frac{\delta_{j'}^+ \delta_{i'}}{\delta_{i'}^+ \delta_{j'}} = |\bar{B}_{j'i'}| > 4\gamma. \tag{55}$$

The next claim finishes the proof. □

**Claim 6.2** For  $i'$  and  $j'$  selected as above, (53) holds.

**Proof**  $(i', j') \in E_{\delta^+, \gamma/(4n)}$  holds by (55). From the above, we have

$$|B_{j'i'}| > 4\gamma \cdot \frac{\delta_{i'}^+ \delta_{j'}}{\delta_{i'} \delta_{j'}^+}.$$

According to Remark 2.17,  $|B_{j'i'}| \leq \gamma$  follows since VERIFY-LIFT(Diag( $\delta$ ) $W, I, \gamma$ ) returned with ‘pass’. We thus have

$$\frac{\delta_{i'}^+ \delta_{j'}}{\delta_{i'} \delta_{j'}^+} < \frac{1}{4}.$$

Lemma 6.1 excludes the scenarios  $i', j' \in N, i', j' \in B$ , and  $i' \in N, j' \in B$ , leaving  $i' \in B$  and  $j' \in N$  as the only possibility. Therefore,  $i' \in J_q \subseteq B$  and  $j' \in J_r \subseteq N$ . We have  $r < q$  since  $i \in I = J_{\geq k}$  and  $j \in [n] \setminus I = J_{< k}$ . □

### 7 Initialization

Our main algorithm (Algorithm 2 in Sect. 3.6), requires an initial solution  $w^0 = (x^0, y^0, s^0) \in \mathcal{N}(\beta)$ . In this section, we remove this assumption by adapting the initialization method of [63] to our setting.

We use the “big- $M$  method”, a standard initialization approach for path-following interior point methods that introduces an auxiliary system whose optimal solutions map back to the optimal solutions of the original system. The primal-dual system we consider is

$$\begin{array}{ll} \min & c^\top x + Me^\top \underline{x} & \max & y^\top b + 2Me^\top z \\ & Ax - A\underline{x} = b & & A^\top y + z + s = c \\ & x + \bar{x} = 2Me & & z + \bar{s} = 0 \\ & x, \bar{x}, \underline{x} \geq 0 & & -A^\top y + \underline{s} = Me \\ & & & s, \bar{s}, \underline{s} \geq 0. \end{array} \tag{Init-LP}$$

The constraint matrix used in this system is

$$\hat{A} = \begin{pmatrix} A & -A & 0 \\ I & 0 & I \end{pmatrix}$$

The next lemma asserts that the  $\bar{\chi}$  condition number of  $\hat{A}$  is not much bigger than that of  $A$  of the original system (LP).

**Lemma 7.1** [63, Lemma 23]  $\bar{\chi}_{\hat{A}} \leq 3\sqrt{2}(\bar{\chi}_A + 1)$ .

We extend this bound for  $\bar{\chi}^*$ .

**Lemma 7.2**  $\bar{\chi}_{\hat{A}}^* \leq 3\sqrt{2}(\bar{\chi}_A^* + 1)$ .

**Proof** Let  $D \in \mathbf{D}_n$  and let  $\hat{D} \in \mathbf{D}_{3n}$  the matrix consisting of three copies of  $D$ , i.e.

$$\hat{D} = \begin{pmatrix} D & 0 & 0 \\ 0 & D & 0 \\ 0 & 0 & D \end{pmatrix}.$$

Then

$$\hat{A}\hat{D} = \begin{pmatrix} AD & -AD & 0 \\ D & 0 & 0 \end{pmatrix}$$

Row-scaling does not change  $\bar{\chi}$  as the kernel of the matrix remains unchanged. Thus, we can rescale the last  $n$  rows of  $\hat{A}\hat{D}$ , to the identity matrix, i.e. multiplying by  $(I, D^{-1})$  from the left hand side. We observe that

$$\bar{\chi}_{\hat{A}\hat{D}} = \bar{\chi} \left( \begin{pmatrix} AD & -AD & 0 \\ I & 0 & I \end{pmatrix} \right) \leq 3\sqrt{2}(\bar{\chi}_{AD} + 1)$$

where the inequality follows from Lemma 7.1. The lemma now readily follows as

$$\bar{\chi}_{\hat{A}}^* = \inf\{\bar{\chi}_{\hat{A}\hat{D}} : D \in \mathbf{D}_{3n}\} \leq \inf\{3\sqrt{2}(\bar{\chi}_{AD} + 1) : D \in \mathbf{D}_n\} = 3\sqrt{2}(\bar{\chi}_A^* + 1).$$

□

We show next that the optimal solutions of the original system are preserved for sufficiently large  $M$ . We let  $d$  be the min-norm solution to  $Ax = b$ , i.e.,  $d = A^\top(AA^\top)^{-1}b$ .

**Proposition 7.3** *Assume both primal and dual of (LP) are feasible, and  $M > \max\{(\bar{\chi}_A + 1)\|c\|, \bar{\chi}_A\|d\|\}$ . Every optimal solution  $(x, y, s)$  to (LP), can be extended to an optimal solution  $(x, \underline{x}, \bar{x}, y, z, s, \underline{s}, \bar{s})$  to (Init-LP); and conversely, from every optimal solution  $(x, \underline{x}, \bar{x}, y, z, s, \underline{s}, \bar{s})$  to (Init-LP), we obtain an optimal solution  $(x, y, s)$  by deleting the auxiliary variables.*

**Proof** If system (LP) is feasible, it admits a basic optimal solution  $(x^*, y^*, s^*)$  with basis  $B$  such that  $A_B x_B^* = b, x^* \geq 0, A_B^\top y^* = c$  and  $A^\top y^* \leq c$ . Using Proposition 2.1(ii) we see that

$$\|x_B^*\| = \|A_B^{-1}b\| = \|A_B^{-1}Ad\| \leq \bar{\chi}_A\|d\| < M, \tag{56}$$

and using that  $\|A\| = \|A^\top\|$  we observe

$$\|A^\top y^*\| = \|A^\top A_B^{-\top}c\| \leq \|A^\top A_B^{-\top}\| \|c\| = \|A_B^{-1}A\| \|c\| \leq \bar{\chi}_A\|c\| < M. \tag{57}$$

We can extend this solution to a solution of system (Init-LP) via setting  $\bar{x}^* = 2Me - x^*, \underline{x}^* = 0, z^* = \bar{s}^* = 0$  and  $\underline{s}^* = Me + A^\top y^*$ . Observe that  $\bar{x}^* > 0$  and  $\underline{s}^* > 0$  by (56) and (57). Furthermore observe that by complementary slackness this extended solution for (Init-LP) is an optimal solution. The property that  $\underline{s}^* > 0$  immediately tells us that  $\underline{x}$  vanishes for all optimal solutions of (Init-LP) and thus all optimal solutions of (LP) coincide with the optimal solutions of (Init-LP), with the auxiliary variables removed. □

The next lemma is from [36, Lemma 4.4]. Recall that  $w = (x, y, s) \in \mathcal{N}(\beta)$  if  $\|xs/\mu(w) - e\| \leq \beta$ .

**Lemma 7.4** Let  $w = (x, y, s) \in \mathcal{P}^{++} \times \mathcal{D}^{++}$ , and let  $v > 0$ . Assume that  $\|xs/v - e\| \leq \tau$ . Then  $(1 - \tau/\sqrt{n})v \leq \mu(w) \leq (1 + \tau/\sqrt{n})v$  and  $w \in \mathcal{N}(\tau/(1 - \tau))$ .

The new system has the advantage that we can easily initialize the system with a feasible solution in close proximity to central path:

**Proposition 7.5** We can initialize system (Init-LP) close to the central path with initial solution  $w^0 = (x^0, y^0, s^0) \in \mathcal{N}(1/8)$  and parameter  $\mu(w^0) \approx M^2$  if  $M > 15 \max\{(\bar{\chi}_A + 1)\|c\|, \bar{\chi}_A\|d\|\}$ .

**Proof** The initialization follows along the lines of [63, Section 10]. We let  $d$  as above, and set

$$\begin{aligned} \bar{x}^0 &= Me, x^0 = Me, \underline{x}^0 = Me - d \\ y^0 &= 0, z^0 = -Me \\ \bar{s}^0 &= Me, s^0 = Me + c, \underline{s}^0 = Me. \end{aligned}$$

This is a feasible primal-dual solution to system (Init-LP) with parameter

$$\mu^0 = (3n)^{-1}(\langle x^0, s^0 \rangle + \langle \underline{x}^0, \underline{s}^0 \rangle + \langle \bar{x}^0, \bar{s}^0 \rangle) = (3n)^{-1}(3nM^2 + Mc^\top e - Md^\top e) \approx M^2.$$

We see that

$$\left\| \frac{1}{M^2} \begin{pmatrix} \bar{x}^0 \bar{s}^0 \\ x^0 s^0 \\ \underline{x}^0 \underline{s}^0 \end{pmatrix} - e \right\|^2 = M^{-2}\|c\|^2 + M^{-2}\|d\|^2 \leq \frac{1}{9^2 \bar{\chi}_A^2} \leq \frac{1}{9^2}.$$

With Lemma 7.4 we conclude that  $w^0 = (x^0, y^0, s^0) \in \mathcal{N}\left(\frac{1/9}{1-1/9}\right) = \mathcal{N}(1/8)$ .  $\square$

**Detecting infeasibility** To use the extended system (Init-LP), we still need to assume that both the primal and dual programs in (LP) are feasible. For arbitrary instances, we first need to check if this is the case, or conclude that the primal or the dual (or both) are infeasible.

This can be done by employing a two-phase method. The first phase decides feasibility by running (Init-LP) with data  $(A, b, 0)$  and  $M > \bar{\chi}_A\|d\|_1$ . The objective value of the optimal primal-dual pair is 0 if and only if (LP) has a feasible solution. If the optimal primal/dual solution  $(x^*, \underline{x}^*, \bar{x}^*, y^*, z^*, s^*, \underline{s}^*, \bar{s}^*)$  has positive objective value, we can extract an infeasibility certificate in the following way.

We can w.l.o.g. assume that  $x^*$  is supported on some basis  $B$  of  $A$ . Note that the objective function of the primal is equivalent to  $\|\underline{x}\|_1$ . Therefore, clearly  $\|\underline{x}^*\|_1 \leq -\sum_{i:d_i < 0} d_i \leq \|d\|_1$  and so  $\|\underline{x}^*\| \leq \|d\|_1$ . Due to the constraint  $Ax^* - A\underline{x}^* = b = Ad$  we get that

$$\|x^*\| = \|B^{-1}A(d + \underline{x}^*)\| \leq \|B^{-1}A\|(\|d\| + \|\underline{x}^*\|) \leq 2\bar{\chi}_A\|d\|_1. \tag{58}$$

Therefore, if  $M > \bar{\chi}_A\|d\|_1$ , then  $\bar{x}^* = 2Me - \|x^*\| > 0$  so by strong duality,  $\bar{s}^* = 0$ . From the dual, we conclude that  $z^* = 0$ , and therefore  $A^\top y^* \leq A^\top y^* + s^* + z^* =$

$c = 0$ . On the other hand, by assumption the objective value of the dual is positive, and so  $(y^*)^\top b \geq (y^*)^\top b + 2Me^\top z^* > 0$ . Hence,  $y^*$  is the desired certificate.

Feasibility of the dual of (LP) can be decided by running (Init-LP) on data  $(A, 0, c)$  and  $M > (\bar{\chi}_A + 1)\|c\|$  with the same argumentation: Either the objective value of the dual is 0 and therefore the dual optimal solution  $(y^*, z^*, s^*, \bar{s}^*)$  corresponds to a feasible dual solution of (LP) or the objective value is negative and we extract a dual infeasibility certificate in the following way: For the optimal corresponding primal solution  $(x^*, \underline{x}^*, \bar{x}^*)$  we have by assumption  $c^\top x^* \leq c^\top x^* + Me^\top \underline{x}^* < 0$ . Furthermore, w.l.o.g. the support of  $s^*$  is contained in a basis which allows us to conclude that  $\bar{s}^* > 0$  and therefore  $\underline{x}^* = 0$ . So we have  $Ax^* = 0 + A\underline{x}^* = 0$ , which together with  $c^\top x^* < 0$  yields the certificate of dual infeasibility.

**Finding the right value of  $M$**  While Algorithm 2 does not require any estimate on  $\bar{\chi}^*$  or  $\bar{\chi}$ , the initialization needs to set  $M \geq \max\{(\bar{\chi}_A + 1)\|c\|, \bar{\chi}_A \|d\|\}$  as in Proposition 7.3.

A straightforward guessing approach (attributed to Renegar in [63]) starts with a constant guess, say  $\bar{\chi}_A = 100$ , constructs the extended system, and runs the algorithm. In case the optimal solution to the extended system does not map to an optimal solution of (LP), we restart with  $\bar{\chi}_A = 100^2$  and try again; we continue squaring the guess until an optimal solution is found.

This would still require a series of  $\log \log \bar{\chi}_A$  guesses, and thus, result in a dependence on  $\bar{\chi}_A$  in the running time. However, if we initially rescale our system using the near-optimal rescaling Theorem 2.5, then we can turn the dependence from  $\bar{\chi}_A$  to  $\bar{\chi}_A^*$ . The overall iteration complexity remains  $O(n^{2.5} \log n \log(\bar{\chi}_A^* + n))$ , since the running time for the final guess on  $\bar{\chi}_A^*$  dominates the total running time of all previous computations due to the repeated squaring.

An alternative approach, that does not rescale the system, is to use Theorem 2.5 to approximate  $\bar{\chi}_A$ . In this case we repeatedly square a guess of  $\bar{\chi}_A^*$  instead of  $\bar{\chi}_A$  which takes  $O(\log \log \bar{\chi}_A^*)$  iterations until our guess corresponds to a valid upper bound for  $\bar{\chi}_A$ .

Note that either guessing technique can handle bad guesses gracefully. For the first phase, if neither a feasible solution to (LP) is returned nor a Farkas' certificate can be extracted, we have proof that the guess was too low by the above paragraph. Similarly, in phase two, when feasibility was decided in the affirmative for primal and dual, an optimal solution to (Init-LP) that corresponds to an infeasible solution to (LP) serves as a certificate that another squaring of the guess is necessary.

**Acknowledgements** The authors are grateful to the anonymous reviewers for their comments that helped to improve the presentation.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Allamigeon, X., Benchimol, P., Gaubert, S., Joswig, M.: Log-barrier interior point methods are not strongly polynomial. *SIAM Journal on Applied Algebra and Geometry* **2**(1), 140–178 (2018)
2. Allamigeon, X., Dadush, D., Loho, G., Natura, B., Véghe, L.A.: Interior point methods are not worse than simplex. In: *Proceedings of the 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pp. 267–277. IEEE (2022)
3. Allamigeon, X., Gaubert, S., Vandame, N.: No self-concordant barrier interior point method is strongly polynomial. In: *Proceedings of the 54th Annual ACM Symposium on Theory of Computing (STOC)*, pp. 515–528 (2022)
4. Ahuja, R.K., Magnanti, T.L., Orlin, J.B.: *Network Flows: Theory, Algorithms, and Applications*. Prentice-Hall Inc., New York (1993)
5. Bubeck, S., Eldan, R.: The entropic barrier: a simple and optimal universal self-concordant barrier. [arXiv:1412.1587](https://arxiv.org/abs/1412.1587) (2014)
6. Chubanov, S.: A polynomial algorithm for linear optimization which is strongly polynomial under certain conditions on optimal solutions. [http://www.optimization-online.org/DB\\_HTML/2014/12/4710.html](http://www.optimization-online.org/DB_HTML/2014/12/4710.html) (2014)
7. Cohen, M.B., Lee, Y.T., Song, Z.: Solving linear programs in the current matrix multiplication time. In: *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pp. 938–942 (2019)
8. Dadush, D., Huiberts, S., Natura, B., Véghe, L.A.: A scaling-invariant algorithm for linear programming whose running time depends only on the constraint matrix. In: *Proceedings of the 52nd Annual ACM Symposium on Theory of Computing (STOC)*, pp. 761–774 (2020)
9. Dikin, I.: Iterative solution of problems of linear and quadratic programming. *Dokl. Akad. Nauk SSSR* **174**(4), 747–748 (1967)
10. Dikin, I.: On the speed of an iterative process. *Upravlyaemye Sistemi* **12**(1), 54–60 (1974)
11. Dadush, D., Koh, Z.K., Natura, B., Véghe, L.A.: On circuit diameter bounds via circuit imbalances. In: *Proceedings of the 23rd Integer Programming and Combinatorial Optimization Conference (IPCO)*, pp. 140–153. Springer (2022)
12. De Loera, J.A., Hemmecke, R., Lee, J.: On augmentation algorithms for linear and integer-linear programming: From edmonds-karp to bland and beyond. *SIAM J. Optim.* **25**(4), 2494–2511 (2015)
13. De Loera, J.A., Kafer, S., Sanita, L.: Pivot rules for circuit-augmentation algorithms in linear optimization. *SIAM J. Optim.* **32**(3), 2156–2179 (2022)
14. Dadush, D., Natura, B., Véghe, L.A.: Revisiting Tardos’s framework for linear programming: faster exact solutions using approximate solvers. In: *Proceedings of the 61st Annual Symposium on Foundations of Computer Science*, pp. 931–942. IEEE (2020)
15. Daitch, S.I., Spielman, D.A.: Faster approximate lossy generalized flow via interior point algorithms. In: *Proceedings of the 40th Annual ACM Symposium on Theory of Computing*, pp. 451–460 (2008)
16. Ekbatani, F., Natura, B., Véghe, A.L.: Circuit imbalance measures and linear programming. In: *Surveys in Combinatorics 2022*, London Mathematical Society Lecture Note Series, pp. 64–114. Cambridge University Press (2022)
17. Frank, A.: *Connections in Combinatorial Optimization*. Number 38 in Oxford Lecture Series in Mathematics and its Applications. Oxford University Press (2011)
18. Gonzaga, C.C., Lara, H.J.: A note on properties of condition numbers. *Linear Algebra Appl.* **261**(1), 269–273 (1997)
19. Goffin, J.-L.: The relaxation method for solving systems of linear inequalities. *Math. Oper. Res.* **5**(3), 388–414 (1980)
20. Gonzaga, C.C.: Path-following methods for linear programming. *SIAM Rev.* **34**(2), 167–224 (1992)
21. Ho, J.C., Tunçel, L.: Reconciliation of various complexity and condition measures for linear programming problems and a generalization of Tardos’ theorem. In: *Foundations of Computational Mathematics*, pp. 93–147. World Scientific (2002)
22. Karmarkar, N.: A new polynomial-time algorithm for linear programming. In: *Proceedings of the 16th Annual ACM Symposium on Theory of Computing (STOC)*, pp. 302–311 (1984)
23. Khachiyan, L.G.: A polynomial algorithm in linear programming. In *Doklady Akademii Nauk SSSR* **244**, 1093–1096 (1979)
24. Kitahara, T., Mizuno, S.: A bound for the number of different basic solutions generated by the simplex method. *Math. Program.* **137**(1–2), 579–586 (2013)

25. Kakahara, S., Ohara, A., Tsuchiya, T.: Information geometry and interior-point algorithms in semidefinite programs and symmetric cone programs. *J. Optim. Theory Appl.* **157**, 749–780 (2013)
26. Kakahara, S., Ohara, A., Tsuchiya, T.: Curvature integrals and iteration complexities in SDP and symmetric cone programs. *Comput. Optim. Appl.* **57**, 623–665 (2014)
27. Kitahara, T., Tsuchiya, T.: A simple variant of the Mizuno-Todd-Ye predictor-corrector algorithm and its objective-function-free complexity. *SIAM J. Optim.* **23**(3), 1890–1903 (2013)
28. Lan, G., Monteiro, R.D., Tsuchiya, T.: A polynomial predictor-corrector trust-region algorithm for linear programming. *SIAM J. Optim.* **19**(4), 1918–1946 (2009)
29. Lee, Y.T., Sidford, A.: Path finding methods for linear programming: solving linear programs in  $\tilde{O}(\sqrt{\text{rank}})$  iterations and faster algorithms for maximum flow. In: *Proceedings of the 55th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pp. 424–433 (2014)
30. Lee, Y.T., Sidford, A.: Efficient inverse maintenance and faster algorithms for linear programming. In: *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*, pp. 230–249 (2015)
31. Lee, Y.T., Sidford, A.: Solving linear programs with  $\tilde{O}(\sqrt{\text{rank}})$  linear system solves. *arXiv preprint arXiv:1910.08033* (2019)
32. Madry, A.: Navigating central path with electrical flows: From flows to matchings, and back. In: *Proceedings of the 54th IEEE Annual Symposium on Foundations of Computer Science*, pp. 253–262. *IEEE* (2013)
33. Megiddo, N.: Towards a genuinely polynomial algorithm for linear programming. *SIAM J. Comput.* **12**(2), 347–353 (1983)
34. Mehrotra, S.: On the implementation of a primal-dual interior point method. *SIAM J. Optim.* **2**(4), 575–601 (1992)
35. Megiddo, N., Mizuno, S., Tsuchiya, T.: A modified layered-step interior-point algorithm for linear programming. *Math. Program.* **82**(3), 339–355 (1998)
36. Monteiro, R.D.C., Tsuchiya, T.: A variant of the Vavasis-Ye layered-step interior-point algorithm for linear programming. *SIAM J. Optim.* **13**(4), 1054–1079 (2003)
37. Monteiro, R.D.C., Tsuchiya, T.: A new iteration-complexity bound for the MTY predictor-corrector algorithm. *SIAM J. Optim.* **15**(2), 319–347 (2005)
38. Monteiro, R.D., Tsuchiya, T.: A strong bound on the integral of the central path curvature and its relationship with the iteration-complexity of primal-dual path-following LP algorithms. *Math. Program.* **115**(1), 105–149 (2008)
39. Mizuno, S., Todd, M., Ye, Y.: On adaptive-step primal-dual interior-point algorithms for linear programming. *Math. Oper. Res. MOR* **18**, 964–981 (1993)
40. O’Leary, D.P.: On bounds for scaled projections and pseudoinverses. *Linear Algebra Appl.* **132**, 115–117 (1990)
41. Olver, N., Végh, L.A.: A simpler and faster strongly polynomial algorithm for generalized flow maximization. *Journal of the ACM (JACM)* **67**(2), 1–26 (2020)
42. Renegar, J.: A polynomial-time algorithm, based on Newton’s method, for linear programming. *Math. Program.* **40**(1–3), 59–93 (1988)
43. Renegar, J.: Is it possible to know a problem instance is ill-posed?: some foundations for a general theory of condition numbers. *J. Complex.* **10**(1), 1–56 (1994)
44. Renegar, J.: Incorporating condition measures into the complexity theory of linear programming. *SIAM J. Optim.* **5**(3), 506–524 (1995)
45. Schrijver, A.: *Combinatorial Optimization—Polyhedra and Efficiency*. Springer, Berlin (2003)
46. Smale, S.: Mathematical problems for the next century. *The Mathematical Intelligencer* **20**, 7–15 (1998)
47. Sonnevend, G., Stoer, J., Zhao, G.: On the complexity of following the central path of linear programs by linear extrapolation II. *Math. Program.* **52**(1–3), 527–553 (1991)
48. Spielman, D.A., Teng, S.-H.: 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)
49. Stewart, G.: On scaled projections and pseudoinverses. *Linear Algebra Appl.* **112**, 189–193 (1989)
50. Tardos, É.: A strongly polynomial minimum cost circulation algorithm. *Combinatorica* **5**(3), 247–255 (1985)
51. Tardos, É.: A strongly polynomial algorithm to solve combinatorial linear programs. *Oper. Res.* **34**: 250–256 (1986)
52. Todd, M.J.: A Dantzig-Wolfe-like variant of Karmarkar’s interior-point linear programming algorithm. *Oper. Res.* **38**(6), 1006–1018 (1990)

53. Todd, M.J., Tunçel, L., Ye, Y.: Characterizations, bounds, and probabilistic analysis of two complexity measures for linear programming problems. *Math. Program.* **90**(1), 59–69 (2001)
54. Tunçel, L.: Approximating the complexity measure of Vavasis-Ye algorithm is NP-hard. *Math. Program.* **86**(1), 219–223 (1999)
55. Vaidya, P.M.: Speeding-up linear programming using fast matrix multiplication. In: *Proceedings of the 30th IEEE Annual Symposium on Foundations of Computer Science*, pp. 332–337 (1989)
56. Vavasis, S.A.: Stable numerical algorithms for equilibrium systems. *SIAM J. Matrix Anal. Appl.* **15**(4), 1108–1131 (1994)
57. van den Brand, J.: A deterministic linear program solver in current matrix multiplication time. In: *Proceedings of the Symposium on Discrete Algorithms (SODA)*, pp. 259–278. SIAM (2020)
58. van den Brand, J., Liu, Y.P., Lee, Y.-T., Saranurak, T., Sidford, A., Song, Z., Wang, D.: Minimum cost flows, MDPs, and L1-regression in nearly linear time for dense instances. In: *STOC (to appear)* (2021)
59. van den Brand, J., Lee, Y.-T., Nanongkai, D., Peng, R., Saranurak, T., Sidford, A., Song, Z., Wang, D.: Bipartite matching in nearly-linear time on moderately dense graphs. In: *IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pp. 919–930 (2020)
60. van den Brand, J., Tat Lee, Y., Sidford, A., Song, Z.: Solving tall dense linear programs in nearly linear time. In: *Proceedings of the 52nd Annual ACM Symposium on Theory of Computing (STOC)*, pp. 775–788 (2020)
61. Végh, L.A.: A strongly polynomial algorithm for generalized flow maximization. *Math. Oper. Res.* **42**(2), 179–211 (2017)
62. Vassilevska Williams, V.: Multiplying matrices faster than coppersmith-winograd. In: *Proceedings of the 44th Annual ACM Symposium on Theory of Computing*, pp. 887–898 (2012)
63. Vavasis, S.A., Ye, Y.: A primal-dual interior point method whose running time depends only on the constraint matrix. *Math. Program.* **74**(1), 79–120 (1996)
64. Ye, Y.: *Interior-Point Algorithms: Theory and Analysis*. John Wiley and Sons, New York (1997)
65. Ye, Y.: A new complexity result on solving the Markov decision problem. *Math. Oper. Res.* **30**(3), 733–749 (2005)
66. Ye, Y.: Improved complexity results on solving real-number linear feasibility problems. *Math. Program.* **106**(2), 339–363 (2006)
67. Ye, Y.: The simplex and policy-iteration methods are strongly polynomial for the Markov decision problem with a fixed discount rate. *Math. Oper. Res.* **36**(4), 593–603 (2011)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.