



# A first order method for linear programming parameterized by circuit imbalance

Richard Cole<sup>1</sup> · Christoph Hertrich<sup>2</sup> · Yixin Tao<sup>3</sup> · László A. Végh<sup>4,5,6</sup>

Received: 17 July 2024 / Accepted: 21 July 2025  
© The Author(s) 2025

## Abstract

Various first order approaches have been proposed in the literature to solve Linear Programming (LP) problems, recently leading to practically efficient solvers for large-scale LPs. From a theoretical perspective, linear convergence rates have been established for first order LP algorithms, despite the fact that the underlying formulations are not strongly convex. However, the convergence rate typically depends on the Hoffman constant of a large matrix that contains the constraint matrix, as well as the right hand side, cost, and capacity vectors. We introduce a first order approach for LP optimization with a convergence rate depending polynomially on the circuit imbalance measure, which is a geometric parameter of the constraint matrix, and depending logarithmically on the right hand side, capacity, and cost vectors. This provides much stronger convergence guarantees. For example, if the constraint matrix is totally unimodular, we obtain polynomial-time algorithms, whereas the convergence guarantees for approaches based on primal-dual formulations may have arbitrarily slow convergence rates for this class. Our approach is based on a fast gradient method due to Necoara, Nesterov, and Glineur (Math. Prog. 2019); this algorithm is called repeatedly in a framework that gradually fixes variables to the boundary. This technique is

✉ Christoph Hertrich  
christoph.hertrich@utn.de

Richard Cole  
cole@cs.nyu.edu

Yixin Tao  
taoyixin@mail.shufe.edu.cn

László A. Végh  
lvegh@uni-bonn.de

- <sup>1</sup> Courant Institute, New York University, NY 10012, USA
- <sup>2</sup> University of Technology Nuremberg, 90461 Nürnberg, Germany
- <sup>3</sup> ITCS, Key Laboratory of Interdisciplinary Research of Computation and Economics, Shanghai University of Finance and Economics, Shanghai 200433, China
- <sup>4</sup> Hertz Chair in Algorithms and Optimization, University of Bonn, 53121 Bonn, Germany
- <sup>5</sup> London School of Economics and Political Science, London WC2A 2AE, UK
- <sup>6</sup> Corvinus Institute for Advanced Studies, Corvinus University, Budapest, 1093 Hungary, Hungary

based on a new approximate version of Tardos's method, that was used to obtain a strongly polynomial algorithm for combinatorial LPs (Oper. Res. 1986).

**Keywords** Linear Programming · First Order Methods · Hoffman Proximity · Circuit Imbalances

**Mathematics Subject Classification** 90C05 · 90C25 · 90C06 · 65K05 · 52B05

## 1 Introduction

In this paper, we develop new first order algorithms for approximately solving the linear program

$$\begin{aligned} \min \quad & \langle c, x \rangle \\ & Ax = b, \\ & \mathbf{0} \leq x \leq u, \end{aligned} \quad (\text{LP}(A, b, c, u))$$

where  $A \in \mathbb{R}^{m \times n}$ ,  $b \in \mathbb{R}^m$ ,  $c, u \in \mathbb{R}^n$ . We assume that  $m \leq n$ . We use the notation  $[\mathbf{0}, u] = \{x \in \mathbb{R}^n \mid \mathbf{0} \leq x \leq u\}$ , and denote the feasible region as  $\mathcal{P}_{A,b,u} := \{x \in \mathbb{R}^n \mid Ax = b, x \in [\mathbf{0}, u]\}$ .

Linear programming (LP) is one of the most fundamental optimization problems with an immense range of applications in applied mathematics, operations research, computer science, and more. While Dantzig's Simplex method works well in practice, its running time may be exponential in the worst case. Breakthrough results in the 1970s and 1980s led to the development of the first polynomial time algorithms, the ellipsoid method [14] and interior point methods (IPMs) [13]. The Simplex method was one of the earliest computations implemented on a computer, and there are highly efficient LP solvers available, based on Simplex and interior point methods.

Linear programming can also be seen as a special case of more general optimization models: it can be captured by various convex programs, saddle point problems, and linear complementarity problems. Due to these connections, the development of new LP algorithms has been an important driving force in the development of optimization theory.

In this paper, we focus on first order methods (FOMs) for LP. The benefit of FOMs is cheap iteration complexity and efficient implementability for large-scale problems. In contrast to IPMs, they do not require careful initialization. FOMs are prevalent in optimization and machine learning, but they are not an obvious choice for LP for two reasons. First, the standard formulation has a complicated polyhedral feasible region, and therefore standard techniques are not directly applicable. Second, FOMs usually do not lead to polynomial running time guarantees: this is in contrast with IPMs that are polynomial and also efficient in practice.

Nevertheless, FOMs turn out to be practically efficient for large-scale LPs. In a recent paper Applegate et al. [1] use a restarted primal-dual hybrid gradient (PDHG) method based on a saddle point formulation. Their implementation outperforms the state-of-the-art commercial Simplex and IPM solvers on standard benchmark instances, and is able to find high accuracy solutions to large-scale PageRank instances.

The number of iterations needed to find an  $\varepsilon$ -approximate solution in standard FOMs is typically  $O(1/\varepsilon)$  or  $O(1/\sqrt{\varepsilon})$ . However, strong convexity properties can yield *linear convergence*, i.e., an  $O(\log(1/\sqrt{\varepsilon}))$  dependence. No strongly convex formulation is known to capture LP. Despite this, there is a long line of work on FOMs that achieve linear convergence guarantees for LP, starting from Eckstein and Bertsekas's alternating direction method from 1990 [6], followed by a variety of other techniques, e.g., [1, 2, 10, 11, 15, 22, 23].

Before discussing these approaches, let us specify the notion of approximate solutions. By a  $\delta$ -feasible solution, we mean an  $x \in [0, u]$  with  $\|Ax - b\|_1 \leq \delta \|A\|_1$ . If  $\text{LP}(A, b, c, u)$  is feasible, we let  $\Phi(A, b, c, u)$  denote the optimum value. A  $\delta$ -optimal solution satisfies  $\langle c, x \rangle \leq \Phi(A, b, c, u) + \delta \|c\|_\infty$ . Our goal will be to find a  $\delta$ -feasible and  $\delta$ -optimal solution for a required accuracy  $\delta > 0$ .<sup>1</sup>

The above mentioned works are able to find  $\delta$ -feasible and  $\delta$ -optimal solutions in running times that depend polynomially on  $\log(1/\delta)$ ,  $n$ , and  $C(A, b, c, u)$ , a constant depending on the problem input. In particular, Applegate, Hinder, Lu, and Lubin [2] give a running time bound  $O(C \log(1/\delta))$  for restarted PDHG, where  $C$  is the Hoffman-constant associated with the primal-dual embedding of the LP. Recently, for the case when  $A$  is a totally unimodular matrix and there are no upper bounds  $u$ , Hinder [11] bounded the running time of restarted PDHG as  $O(Hn^{2.5} \sqrt{\text{nzz}(A)} \log(Hm/\delta))$ , where  $\text{nzz}(A)$  is the number of nonzero entries of  $A$ , and  $b, c$  are integer vectors with  $\|b\|_\infty, \|c\|_\infty \leq H$ .

However, the constants involved in the running time bound are typically not polynomial in the binary encoding length of the input. In this paper, we give the first FOM-based algorithm with polynomial dependence on  $\log(1/\delta)$ ,  $n$ , a constant  $\bar{\kappa}(\mathcal{X}_A)$ , and  $\log \|b\|$ ,  $\log \|c\|$ , and  $\log \|u\|$ , as stated in Theorem 1.1 below. The constant  $\bar{\kappa}(\mathcal{X}_A)$  is the *max circuit imbalance measure* defined and discussed below. In particular, it is upper bounded by the maximal subdeterminant  $\Delta(A)$ , but it is often much smaller than  $\Delta(A)$ . For totally unimodular matrices, we have  $\bar{\kappa}(\mathcal{X}_A) = 1$ . Note that the running time depends polynomially on the logarithms of the capacity and cost vectors. In contrast, the bound in [11] only applies for the totally unimodular case and the running time is linear in  $\|b\|_\infty + \|c\|_\infty$ .

For technical convenience, we will assume throughout that  $\|A\|_1 \geq 1$ . In general, our results still hold by replacing  $\|A\|_1$  with  $\max\{\|A\|_1, 1\}$  in the statement of the following theorem.<sup>2</sup> Note that since the system  $\text{LP}(A, b, c, u)$  is bounded, whenever feasible, it has an optimal solution.

**Theorem 1.1** *Consider an instance of  $\text{LP}(A, b, c, u)$  with  $\|A\|_1 \geq 1$ . There is an FOM-based algorithm that obtains a solution  $x$  that is  $\delta$ -feasible and  $\delta$ -optimal, or concludes that no feasible solution exists, and whose runtime is dominated by the cost of performing  $O(n^{1.5} m^2 \|A\|_1^2 \cdot \bar{\kappa}^3(\mathcal{X}_A) \log^3((\|u\|_1 + \|b\|_1)nm \cdot \kappa(\mathcal{X}_A) \|A\|_1/\delta))$  gradient descent updates. Additionally, our algorithm returns a dual solution certifying*

<sup>1</sup> Different papers may use different norms and normalizations in their accuracy requirement, but these can be easily translated to each other. Such a change of the norms might incur polynomial factors, but these do not impact the asymptotic running time due to linear convergence.

<sup>2</sup> We could also scale  $A$  and  $b$  such that  $\|A\|_1 = 1$ . However, we do not make this assumption since multiplying  $A$  by a constant may change  $\bar{\kappa}(\mathcal{X}_A)$ .

approximate optimality of the solution in  $O(m\|A\|_2 \cdot \bar{\kappa}(\mathcal{X}_A) \cdot \log(n\|c\|_1/\delta))$  gradient descent updates.

### Hoffman bounds and quadratic function growth.

The main underlying tool for proving linear convergence bounds is Hoffman-proximity theory, introduced by Hoffman in 1952 [12]. Let  $A \in \mathbb{R}^{m \times n}$ , let  $\|\cdot\|_\alpha$  be a norm in  $\mathbb{R}^m$  and  $\|\cdot\|_\beta$  be a norm in  $\mathbb{R}^n$ . Then there exists a constant  $\theta_{\alpha,\beta}(A)$  such that for any  $x \in [\mathbf{0}, u]$ , and any  $b \in \mathbb{R}^m$ , whenever  $\mathcal{P}_{A,b,u}$  is nonempty, there exists an  $\bar{x} \in \mathcal{P}_{A,b,u}$  such that

$$\|\bar{x} - x\|_\beta \leq \theta_{\alpha,\beta}(A) \|Ax - b\|_\alpha.$$

To see how such bounds can lead to linear convergence, let us first focus on finding a feasible solution in  $\mathcal{P}_{A,b,u}$ . This can be formulated as a convex quadratic minimization problem:

$$\min \frac{1}{2} \|Ax - b\|^2 \quad \text{s.t. } x \in [\mathbf{0}, u]. \quad (1)$$

This is a smooth objective function, but not strongly convex. Nevertheless, Hoffman-proximity guarantees that for any  $x \in [\mathbf{0}, u]$  where  $f(x) := \frac{1}{2} \|Ax - b\|^2$  is close to the optimum value, there exists *some* optimal solution  $\bar{x}$  nearby. Necoara, Nesterov, and Glineur [15] introduce various relaxations of strong convexity, including the notion of  $\mu_f$ -quadratic growth (Definition 2.3), and show that these weaker properties suffice for linear convergence.

Hence, [15] implies that a  $\delta$ -feasible LP solution can be found by a Fast Gradient Method with Restart (R-FGM) in  $O(\|A\|_2 \cdot \theta_{2,2}(A) \log(m\|b\|_1/\delta))$  iterations. If  $A$  is a totally unimodular (TU) matrix, the dependence is given by  $\theta_{2,2}(A) \leq m$  (see Lemma 3.6).

To solve  $\text{LP}(A, b, c, u)$ , [15] in effect uses the standard reduction from optimization to feasibility by writing the primal and dual systems together. By strong duality, if  $\text{LP}(A, b, c, u)$  is feasible and bounded, then  $x$  is a primal and  $(\pi, w^+, w^-)$  is a dual optimal solution if and only if

$$Ax = b, \quad A^\top \pi + w^- - w^+ = c, \quad \langle c, x \rangle - \langle b, \pi \rangle + \langle u, w^+ \rangle = 0, \quad x, w^-, w^+ \geq 0. \quad (2)$$

We can use R-FGM for this larger feasibility problem. However, the constraint matrix  $M$  now also includes the vectors  $c$ ,  $b$ , and  $u$ . In particular, while  $\theta_{2,2}(A)$  is small for a TU matrix,  $\theta_{2,2}(M)$  may be unbounded, as shown in Sect. 10.

Other previous works obtain linear convergence bounds using different approaches, but share the above characteristics: their running time includes a constant term  $C(A, b, c, u)$ . For example, [6] and [22] use an alternating direction method based on an augmented Lagrangian, and [2] and [11] use restart PDHG. The convergence bounds depend not only on the Hoffman-constant of the system, but also linearly on the maximum possible norm of the primal and dual iterates seen during the algorithm.

## 1.1 Our approach

We present an algorithm in the FOM family with polynomial dependence on  $\log(1/\delta)$ ,  $n$ ,  $m$ ,  $\log \|u\|$ ,  $\log \|b\|$ ,  $\log \|c\|$  and a constant  $C(A)$  only dependent on  $A$ . Our algo-

rithm repeatedly calls R-FGM, described in [15], on a potential function of the form

$$F_{\tau}(x) := \frac{1}{2}(\max\{0, \langle \hat{c}, x \rangle - \tau\})^2 + \frac{1}{2\|A\|_1^2} \|Ax - b\|_2^2, \quad (3)$$

for a suitably chosen parameter  $\tau \in \mathbb{R}$ , and a modified cost function  $\hat{c}$ . If we use  $\hat{c} = c/\|c\|_{\infty}$ , and  $\tau$  is slightly below the optimum value, then one can show that a near-minimizer  $x$  of  $F_{\tau}(x)$  is a near optimal primal solution to the original LP, and moreover, we can extract a near-optimal dual solution from the gradient  $\nabla F_{\tau}(x)$ .

Thus, one could find a  $\delta$ -approximate and  $\delta$ -optimal solution to  $\text{LP}(A, b, c, u)$  with  $\log(1/\delta)$  dependence by doing a binary search over the possible values of  $\tau$ , and running R-FGM for each guess. This already improves on the parameter dependence, however, it still involves a constant  $C(A, c)$ . This is because the minimization of  $F_{\tau}(x)$  can be casted in the form (1) with the matrix  $\begin{pmatrix} A & 0 \\ \|A\|_1 c^{\top} & \|A\|_1 \end{pmatrix}$ . The resulting Hoffman constant can be arbitrarily worse than the one for the original system.

To overcome this issue, we instead define  $\hat{c}$  as an  $\varepsilon$ -discretization of  $c/\|c\|_{\infty}$ . We show that the Hoffman constant remains bounded in terms of the Hoffman constant of the feasibility system and a suitably chosen  $\varepsilon > 0$ . Now, for the appropriate choice of  $\tau$ , a near-minimizer of  $F_{\tau}(x)$  only gives a crude approximation to the original LP: the error depends on the discretization parameter  $\varepsilon$ , and to keep the Hoffman constant under control we cannot choose  $\varepsilon$  very small. Nonetheless, the dual solution obtained from the gradient contains valuable information. For certain indices  $i \in N$ , using primal-dual slackness, one can conclude  $x_i^* \approx 0$  or  $x_i^* \approx u_i$  for an optimal solution  $x^*$  to the original LP. We fix all such  $x_i$  to 0 or  $u_i$ , respectively, and recurse. Even if we not find any such  $x_i$ , we make progress by replacing our cost function by an equivalent reduced cost with the  $\ell_{\infty}$  norm decreasing by at least a factor two.

To summarize: our overall algorithm has an outer loop that gradually fixes the variables to the upper and lower bounds, and repeatedly replaces the cost by a reduced cost. In the inner loop, we call R-FGM in a binary search framework that guesses the parameter  $\tau$ . We note that while R-FGM is run on a number of systems, the total number of these systems is logarithmically bounded in  $1/\delta$  and the input parameters. Moreover, besides the first order updates, we only perform simple arithmetic operations: based on the gradient, we eliminate a subset of variables and shift the cost function. On a high level, our algorithm is a repeatedly applied FOM, where after each run, we ‘zoom in’ to a ‘critical’ part of the problem based on what we learned from the previous iteration.

### Circuit imbalance measures and proximity

The key condition numbers for our algorithm are *circuit imbalance measures*. For a linear space  $W \subseteq \mathbb{R}^n$ , an elementary vector is a support minimal nonzero vector in  $W$ . The (fractional) circuit imbalance measure  $\kappa(W)$  is the largest ratio between the absolute values of two entries of an elementary vector. If  $W$  is a rational space, then every elementary vector can be rescaled to have integer entries; and the *max circuit imbalance measure*  $\bar{\kappa}(W)$  is the smallest integer  $k$  such that all elementary vectors can be scaled to have integer entries between  $-k$  and  $k$ . Note that  $\kappa(W) \leq \bar{\kappa}(W)$ . For a matrix  $A$ , we also use  $\kappa(A) = \kappa(\ker(A))$  and  $\bar{\kappa}(A) = \bar{\kappa}(\ker(A))$ . We give

a more detailed introduction to these measures in Sect. 3. We define the subspace  $\mathcal{X}_A = \ker(A| - I_m)$ ; thus,  $(v, -Av) \in \mathcal{X}_A$  for any  $v \in \mathbb{R}^n$ .

Circuit imbalances play two roles in our algorithm. First, they are used to bound the number of iterations of R-FGM. The circuit imbalance measure of  $\mathcal{X}_A$  gives the Hoffman-proximity bound  $\theta_{1,\infty}(A) \leq \kappa(\mathcal{X}_A)$  (see Lemma 3.6). To bound the number of iterations in R-FGM, we need a Hoffman bound—equivalently, a circuit imbalance bound—for the matrix  $B = \begin{pmatrix} A & 0 \\ \|A\|_1 \hat{c}^\top & \|A\|_1 \end{pmatrix}$ , where  $\hat{c}$  is the  $\varepsilon$ -discretization of  $c/\|c\|_\infty$ .  $\kappa(\mathcal{X}_B)$  can be bounded in terms of the max circuit imbalance measure  $\bar{\kappa}(\mathcal{X}_A)$ , namely,  $\kappa(\mathcal{X}_B) \leq 2m \cdot \bar{\kappa}^2(\mathcal{X}_A)/\varepsilon$ .

The second role of  $\kappa(\mathcal{X}_A)$  is for the variable fixing argument in the outer loop of the algorithm. Recall that the inner loop returns a near optimal primal solution with respect to the rounded cost, as well as a near optimal dual solution derived from the gradient of the potential function. We would like to infer that variables with a large positive or negative dual slack can be rounded to the lower or upper bounds. To make such an inference, the rounding accuracy  $\varepsilon$  needs to be calibrated to  $\kappa(\mathcal{X}_A)$ . The larger  $\kappa(\mathcal{X}_A)$  is, the more refined the rounding is needed to obtain such guarantees.

### Guessing the condition numbers.

Our algorithm requires explicit bounds on the circuit imbalance measures both in the inner and outer loops. However, the circuit imbalance measures cannot be approximated even within an exponential factor unless  $P = NP$  (see Sect. 3.1). Similar issues arise in several algorithms that rely on condition numbers. In particular, R-FGM in [15] explicitly requires a bound on the Hoffman constant to determine the step-length; it does not address how such a bound could be obtained.<sup>3</sup>

We circumvent this problem by a standard doubling guessing procedure. We first run the algorithm with the initial guess  $\hat{\kappa} = 1$ . Either it succeeds, or otherwise we restart after doubling the guess  $\hat{\kappa}$ ; and so forth, until the algorithm first succeeds. The asymptotic running time is the same as when the circuit imbalance value is known. The only nontrivial issue is checking whether we succeeded; this can be done by running a final dual feasibility algorithm.

We note that the algorithm may succeed even if  $\hat{\kappa}$  is much better than the actual  $\bar{\kappa}(\mathcal{X}_A)$  value. In fact, the guessing procedure is a natural heuristic. We initially start with a crude discretization strategy to guess variable fixings from the outputs of R-FGM. In the event this leads to an infeasible or suboptimal solution, we restart after increasing the accuracy.

<sup>3</sup> However, the same paper shows that standard projected gradient also converges linearly albeit at a slower rate, and does not require knowing this constant.

## 1.2 Related work

We recall that an LP algorithm is strongly polynomial if it only uses basic arithmetic operations ( $+$ ,  $-$ ,  $\times$ ,  $/$ ) and comparisons, and the number of such operations is polynomial in the number of variables and constraints. Further, the algorithm must be in PSPACE, that is, the size of the numbers appearing in the computation must remain bounded in the input size. The existence of a strongly polynomial algorithm for LP is on Smale's list of main challenges for 21st century mathematics [17].

The variable fixing idea in our algorithm traces its roots to Tardos's strongly polynomial algorithm for minimum-cost circulations. The same idea was extended by Tardos [19] to obtain a  $\text{poly}(n, \log \Delta(A))$  time algorithm for finding an exact solution to  $\text{LP}(A, b, c, u)$  for an integer constraint matrix  $A \in \mathbb{Z}^{m \times n}$  with largest subdeterminant  $\Delta(A)$ . This running time bound is strongly polynomial for 'combinatorial LPs', that is, LPs with all entries being integers of absolute value  $\text{poly}(n)$ .

We note that  $\kappa(A) \leq \kappa(\mathcal{X}_A) \leq \bar{\kappa}(\mathcal{X}_A) \leq \Delta(A)$  for an integer matrix  $A$ . Dadush et al. [4] strengthened Tardos' result by replacing  $\Delta(A)$  by  $\kappa(A)$ , and removing all integrality-based arguments, and obtained a  $\text{poly}(n, \log \kappa(A))$  running time bound. The algorithm is of black-box nature, and can use any LP solver; an exact optimal solution can be found by running  $nm$  LP-solvers to accuracy  $\delta = 1/\text{poly}(n, \log \kappa(A))$ .

Our algorithm uses variable fixing in a different manner, giving a robust extension to the approximate setting. Our end goal is not an exact optimal solution, but rather an approximate one. The approximate solution obtained from the FOM in the inner loop has weaker guarantees. Tardos [19] also uses subproblems with a similarly rounded cost function, but requires exact feasibility, which cannot be obtained from an FOM.

For this reason, we obtain weaker guarantees, e.g., we may fix variables to 0 that are small but positive in all optimal solutions. However, this is acceptable if we are only aiming for an approximate solution. On the positive side, we only need a logarithmic number of executions of the outer loop, in contrast to  $nm$  in [4, 19]. This is because for us it is already sufficient progress to decrease the norm of the reduced cost, even if we cannot fix any variables.

A  $\text{poly}(n, \log \kappa(A))$  running time for LP can also be achieved by a special class of 'combinatorial' interior point methods, called *Layered Least Squares (LLS) IPMs*. This class was introduced by Vavasis and Ye [21]. The parameter dependence was on the Dikin–Stewart–Todd condition measure  $\bar{\chi}(A)$ , but [5] observed that the two condition numbers are close to each other. Further, they gave a stronger LLS IPM with running time dependent on the optimal value  $\kappa^*(A)$  of  $\kappa(A)$  achievable by column rescaling. We refer the reader to the survey [7] for further results related to circuit imbalances and their uses in LP, including also diameter and circuit diameter bounds.

We also note that Fujishige et al. [8] recently gave a  $\text{poly}(n, \kappa(A))$  algorithm for the minimum norm point problem (1) by combining FOMs and active set methods. Their algorithm terminates with an exact solution; on the other hand, it also uses projection steps that involve solving a system of linear equations. Thus, it is not an FOM; moreover, it is not applicable for optimization LP.

### Notation.

We let  $[n] = \{1, 2, \dots, n\}$ . For a vector  $x \in \mathbb{R}^n$ , let  $\text{supp}(x) = \{i \in [n] \mid x_i \neq 0\}$  denote its support. For  $A \in \mathbb{R}^{m \times n}$ , let  $A_i$  denote the  $i$ -th column of  $A$ . We use the norm  $\|A\|_1 = \max_{j \in [n]} \sum_{i=1}^m |a_{ij}|$ , and the spectral norm  $\|A\|_2$ . We normalize the input matrix so that  $\|A\|_1 \geq 1$ . For a linear subspace  $W \subseteq \mathbb{R}^n$ , we let  $W^\perp \subseteq \mathbb{R}^n$  denote the orthogonal complement.

### Overview.

The remainder of the paper is structured as follows. Sect. 2 is on preliminaries regarding the convergence guarantees of R-FGM [15]. Sect. 3 discusses further preliminaries regarding circuit imbalances, proximity and their algorithmic uses. Sect. 4 is a more detailed overview of our main ideas including formal statements; the algorithm is formally presented Sect. 5, where the main result Theorem 1.1 is also proved. Sect. 6 contains proofs related to the crucial proximity results of Sect. 4. The main statements for the outer and inner routines are proved in Sects. 7 and 8, respectively. The proofs related to dual certification are in Sect. 9. Finally, Sect. 10 provides an example showing that a simple self-dual embedding can blow up the Hoffman constant, serving as a motivation for our work.

## 2 Linear convergence for functions with quadratic growth

Assume we are interested in finding a  $\delta$ -feasible solution to  $\text{LP}(A, b, c, u)$  or concluding that the system is infeasible. We consider the convex formulation

$$\begin{aligned} \min \quad & \frac{1}{2} \|Ax - b\|_2^2 \\ & \mathbf{0} \leq x \leq u. \end{aligned} \quad (4)$$

We now outline the running time bounds obtained by Neocara, Nesterov, and Glineur [15].

**Definition 2.1** The function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is  $L_f$ -smooth or has  $L_f$ -Lipschitz continuous gradient if  $\|\nabla f(x) - \nabla f(y)\|_2 \leq L_f \cdot \|x - y\|_2$  for any  $x, y \in \text{dom}(f)$ .

**Lemma 2.2** The function  $f(x) = \frac{1}{2} \|Ax - b\|_2^2$  is  $\|A\|_2^2$ -smooth.

**Proof** This follows as  $\|\nabla f(x) - \nabla f(y)\|_2 = \|A^\top A(x - y)\|_2 \leq \|A^\top A\|_2 \cdot \|x - y\|_2 = \|A\|_2^2 \cdot \|x - y\|_2$ .  $\square$

**Definition 2.3** Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  be continuously differentiable, let  $X \subseteq \text{dom}(f)$  be a closed convex set, and  $f^* = \min_{x \in X} f(x)$  the minimum value. Then  $f$  has  $\mu_f$ -quadratic growth on  $X$  if, for any  $x \in X$ , there exists an optimal solution  $\bar{x}$  (that is,  $f(\bar{x}) = f^*$ ) such that  $f(x) - f^* \geq \frac{\mu_f}{2} \|x - \bar{x}\|_2^2$ .

**Lemma 2.4** ([15, Theorem 8]) The function  $f(x) = \frac{1}{2} \|Ax - b\|_2^2$  has  $1/\theta_{2,2}^2(A)$ -quadratic growth.

The R-FGM method proposed in [15] optimizes a convex function by iteratively applying the standard accelerated projected gradient descent algorithm. R-FGM starts with  $x^0$  as the initial point and then performs  $h_R$  iterations of accelerated projected gradient descent to obtain  $x^{h_R}$ , for a suitable  $h_R$ . R-FGM then uses  $x^{h_R}$  as the new starting point and repeats the process, performing another  $h_R$  iterations of accelerated projected gradient descent. This process is repeated multiple times. For a convex function which is  $L_f$ -smooth and has  $\mu_f$ -quadratic growth, [15] shows that, for every  $h_R = O(\sqrt{L_f/\mu_f})$  iterations, the difference between the current function value and the optimum is reduced by a factor of  $e^2$ :

**Theorem 2.5** *Suppose function  $f$  is  $L_f$ -smooth and has  $\mu_f$ -quadratic growth. Let  $h_R = \lceil 2e\sqrt{L_f/\mu_f} \rceil$  and  $x^0$  is the starting point. Then, after  $k \cdot h_R$  iterations, the R-FGM method outputs  $x$  such that*

$$f(x) - f^* \leq e^{-2k} (f(x^0) - f^*).$$

### 3 Circuit imbalances and proximity

For a linear space  $W \subset \mathbb{R}^n$ ,  $g \in W$  is an *elementary vector* if  $g$  is a support minimal nonzero vector in  $W$ , that is, no  $h \in W \setminus \{0\}$  exists such that  $\text{supp}(h) \subsetneq \text{supp}(g)$ . A *circuit* in  $W$  is the support of some elementary vector. We let  $\mathcal{F}(W) \subseteq W$  denote the set of elementary vectors in the space  $W$ .<sup>4</sup>

The subspaces  $W = \{0\}$  and  $W = \mathbb{R}^n$  are called trivial subspaces; all other subspaces are nontrivial. We define the *fractional circuit imbalance measure*

$$\kappa(W) := \max \left\{ \left| \frac{g_j}{g_i} \right| : g \in \mathcal{F}(W), i, j \in \text{supp}(g) \right\}$$

for nontrivial subspaces, and  $\kappa(W) := 1$  for trivial subspaces.

Further, if  $W$  is a rational linear space, we let  $\tilde{\mathcal{F}}(W) \subseteq \mathcal{F}(W)$  denote the set of integer elementary vectors  $g \in \mathbb{Z}^n \cap \mathcal{F}(W)$  such that the largest common divisor of the entries is 1. We define the *max circuit imbalance measure* as

$$\bar{\kappa}(W) := \max \{ \|g\|_\infty : g \in \tilde{\mathcal{F}}(W) \}.$$

When using the term ‘circuit imbalance measure’ without any specification, it will refer to the fractional version. Note that  $\kappa(W) \leq \bar{\kappa}(W)$  but they may not be equal. For example, if the single elementary vector up to scaling is  $(4, 7, 8)$ , then  $\kappa(W) = 2$  but  $\bar{\kappa}(W) = 8$ .

Let  $A \in \mathbb{R}^{m \times n}$  be a matrix, and let  $W = \ker(A)$  be the kernel space of  $A$ . We let  $\mathcal{F}(A)$ ,  $\kappa(A)$ ,  $\bar{\kappa}(A)$  denote  $\mathcal{F}(W)$ ,  $\kappa(W)$ ,  $\bar{\kappa}(W)$ , respectively, for the kernel space

<sup>4</sup> Let  $M \in \mathbb{R}^{m \times n}$  be a matroid such that  $W = \ker(M)$ . One can associate a linear matroid with  $M$  (and  $W$ ), where a subset of  $[n]$  is independent if the corresponding column vectors are linearly independent. Circuits defined as above are precisely the circuits, i.e., minimal non-independent sets in this matroid.

$W = \ker(A)$ . We refer the reader to the survey [7] for properties and applications of circuit imbalances. Below, we mention some basic properties.

Recall that a matrix is *totally unimodular* (TU) if the determinant of every square submatrix is 0, +1, or -1. We note that  $\kappa(W) = \bar{\kappa}(W) = 1$  if and only if there exists a TU matrix  $A \in \mathbb{R}^{m \times n}$  such that  $W = \ker(A)$ . This follows by a 1957 result of Cederbaum. Further, it is easy to verify that for an integer matrix  $A \in \mathbb{Z}^{m \times n}$ , the inequality  $\bar{\kappa}(A) \leq \Delta(A)$  holds, where  $\Delta(A)$  is the largest absolute value of a subdeterminant of  $A$ . However,  $\bar{\kappa}(A)$  can be arbitrarily smaller:  $\bar{\kappa}(A) = 2$  for the node-edge incidence matrix of any undirected graph, whereas  $\Delta(A)$  can be exponentially large. See [7, Section 3.1] for the above results. We will also use the following important self-duality of  $\kappa$ :

**Lemma 3.1** ([5]) *Let  $W \subseteq \mathbb{R}^n$  be a linear subspace. Then  $\kappa(W) = \kappa(W^\perp)$ .*

### Conformal circuit decompositions

We say that the vector  $y \in \mathbb{R}^n$  *conforms to*  $x \in \mathbb{R}^n$  if  $x_i y_i > 0$  whenever  $y_i \neq 0$ . Given a subspace  $W \subseteq \mathbb{R}^n$ , a *conformal circuit decomposition* of a vector  $z \in W$  is a decomposition  $z = \sum_{k=1}^h g^k$ , where  $h \leq n$  and  $g^1, g^2, \dots, g^h \in \mathcal{F}(W)$  are elementary vectors that conform to  $z$ . Further, for each  $i = 1, 2, \dots, h-1$ ,  $\text{supp}(g^i) \setminus \bigcup_{j=i+1}^h \text{supp}(g^j) \neq \emptyset$ . A fundamental result on elementary vectors asserts the existence of a conformal circuit decomposition, see e.g. [9, 16]. Note that there may be multiple conformal circuit decompositions of the same vector.

**Lemma 3.2** *For every subspace  $W \subseteq \mathbb{R}^n$ , every  $z \in W$  admits a conformal circuit decomposition.*

Given  $A \in \mathbb{R}^{m \times n}$ , we define the extended subspace  $\mathcal{X}_A \subset \mathbb{R}^{n+m}$  as  $\mathcal{X}_A := \ker(A \mid -I_m)$ . Hence, for every  $z \in \mathbb{R}^n$ ,  $(z, Az) \in \mathcal{X}_A$ . For  $z \in \mathbb{R}^n$ , a *generalized path-circuit decomposition of  $z$  with respect to  $A$*  is a decomposition  $z = \sum_{k=1}^h g^k$ , where  $h \leq n+m$ , and for each  $k \in [h]$ ,  $(g^k, Ag^k) \in \mathbb{R}^{n+m}$  is an elementary vector in  $\mathcal{X}_A$  that conforms to  $(z, Az)$ . Note that this corresponds to a conformal circuit decomposition of  $(z, Az)$  in  $\mathcal{X}_A$ . We say that  $g^k$  is an *inner vector* in the decomposition if  $Ag^k = 0$  and an *outer vector* otherwise.

**Definition 3.3** We say that  $z \in \mathbb{R}^n$  is *cycle-free with respect to  $A$* , if no  $y \in \ker(A)$ ,  $y \neq 0$  exists that conforms to  $z$ .

Note that being cycle-free is equivalent to the property that all generalized path-circuit decompositions of  $z$  contain outer vectors only. The following lemma will play a key role in analyzing our algorithms.

**Lemma 3.4** *Let  $A \in \mathbb{R}^{m \times n}$  and let  $z \in \mathbb{R}^n$  be cycle-free with respect to  $A$ . Then*

$$\|z\|_\infty \leq \kappa(\mathcal{X}_A) \cdot \|Az\|_1 \quad \text{and} \quad \|z\|_2 \leq m \cdot \kappa(\mathcal{X}_A) \cdot \|Az\|_2.$$

**Proof** Consider a generalized path-circuit decomposition  $z = \sum_{k=1}^h g^k$ . Since  $z$  is cycle-free, for each  $g^k$ ,  $Ag^k \neq 0$ , and therefore  $|g_j^k| \leq \kappa(\mathcal{X}_A) \cdot |(Ag^k)_i|$  for any

$j \in \text{supp}(g^k)$  and  $i \in \text{supp}(Ag^k)$ . By the conformity property,  $|z_j| = \sum_{k=1}^h |g_j^k|$  for  $j \in [n]$  and  $|(Az)_i| = \sum_{k=1}^h |(Ag^k)_i|$  for  $i \in [m]$ . Thus, for any  $j \in [n]$ ,

$$|z_j| = \sum_{k=1}^h |g_j^k| \leq \kappa(\mathcal{X}_A) \cdot \sum_{i=1}^m \sum_{k=1}^h |(Ag^k)_i| = \kappa(\mathcal{X}_A) \cdot \sum_{i=1}^m |(Az)_i| = \kappa(\mathcal{X}_A) \cdot \|Az\|_1.$$

For the second inequality, note that  $\|g^k\|_2 \leq \sqrt{m} \cdot \kappa(\mathcal{X}_A) |(Ag^k)_i|$  for any  $k \in [h]$  and  $i \in \text{supp}(Ag^k)$ , since for any elementary vector  $(g^k, Ag^k) \in \mathcal{X}_A$  with  $\text{supp}(Ag^k) \neq \emptyset$ , the columns in  $\text{supp}(g^k)$  must be linearly independent, and therefore  $|\text{supp}(g^k)| \leq m$ . This implies

$$\|z\|_2 \leq \sum_{k=1}^h \|g^k\|_2 \leq \sqrt{m} \cdot \kappa(\mathcal{X}_A) \cdot \|Az\|_1 \leq m \cdot \kappa(\mathcal{X}_A) \cdot \|Az\|_2.$$

□

The following lemma is trivial for the input matrix since we assume  $\|A\|_1 \geq 1$ . However, we also need this guarantee for its column submatrices in the recursive calls.

**Lemma 3.5** *For any non-zero matrix  $A \in \mathbb{R}^{m \times n}$ ,  $\kappa(\mathcal{X}_A) \|A\|_1 \geq 1$ .*

**Proof** Let  $e_i \in \mathbb{R}^n$  be the standard basis vector with a 1 in the  $i$ -th position and zeros elsewhere. Then,  $(e_i, Ae_i)$  is an elementary vector in  $\mathcal{X}_A$ , which implies  $\kappa(\mathcal{X}_A) \geq 1/|A_{ij}|$  for every  $j$  with  $A_{ij} \neq 0$ . Consequently, we have  $\kappa(\mathcal{X}_A) \|A\|_1 \geq 1$ . □

When  $A$  is clear from the context, we simply use  $\kappa = \kappa(\mathcal{X}_A)$ . If  $A$  is a node-arc incidence matrix of a directed graph, then  $A$ , and consequently also  $(A \mid -I_m)$  is a TU matrix, implying  $\bar{\kappa}(\mathcal{X}_A) = 1$ . For undirected graph incidence matrices, one can show  $\bar{\kappa}(\mathcal{X}_A) \leq 2$ .

**Lemma 3.6** *Let  $A \in \mathbb{R}^{n \times m}$ . Then  $\theta_{1,\infty}(A) \leq \kappa(\mathcal{X}_A)$  and  $\theta_{2,2}(A) \leq m \cdot \kappa(\mathcal{X}_A)$ .*

**Proof** We need to show that for any  $x \in [0, u]$ , and any  $b \in \mathbb{R}^m$ , whenever  $\mathcal{P}_{A,b,u}$  is nonempty, there exists an  $\bar{x} \in \mathcal{P}_{A,b,u}$  such that  $\|\bar{x} - x\|_\infty \leq \kappa(\mathcal{X}_A) \|Ax - b\|_1$  and  $\|\bar{x} - x\|_2 \leq m \cdot \kappa(\mathcal{X}_A) \|Ax - b\|_2$ . Let us select  $\bar{x} \in \mathcal{P}_{A,b,u}$  as the nearest feasible point to  $x$  in  $\ell_2$ -norm; in particular,  $A\bar{x} = b$ . We claim that  $\bar{x} - x$  is cycle-free with respect to  $A$ . Indeed, if a generalized path-circuit decomposition of  $\bar{x} - x$  contained an inner vector  $g^k$ , then  $\bar{x}' = x + g^k$  would also be feasible, with  $\|\bar{x}' - x\|_2 < \|\bar{x} - x\|_2$ . Thus, Lemma 3.4 can be applied with  $z = x - \bar{x}$  and the claims follow. □

The first inequality may be tight. Assume there exists an elementary vector  $(g, Ag)$  in  $\mathcal{X}_A$  such that  $|g_i| = \kappa(\mathcal{X}_A)$  for all  $i \in \text{supp}(g)$ ,  $|\text{supp}(Ag)| = 1$ , and  $(Ag)_i = 1$  for the nonzero component. Further, let  $b = 0$ , and let  $u_i = 0$  for all  $i \notin \text{supp}(g)$ . Since  $(g, Ag)$  is a support minimal nonzero vector in  $\mathcal{X}_A$ , it follows that the only feasible solution to  $A\bar{x} = 0$ ,  $\bar{x} \in [0, u]$  is  $\bar{x} = 0$ . Thus, we get a tight example with  $\theta_{1,\infty}(A) = \kappa(\mathcal{X}_A)$ . The same example shows tightness of the second inequality up to a factor  $\sqrt{m}$ .

Our algorithm will also find dual certificates. The next lemma shows that the corresponding dual systems have the same circuit imbalances.

**Lemma 3.7** *For any matrix  $A \in \mathbb{R}^{m \times n}$ ,  $\kappa(A^\top | I_n) = \kappa(\mathcal{X}_A)$ .*

**Proof** Recall that  $\mathcal{X}_A = \ker(A | -I_m)$ . It is easy to verify that  $\ker(I_n | A^\top)$  is the orthogonal complement of  $\mathcal{X}_A$ . The statement then follows by Lemma 3.1, noting also that reordering the columns does not change the circuit imbalances.  $\square$

### 3.1 Guessing the circuit imbalances

Given a matrix  $A \in \mathbb{R}^{m \times n}$ , there is no hope of getting any reasonable approximation of the circuit imbalance values. It is NP-hard to approximate  $\kappa(A)$  within a factor  $2^{\text{poly}(m)}$  for  $A \in \mathbb{R}^{m \times n}$ , see [5], using a result of Tunçel [20] on the related condition number  $\bar{\chi}(A)$ . However, our algorithms make use of the values  $\kappa(A)$  and  $\bar{\kappa}(A)$ .

Nevertheless, one can use a guessing procedure that guarantees the same asymptotic running times without knowing these values. First, note that upper bounds rather than exact values suffice throughout. Algorithm 1 below makes recursive calls to column submatrices of the original input matrix  $A$ . If  $\bar{\kappa}(A) \leq \hat{\kappa}$  for the input matrix  $A$ , then  $\hat{\kappa}$  is an upper bound on all the circuit imbalance values seen in the recursive calls.

As our initial estimate, we set  $\hat{\kappa} = 1$ . When run with a correct guess  $\hat{\kappa} \geq \bar{\kappa}(A)$ , Algorithm 1 (see Sect. 5) returns approximately optimal primal and dual solutions to  $\text{LP}(A, b, c, u)$ . Running it with an incorrect guess may lead to a failure while running the algorithm: either R-FGM does not find a solution of the required accuracy within the given number of steps, or the final primal and dual solutions do not satisfy the required approximate feasibility and complementarity properties. We can easily detect both kind of failures. If no failure is detected, then the primal and dual solutions certify approximate optimality for each other; this may happen even when  $\hat{\kappa} < \bar{\kappa}(A)$ . Each time we detect a failure, we double the estimate  $\hat{\kappa}$  and restart the algorithm.

The overall running time bound in Theorem 1.1 is also the total running time of this process, because at each call we have  $\kappa \leq 2\bar{\kappa}(A)$ , and the running time bound of the final run dominates the running time bound of all previous runs.

## 4 Main ideas and key statements

Before describing the algorithm in Sect. 5, we highlight the key ideas and formulate the main underlying proximity results. We gradually reduce  $\text{LP}(A, b, c, u)$  by fixing some variables to their upper or lower bounds, and replacing the cost vector by an equivalent one of smaller norm. We first start by describing the simpler feasibility algorithm. The optimization algorithm has two components: the outer loop and the inner loop.

### 4.1 The feasibility algorithm

We first show how the R-FGM algorithm from [15] leads to a simple algorithm for finding a  $\delta$ -feasible solution. Here, we assume that the LP is feasible. In the proof

of Theorem 1.1 in Sect. 5.3, we explain how this assumption can be removed in general.

**Theorem 4.1** *There is an algorithm  $\text{Feasible}(A, b, u, \delta)$ , which, on input  $A \in \mathbb{R}^{m \times n}$ ,  $\|A\|_1 \geq 1$ ,  $b \in \mathbb{R}^m$ ,  $u \in \mathbb{R}^n$ , supposing the system  $Ax = b$ ,  $x \in [\mathbf{0}, u]$  is feasible, finds a  $\delta$ -feasible solution using  $O(m\|A\|_2 \cdot \kappa(\mathcal{X}_A) \cdot \log(m\|b\|_1/\delta))$  iterations of R-FGM.*

**Proof** Let  $f(x) = \frac{1}{2}\|Ax - b\|^2$ . We use the R-FGM algorithm from [15] to find an  $\varepsilon$ -approximate minimizer of  $f(x)$  over  $x \in [\mathbf{0}, u]$ , i.e., the system (1), where  $\varepsilon := \|A\|_1^2 \cdot \delta^2/(2m)$ . We choose  $x = 0$  as the starting point. By Theorem 2.5,  $O(\|A\|_2 \cdot \theta_{2,2}(A) \cdot \log(m\|b\|_1/\delta))$  iterations suffice (using  $\|A\|_1 \geq 1$ ). The bound on the number of iterations follows, for by Lemma 3.6,  $\theta_{2,2}(A) \leq m \cdot \kappa(\mathcal{X}_A)$ .

By the assumption that the system is feasible, the optimum value is 0. Thus, an  $\varepsilon$ -approximate solution has  $f(x) \leq \varepsilon$ , which yields a  $\delta$ -feasible solution.  $\square$

## 4.2 The outer loop

In the outer loop, our goal is to find a  $\delta^{\text{feas}}$ -feasible and  $\delta^{\text{opt}}$ -optimal solution to  $\text{LP}(A, b, c, u)$ . We distinguish these two accuracy parameters for the sake of the recursive algorithm, where the required feasibility and optimality accuracies need to be changed differently in the recursive calls.

### Primal-dual optimality and cost shifting.

We use primal-dual arguments, making variable fixing decisions based on approximate complementarity conditions. The dual to  $\text{LP}(A, b, c, u)$  can be written as

$$\begin{aligned} \max \quad & \langle b, \pi \rangle - \langle u, w^+ \rangle \\ A^\top \pi + w^- - w^+ &= c \\ w^-, w^+ &\geq 0. \end{aligned} \quad (\text{Dual}(A, b, c, u))$$

Note that given  $\pi \in \mathbb{R}^m$ , the unique best choice of the variables  $w^-$  and  $w^+$  is  $w^- = \max\{c - A^\top \pi, 0\}$  and  $w^+ = \max\{A^\top \pi - c, 0\}$ . When we speak of a dual solution  $\pi \in \mathbb{R}^m$ , we mean its extension with these variables. Recall the primal-dual optimality conditions:  $x^* \in \mathcal{P}_{A,b,u}$  and  $\pi \in \mathbb{R}^m$  are optimal respectively to  $\text{LP}(A, b, c, u)$  and to  $\text{Dual}(A, b, c, u)$  if and only if the following holds:

$$\text{if } A_i^\top \pi < c_i \text{ then } x_i = 0, \text{ and if } A_i^\top \pi > c_i \text{ then } x_i = u_i \text{ for every } i \in [n]. \quad (5)$$

Also note that we can naturally shift the cost function for any  $\pi \in \mathbb{R}^m$  as stated in the next lemma.

**Lemma 4.2**  $\text{LP}(A, b, c, u)$  has exactly the same solutions and the same optimum value as the following linear program:

$$\min \left\langle c - A^\top \pi, x \right\rangle + \langle b, \pi \rangle \quad \text{s.t.} \quad Ax = b, \quad x \in [\mathbf{0}, u]. \quad (6)$$

### Approximate complementarity and proximity.

Assume now that we have a pair of primal and dual solutions  $x$  and  $\pi$  that do not satisfy complementarity, but we have a quantitative bound on the violation. Namely, for a suitably chosen threshold  $\sigma \geq 0$ , let

$$\theta(x, \pi, \sigma) := \sum_{c_i - A_i^\top \pi > \sigma} x_i + \sum_{c_i - A_i^\top \pi < -\sigma} (u_i - x_i), \quad \text{and} \\ J(\pi, \sigma) := \left\{ i \in [n] : |c_i - A_i^\top \pi| > n \cdot \lceil \kappa(\mathcal{X}_A) \rceil \cdot \sigma \right\}.$$

Note that if  $x$  and  $\pi$  are primal and dual optimal, then the primal-dual complementarity constraints (5) imply  $\theta(x, \pi, 0) = 0$ . Let us assume that for some  $\sigma > 0$ , this quantity is still small. Note also that  $J(\pi, \sigma)$  is the set of indices where the absolute value of the slack is much higher than the threshold  $\sigma$ . In particular,  $\min\{x_i, u_i - x_i\} \leq \theta(x, \pi, \sigma)$  on these indices. Our key proximity result asserts that there exists an optimal solution that is close to the current solution on these indices. The proof is deferred to Sect. 6.

**Lemma 4.3** *Let  $x \in \mathcal{P}_{A,b,u}$  be a feasible solution. Then there exists an optimal solution  $x^*$  for  $LP(A, b, c, u)$  such that*

$$|x_i - x_i^*| \leq \kappa(\mathcal{X}_A) \cdot \theta(x, \pi, \sigma)$$

for all  $i \in J(\pi, \sigma)$ .

### Variable fixing.

Assume that from the inner loop of the algorithm we get  $x \in [0, u]$  and  $\pi \in \mathbb{R}^m$  such that the feasibility violation  $\|Ax - b\|_1$  and the complementarity violation  $\theta(x, \pi, \sigma)$  are both tiny for the choice  $\sigma := \|c\|_\infty / (4n \lceil \kappa(\mathcal{X}_A) \rceil)$ . Note that for this choice, the threshold in the definition of  $J(\pi, \sigma)$  becomes  $\|c\|_\infty / 4$ . We partition  $J(\pi, \sigma)$  into

$$J_1 := \left\{ i \in J(\pi, \sigma) \mid c_i - A_i^\top \pi < -\frac{\|c\|_\infty}{4} \right\}, \quad \text{and} \\ J_2 := \left\{ i \in J(\pi, \sigma) \mid c_i - A_i^\top \pi > \frac{\|c\|_\infty}{4} \right\},$$

We apply Lemma 4.3 to the problem with the modified right hand side  $b' = Ax$ . By ensuring that  $\theta(x, \pi, \sigma)$  is sufficiently small, we will see that there is an optimal solution  $x^*$  with  $x_i^* \approx 0$  for  $i \in J_1$  and  $x_i^* \approx u_i$  for  $i \in J_2$ .

We fix these variables to the lower and upper bounds, respectively, and shift the cost function according to  $\pi$ . Thus, we specify the following new LP. Let  $N := [n] \setminus (J_1 \cup J_2)$  and  $\bar{b} := A_N x_N$ .

$$\begin{aligned} \min \quad & \langle c_N - A_N^\top \pi, z \rangle \\ & A_N z = \bar{b} \\ & 0_N \leq z \leq u_N. \end{aligned} \quad (\text{LP}(A_N, \bar{b}, c_N - A_N^\top \pi, u_N))$$

We show the following result, which says the optimal solution of  $\text{LP}(A_N, \bar{b}, c_N - A_N^\top \pi, u_N)$  provides an approximately feasible and optimal solution to  $\text{LP}(A, b, c, u)$ . The approximation is in terms of  $\theta(x, \pi, \sigma)$  and  $\|Ax - b\|_1$ . Recall that  $\Phi(A, b, c, u)$  denotes the optimal value, the value achieved by the solution to  $\text{LP}(A, b, c, u)$ . The proof is given in Sect. 6.

**Theorem 4.4** *For  $A \in \mathbb{R}^{m \times n}$ ,  $b \in \mathbb{R}^m$ , and  $c, u \in \mathbb{R}^n$  such that  $\text{LP}(A, b, c, u)$  is feasible, let  $\sigma := \|c\|_\infty / (4n \cdot \lceil \kappa(\mathcal{X}_A) \rceil)$ , and let  $x \in [\mathbf{0}, u]$  and  $\pi \in \mathbb{R}^m$  be a pair of (not necessarily feasible) primal and dual solutions. Then,  $\text{LP}(A_N, \bar{b}, c_N - A_N^\top \pi, u_N)$  is feasible and, in addition satisfies the following:*

- *feasibility condition:*

$$\|b - \bar{b} - A_{J_2} u_{J_2}\|_1 \leq \theta(x, \pi, \sigma) \cdot \|A\|_1 + \|Ax - b\|_1, \quad (7)$$

- *optimality condition:*

$$\begin{aligned} & |\Phi(A_N, \bar{b}, c_N - A_N^\top \pi, u_N) + \langle \bar{b}, \pi \rangle + \langle c_{J_2}, u_{J_2} \rangle - \Phi(A, b, c, u)| \\ &= |\Phi(A_N, \bar{b}, c_N, u_N) + \langle c_{J_2}, u_{J_2} \rangle - \Phi(A, b, c, u)| \\ &\leq \kappa(\mathcal{X}_A) \cdot \|c\|_1 \cdot \|Ax - b\|_1 \\ &\quad + |J_1 \cup J_2| \cdot \kappa(\mathcal{X}_A) \cdot \|c\|_1 \cdot (2 + \kappa(\mathcal{X}_A) \|A\|_1) \cdot \theta(x, \pi, \sigma), \end{aligned} \quad (8)$$

- *cost reduction:*  $\|c_N - A_N^\top \pi\|_\infty \leq \|c\|_\infty / 4$ .

With this theorem, if one can find a pair  $(x, \pi)$  such that the right hand sides of inequalities (7) and (8) are tiny, then  $\text{LP}(A, b, c, u)$  can be reduced to  $\text{LP}(A_N, \bar{b}, c_N - A_N^\top \pi, u_N)$  with a tiny loss on feasibility and optimality. Moreover, each iteration reduces the  $\ell_\infty$ -cost on the remaining variables by a factor 4. One can repeat this procedure and ultimately reduce the original problem to one with an extremely small objective function value and possibly with fewer variables. Solving this problem will give a good enough solution to the original  $\text{LP}(A, b, c, u)$ , after restoring any variables fixed to the lower or upper bounds.

It is possible that both  $J_1$  and  $J_2$  are empty. This means that  $\|c - A^\top \pi\|_\infty \leq \|c\|_\infty / 4$ ; we can simply recurse with the same  $b$  but improved cost function. Note that we could make progress more aggressively by a preprocessing step that projects  $c$  to the kernel of  $A$ ; this gets a cost vector of the form  $c' = c - A^\top \pi$  with the smallest possible  $\ell_2$ -norm—such a preprocessing is used in the strongly polynomial algorithms [4, 18, 19]. Setting a slightly smaller  $\sigma$  would then guarantee variable fixing in every iteration. However, the projection amounts to solving a system of linear equations that is computationally more expensive. We instead proceed with lazier updates as above.

### 4.3 The inner loop

Next, we describe our approach for obtaining a pair  $(x, \pi)$  such that the right hand sides of inequalities (7) and (8) are tiny, which is the purpose of the inner loop. For

this, we need to guarantee that  $\theta(x, \pi, \sigma)$  and  $\|Ax - b\|_1$  are sufficiently small. We use a potential function  $F_\tau(x)$  of the form (3) for a modified cost function  $\hat{c}$ .

As noted in the introduction, if  $\hat{c} = c/\|c\|_\infty$ , and  $\tau$  is within  $\delta/2$  of the optimum value of  $\text{LP}(A, b, c, u)$ , then a  $\delta'$ -approximate minimizer to  $F_\tau(x)$  for a suitably chosen  $\delta'$  would immediately give a  $\delta$ -approximate and  $\delta$ -feasible solution to  $\text{LP}(A, b, c, u)$ . Thus, we would not need the outer loops; a binary search on  $\tau$  and using the feasibility algorithm on this system would already give the desired solution, without the need for variable fixings in the outer loop.

However, the Hoffman-constant corresponding to the function (3) with  $\hat{c} = c/\|c\|_\infty$  could be unbounded in terms of  $\bar{\kappa}(\mathcal{X}_A)$  if  $c$  can be arbitrary, as discussed in Sect. 10. To circumvent this problem, we discretize  $c/\|c\|_\infty$  into integer multiples of  $\varepsilon = 1/(8n \cdot \lceil \kappa(\mathcal{X}_A) \rceil) = \sigma/(2\|c\|_\infty)$ .

Using the discretized  $\hat{c}$ , for a suitable choice of  $\tau$ , we can guarantee (7) and (8), that is, bound  $\|Ax - b\|_1$  and  $\theta(x, \pi, \sigma)$ , where the dual solution  $\pi$  is defined based on the gradient of  $F_\tau(x)$  as  $\pi := \frac{\|c\|_\infty}{\|A\|_1^2 \alpha} (b - Ax)$  for  $\alpha := \max\{0, \langle \hat{c}, x \rangle - \tau\}$ .

To bound the infeasibility  $\|Ax - b\|_1$ , we need to find a solution  $x$  where  $F_\tau(x)$  is small, since  $\|Ax - b\|_1 \leq (2m\|A\|_1^2 F_\tau(x))^{1/2}$ . Therefore,  $\tau$  should not be much smaller than the optimum value  $\Phi(A, b, \hat{c}, u)$ . The bound on  $\theta(x, \pi, \sigma)$  can be shown by arguing that the improving directions of the gradient are small at an approximately optimal solution  $x$ :  $x_i \approx 0$  if  $\nabla_i F_\tau(x) \gg 0$  and  $x_i \approx u_i$  if  $\nabla_i F_\tau(x) \ll 0$ , and that  $|c_i/\|c\|_\infty - \hat{c}_i| \cdot \|c\|_\infty \leq \|c\|_\infty \cdot \varepsilon = \sigma/2$ . We also need that  $\alpha > 0$  and is not too small. Based on these requirements, we can establish a narrow (but not too narrow) interval of  $\tau$  where a sufficiently accurate approximate solution to  $F_\tau(x)$  exists. Using that  $F_\tau^* := \min\{F_\tau(x) \mid x \in [\mathbf{0}, u]\}$  is a Lipschitz-continuous and non-increasing continuous function in  $\tau$ , we can find a suitable  $\tau$  by binary search.

#### 4.4 Dual certificates

As discussed above, our goal is not just to find a  $\delta^{\text{feas}}$ -feasible and  $\delta^{\text{opt}}$ -optimal solution, but also a dual certificate for the latter property. This is important as it enables us to verify the correctness of the solution, which is only guaranteed when using an estimate  $\kappa \geq \bar{\kappa}$ .

**Definition 4.5** Let  $A \in \mathbb{R}^{m \times n}$ ,  $b \in \mathbb{R}^m$ ,  $c, u \in \mathbb{R}^n$  and  $\delta \geq 0$ , and let  $x \in [\mathbf{0}, u]$ . We say that  $(\pi, w^+, w^-) \in \mathbb{R}^{m \times n \times n}$  is a  $\delta$ -certificate for  $x$ , if

- (i)  $A^\top \pi + w^- - w^+ = c$ ,
- (ii)  $0 \leq w_i^- \leq 2\delta\|c\|_\infty/x_i$ ,  $0 \leq w_i^+ \leq 2\delta\|c\|_\infty/(u_i - x_i)$  for all  $i \in [n]$ , and
- (iii)  $\|\pi\|_\infty \leq 2\delta\|c\|_\infty/\|Ax - b\|_1$ .

Our next lemma shows that  $\delta$ -certificates indeed certify approximate optimality, and conversely, for every approximately optimal solution, such a certificate can be found. The proof is given in Sect. 9.

**Lemma 4.6** Let  $A \in \mathbb{R}^{m \times n}$ ,  $b \in \mathbb{R}^m$ ,  $c, u \in \mathbb{R}^n$ , and let  $x \in [\mathbf{0}, u]$ .

- (i) If there is a  $\delta$ -certificate for  $x$  for some  $\delta \geq 0$ , then  $x$  is  $(4n + 2)\delta$ -optimal.

(ii) Suppose  $0 \leq \delta^{\text{feas}} \cdot n \cdot \kappa(\mathcal{X}_A) \cdot \|A\|_1 \leq \delta^{\text{opt}}$ . If  $x$  is a  $\delta^{\text{feas}}$ -feasible and  $\delta^{\text{opt}}$ -optimal solution, then there exists a  $\delta^{\text{opt}}$ -certificate for  $x$ .

Our next theorem justifies this concept and provides certificates efficiently. The proof is also deferred to Sect. 9.

**Theorem 4.7** Suppose  $A \in \mathbb{R}^{m \times n}$ ,  $b \in \mathbb{R}^m$ ,  $x, c, u \in \mathbb{R}^n$ ,  $0 \leq \delta^{\text{feas}} \cdot n \cdot \kappa(\mathcal{X}_A) \cdot \|A\|_1 \leq \delta^{\text{opt}}$ ,  $\|A\|_1 \geq 1$  and  $x \in [0, u]$  is both  $\delta^{\text{feas}}$ -feasible and  $\delta^{\text{opt}}$ -optimal. Then there is an algorithm *Dual-Certificate*( $x, A, b, c, u, \delta^{\text{feas}}, \delta^{\text{opt}}$ ) which on such inputs finds a  $2 \cdot \delta^{\text{opt}}$ -certificate for  $x$  in  $O(m\sqrt{n} \cdot \|A\|_1 \cdot \kappa(\mathcal{X}_A) \cdot \log(n\|c\|_1/\delta^{\text{opt}}))$  iterations of *R-FGM*.

## 5 The algorithm

We describe our algorithm under the simplifying assumption that the exact values of  $\kappa(A)$  and  $\bar{\kappa}(A)$  are known for the input matrix as well as all submatrices obtained by column deletions. As discussed in Sect. 3.1, even though these quantities cannot be computed, one can obtain the same asymptotic running time bounds by repeatedly guessing an estimate  $\hat{\kappa}$ .

### 5.1 The outer loop: variable fixing

Algorithm 1 takes as input  $(A, b, c, u)$  such that  $\text{LP}(A, b, c, u)$  is feasible, and accuracy parameters  $\delta^{\text{feas}}$  and  $\delta^{\text{opt}}$  such that  $0 \leq \delta^{\text{feas}} \cdot 8n\sqrt{m} \cdot \kappa(\mathcal{X}_A) \cdot \|A\|_1 \leq \delta^{\text{opt}}$ .

No feasibility assumption is made in Theorem 1.1; we prove this theorem in Sect. 5.3. We handle feasibility in the same spirit as in Simplex, using a two-stage approach. Algorithm 1 uses a subroutine *GetPrimalDualPair*( $A, b, c, u, \delta^{\text{feas}}, \delta^{\text{opt}}$ ) specified as follows.

Subroutine *GetPrimalDualPair*  
**Input:**  $A \in \mathbb{R}^{m \times n}$ ,  $b \in \mathbb{R}^m$ ,  $c, u \in \mathbb{R}^n$ ,  $\delta^{\text{feas}}, \delta^{\text{opt}} > 0$  such that  $(A, b, c, u)$  is feasible,  $\delta^{\text{opt}} \leq \|u\|_1$ , and  $\delta^{\text{feas}} \|A\|_1 \leq \delta^{\text{opt}}$ .  
**Output:**  $(x, \pi)$ ,  $x \in [0, u]$ ,  $\pi \in \mathbb{R}^m$  such that

- The right-hand side of the feasibility bound (7) in Theorem 4.4 is at most  $\delta^{\text{feas}} \|A\|_1/n$ .
- The right-hand side of the optimality bound (8) in Theorem 4.4 is at most  $\delta^{\text{opt}} \|c\|_\infty/n$ .
- $\|\pi\|_\infty \leq 4n\sqrt{m} \cdot \kappa(\mathcal{X}_A) \cdot \|c\|_\infty$ .

The following theorem provides the analysis of the outer loop. The proof is deferred to Sect. 7.

**Theorem 5.1** If  $\text{LP}(A, b, c, u)$  is feasible, then Algorithm 1 returns a  $\delta^{\text{feas}}$ -feasible solution that is  $\delta^{\text{opt}}$ -optimal along with a  $2\delta^{\text{opt}}$ -certificate. It makes at most  $\log_2(n\|u\|_1/\delta^{\text{opt}})$  many recursive calls.

**Algorithm 1:** SolveLP

---

**Input** :  $A \in \mathbb{R}^{m \times n}$ ,  $b \in \mathbb{R}^m$ ,  $c, u \in \mathbb{R}^n$ , such that  $(A, b, c, u)$  is feasible,  
 $0 < \delta^{\text{feas}}(8n\sqrt{m} \cdot \kappa(\mathcal{X}_A)\|A\|_1) \leq \delta^{\text{opt}}$ .

**Output** : A  $\delta^{\text{feas}}$ -feasible and  $\delta^{\text{opt}}$ -optimal solution to  $\text{LP}(A, b, c, u)$  along with a  $2\delta^{\text{opt}}$ -certificate

```

1 if  $\delta^{\text{opt}} \geq \|u\|_1$  then
2    $\bar{x} \leftarrow \text{Feasible}(A, b, c, u, \delta^{\text{feas}})$  (See Theorem 4.1.);
3   return  $\bar{x}$ 
4 else
5    $(x, \pi) \leftarrow \text{GetPrimalDualPair}(A, b, c, u, \delta^{\text{feas}}, \delta^{\text{opt}})$ ;
6   Define  $J_1, J_2, N, \bar{b}$  as for Theorem 4.4;
7    $c^{\text{new}} \leftarrow c - A^\top \pi$ ;
8    $\lambda \leftarrow \frac{\|c_N\|_\infty}{2\|c_N^{\text{new}}\|_\infty}$ ;
9   if  $J_1 \cup J_2 = \emptyset$  then
10     $x^{\text{out}} \leftarrow \text{SolveLP}(A, b, c^{\text{new}}, u, \delta^{\text{feas}}, \lambda\delta^{\text{opt}})$ ;
11     $\bar{\pi} \leftarrow \text{Dual-Certificate}(x^{\text{out}}, A, b, c, u, \delta^{\text{feas}}, \delta^{\text{opt}})$  (See Theorem 4.7.);
12  else
13     $x_{J_1}^{\text{out}} \leftarrow 0$ ;  $x_{J_2}^{\text{out}} \leftarrow u_{J_2}$ ;
14     $\bar{b} \leftarrow A_N x_N$ ;
15     $x_N^{\text{out}} \leftarrow \text{SolveLP}(A_N, \bar{b}, c_N^{\text{new}}, u_N, \delta^{\text{feas}} \cdot |N|/n, \lambda\delta^{\text{opt}} \cdot |N|/n)$ ;
16     $\bar{\pi} \leftarrow \text{Dual-Certificate}(x^{\text{out}}, A, b, c, u, \delta^{\text{feas}}, \delta^{\text{opt}})$  (See Theorem 4.7.);
17  return  $(x^{\text{out}}, \bar{\pi})$ 

```

---

**5.2 The inner loop: fast gradient with binary search**

We now describe  $\text{GetPrimalDualPair}(A, b, c, u, \delta^{\text{feas}}, \delta^{\text{opt}})$ , introduced at the beginning of Sect. 5.1. The subroutine needs to output primal and dual vectors  $(x, \pi)$  satisfying the feasibility and optimality bounds in Theorem 4.4. In particular, we need to bound  $\theta(x, \pi, \sigma) = \sum_{i:c_i - A_i^\top \pi > \sigma} x_i + \sum_{i:c_i - A_i^\top \pi < -\sigma} (u_i - x_i)$  and  $\|Ax - b\|_1$  for  $\sigma = \|c\|_\infty / (4n \cdot \lceil \kappa(\mathcal{X}_A) \rceil)$ . Let us define the accuracy parameter

$$\varepsilon := \frac{1}{8n \cdot \lceil \kappa(\mathcal{X}_A) \rceil} = \frac{\sigma}{2\|c\|_\infty}. \quad (9)$$

Thus,  $1/\varepsilon$  is integer. For some parameter  $\tau \in \mathbb{R}$ , we use the potential function

$$F_\tau(x) := \frac{1}{2}(\max\{0, \langle \hat{c}, x \rangle - \tau\})^2 + \frac{1}{2\|A\|_1^2} \|Ax - b\|_2^2, \quad (10)$$

where the rounded cost function  $\hat{c}$  is defined by taking the normalized vector  $c/\|c\|_\infty$ , and rounding each entry to the nearest integer multiple of  $\varepsilon$  in the direction of the 0 value (i.e., rounding down the positive entries and rounding up the negative entries). Recalling that  $1/\varepsilon$  is an integer, we have  $\|\hat{c}\|_\infty = 1$ .

Let  $F_\tau^* := \min\{F_\tau(x) \mid x \in [0, u]\}$  denote the optimum value. We say that  $x \in [0, u]$  is a  $\zeta$ -approximate minimizer of  $F_\tau$  if  $F(x) \leq F_\tau^* + \zeta$ . The following proposition is immediate.

**Proposition 5.2**  $F_\tau^*$  is a non-increasing continuous function of  $\tau$ . If  $LP(A, b, \hat{c}, u)$  is feasible, then  $\Phi(A, b, \hat{c}, u)$  is the smallest value of  $\tau$  such that  $F_\tau^* = 0$ .

The main driver of our algorithm is Necoara, Nesterov, and Glineur's R-FGM algorithm, applied to  $F_\tau$ . We specify this subroutine as follows.

**Subroutine R-FGM**

**Input:**  $A \in \mathbb{R}^{m \times n}$ ,  $b \in \mathbb{R}^m$ ,  $c, u \in \mathbb{R}^n$ ,  $\tau \in \mathbb{R}$ ,  $\zeta > 0$ .

**Output:** A  $\zeta$ -approximate minimizer  $x \in [0, u]$  of  $F_\tau$ .

The purpose of `GetPrimalDualPair` is to identify a value  $\tau$  by binary search that is slightly below  $\Phi(A, b, \hat{c}, u)$ . We show that there is a suitable  $\tau$  such that a sufficiently accurate approximate minimizer of  $F_\tau$  returns the required primal solution  $x$ . Moreover, we can also construct the dual  $\pi$  from the gradient of  $F_\tau$  at this point.

We define some further parameters to calibrate the accuracy used in the algorithm.

$$\begin{aligned} \mathcal{C} &:= n\sqrt{m} \cdot \kappa(\mathcal{X}_A) \cdot \|A\|_1, \\ \bar{\mathcal{C}} &:= 64n\mathcal{C} \cdot \kappa(\mathcal{X}_A) = 64n^2\sqrt{m} \cdot \kappa^2(\mathcal{X}_A) \cdot \|A\|_1, \\ \zeta &:= \left( \frac{\delta^{\text{feas}}}{4\kappa^2(\mathcal{X}_A)n^4\bar{\mathcal{C}}\sqrt{m}} \right)^2. \end{aligned} \quad (11)$$

These admit the following simple lower bounds.

**Lemma 5.3**  $\mathcal{C} \geq n\sqrt{m}$ ,  $\bar{\mathcal{C}} \geq 64n\sqrt{m}$ , and  $\bar{\mathcal{C}}\sqrt{\zeta} = \frac{\delta^{\text{feas}}}{4n^4\sqrt{m}\kappa^2(\mathcal{X}_A)}$ .

**Proof** The first bound follows by Lemma 3.5, and the second bound from the first and using  $\kappa(\mathcal{X}_A) \geq 1$ . The third bound is immediate from the definition.  $\square$

Algorithm `GetPrimalDualPair` is shown in Algorithm 2.

---

#### Algorithm 2: `GetPrimalDualPair`

---

**Input** :  $A \in \mathbb{R}^{m \times n}$ ,  $b \in \mathbb{R}^m$ ,  $c, u \in \mathbb{R}^n$ , and  $\delta^{\text{feas}}, \delta^{\text{opt}} > 0$  such that  $(A, b, c, u)$  is feasible, and  $\delta^{\text{feas}} \|A\|_1 \leq \delta^{\text{opt}} < \|u\|_1$ .

**Output** :  $x \in [0, u]$  and  $\pi \in \mathbb{R}^n$ .

1  $\tau^+ \leftarrow \|u\|_1$  and  $\tau^- \leftarrow -\|u\|_1 - 2\bar{\mathcal{C}}\sqrt{\zeta}$ ;

2 **Repeat**

3  $\tau \leftarrow \frac{\tau^+ + \tau^-}{2}$ ;

4  $x \leftarrow \text{R-FGM}(A, b, c, u, \tau, \zeta)$ ;

5 **if**  $F_\tau(x) > 2\bar{\mathcal{C}}^2\zeta$  **then**  $\tau^- \leftarrow \tau$ ;

6 **if**  $F_\tau(x) < \bar{\mathcal{C}}^2\zeta$  **then**  $\tau^+ \leftarrow \tau$ ;

7 **if**  $F_\tau(x) \in [\bar{\mathcal{C}}^2\zeta, 2\bar{\mathcal{C}}^2\zeta]$  **then**

8  $\alpha \leftarrow \max\{0, \langle \hat{c}, x \rangle - \tau\}$ ;

9  $\pi \leftarrow \frac{\|c\|_\infty}{\|A\|_1^2 \alpha} (b - Ax)$ ;

10 **return**  $(x, \pi)$

---

**Theorem 5.4** Assume  $LP(A, b, c, u)$  is feasible. Algorithm 2 makes  $O(\log[\|u\|_1 nm \cdot \kappa(\mathcal{X}_A)/\delta^{\text{feas}}])$  calls to  $R\text{-FGM}$ , and altogether these calls use  $O(n^{1.5} m^2 \|A\|_1^2 \cdot \bar{\kappa}^3(\mathcal{X}_A) \log^2[\|u\|_1 nm \cdot \|A\|_1 \kappa(\mathcal{X}_A)/\delta^{\text{feas}}])$  iterations. On terminating, it outputs  $(x, \pi)$  satisfying:

- (i)  $\theta(x, \pi, \sigma) \cdot \|A\|_1 + \|Ax - b\|_1 \leq \delta^{\text{feas}} \|A\|_1/n.$
- (ii)  $\kappa(\mathcal{X}_A) \cdot \|c\|_1 \cdot \|Ax - b\|_1 + |J_1 \cup J_2| \cdot \kappa(\mathcal{X}_A) \cdot \|c\|_1 \cdot (2 + \kappa(\mathcal{X}_A) \|A\|_1) \cdot \theta(x, \pi, \sigma) \leq \delta^{\text{opt}} \|c\|_\infty/n.$
- (iii)  $\|\pi\|_\infty \leq 4n\sqrt{m} \cdot \kappa(\mathcal{X}_A) \cdot \|c\|_\infty.$

### 5.3 Putting everything together

We now combine the above ingredients to prove Theorem 1.1.

**Proof of Theorem 1.1** First, let us assume that  $LP(A, b, c, u)$  is feasible. We set  $\delta^{\text{opt}} = \delta/(8n + 4)$  and start with the guess  $\hat{\kappa} = 1$ . Then, we set  $\delta^{\text{feas}} = \delta^{\text{opt}}/(8n\sqrt{m}\hat{\kappa}\|A\|_1)$ , and run the algorithm  $\text{SolveLP}(A, b, c, u, \delta^{\text{feas}}, \delta^{\text{opt}})$ . If it succeeds and outputs a primal solution  $x$  and a certificate  $\pi$ , then we can check if the solution is a  $\delta^{\text{feas}}$ -feasible solution by checking the constraint  $\|Ax - b\|_1 \leq \delta^{\text{feas}} \cdot \|A\|_1$ ; and check if  $\pi$  is a  $2\delta^{\text{opt}}$ -certificate by checking the constraints in Definition 4.5. If  $x$  is a  $\delta^{\text{feas}}$ -feasible solution and  $\pi$  is a  $2\delta^{\text{opt}}$ -certificate, then we output the  $x$ . Otherwise, we double the value of  $\hat{\kappa}$  and restart this procedure.

Assuming  $\hat{\kappa} \geq \bar{\kappa}(\mathcal{X}_A)$ ,  $\text{SolveLP}(A, b, c, u, \delta^{\text{feas}}, \delta^{\text{opt}})$  returns a  $\delta^{\text{feas}}$ -feasible and  $\delta^{\text{opt}}$ -optimal solution along with a  $2\delta^{\text{opt}}$ -certificate. These provide a  $\delta$ -feasible and  $\delta$ -optimal solution, along with a dual certificate of  $\delta$ -optimality, in accordance with Lemma 4.6(i).

For the running time,  $\text{SolveLP}(A, b, c, u, \delta^{\text{feas}}, \delta^{\text{opt}})$  makes at most  $\log_2(n\|u\|_1/\delta^{\text{opt}})$  recursive calls (Theorem 5.1), and, for each recursive call,  $\text{GetPrimalDualPair}$  uses at most  $O(n^{1.5} m^2 \|A\|_1^2 \hat{\kappa}^3 \log^2(\|u\|_1 nm \hat{\kappa}/\delta^{\text{feas}}))$  iterations (Theorem 5.4). Note that  $\hat{\kappa}$  will stop doubling no later than the first time  $\hat{\kappa} \geq \bar{\kappa}(\mathcal{X}_A)$ . Therefore, the total number of iterations is at most  $O(n^{1.5} m^2 \|A\|_1^2 \cdot \bar{\kappa}(\mathcal{X}_A)^3 \log^3(\|u\|_1 nm \|A\|_1 \cdot \bar{\kappa}(\mathcal{X}_A)/\delta))$  as  $\delta = \delta^{\text{feas}}/[8n\sqrt{m}\hat{\kappa}\|A\|_1(8n + 4)]$ .

We now remove the feasibility assumption. We use a two-stage approach, similarly to Simplex. We consider the following extended system;  $\mathbf{1} \in \mathbb{R}^m$  denotes the all 1's vector.

$$\begin{aligned} & \min \langle \mathbf{1}, s' \rangle + \langle \mathbf{1}, s'' \rangle \\ & Ax + s' - s'' = b \\ & \mathbf{0} \leq x \leq u \\ & \mathbf{0} \leq s', s'' \leq \|b\|_\infty \end{aligned}$$

This system is trivially feasible with the solution  $s'_i = \max\{b_i, 0\}$  and  $s''_i = \max\{-b_i, 0\}$ . Moreover, denoting the constraint matrix as  $B = (A \mid I_m \mid -I_m)$ , note that  $\kappa(\mathcal{X}_B) = \kappa(\mathcal{X}_A)$  and  $\bar{\kappa}(\mathcal{X}_B) = \bar{\kappa}(\mathcal{X}_A)$ .

We obtain a  $\delta/4$ -feasible and  $\delta/4$ -optimal solution  $(\bar{x}, s', s'')$  for this system by applying  $\text{SolveLP}$ . If the original system was feasible, then  $(\bar{x}, s', s'')$  provides a solution to the new LP with objective value of at most  $\delta/4$ . As the new cost vector is

the all ones vector, we see that  $\|s'\|_1 + \|s''\|_1 < \delta/4$ . As we show next, this implies that the returned solution  $\bar{x}$  will be  $\delta/2$ -feasible for the original system:

$$\|A\bar{x} - b\|_1 \leq \|A\bar{x} + s' - s'' - b\|_1 + \|s'\|_1 + \|s''\|_1 \leq (\delta/4) \cdot \|A\|_1 + \delta/4 \leq (\delta/2) \cdot \|A\|_1.$$

Both the second and third inequalities use  $\|A\|_1 \geq 1$ . We now run the above algorithm with the modified right hand side  $\bar{b} = A\bar{x}$ . Now,  $\text{LP}(A, \bar{b}, c, u)$  is feasible, and a  $\delta^{\text{feas}}$ -feasible solution for this system will also be  $\delta$ -feasible for the original system, since  $\delta^{\text{feas}} + \delta/2 < \delta$ .  $\square$

## 6 Proofs of the proximity statements

In this section, we prove Lemma 4.3 and Theorem 4.4.

**Lemma 4.3** *Let  $x \in \mathcal{P}_{A,b,u}$  be a feasible solution. Then there exists an optimal solution  $x^*$  for  $\text{LP}(A, b, c, u)$  such that*

$$|x_i - x_i^*| \leq \kappa(\mathcal{X}_A) \cdot \theta(x, \pi, \sigma)$$

for all  $i \in J(\pi, \sigma)$ .

**Proof** We define the following modified capacities. For  $i \in [n]$ , let

$$\begin{aligned} \bar{\ell}_i &:= \begin{cases} 0 & \text{if } c_i - A_i^\top \pi \leq \sigma, \\ x_i & \text{if } c_i - A_i^\top \pi > \sigma. \end{cases} \\ \bar{u}_i &:= \begin{cases} u_i & \text{if } c_i - A_i^\top \pi \geq -\sigma, \\ x_i & \text{if } c_i - A_i^\top \pi < -\sigma. \end{cases} \end{aligned}$$

We now consider the optimization problem

$$\begin{aligned} \min \quad & \langle c, x \rangle \\ & Ax = b \\ & \bar{\ell} \leq x \leq \bar{u}. \end{aligned} \tag{12}$$

This problem is feasible since  $x$  is a feasible solution; let  $\bar{x}$  be an optimal solution.

**Claim 6.1**  $\bar{x}_i = x_i$  for every  $i \in J(\pi, \sigma)$ .

**Proof** Consider a generalized path-circuit decomposition  $\bar{x} - x = \sum_{k=1}^h g^k$ . Since  $A(\bar{x} - x) = \mathbf{0}$ , any conformal circuit decomposition may contain only inner vectors:  $Ag^k = \mathbf{0}$  for all  $k \leq h$ . We claim that  $\text{supp}(g^k) \cap J(\pi, \sigma) = \emptyset$  for all  $k \in [h]$ ; this implies the statement. For a contradiction, assume there exists a  $j \in \text{supp}(g^k) \cap J(\pi, \sigma)$  for some  $k \in [h]$ .

The optimality of  $\bar{x}$  implies that  $0 \geq \langle c, g^k \rangle = \langle c - A^\top \pi, g^k \rangle$ , as  $Ag^k = \mathbf{0}$ . Also, as we argue next, the definition of  $\bar{\ell}$  implies that  $g_i^k \geq 0$  for all  $i \in [n]$  with  $c_i - A_i^\top \pi > \sigma$ ;

for in this case, for every such index  $i$ , by definition,  $\bar{\ell}_i = x_i$ , and thus  $\bar{x}_i - x_i \geq 0$ , that is,  $g_i^k \geq 0$ . Similarly,  $g_i^k \leq 0$  for all  $i \in [n]$  with  $c_i - A_i^\top \pi < -\sigma$ . Thus  $(c_i - A_i^\top \pi)g_i^k \geq 0$  whenever  $|c_i - A_i^\top \pi| > \sigma$ .

Let  $S \subseteq \text{supp}(g^k)$  denote the set of indices  $i$  with  $(c_i - A_i^\top \pi)g_i^k < 0$ , and let  $\alpha := \|g_S^k\|_\infty$ . By the above,  $j \notin S$ , and  $(c_i - A_i^\top \pi)g_i^k \geq -\sigma\alpha$  for all  $i \in S$ .

If  $S = \emptyset$ , then  $g_i^k = 0$  must hold for all  $i \in [n]$  with  $|c_i - A_i^\top \pi| > \sigma$ , implying  $\bar{x}_i = x_i$  for  $i \in J(\pi, \sigma)$ , completing the proof. Assume that  $S \neq \emptyset$ . The definition of  $\kappa(A) \leq \kappa(\mathcal{X}_A)$  implies that  $|g_j^k| \geq \alpha/\kappa(\mathcal{X}_A)$ . Hence,

$$\begin{aligned} 0 &\geq \langle c, g^k \rangle = \langle c - A^\top \pi, g^k \rangle \\ &= \sum_{i \in \text{supp}(g^k) \setminus S} (c_i - A_i^\top \pi)g_i^k + \sum_{i \in S} (c_i - A_i^\top \pi)g_i^k \\ &\geq (c_j - A_j^\top \pi)g_j^k - (n-1)\sigma\alpha \\ &\geq n \cdot \kappa(\mathcal{X}_A) \cdot \sigma \cdot \frac{\alpha}{\kappa(\mathcal{X}_A)} - (n-1)\sigma\alpha > 0, \end{aligned}$$

a contradiction.  $\square$

The following claim completes the proof of Lemma 4.3.

**Claim 6.2** Consider an optimal solution  $x^*$  to  $LP(A, b, c, u)$  with  $\|\bar{x} - x^*\|_1$  minimal. Then,  $\|\bar{x} - x^*\|_\infty \leq \kappa(\mathcal{X}_A) \cdot \theta(x, \pi, \sigma)$ .

**Proof** Consider a generalized path-circuit decomposition  $x^* - \bar{x} = \sum_{k=1}^h g^k$ . This may only contain inner vectors, since  $A(x^* - \bar{x}) = \mathbf{0}$ . By the choice of  $x^*$ , we must have  $\langle c, g^k \rangle < 0$  for every  $k$ . Hence,  $\bar{x} + \lambda g^k$  is not feasible to (12) for any  $\lambda > 0$ , as otherwise we get a contradiction to the optimality of  $\bar{x}$  to (12).

Therefore, for each  $k$ , there exists an  $i_k \in \text{supp}(g^k)$  with  $\bar{\ell}_{i_k} = \bar{x}_{i_k} > 0$  or  $\bar{u}_{i_k} = \bar{x}_{i_k} < u_{i_k}$ . By construction,  $\bar{x}_{i_k} = x_{i_k}$  for these  $i$ . Using the definition of  $\kappa(A) \leq \kappa(\mathcal{X}_A)$  and the conformity of the decomposition, we get

$$\begin{aligned} \|\bar{x} - x^*\|_\infty &\leq \max_j \sum_{k=1}^h |g_j^k| \leq \sum_{k=1}^h \kappa(A) \cdot |g_{i_k}^k| \\ &\leq \kappa(A) \left[ \sum_{c_i - A_i^\top \pi \geq \sigma} x_i + \sum_{c_i - A_i^\top \pi < -\sigma} (u_i - x_i) \right] \\ &\leq \kappa(\mathcal{X}_A) \cdot \theta(x, \pi, \sigma), \end{aligned}$$

which is the claimed bound.  $\square$

The lemma by from combining the previous two claims.  $\square$

The next lemma bounds the difference in the optimum value if the right hand side changes; we will use it in the proof of Theorem 4.4.

**Lemma 6.3** Let  $A \in \mathbb{R}^{m \times n}$ ,  $c, u \in \mathbb{R}^n$ , and  $b, \bar{b} \in \mathbb{R}^m$ . If both LPs are feasible, then

$$|\Phi(A, b, c, u) - \Phi(A, \bar{b}, c, u)| \leq \kappa(\mathcal{X}_A) \cdot \|c\|_1 \cdot \|b - \bar{b}\|_1.$$

**Proof** Let  $x$  be an optimal solution to  $\text{LP}(A, b, c, u)$  and  $\bar{x}$  an optimal solution to  $\text{LP}(A, \bar{b}, c, u)$  such that  $\|x - \bar{x}\|_1$  is minimal, and consider a generalized path-circuit decomposition  $\bar{x} - x = \sum_{k=1}^h g^k$ . By the choice of  $x$  and  $\bar{x}$ , the decomposition contains no inner circuits. Thus, by Lemma 3.4,  $\|x - \bar{x}\|_\infty \leq \kappa(\mathcal{X}_A) \cdot \|b - \bar{b}\|_1$ . Therefore,

$$\begin{aligned} & |\Phi(A, b, c, u) - \Phi(A, \bar{b}, c, u)| \\ &= |\langle c, x - \bar{x} \rangle| \leq \|c\|_1 \cdot \|x - \bar{x}\|_\infty \leq \kappa(\mathcal{X}_A) \cdot \|c\|_1 \cdot \|b - \bar{b}\|_1, \end{aligned}$$

proving the claim.  $\square$

We now restate and prove Theorem 4.4.

**Theorem 4.4** For  $A \in \mathbb{R}^{m \times n}$ ,  $b \in \mathbb{R}^m$ , and  $c, u \in \mathbb{R}^n$  such that  $\text{LP}(A, b, c, u)$  is feasible, let  $\sigma := \|c\|_\infty / (4n \cdot \lceil \kappa(\mathcal{X}_A) \rceil)$ , and let  $x \in [\mathbf{0}, u]$  and  $\pi \in \mathbb{R}^m$  be a pair of (not necessarily feasible) primal and dual solutions. Then,  $\text{LP}(A_N, \bar{b}, c_N - A_N^\top \pi, u_N)$  is feasible and, in addition satisfies the following:

- *feasibility condition:*

$$\|b - \bar{b} - A_{J_2} u_{J_2}\|_1 \leq \theta(x, \pi, \sigma) \cdot \|A\|_1 + \|Ax - b\|_1, \quad (7)$$

- *optimality condition:*

$$\begin{aligned} & |\Phi(A_N, \bar{b}, c_N - A_N^\top \pi, u_N) + \langle \bar{b}, \pi \rangle + \langle c_{J_2}, u_{J_2} \rangle - \Phi(A, b, c, u)| \\ &= |\Phi(A_N, \bar{b}, c_N, u_N) + \langle c_{J_2}, u_{J_2} \rangle - \Phi(A, b, c, u)| \\ &\leq \kappa(\mathcal{X}_A) \cdot \|c\|_1 \cdot \|Ax - b\|_1 \\ &\quad + |J_1 \cup J_2| \cdot \kappa(\mathcal{X}_A) \cdot \|c\|_1 \cdot (2 + \kappa(\mathcal{X}_A) \|A\|_1) \cdot \theta(x, \pi, \sigma), \end{aligned} \quad (8)$$

- *cost reduction:*  $\|c_N - A_N^\top \pi\|_\infty \leq \|c\|_\infty / 4$ .

**Proof** Feasibility is trivial, since  $x_N$  is a feasible solution. By the definition of  $\theta(x, \pi, \sigma)$ ,

$$\begin{aligned} \|b - \bar{b} - A_{J_2} u_{J_2}\|_1 &\leq \|b - Ax\|_1 + \|Ax - A_N x_N - A_{J_2} u_{J_2}\|_1 \\ &= \|b - Ax\|_1 + \|A_{J_1} x_{J_1} - A_{J_2} (u_{J_2} - x_{J_2})\|_1 \\ &\leq \theta(x, \pi, \sigma) \cdot \|A\|_1 + \|Ax - b\|_1, \end{aligned}$$

which finishes the proof of the feasibility condition.

To show the optimality condition, we will apply Lemmas 4.3 and 6.3 as follows. Consider the linear program  $\text{LP}(A, Ax, c, u)$ ;  $x$  is a feasible solution. Let  $x^*$  be an optimal solution to this LP as in Lemma 4.3. Let us define  $y^* := x^*|_N$  and  $b^* := A_N y^*$ . Thus,  $y^*$  is an optimal solution to  $\text{LP}(A_N, b^*, c_N, u_N)$ .

For convenience, let  $\rho := 2 \cdot \kappa(\mathcal{X}_A) \cdot \theta(x, \pi, \sigma)$ ,  $J := J_1 \cup J_2$ , and  $k := |J|$ . For each  $i \in J_2$ , we have that  $x_i^* \leq x_i + \kappa(\mathcal{X}_A) \cdot \theta(x, \pi, \sigma) \leq \rho$ , where the first inequality follows from Lemma 4.3 and the second one by the definition of  $\theta(x, \pi, \sigma)$ . Similarly, for each  $i \in J_1$ , we obtain  $x_i^* \geq u_i - \rho$ . Recalling that  $\Phi(A, Ax, c, u) = \langle c, x^* \rangle$ , this yields the bound

$$\begin{aligned} & \left| \Phi(A_N, b^*, c_N, u_N) + \langle c_{J_2}, u_{J_2} \rangle - \Phi(A, Ax, c, u) \right| \\ &= \left| \langle c_N, y^* \rangle + \langle c_{J_2}, u_{J_2} \rangle - \langle c, x^* \rangle \right| \leq k\rho \|c\|_\infty. \end{aligned} \quad (13)$$

To finish the proof we need to bound the differences  $|\Phi(A_N, \bar{b}, c_N, u_N) - \Phi(A_N, b^*, c_N, u_N)|$  and  $|\Phi(A, b, c, u) - \Phi(A, Ax, c, u)|$ , which we will do by applying Lemma 6.3. This yields

$$\left| \Phi(A, b, c, u) - \Phi(A, Ax, c, u) \right| \leq \kappa(\mathcal{X}_A) \cdot \|c\|_1 \cdot \|Ax - b\|_1 \quad (14)$$

and

$$\left| \Phi(A_N, \bar{b}, c_N, u_N) - \Phi(A_N, b^*, c_N, u_N) \right| \leq \kappa(\mathcal{X}_A) \cdot \|c\|_1 \cdot \|\bar{b} - b^*\|_1. \quad (15)$$

We further bound  $\|\bar{b} - b^*\|_1$  as follows:

$$\begin{aligned} \|\bar{b} - b^*\|_1 &= \|A_N x_N - A_N y^*\|_1 \\ &= \|Ax - Ax^* - A_J(x_J - x_J^*)\|_1 \\ &= \|A_J(x_J - x_J^*)\|_1 \\ &\leq k \cdot \|A\|_1 \cdot \kappa(\mathcal{X}_A) \cdot \theta(x, \pi, \sigma), \end{aligned} \quad (16)$$

where the inequality in the last line follows by how we obtained  $x^*$  from Lemma 4.3. Finally, we conclude by combining (13), (14), (15), and (16):

$$\begin{aligned} & \left| \Phi(A_N, \bar{b}, c_N, u_N) + \langle c_{J_2}, u_{J_2} \rangle - \Phi(A, b, c, u) \right| \\ &\leq \left| \Phi(A_N, b^*, c_N, u_N) + \langle c_{J_2}, u_{J_2} \rangle - \Phi(A, Ax, c, u) \right| \\ &\quad + \left| \Phi(A, Ax, c, u) - \Phi(A, b, c, u) \right| \\ &\quad + \left| \Phi(A_N, \bar{b}, c_N, u_N) - \Phi(A_N, b^*, c_N, u_N) \right| \\ &\leq k \cdot \rho \cdot \|c\|_\infty + \kappa(\mathcal{X}_A) \cdot \|c\|_1 \cdot \|Ax - b\|_1 + \kappa(\mathcal{X}_A) \cdot \|c\|_1 \cdot \|\bar{b} - b^*\|_1 \\ &\leq k \cdot \rho \cdot \|c\|_\infty + \kappa(\mathcal{X}_A) \cdot \|c\|_1 \cdot \|Ax - b\|_1 \\ &\quad + k \cdot \kappa(\mathcal{X}_A)^2 \cdot \|c\|_1 \cdot \|A\|_1 \cdot \theta(x, \pi, \sigma) \\ &= \kappa(\mathcal{X}_A) \cdot \|c\|_1 \cdot \|Ax - b\|_1 + \left[ 2k \cdot \kappa(\mathcal{X}_A) \cdot \|c\|_\infty \right. \\ &\quad \left. + k \cdot \kappa(\mathcal{X}_A)^2 \cdot \|c\|_1 \cdot \|A\|_1 \right] \cdot \theta(x, \pi, \sigma) \\ &\leq \kappa(\mathcal{X}_A) \cdot \|c\|_1 \cdot \|Ax - b\|_1 + k \cdot \kappa(\mathcal{X}_A) \cdot \|c\|_1 \\ &\quad \cdot (2 + \kappa(\mathcal{X}_A) \cdot \|A\|_1) \cdot \theta(x, \pi, \sigma). \end{aligned}$$

This proves the optimality condition. The third statement, that is, the cost reduction, directly follows from the definitions of  $J_1$ ,  $J_2$ ,  $N$  and  $\sigma$ .  $\square$

## 7 Analysis of the outer loop

We now prove the main statement for the outer loop.

**Theorem 5.1** *If  $LP(A, b, c, u)$  is feasible, then Algorithm 1 returns a  $\delta^{\text{feas}}$ -feasible solution that is  $\delta^{\text{opt}}$ -optimal along with a  $2\delta^{\text{opt}}$ -certificate. It makes at most  $\log_2(n\|u\|_1/\delta^{\text{opt}})$  many recursive calls.*

**Proof** We first show that the algorithm terminates with recursion depth at most  $\log_2(n\|u\|_1/\delta^{\text{opt}})$ . By the cost reduction condition in Theorem 4.4,  $\lambda \geq 2$  holds in each recursive call in Algorithm 1. If  $\delta_0^{\text{opt}}$  denotes the initial  $\delta^{\text{opt}}$  and  $\delta_j^{\text{opt}}$  denotes the  $\delta^{\text{opt}}$  of the  $j$ -th recursive call, then it follows that  $\delta_j^{\text{opt}} \geq 2^j \delta_0^{\text{opt}}/n$ . Thus, when  $j \geq \log_2(n\|u\|_1/\delta_0^{\text{opt}})$ , we have  $\delta_j^{\text{opt}} \geq \|u\|_1$  and the algorithm terminates.

Now we prove the correctness of the algorithm by induction on the recursion depth. For the induction start, we assume that  $\delta^{\text{opt}} \geq \|u\|_1$ , in which case an arbitrary  $\delta^{\text{feas}}$ -feasible solution  $x$  is returned. Let  $x^*$  be an optimal solution to  $LP(A, b, c, u)$ . Since  $0 \leq x, x^* \leq u$ , it follows that  $\langle c, x \rangle - \Phi(A, b, c, u) = \langle c, x - x^* \rangle \leq \|u\|_1 \|c\|_\infty \leq \delta^{\text{opt}} \|c\|_\infty$ , proving  $\delta^{\text{opt}}$ -optimality.

For the inductive step, assume  $\delta^{\text{opt}} < \|u\|_1$ . Then the algorithm invokes the subroutine `GetPrimalDualPair` followed by a recursive call to itself. Recall that the subroutine `GetPrimalDualPair` returns  $(x, \pi)$  that satisfy the three bounds stated as part of the subroutine. We distinguish two cases depending on whether  $J_1 \cup J_2$  is empty or not.

If  $J_1 \cup J_2 = \emptyset$ , then the induction hypothesis applied to the call in line 10 implies that  $x^{\text{out}}$  is  $\delta^{\text{feas}}$ -feasible for the original system. Moreover, it is a  $\lambda\delta^{\text{opt}}$ -optimal solution to  $LP(A, b, c^{\text{new}}, u)$ , with  $\lambda = \|c\|_\infty/(2\|c^{\text{new}}\|_\infty)$  in this case. Note that  $\Phi(A, b, c^{\text{new}}, u) = \Phi(A, b, c, u) - \langle \pi, b \rangle$ . This implies:

$$\begin{aligned} |\langle c, x^{\text{out}} \rangle - \Phi(A, b, c, u)| &\leq |\langle c - c^{\text{new}}, x^{\text{out}} \rangle - \langle \pi, b \rangle| \\ &\quad + |\langle c^{\text{new}}, x^{\text{out}} \rangle - \Phi(A, b, c^{\text{new}}, u)| \\ &\leq |\langle \pi, Ax^{\text{out}} - b \rangle| + \lambda\delta^{\text{opt}}\|c^{\text{new}}\|_\infty \quad (\text{as } x^{\text{out}} \text{ is } \lambda\delta^{\text{opt}}\text{-optimal}) \\ &\leq \|\pi\|_\infty \cdot \delta^{\text{feas}}\|A\|_1 + \lambda\delta^{\text{opt}}\|c^{\text{new}}\|_\infty \quad (\text{as } x^{\text{out}} \text{ is } \delta^{\text{feas}}\text{-feasible}) \\ &\leq \frac{\delta^{\text{opt}}\|c\|_\infty}{2\delta^{\text{feas}}\|A\|_1} \cdot \delta^{\text{feas}}\|A\|_1 + \lambda\delta^{\text{opt}}\|c^{\text{new}}\|_\infty \\ &\quad (\text{using the upper bound on } \|\pi\|_\infty \text{ in } \text{GetPrimalDualPair} \\ &\quad \text{and the lower bound on } \delta^{\text{opt}}/\delta^{\text{feas}} \text{ in } \text{SolveLP}) \\ &\leq \frac{\delta^{\text{opt}}}{2}\|c\|_\infty + \frac{\delta^{\text{opt}}}{2}\|c\|_\infty = \delta^{\text{opt}}\|c\|_\infty, \end{aligned}$$

finishing the first case.

Otherwise, if  $J_1 \cup J_2 \neq \emptyset$ , note that  $\text{LP}(A_N, \bar{b}, c_N^{\text{new}}, u_N)$  is feasible because  $x_N$  is a feasible solution. Thus, we can apply the induction hypothesis to the recursive call, and obtain that  $x_N^{\text{out}}$  is  $(\delta^{\text{feas}} \cdot |N|/n)$ -feasible and  $(\lambda\delta^{\text{opt}} \cdot |N|/n)$ -optimal for  $\text{LP}(A_N, \bar{b}, c_N^{\text{new}}, u_N)$ . We combine this with Theorem 4.4 and the properties of the subroutine to prove first the  $\delta^{\text{feas}}$ -feasibility and then the  $\delta^{\text{opt}}$ -optimality of  $x^{\text{out}}$  for the original system.

With respect to feasibility, we obtain:

$$\begin{aligned} \|Ax^{\text{out}} - b\|_1 &\leq \|A_N x_N^{\text{out}} - \bar{b}\|_1 + \|\bar{b} + A_{J_2} u_{J_2} - b\|_1 \quad (\text{as } Ax^{\text{out}} = A_N x_N^{\text{out}} + A_{J_2} u_{J_2}) \\ &\leq (\delta^{\text{feas}} \cdot |N|/n) \|A_N\|_1 + \theta(x, \pi, \sigma) \cdot \|A\|_1 \\ &\quad + \|Ax - b\|_1 \quad (\text{explanation below}) \\ &\leq (\delta^{\text{feas}} \cdot |N|/n) \|A_N\|_1 + \delta^{\text{feas}} \|A\|_1/n \quad (\text{explanation below}) \\ &\leq \delta^{\text{feas}} \|A\|_1. \end{aligned}$$

The second inequality uses induction for the first term and Theorem 4.4 for the second term. The third inequality follows because `GetPrimalDualPair` returns a solution bounding the second and third terms in the second line, the RHS of (7), by the second term in the third line.

To prove approximate optimality, we first focus on the coordinates in  $N$  and obtain the following bound. (See below for explanations.)

$$\begin{aligned} &|\langle c_N, x_N^{\text{out}} \rangle - \Phi(A_N, \bar{b}, c_N^{\text{new}}, u_N) - \langle \pi, \bar{b} \rangle| \\ &\leq |\langle c_N - c_N^{\text{new}}, x_N^{\text{out}} \rangle - \langle \pi, \bar{b} \rangle| + |\langle c_N^{\text{new}}, x_N^{\text{out}} \rangle - \Phi(A_N, \bar{b}, c_N^{\text{new}}, u_N)| \\ &\leq |\langle \pi, A_N x_N^{\text{out}} - \bar{b} \rangle| + (\lambda\delta^{\text{opt}} \cdot |N|/n) \cdot \|c^{\text{new}}\|_\infty \\ &\leq \|\pi\|_\infty \|A_N\|_1 \delta^{\text{feas}} \cdot \frac{|N|}{n} + \frac{\delta^{\text{opt}} |N|}{2n} \cdot \|c\|_\infty \\ &\leq \frac{\delta^{\text{opt}} \|c\|_\infty}{2\delta^{\text{feas}} \|A\|_1} \|A_N\|_1 \delta^{\text{feas}} \cdot \frac{|N|}{n} + \frac{\delta^{\text{opt}} |N|}{2n} \|c\|_\infty \\ &\leq (\delta^{\text{opt}} \cdot |N|/n) \|c\|_\infty. \end{aligned} \tag{17}$$

The bound  $\|\pi\|_\infty \|A_N\|_1 \delta^{\text{feas}} \cdot |N|/n$  on the first term in the third line follows as  $x_N^{\text{out}}$  is a  $\delta^{\text{feas}} |N|/n$  feasible solution. In turn, the bound on this term follows because of the bound on  $\|\pi\|_\infty$  in `GetPrimalDualPair` and  $0 < \delta^{\text{feas}} (8n\sqrt{m} \cdot \kappa(\mathcal{X}_A) \|A\|_1) \leq \delta^{\text{opt}} \leq \|u\|_1$  (the second inequality is an input constraint for `SolveLP`, and the final inequality arises due to computing  $N$  in `SolveLP` only when  $\delta^{\text{opt}} < \|u\|_1$ ). Finally,

$$\begin{aligned} &|\langle c, x^{\text{out}} \rangle - \Phi(A, b, c, u)| \leq |\langle c_N, x_N^{\text{out}} \rangle - \Phi(A_N, \bar{b}, c_N^{\text{new}}, u_N) - \langle \pi, \bar{b} \rangle| \\ &\quad + |\Phi(A_N, \bar{b}, c_N^{\text{new}}, u_N) + \langle \pi, \bar{b} \rangle + \langle c_{J_2}, u_{J_2} \rangle - \Phi(A, b, c, u)| \\ &\leq (\delta^{\text{opt}} \cdot |N|/n) \|c\|_\infty + \delta^{\text{opt}} \|c\|_\infty/n \\ &\leq \delta^{\text{opt}} \|c\|_\infty. \end{aligned}$$

To obtain the first inequality, recall that  $x^{\text{out}}$  is a combination of  $x_N^{\text{out}}$ ,  $x_{j_1}^{\text{out}} = 0$ , and  $x_{j_1}^{\text{out}} = u_{j_2}$ . For the next inequality, note that the first term on line 2 is bounded by  $(\delta^{\text{opt}} \cdot |N|/n) \|c\|_\infty$  using (17), and the second term is bounded by the LHS of (8), which is bounded by  $\delta^{\text{opt}} \|c\|_\infty / n$  according to the second condition of `GetPrimalDualPair`. This concludes the second case.  $\square$

## 8 Analysis of the inner loop

In this section, we prove Theorem 5.4 on the correctness and running time of the inner loop. The proof uses three main lemmas. The first one bounds the number of iterations in R-FGM, using Theorem 2.5. This lemma will be proved in Sect. 8.2. Recall that the approximation factor  $\varepsilon$  was chosen to be  $\varepsilon = 1/(8n \cdot \lceil \kappa(\mathcal{X}_A) \rceil)$ .

**Lemma 8.1** *After  $O(k\sqrt{nm}^2 \|A\|_1^2 \cdot \bar{\kappa}^2(\mathcal{X}_A)/\varepsilon)$  iterations, R-FGM returns an  $e^{-2k} \|b\|_2^2 / (2\|A\|_1^2)$ -approximate minimizer of  $F_\tau$ .*

Recall that  $F_\tau^* = \min\{F_\tau(x) \mid x \in [\mathbf{0}, u]\}$  denotes the optimum value of  $F_\tau$  defined in (10). The next lemma strengthens Proposition 5.2, by asserting a Lipschitz property of  $F_\tau^*$  as a function of  $\tau$ .

**Lemma 8.2**  *$F_\tau^*$  is a non-increasing and continuous function of  $\tau$ . In addition, If  $F_\tau^* \leq \bar{C}^2 \zeta$ , then  $F_{\tau - \bar{C}\sqrt{\zeta}/2}^* \leq (2\bar{C}^2 - 1)\zeta$ .*

**Proof** The first part is simple and was already stated as Proposition 5.2. Let  $x^*$  be the optimal solution of  $F_\tau^*$ . Let

$$\alpha := \max\{0, \langle \hat{c}, x^* \rangle - \tau\} \quad \text{and} \quad \Delta := \bar{C}\sqrt{\zeta}/2.$$

The lemma follows by showing that

$$F_{\tau - \Delta}(x^*) - F_\tau^* \leq (\bar{C}^2 - 1)\zeta. \quad (18)$$

Since  $F_\tau^* \leq \bar{C}^2 \zeta$  and  $F_\tau^* \geq \frac{1}{2}\alpha^2$ , we have  $\alpha \leq \bar{C}\sqrt{2\zeta} = 2\sqrt{2}\Delta$ . Therefore,

$$\begin{aligned} F_{\tau - \Delta}(x^*) - F_\tau^* &= \frac{1}{2}(\max\{0, \langle \hat{c}, x^* \rangle - \tau + \Delta\})^2 - \frac{1}{2}(\max\{0, \langle \hat{c}, x^* \rangle - \tau\})^2 \\ &\leq \frac{1}{2}(\alpha + \Delta)^2 - \frac{1}{2}\alpha^2 \\ &= \alpha\Delta + \frac{1}{2}\Delta^2 \leq \left(2\sqrt{2} + \frac{1}{2}\right)\Delta^2. \end{aligned}$$

Using the above bound, we get (18) from

$$F_{\tau - \Delta}(x^*) - F_\tau^* \leq \left(2\sqrt{2} + \frac{1}{2}\right)\Delta^2 = \left(\frac{1}{\sqrt{2}} + \frac{1}{8}\right)\bar{C}^2 \zeta \leq (\bar{C}^2 - 1)\zeta.$$

The last inequality follows by Lemma 5.3.  $\square$

The third lemma shows that if  $F_\tau(x)$  lies in the appropriate interval, then the output satisfies the desired properties of the algorithm. The proof is given in Sect. 8.1.

**Lemma 8.3** *Let  $x$  be a  $\zeta$ -approximate minimizer and  $\pi$  as defined in the algorithm. Assume that  $F_\tau(x) \in [\bar{C}^2\zeta, 2\bar{C}^2\zeta]$ . Then*

- (i)  $\theta(x, \pi, \sigma) \leq n\sqrt{\zeta}/2$ ,
- (ii)  $\|Ax - b\|_1 \leq 128n^2m \cdot \kappa^2(\mathcal{X}_A) \cdot \|A\|_1^2\sqrt{\zeta}$ , and
- (iii)  $\|\pi\|_\infty \leq 4n\sqrt{m} \cdot \kappa(\mathcal{X}_A) \cdot \|c\|_\infty$ .

We now give the proof of Theorem 5.4 using these three lemmas.

**Theorem 5.4** *Assume  $LP(A, b, c, u)$  is feasible. Algorithm 2 makes  $O(\log[\|u\|_1nm \cdot \kappa(\mathcal{X}_A)/\delta^{\text{feas}}])$  calls to R-FGM, and altogether these calls use  $O(n^{1.5}m^2\|A\|_1^2 \cdot \bar{\kappa}^3(\mathcal{X}_A) \log^2[\|u\|_1nm \cdot \|A\|_1\kappa(\mathcal{X}_A)/\delta^{\text{feas}}])$  iterations. On terminating, it outputs  $(x, \pi)$  satisfying:*

- (i)  $\theta(x, \pi, \sigma) \cdot \|A\|_1 + \|Ax - b\|_1 \leq \delta^{\text{feas}}\|A\|_1/n$ .
- (ii)  $\kappa(\mathcal{X}_A) \cdot \|c\|_1 \cdot \|Ax - b\|_1 + |J_1 \cup J_2| \cdot \kappa(\mathcal{X}_A) \cdot \|c\|_1 \cdot (2 + \kappa(\mathcal{X}_A)\|A\|_1) \cdot \theta(x, \pi, \sigma) \leq \delta^{\text{opt}}\|c\|_\infty/n$ .
- (iii)  $\|\pi\|_\infty \leq 4n\sqrt{m} \cdot \kappa(\mathcal{X}_A) \cdot \|c\|_\infty$ .

**Proof** First of all, we need to show that the algorithm eventually outputs a primal dual pair  $(x, \pi)$ , i.e., the condition  $F_\tau(x) \in [\bar{C}^2\zeta, 2\bar{C}^2\zeta]$  will be met, and the claimed bound on the number of calls to R-FGM is applicable. This follows by the next claim.

**Claim 8.4**  $\tau^+ - \tau^- \geq \bar{C}\sqrt{\zeta}/2$  throughout Algorithm 2.

**Proof** We show the following invariant property of the algorithm:

$$F_{\tau^-}^* > (2\bar{C}^2 - 1)\zeta \text{ and } F_{\tau^+}^* < \bar{C}^2\zeta.$$

Initially, for any feasible  $x$  to  $LP(A, b, c, u)$ , we have  $F_{\tau^+}(x) = 0$ , and therefore  $F_{\tau^+}^* = 0$ . Also,  $F_{\tau^-}^* \geq 2\bar{C}^2\zeta$  as  $F_{\tau^-} \geq \frac{1}{2}(\max\{0, \langle \hat{c}, x \rangle - \tau^-\})^2 \geq \frac{1}{2}(2\bar{C}\sqrt{\zeta})^2 = 2\bar{C}^2\zeta$ , noting that  $\langle \hat{c}, x \rangle \geq -\|\hat{c}\|_\infty\|u\|_1 \geq -\|u\|_1$ . During the algorithm, each computed  $x$  is a  $\zeta$ -approximate minimizer of  $F_\tau$ , and therefore the updating of  $\tau^-$  and  $\tau^+$  in steps 5–8 maintains the invariant.

For a contradiction, assume  $\tau^+ - \tau^- \leq \bar{C}\sqrt{\zeta}/2$  at some point. Then,  $F_{\tau^+}^* < \bar{C}^2\zeta$  and Lemma 8.2 imply  $F_{\tau^-}^* \leq (2\bar{C}^2 - 1)\zeta$ , a contradiction.  $\square$

To bound the number of calls to R-FGM, note that initially  $\tau^+ - \tau^- = 2\|u\|_1 + 2\bar{C}\sqrt{\zeta} \leq 4\|u\|_1$  as  $\|u\|_1 \geq \delta^{\text{feas}}$ , as required for the input of `GetPrimalDualPair`; and  $\tau^+ - \tau^-$  is halved in every iteration. According to the above claim, it never goes below  $\bar{C}\sqrt{\zeta}/2$ , and applying Lemma 5.3 yields the following bound on the number of calls:

$$O\left(\log\left(\|u\|_1/(\bar{C}\sqrt{\zeta})\right)\right) = O\left(\log\left(\|u\|_1nm \cdot \kappa(\mathcal{X}_A)/\delta^{\text{feas}}\right)\right).$$

In each R-FGM call, by Lemma 8.1, to obtain a  $\zeta$ -approximate minimizer takes at most  $O(k\sqrt{nm}^2\|A\|_1^2 \cdot \bar{\kappa}^2(\mathcal{X}_A)/\varepsilon)$  steps, where  $k = O(\log(\|b\|_2^2/(\|A\|_1^2\zeta)))$  and  $\varepsilon = 1/[8n \cdot \kappa(\mathcal{X}_A)]$ . The claimed bound follows as  $\|b\|_2^2 \leq \|b\|_1^2 \leq \|A\|_1^2 \cdot \|u\|_1^2$ , which follows from the feasibility assumption.

The bounds on  $x$  and  $\pi$  asserted in the theorem follow by simple calculation from the bounds in Lemma 8.3 and the definition of  $\zeta$  in (11).  $\square$

## 8.1 Optimality bounds from the potential function

Recall the parameters  $\varepsilon = 1/(8n \cdot \lceil \kappa(\mathcal{X}_A) \rceil)$ ,  $\sigma = \|c\|_\infty/(4n \cdot \lceil \kappa(\mathcal{X}_A) \rceil) = 2\varepsilon\|c\|_\infty$ . Recall that in Algorithm 2, we defined

$$\alpha(x) := \alpha = \max\{0, \langle \hat{c}, x \rangle - \tau\}, \quad \text{and} \quad \pi(x) := \pi = \frac{\|c\|_\infty}{\|A\|_1^2 \alpha(x)}(b - Ax)$$

Let us further define

$$\beta(x) := \frac{\|Ax - b\|_2}{\|A\|_1}. \quad (19)$$

With this notation, we can write

$$F_\tau(x) = \frac{1}{2}\alpha(x)^2 + \frac{1}{2}\beta(x)^2.$$

We will also often use the gradient, which can be expressed as

$$\nabla F_\tau(x) = \alpha(x) \left( \hat{c} - \frac{A^\top \pi(x)}{\|c\|_\infty} \right). \quad (20)$$

When  $x$  is clear from the context, we simply write  $\alpha$ ,  $\beta$ , and  $\pi$ . We will use a simple convexity statement, formulated in the following general form.

**Proposition 8.5** *Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  be a continuously differentiable convex function with  $[\mathbf{0}, u] \subseteq \text{dom}(f)$ , such that for some  $M$ ,  $f$  satisfies the following smoothness property: for every index  $i \in [n]$ , for every pair  $x, y \in \mathbb{R}^n$  such that  $x_j = y_j$  for all  $j \neq i$ ,*

$$|\nabla_i f(x) - \nabla_i f(y)| \leq M|x_i - y_i|. \quad (21)$$

*Also, suppose that for some  $\eta > 0$ ,  $x$  is an  $\eta$ -approximate minimizer to the program*

$$\min f(x) \quad \text{s.t.} \quad x \in [\mathbf{0}, u].$$

*Then, for any  $i \in [n]$ , the following hold:*

- (i) *If  $\nabla_i f(x) \geq 2\sqrt{M\eta}$ , then  $x_i \leq \sqrt{\eta/M}$ .*
- (ii) *If  $\nabla_i f(x) \leq -2\sqrt{M\eta}$ , then  $x_i \geq u_i - \sqrt{\eta/M}$ .*

**Proof** We only prove part (i); part (ii) follows analogously. For a contradiction, assume that for some  $i \in [n]$ ,  $\nabla_i f(x) \geq 2\sqrt{M\eta}$  and  $x_i > \sqrt{\eta/M}$ . Let us define  $z \in \mathbb{R}^n$  with  $z_i := x_i - \sqrt{\eta/M} > 0$  and  $z_j := x_j$  for  $j \neq i$ . Thus,  $z \in [0, u]$ , and by the smoothness property,  $\nabla_i f(z) \geq \nabla_i f(x) - \sqrt{M\eta} \geq \sqrt{M\eta}$ . By convexity,

$$f(x) \geq f(z) + \langle \nabla f(z), x - z \rangle = f(z) + \nabla_i f(z) \cdot \sqrt{\eta/M} \geq f(z) + \sqrt{M\eta} \cdot \sqrt{\eta/M} > f(z) + \eta,$$

using  $M \geq 1$ . This is a contradiction to the assumption that  $x$  is an  $\eta$ -approximate minimizer.  $\square$

The next lemma is used to prove the first key ingredient of Lemma 8.3, an upper bound on  $\theta(x, \pi, \sigma)$ , provided that  $\alpha(x)$  is sufficiently large.

**Lemma 8.6** *Suppose  $x$  is a  $\zeta$ -approximate minimizer of  $F_\tau(x)$  satisfying  $\varepsilon \cdot \alpha(x) \geq 4\sqrt{\zeta}$ , and  $\pi = \pi(x)$ . Then  $\theta(x, \pi, \sigma) \leq n\sqrt{\zeta}/2$ .*

**Proof** We use  $\alpha = \alpha(x)$ ,  $\beta = \beta(x)$ ,  $\pi = \pi(x)$  throughout. Recall that  $\theta(x, \pi, \sigma) = \sum_{i: c_i - A_i^\top \pi > \sigma} x_i + \sum_{i: c_i - A_i^\top \pi < -\sigma} (u_i - x_i)$ , and that the vector  $\hat{c}$  was obtained by normalizing as  $c/\|c\|_\infty$ , and then rounding down each positive entry to the nearest integer multiple of  $\varepsilon$ , and rounding up each negative entry to the nearest integer multiple of  $\varepsilon$ . Also noting that  $\sigma = 2\varepsilon\|c\|_\infty$ , we obtain the following upper bound in terms of  $\hat{c}$ .

$$\theta(x, \pi, \sigma) \leq \sum_{i: \hat{c}_i - A_i^\top \pi / \|c\|_\infty > \varepsilon} x_i + \sum_{i: \hat{c}_i - A_i^\top \pi / \|c\|_\infty < -\varepsilon} (u_i - x_i). \quad (22)$$

We will show each of the terms  $x_i$  and  $u_i - x_i$  in these two sums is bounded by  $\frac{\sqrt{\zeta}}{2}$ , and then the result is immediate. Recall from (20) that

$$\hat{c}_i - A_i^\top \pi / \|c\|_\infty = \nabla_i F_\tau(x) / \alpha.$$

We apply Proposition 8.5 to  $F_\tau(x)$  and  $\eta = \zeta$ . From (20), a simple calculation shows that  $f(x) = F_\tau(x)$  satisfies the smoothness bound (21) with  $M = 2$ . Namely, for any  $i \in [n]$  and for any  $x, y \in \mathbb{R}^n$  such that  $x_j = y_j$  for  $j \neq i$ ,

$$|\nabla_i F_\tau(x) - \nabla_i F_\tau(y)| \leq \|\hat{c}\|_\infty \|x - y\|_1 + |A_i^\top A_i| \cdot \|(x - y)\|_1 / \|A\|_1^2 \leq 4|x_i - y_i|.$$

$\hat{c}_i - A_i^\top \pi / \|c\|_\infty > \varepsilon$  is equivalent to  $\nabla_i F_\tau(x) \geq \varepsilon\alpha$ . By the assumption of the lemma,  $\varepsilon\alpha \geq 4\sqrt{\zeta} = 2\sqrt{M\zeta}$ .

Thus, Proposition 8.5(i) implies that whenever  $\hat{c}_i - A_i^\top \pi / \|c\|_\infty > \varepsilon$ , we must have  $x_i \leq \sqrt{\zeta}/2$ . Similarly, Proposition 8.5(ii) implies that whenever  $\hat{c}_i - A_i^\top \pi / \|c\|_\infty < -\varepsilon$ , we have  $u_i - x_i \leq \sqrt{\zeta}/2$ . This completes the proof.  $\square$

Note the previous lemma requires a lower bound on  $\alpha$ . Our second lemma shows that this requirement is satisfied if one get a good approximate minimizer with a sufficiently large function value.

**Lemma 8.7** Suppose  $x$  is a  $\zeta$ -approximate minimizer of  $F_\tau(x)$ , satisfying  $F_\tau(x) \geq 10\mathcal{C}^2\zeta$ . Then,  $\alpha(x) \geq \frac{1}{2\mathcal{C}}\sqrt{F_\tau(x)}$ .

The proof relies on the following statement:

**Proposition 8.8** Assume  $LP(A, b, c, u)$  is feasible, and let  $x \in [\mathbf{0}, u]$  with  $\beta(x) > 0$ . Then, for any  $0 \leq \mu \leq 1$ , there exists a solution  $x' \in [\mathbf{0}, u]$  such that  $\alpha(x') \leq \alpha(x) + \mu\mathcal{C}\beta(x)$  and  $\beta(x') \leq (1 - \mu)\beta(x)$ .

**Proof** Let  $z \in \mathbb{R}^n$  be chosen as

$$z := \arg \min\{\|z - x\|^2 \mid Az = b, z \in [\mathbf{0}, u]\} \quad \text{and set} \quad x' := \mu z + (1 - \mu)x.$$

Clearly,  $\beta(x') = (1 - \mu)\beta(x)$ . By Lemma 3.4,  $\|z - x\|_\infty \leq \sqrt{m} \cdot \kappa(\mathcal{X}_A) \cdot \|Ax - b\|_2$ . Therefore,  $\alpha$  increases by at most

$$\begin{aligned} \langle \hat{c}, (\mu z + (1 - \mu)x) - x \rangle &\leq \mu \cdot \|\hat{c}\|_1 \|z - x\|_\infty \leq \mu n \cdot \sqrt{m} \cdot \kappa(\mathcal{X}_A) \cdot \|Ax - b\|_2 \\ &= \mu n \cdot \sqrt{m} \cdot \kappa(\mathcal{X}_A) \cdot \|A\|_1 \beta(x) \leq \mu\mathcal{C}\beta(x), \end{aligned}$$

recalling the definition of  $\mathcal{C}$  in (11). □

**Proof of Lemma 8.7** We distinguish two cases.

**Case i:**  $2\mathcal{C}\alpha(x) \geq \beta(x)$ .

In this case,

$$F_\tau(x) = \frac{1}{2}\alpha(x)^2 + \frac{1}{2}\beta(x)^2 \leq \frac{4\mathcal{C}^2 + 1}{2}\alpha(x)^2.$$

Thus, the claimed  $\alpha(x) \geq \frac{1}{2\mathcal{C}}\sqrt{F_\tau(x)}$  follows by recalling  $\mathcal{C} \geq 1$  (Lemma 5.3).

**Case ii:**  $2\mathcal{C}\alpha(x) < \beta(x)$ .

Let us use Proposition 8.8 with

$$\mu := \frac{\beta(x) - \mathcal{C}\alpha(x)}{(\mathcal{C}^2 + 1)\beta(x)}.$$

Clearly,  $\mu \in [0, 1]$ . Thus, there exists  $x' \in [\mathbf{0}, u]$  such that  $\alpha(x') \leq \alpha(x) + \mu\mathcal{C}\beta(x)$  and  $\beta(x') \leq (1 - \mu)\beta(x)$ . Therefore,

$$F_\tau(x') = \frac{1}{2}\alpha(x')^2 + \frac{1}{2}\beta(x')^2 \leq \frac{1}{2}(\alpha(x) + \mu\mathcal{C}\beta(x))^2 + \frac{1}{2}(1 - \mu)^2\beta(x)^2.$$

Above, we picked  $\mu$  to minimize this expression, and some calculation yields

$$F_\tau(x') \leq \frac{(\alpha(x) + \mathcal{C}\beta(x))^2}{2(\mathcal{C}^2 + 1)}.$$

Further calculation yields

$$F_{\tau}(x) - F_{\tau}(x') \geq \frac{(\mathcal{C}\alpha(x) - \beta(x))^2}{2(\mathcal{C}^2 + 1)} > \frac{\beta(x)^2}{8(\mathcal{C}^2 + 1)},$$

where the last inequality uses the assumption  $2\mathcal{C}\alpha(x) < \beta(x)$ . The same condition also implies that  $F_{\tau}(x) = \frac{1}{2}\alpha(x)^2 + \frac{1}{2}\beta(x)^2 \leq \frac{(4\mathcal{C}^2+1)\beta(x)^2}{8\mathcal{C}^2}$ . Using  $\mathcal{C} \geq 1$ , these two bounds in turn, and also the lower bound on  $F_{\tau}(x)$  assumed in the lemma, we obtain

$$F_{\tau}(x) - F_{\tau}(x') > \frac{\beta^2(x)}{8(\mathcal{C}^2 + 1)} \geq \frac{\mathcal{C}^2}{(4\mathcal{C}^2 + 1)(\mathcal{C}^2 + 1)} F_{\tau}(x) \geq \frac{1}{10\mathcal{C}^2} F_{\tau}(x) \geq \zeta,$$

a contradiction to the assumption that  $x$  is a  $\zeta$ -approximate minimizer of  $F_{\tau}(x)$ .  $\square$

We are now ready to prove Lemma 8.3.

**Proof of Lemma 8.3** By Lemma 8.7, as  $F_{\tau}(x) \geq \bar{\mathcal{C}}^2 \zeta \geq 10\mathcal{C}^2 \zeta$ ,

$$\alpha \geq \frac{\sqrt{F_{\tau}(x)}}{2\mathcal{C}} \geq \frac{\bar{\mathcal{C}}\sqrt{\zeta}}{2\mathcal{C}} \geq 32n \cdot \kappa(\mathcal{X}_A) \cdot \sqrt{\zeta}.$$

Thus,

$$\varepsilon\alpha = \frac{\alpha}{8n \cdot \kappa(\mathcal{X}_A)} \geq 4\sqrt{\zeta}.$$

Lemma 8.6 now gives  $\theta(x, \pi, \sigma) \leq n\sqrt{\zeta}/2$ , proving (i).

In addition, the definition of  $F_{\tau}(x)$  implies that  $\|Ax - b\|_2^2 \leq 2\|A\|_1^2 \cdot F_{\tau}(x)$ . The assumption  $F_{\tau}(x) \leq 2\bar{\mathcal{C}}^2 \zeta$  implies that

$$\|Ax - b\|_1 \leq \sqrt{m} \cdot \|Ax - b\|_2 \leq \sqrt{2m}\|A\|_1 \cdot \sqrt{F_{\tau}(x)} \leq 2\|A\|_1 \cdot \bar{\mathcal{C}}\sqrt{m\zeta},$$

proving (ii).

Finally,

$$\begin{aligned} \|\pi\|_{\infty} &= \frac{\|c\|_{\infty}}{\alpha\|A\|_1^2} \|Ax - b\|_{\infty} \leq \frac{2\mathcal{C}\|c\|_{\infty}}{\bar{\mathcal{C}}\sqrt{\zeta}\|A\|_1^2} \|Ax - b\|_2 \\ &\leq 4\mathcal{C}\|c\|_{\infty}/\|A\|_1 = 4n\sqrt{m} \cdot \kappa(\mathcal{X}_A) \cdot \|c\|_{\infty}, \end{aligned}$$

proving (iii).  $\square$

## 8.2 Convergence speed of R-FGM

Let us define  $B \in \mathbb{R}^{(m+1) \times (n+1)}$ ,  $\tilde{u}, \tilde{b} \in \mathbb{R}^{m+1}$  as

$$B := \begin{pmatrix} A & 0 \\ \|A\|_1 \hat{c}^{\top} & \|A\|_1 \end{pmatrix}, \quad \tilde{u} := \begin{pmatrix} u \\ M \end{pmatrix} \quad \tilde{b} := \begin{pmatrix} b \\ \|A\|_1 \tau \end{pmatrix}$$

for sufficiently large  $M$ . With this notation, minimizing  $F_\tau(x)$  over  $0 \leq x \leq u$  can be written in the form

$$\min \frac{1}{2\|A\|_1^2} \cdot \left\| B \begin{pmatrix} x \\ t \end{pmatrix} - \tilde{b} \right\|_2^2 \quad (23)$$

$$0 \leq \begin{pmatrix} x \\ t \end{pmatrix} \leq \tilde{u}.$$

We let  $\tilde{F}_\tau(x, t)$  denote the objective function. We restate Lemma 8.1 for convenience.

**Lemma 8.1** *After  $O(k\sqrt{nm}^2\|A\|_1^2 \cdot \bar{\kappa}^2(\mathcal{X}_A)/\varepsilon)$  iterations, R-FGM returns an  $e^{-2k}\|b\|_2^2/(2\|A\|_1^2)$ -approximate minimizer of  $F_\tau$ .*

We use the starting point  $(x^0, t^0) = (0, \tau)$ . The bound follows from Theorem 2.5, with the smoothness and quadratic growth bounds as below.

**Lemma 8.9** *The function  $\tilde{F}_\tau(x, t)$  is  $(2n + 2)$ -smooth.*

**Proof** We can bound the smoothness parameter as  $\|B\|_2^2/\|A\|_1^2 \leq (n+1)\|B\|_1^2/\|A\|_1^2 \leq (n+1)(\max_i(A_i + \|A\|_1\hat{c}_i))^2/\|A\|_1^2 \leq 2(n+1)$ .  $\square$

**Lemma 8.10** (Quadratic growth)

$$\kappa(\mathcal{X}_B) \leq 2(m+1)\|A\|_1 \cdot \bar{\kappa}^2(\mathcal{X}_A) \cdot \frac{1}{\varepsilon}.$$

Consequently, the function  $\tilde{F}_\tau(x, t)$  has  $\varepsilon^2/[64m^4 \cdot \|A\|_1^4 \cdot \bar{\kappa}^4(\mathcal{X}_A)]$ -quadratic growth.

The proof is based on the following lemma.

**Lemma 8.11** *For a matrix  $A \in \mathbb{Q}^{m \times n}$  and a vector  $d \in \mathbb{R}^n$ , let  $K = \begin{pmatrix} A \\ d^\top \end{pmatrix}$ . Then, every elementary vector in  $\mathcal{F}(K)$  is either an elementary vector in  $\mathcal{F}(A)$ , or the sum of two conformal elementary vectors in  $\mathcal{F}(A)$ . Further, if  $d \in \mathbb{Z}^n$ ,  $d \neq \mathbf{0}$ , then*

$$\kappa(K) \leq 2(m+1) \cdot \|d\|_\infty \cdot \bar{\kappa}^2(A).$$

**Proof** Let  $z$  be an elementary vector in  $\mathcal{F}(K)$ , and consider a generalized circuit-path decomposition of  $z$  w.r.t.  $A$ ,

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

where  $h \leq n$  and  $g^1, g^2, \dots, g^h \in \mathcal{F}(A)$  are elementary vectors that conform to  $z$ . Further, for each  $i = 1, 2, \dots, h-1$ ,  $\text{supp}(g^i) \setminus \cup_{j=i+1}^h \text{supp}(g^j) \neq \emptyset$ .

The first statement follows by showing  $h \leq 2$ . The proof is by contradiction: suppose  $h > 2$ , and consider  $g^2$  and  $g^3$ .

First, we observe that  $\text{supp}(g^2) \cup \text{supp}(g^3) \subsetneq \text{supp}(z)$ , because  $\text{supp}(g^1) \setminus \cup_{j=2}^h \text{supp}(g^j) \neq \emptyset$ . Next, we show that  $\langle c, g^2 \rangle \neq 0$  and  $\langle c, g^3 \rangle \neq 0$ . For if one of them equals 0, for example,  $\langle c, g^2 \rangle = 0$ , then as  $g^2$  is elementary,  $Ag^2 = 0$  also,

which implies  $z$  is not an elementary vector of  $\mathcal{F}(K)$ , given that  $g_2$  has a strictly smaller support than  $z$ .

To obtain a contradiction, we consider  $g^{23} = \langle c, g^2 \rangle g^3 - \langle c, g^3 \rangle g^2$ , which is non-zero as  $\text{supp}(g^2) \setminus \cup_{j=3}^h \text{supp}(g^j) \neq \emptyset$  and  $\langle c, g^3 \rangle \neq 0$ .  $g^{23}$  also has a strictly smaller support than  $z$ , as  $\text{supp}(g_2) \cup \text{supp}(g_3) \subsetneq \text{supp}(z)$ . In addition,  $\langle c, g^{23} \rangle = 0$  and  $Ag^{23} = 0$ . This implies  $z$  is not an elementary vector of  $\mathcal{F}(K)$ , which provides a contradiction.

Let us now turn to the second statement: assume that  $d \in \mathbb{Z}^n$ . Take an elementary vector  $z$  in  $\mathcal{F}(K)$  such that  $\kappa(K) = |z_i/z_j|$  for some  $i, j \in \text{supp}(z)$ . If  $z \in \mathcal{F}(A)$ , then  $\kappa(K) \leq \kappa(A)$ , and hence the bound follows.

Otherwise,  $z$  is the sum of two elementary vectors in  $\mathcal{F}(A)$ . After appropriately scaling  $z$ , it can be written in the form  $z = \lambda_1 g^1 + \lambda_2 g^2$  with  $g^1, g^2 \in \mathcal{F}(A)$ , i.e., they are integer vectors such that the largest common divisor of their entries is 1. Also,  $\lambda_1, \lambda_2 \neq 0$ . Further, we must have  $0 = \langle d, z \rangle = \lambda_1 \langle d, g^1 \rangle + \lambda_2 \langle d, g^2 \rangle$ . After possibly another scaling of  $z$ , we get  $\lambda_1 = \langle d, g^2 \rangle$  and  $\lambda_2 = -\langle d, g^1 \rangle$ , that is,

$$z = \langle d, g^2 \rangle g^1 - \langle d, g^1 \rangle g^2.$$

By the definition of  $\bar{\kappa}(A)$ ,  $\|g^1\|_\infty, \|g^2\|_\infty \leq \bar{\kappa}(A)$ . By the integrality of  $g^1$  and  $g^2$ ,

$$1 \leq |\langle d, g^2 \rangle|, |\langle d, g^1 \rangle| \leq (m+1) \cdot \|d\|_\infty \cdot \bar{\kappa}(A).$$

Thus, all nonzero entries of  $z$  have  $1 \leq |z_i| \leq 2(m+1) \cdot \|d\|_\infty \cdot \bar{\kappa}(A) \cdot \kappa(A)$ , implying the claim since  $\kappa(A) \leq \bar{\kappa}(A)$ .  $\square$

**Proof of Lemma 8.10** The bound on the quadratic growth parameter follows from the circuit imbalance bound: By Lemma 2.4,  $\tilde{F}_\tau(x, t)$  has  $1/\theta_{2,2}^2(B)$ -quadratic growth, and, by Lemma 3.6,  $\theta_{2,2}(B) \leq (m+1) \cdot \kappa(\mathcal{X}_B)$ .

Let us now show the circuit imbalance bound. Recall that  $\mathcal{X}_A = \ker(A - I_m)$ . Since arbitrarily scaling the rows of a matrix does not change the kernel and thus does not affect the circuit imbalances, we can write  $\mathcal{X}_B = \ker(B - I_m) = \ker(H)$  for

$$H := \begin{pmatrix} A & 0 & -I_m & 0 \\ \frac{1}{\varepsilon} \hat{c}^\top & \frac{1}{\varepsilon} & 0 & -\frac{1}{\|A\|_1 \cdot \varepsilon} \end{pmatrix}.$$

Let  $H'$  be the matrix arising from  $H$  by deleting the last column, and let us also define  $D = (A \ 0 \ -I_m)$ .

Clearly,  $\bar{\kappa}(\mathcal{X}_A) = \bar{\kappa}(D)$ . Recall from (9) that  $1/\varepsilon$  is defined to be an integer. Hence, Lemma 8.11 is applicable to the matrix  $H'$ . Note that the last row of this matrix has  $\ell_\infty$  norm  $1/\varepsilon$ , since  $\|\hat{c}\|_\infty = 1$ . Therefore,  $\kappa(H') \leq 2(m+1) \cdot \bar{\kappa}(\mathcal{X}_A) \cdot \frac{1}{\varepsilon}$ .

We show that  $\kappa(H) \leq \|A\|_1 \cdot \kappa(H')$ ; this implies the statement. Indeed,  $H$  arises from  $H'$  by duplicating one of its columns and scaling it by  $1/\|A\|_1$ . Duplicating columns does not affect the circuit imbalance, whereas multiplying a column by any constant  $1/\alpha$  may increase it by at most a factor  $\alpha$ . This completes the proof.  $\square$

## 9 Analysis of dual certificates

We now present the proof of Lemma 4.6 and Theorem 4.7 on  $\delta$ -certificates.

**Lemma 4.6** *Let  $A \in \mathbb{R}^{m \times n}$ ,  $b \in \mathbb{R}^m$ ,  $c, u \in \mathbb{R}^n$ , and let  $x \in [\mathbf{0}, u]$ .*

- (i) *If there is a  $\delta$ -certificate for  $x$  for some  $\delta \geq 0$ , then  $x$  is  $(4n + 2)\delta$ -optimal.*
- (ii) *Suppose  $0 \leq \delta^{\text{feas}} \cdot n \cdot \kappa(\mathcal{X}_A) \cdot \|A\|_1 \leq \delta^{\text{opt}}$ . If  $x$  is a  $\delta^{\text{feas}}$ -feasible and  $\delta^{\text{opt}}$ -optimal solution, then there exists a  $\delta^{\text{opt}}$ -certificate for  $x$ .*

**Proof Part (i)** Let  $(\pi, w^-, w^+)$  denote the  $\delta$ -certificate. This is a feasible solution to  $\text{Dual}(A, b, c, u)$ , and therefore  $\Phi(A, b, c, u) \geq \langle b, \pi \rangle - \langle u, w^+ \rangle$ . Thus, using the properties of  $\delta$ -certificates, we get the bound

$$\begin{aligned} \langle c, x \rangle - \Phi(A, b, c, u) &\leq \langle c, x \rangle - \langle b, \pi \rangle + \langle u, w^+ \rangle \\ &= \langle c - A^\top \pi + w^+, x \rangle + \langle Ax - b, \pi \rangle + \langle u - x, w^+ \rangle \\ &= \langle w^-, x \rangle + \langle w^+, u - x \rangle + \langle Ax - b, \pi \rangle \\ &\leq (4n + 2)\delta \|c\|_\infty. \end{aligned}$$

**Part (ii)**

Let  $\bar{b} = Ax$ . By Lemma 6.3, and the assumption on  $\delta^{\text{feas}}$ , it follows that

$$\begin{aligned} |\Phi(A, b, c, u) - \Phi(A, \bar{b}, c, u)| &\leq \kappa(\mathcal{X}_A) \cdot \|c\|_1 \cdot \|b - \bar{b}\|_1 \leq n \\ &\cdot \kappa(\mathcal{X}_A) \cdot \|c\|_\infty \cdot \|A\|_1 \cdot \delta^{\text{feas}} \leq \delta^{\text{opt}} \cdot \|c\|_\infty, \end{aligned}$$

Thus,  $x$  is  $2\delta^{\text{opt}}$ -optimal for  $\text{LP}(A, \bar{b}, c, u)$ . Let us select an optimal dual solution  $(\bar{\pi}, \bar{w}^-, \bar{w}^+)$  to  $\text{Dual}(A, \bar{b}, c, u)$ . Thus,  $\langle \bar{b}, \bar{\pi} \rangle - \langle u, \bar{w}^+ \rangle = \Phi(A, \bar{b}, c, u)$ . As in part (i), we get

$$\begin{aligned} 2\delta^{\text{opt}} \|c\|_\infty &\geq \langle c, x \rangle - \Phi(A, \bar{b}, c, u) = \langle c, x \rangle - \langle \bar{b}, \bar{\pi} \rangle + \langle u, \bar{w}^+ \rangle \\ &= \langle \bar{w}^-, x \rangle + \langle \bar{w}^+, u - x \rangle + \langle Ax - \bar{b}, \bar{\pi} \rangle. \end{aligned}$$

Since all terms here are nonnegative, we get that  $(\bar{\pi}, \bar{w}^-, \bar{w}^+)$  is a feasible solution to the LP

$$\begin{aligned} A^\top \pi + w^- - w^+ &= c \\ 0 \leq w_i^- &\leq \frac{2\delta^{\text{opt}} \|c\|_\infty}{x_i}, \quad \forall i \in [n] \\ 0 \leq w_i^+ &\leq \frac{2\delta^{\text{opt}} \|c\|_\infty}{u_i - x_i}, \quad \forall i \in [n]. \end{aligned} \tag{24}$$

We now show that (24) has a feasible solution  $(\pi, w^-, w^+)$  with

$$\|\pi\|_\infty \leq \kappa(\mathcal{X}_A) \cdot \|c\|_1. \tag{25}$$

This implies that  $(\pi, w^-, w^+)$  also satisfies requirement (iii) in the definition of  $\delta^{\text{opt}}$ -certificates (Definition 4.5), since

$$\begin{aligned} \kappa(\mathcal{X}_A) \cdot \|c\|_1 &\leq \|c\|_1 \cdot \delta^{\text{opt}} / [\delta^{\text{feas}} \cdot n \cdot \|A\|_1] \leq (\|c\|_1 / n) \cdot \delta^{\text{opt}} / \|Ax - b\|_1 \\ &\leq \delta^{\text{opt}} \|c\|_\infty / \|Ax - b\|_1, \end{aligned}$$

where the first inequality uses the assumption that  $\delta^{\text{feas}} \cdot n \cdot \kappa(\mathcal{X}_A) \cdot \|A\|_1 \leq \delta^{\text{opt}}$ , and the second inequality uses the fact that  $x$  is  $\delta^{\text{feas}}$ .

We now show the existence of such a solution to (24). Let  $c' := \max\{0, c\}$  and  $c'' := \max\{0, -c\}$ . Thus,  $(0, c', c'')$  satisfies all inequalities in (24) except the upper bounds. Now, let  $(\pi, w^-, w^+)$  be a feasible solution to (24) such that the distance  $\|(0, c', c'') - (\pi, w^-, w^+)\|_2$  is minimal.

Let  $\mathcal{Y} := \ker(A^\top |I_n| - I_n)$ . By Lemma 3.7, and noting that duplicating columns does not affect the circuit imbalances, we see that  $\kappa(\mathcal{Y}) = \kappa(\mathcal{X}_A)$ .

Note that  $(0, c', c'') - (\pi, w^-, w^+) \in \mathcal{Y}$ . Let  $\sum_{k=1}^h g^k$  be a conformal circuit decomposition of the difference of these vectors, where  $g^k \in \mathbb{R}^{m \times n \times n}$ . By the choice of  $(\pi, w^-, w^+)$ , the support of each  $g^k$  must contain at least one coordinate in the second block with  $c'_i > \frac{2\delta^{\text{opt}}\|c\|_\infty}{x_i}$ , or in the third block with  $c''_j > \frac{2\delta^{\text{opt}}\|c\|_\infty}{u_j - x_j}$ , and the corresponding component of  $g^k$  must be negative.

By the conformity of the circuit decomposition, and the definition of  $\kappa(\mathcal{Y})$ , it follows that  $\|\pi\|_\infty \leq \kappa(\mathcal{Y}) \cdot (\|c'\|_1 + \|c''\|_1) = \kappa(\mathcal{X}_A) \cdot \|c\|_1$ ; thus, (25) holds. Therefore,  $(\pi, w^-, w^+)$  is a  $\delta^{\text{opt}}$ -certificate for  $x$ .  $\square$

**Theorem 4.7** Suppose  $A \in \mathbb{R}^{m \times n}$ ,  $b \in \mathbb{R}^m$ ,  $x, c, u \in \mathbb{R}^n$ ,  $0 \leq \delta^{\text{feas}} \cdot n \cdot \kappa(\mathcal{X}_A) \cdot \|A\|_1 \leq \delta^{\text{opt}}$ ,  $\|A\|_1 \geq 1$  and  $x \in [0, u]$  is both  $\delta^{\text{feas}}$ -feasible and  $\delta^{\text{opt}}$ -optimal. Then there is an algorithm *Dual-Certificate*( $x, A, b, c, u, \delta^{\text{feas}}, \delta^{\text{opt}}$ ) which on such inputs finds a  $2 \cdot \delta^{\text{opt}}$ -certificate for  $x$  in  $O(m\sqrt{n} \cdot \|A\|_1 \cdot \kappa(\mathcal{X}_A) \cdot \log(n\|c\|_1/\delta^{\text{opt}}))$  iterations of R-FGM.

**Proof** Similarly to the feasibility algorithm in Theorem 4.1, we use R-FGM on a convex quadratic minimization problem. Namely, we consider the problem

$$\begin{aligned} \min \quad & \frac{1}{2} \|A^\top \pi + w^- - w^+ - c\|^2 \\ & 0 \leq w_i^- \leq \frac{2\delta^{\text{opt}}\|c\|_\infty}{x_i}, \quad \forall i \in [n], \\ & 0 \leq w_i^+ \leq \frac{2\delta^{\text{opt}}\|c\|_\infty}{u_i - x_i}, \quad \forall i \in [n], \\ & -\frac{2\delta^{\text{opt}}\|c\|_\infty}{\|Ax - b\|_1} \leq \pi_i \leq \frac{2\delta^{\text{opt}}\|c\|_\infty}{\|Ax - b\|_1}, \quad \forall i \in [n]. \end{aligned}$$

Lemma 4.6(ii) guarantees the existence of a  $\delta^{\text{opt}}$ -certificate, and using Definition 4.5(i), we can deduce that the optimal value for this program is 0.

We run R-FGM, starting from the all zero solution to find an  $\varepsilon$ -approximate solution, where  $\varepsilon := \frac{1}{2} \cdot \left( \frac{2\delta^{\text{opt}}\|c\|_\infty}{\|u\|_1} \right)^2$ . Note that  $\mu := \sqrt{2\varepsilon}$  is smaller than any of the bounds

in the box constraints above. Hence, we can modify the solution to  $(\pi, \tilde{w}^-, \tilde{w}^+)$  such that  $A^\top \pi + \tilde{w}^- - \tilde{w}^+ = c$ , and this solution violates the box constraints by at most a factor 2. Thus,  $(\pi, \tilde{w}^-, \tilde{w}^+)$  is a  $2\delta^{\text{opt}}$ -certificate.

By Lemma 2.4, the quadratic growth parameter is  $1/\theta_{2,2}^2(A^\top |I_n| - I_n)$ , which is larger than  $1/m^2\kappa^2(\mathcal{X}_A)$ . This bound follows by Lemma 3.7 and the fact that duplicating columns does not affect the circuit imbalances. In addition, by Lemma 2.2, the smoothness parameter is  $(\|A\|_2 + 1)^2$  as

$$\begin{aligned} \|(A^\top |I_n| - I_n)\|_2 &= \max_{p,q,r} \frac{\sqrt{\|A^\top p\|_2^2 + \|q\|_2^2 + \|r\|_2^2}}{\sqrt{\|p\|_2^2 + \|q\|_2^2 + \|r\|_2^2}} \\ &\leq \max_{p,q,r} \frac{\sqrt{\|A^\top p\|_2^2}}{\sqrt{\|p\|_2^2}} + \frac{\sqrt{\|q\|_2^2 + \|r\|_2^2}}{\sqrt{\|q\|_2^2 + \|r\|_2^2}} \leq \|A\|_2 + 1. \end{aligned}$$

Given this, by Theorem 2.5, the total number of iterations of R-FGM is at most  $O(m\sqrt{n}\|A\|_1 \cdot \kappa(\mathcal{X}_A) \cdot \log(n\|u\|_1/\delta^{\text{opt}}))$ .  $\square$

## 10 Hoffman constant example for the self-dual embedding

We show that for the self-dual embedding (2), the Hoffman constant of the corresponding matrix can be unbounded.

**Lemma 10.1** *Let  $A \in \mathbb{R}^{m \times n}$ ,  $b \in \mathbb{R}^m$ ,  $u \in \mathbb{R}^n$  with  $\mathcal{P}_{A,b,u} \neq \emptyset$  being at least 2-dimensional. Then, for any  $M > 0$ , there exist a cost function  $c \in \mathbb{R}^n$  such that the Hoffman constant is  $\theta_{2,2}(\mathcal{S}_c) \geq M$  for the system*

$$\mathcal{S}_c = \left\{ (x, \pi, w^-, w^+) \mid \begin{array}{l} x \in \mathcal{P}_{A,b,u}; A^\top \pi + w^- - w^+ = c; \\ \langle c, x \rangle - \langle b, \pi \rangle + \langle u, w^+ \rangle = 0; w^-, w^+ \geq 0 \end{array} \right\}.$$

**Proof** Let us denote  $\mathcal{P} = \mathcal{P}_{A,b,u}$ , and consider any facet  $\mathcal{F}$  of  $\mathcal{P}_{A,b}$  corresponding to the linear equation  $\langle r, x \rangle = v$  for some  $r \in \mathbb{R}^n$  and  $v \in \mathbb{R}$ . We define  $c \approx r$  as a perturbation of  $r$  with  $\|c - r\|_2 \leq 1/M$ , such that  $\max \langle c, x \rangle$  s.t.  $x \in \mathcal{P}$  has a unique optimal solution  $\bar{x} \in \mathcal{F}$ . Let  $x' \in \mathcal{F} \setminus \{\bar{x}\}$  be another extreme point. Let  $(\pi, w^+, w^-)$  be an optimal solution to the dual of  $\max \langle c, x \rangle$  s.t.  $x \in \mathcal{P}$ . Then, for the primal-dual pair  $(x', \pi, w^+, w^-)$ , we have

- (1)  $x' \in \mathcal{P}$ ;
- (2)  $(\pi, w^+, w^-)$  is feasible:  $A^\top \pi + w^- - w^+ = c$  and  $w^+, w^- \geq 0$ ;
- (3) the optimality gap is tiny but strictly positive, i.e.,  $0 < \langle c, x' \rangle - \langle b, \pi \rangle + \langle u, w^+ \rangle \leq \langle c, x' - \bar{x} \rangle \leq \|c - r\|_2 \cdot \|x' - \bar{x}\|_2 \leq \frac{1}{M} \|x' - \bar{x}\|_2$ ;
- (4) the distance to the nearest point in  $\mathcal{S}_c$  is  $\|x' - \bar{x}\|_2$ .

In this case, the Hoffman constant  $\theta_{2,2}(\mathcal{S}_c)$  is at least the distance to the nearest point in  $\mathcal{S}_c$  divided by the optimality gap, which is at least  $M$ .  $\square$

**Acknowledgements** A preliminary conference version of this article appeared in the Proceedings of the *25th Conference on Integer Programming and Combinatorial Optimization* [3]. This work was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. ScaleOpt-757481). C. Hertrich also acknowledges funding through the ERC grant ForEFront-615640 and the Marie Skłodowska-Curie grant 101153187-NeurExCo. Y. Tao also acknowledges Grant 2023110522 from SUFE, National Key R&D Program of China (2023YFA1009500), NSFC grant 61932002. Part of the work was done while L. Végh was visiting the Corvinus Institute for Advanced Studies, Corvinus University, Budapest, Hungary, and while C. Hertrich was affiliated with London School of Economics and Political Science, UK, with Université Libre de Bruxelles, Belgium, and with Goethe-Universität Frankfurt, Germany.

**Funding** Open Access funding enabled and organized by Projekt DEAL.

**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. Applegate, D., Díaz, M., Hinder, O., et al.: Practical large-scale linear programming using primal-dual hybrid gradient. *Adv. Neural. Inf. Process. Syst.* **34**, 20243–20257 (2021)
2. Applegate, D., Hinder, O., Lu, H., et al.: Faster first-order primal-dual methods for linear programming using restarts and sharpness. *Math. Program.* **201**(1), 133–184 (2023)
3. Cole, R., Hertrich, C., Tao, Y., et al.: A first order method for linear programming parameterized by circuit imbalance. In: Vygen, J., Byrka, J. (eds.) *Integer Programming and Combinatorial Optimization*, pp. 57–70. Springer Nature Switzerland, Cham (2024)
4. Dadush, D., Natus, B., Végh, L.A.: Revisiting Tardos's framework for linear programming: Faster exact solutions using approximate solvers. In: *Proceedings of the 61st Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pp 931–942 (2020)
5. Dadush, D., Huijberts, S., Natus, B., et al.: A scaling-invariant algorithm for linear programming whose running time depends only on the constraint matrix. *Math. Program.* **204**, 135–206 (2024)
6. Eckstein, J., Bertsekas, D.P., et al.: An alternating direction method for linear programming. *Tech. Rep. LIDS-P-1967* (1990)
7. Ekbatani, F., Natus, B., Végh, L.A.: Circuit imbalance measures and linear programming. In: *Surveys in Combinatorics 2022*, London Mathematical Society Lecture Note Series. Cambridge University Press, p 64–114 (2022)
8. Fujishige, S., Kitahara, T., Végh, L.A.: An update-and-stabilize framework for the minimum-norm-point problem. *Math. Program.* **210**, 281–311 (2025)
9. Fulkerson, D.: Networks, frames, blocking systems. *Mathematics of the Decision Sciences, Part I, Lectures in Applied Mathematics* **2**, 303–334 (1968)
10. Gilpin, A., Peña, J., Sandholm, T.: First-order algorithm with convergence for-equilibrium in two-person zero-sum games. *Math. Program.* **133**(1–2), 279–298 (2012)
11. Hinder, O.: Worst-case analysis of restarted primal-dual hybrid gradient on totally unimodular linear programs. *Oper. Res. Lett.* **57**, 107199 (2024)
12. Hoffman, A.J.: On approximate solutions of systems of linear inequalities. *J. Res. Natl. Bur. Stand.* **49**(4), 263–265 (1952)
13. 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)
14. Khachiyan, L.G.: A polynomial algorithm in linear programming. In: *Doklady Akademii Nauk SSSR*, pp 1093–1096 (1979)

15. Necoara, I., Nesterov, Y., Glineur, F.: Linear convergence of first order methods for non-strongly convex optimization. *Math. Program.* **175**, 69–107 (2019)
16. Rockafellar, R.T.: The elementary vectors of a subspace of  $R^N$ . In: *Combinatorial Mathematics and Its Applications: Proceedings North Carolina Conference, Chapel Hill, 1967*. The University of North Carolina Press, pp 104–127 (1969)
17. Smale, S.: Mathematical problems for the next century. *The Mathematical Intelligencer* **20**, 7–15 (1998)
18. Tardos, É.: A strongly polynomial minimum cost circulation algorithm. *Combinatorica* **5**(3), 247–255 (1985)
19. Tardos, É.: A strongly polynomial algorithm to solve combinatorial linear programs. *Operations Research* pp 250–256 (1986)
20. Tunçel, L.: Approximating the complexity measure of Vavasis-Ye algorithm is NP-hard. *Math. Program.* **86**(1), 219–223 (1999)
21. 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)
22. Wang, S., Shroff, N.: A new alternating direction method for linear programming. *Advances in Neural Information Processing Systems* **30** (2017)
23. Yang, T., Lin, Q.: RSG: Beating subgradient method without smoothness and strong convexity. *The Journal of Machine Learning Research* **19**(1), 236–268 (2018)

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