

# Interior point methods are not worse than Simplex

Xavier Allamigeon\*    Daniel Dadush†    Georg Loho‡    Bento Natura§  
László A. Végh¶

## Abstract

We develop a new ‘subspace layered least squares’ interior point method (IPM) for solving linear programs. Applied to an  $n$ -variable linear program in standard form, the iteration complexity of our IPM is up to an  $O(n^{1.5} \log n)$  factor upper bounded by the *straight line complexity* (SLC) of the linear program. This term refers to the minimum number of segments of any piecewise linear curve that traverses the *wide neighborhood* of the central path, a lower bound on the iteration complexity of any IPM that follows a piecewise linear trajectory along a path induced by a self-concordant barrier. In particular, our algorithm matches the number of iterations of any such IPM up to the same factor  $O(n^{1.5} \log n)$ .

As our second contribution, we show that the SLC of any linear program is upper bounded by  $2^{n(1+o(1))}$ , which implies that our IPM’s iteration complexity is at most exponential. This is in contrast to existing iteration complexity bounds that depend on either bit-complexity or condition measures; these can be unbounded in the problem dimension. We achieve our upper bound by showing that the central path is well-approximated by a combinatorial proxy we call the *max central path*, which consists of  $2n$  shadow vertex simplex paths. Our upper bound complements the lower bounds of Allamigeon, Benchimol, Gaubert, and Joswig (SIAGA 2018), and Allamigeon, Gaubert, and Vandame (STOC 2022), who constructed linear programs with exponential SLC.

Finally, we show that each iteration of our IPM can be implemented in strongly polynomial time. Along the way, we develop a deterministic algorithm that approximates the singular value decomposition of a matrix in strongly polynomial time to high accuracy, which may be of independent interest.

---

\*This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme: DD from grant agreement no. 805241-QIP, BN and LAV from grant agreements no. 757481-ScaleOpt. We thank the 2021 Hausdorff Research Institute for Mathematics Program Discrete Optimization during which part of this work was developed.

\*Inria, CMAP, CNRS, Ecole Polytechnique, IP Paris, France

†Centrum Wiskunde & Informatica, Amsterdam, The Netherlands

‡University of Twente, The Netherlands

§Georgia Institute of Technology, USA

¶London School of Economics and Political Science, UK

# Contents

|           |   |           |
|-----------|---|-----------|
| <b>1</b>  | <b>Introduction</b>   | <b>3</b>  |
| 1.1       | Contributions . . . . .   | 4         |
| 1.2       | Techniques . . . . .  | 8         |
| 1.2.1     | Polarization of the Central Path . . . . .                                  | 8         |
| 1.2.2     | Traversing a Polarized Segment . . . . .                                    | 9         |
| 1.3       | Related Work . . . . .  | 11        |
| 1.4       | Organization of the Paper . . . . .   | 12        |
| <b>2</b>  | <b>Preliminaries</b>  | <b>12</b> |
| 2.1       | Notation . . . . .  | 12        |
| 2.2       | Models of Computation . . . . .   | 13        |
| 2.3       | Linear Algebra Preliminaries . . . . .                                      | 14        |
| 2.3.1     | Approximate Singular Subspaces . . . . .                                    | 18        |
| 2.4       | Preliminaries on Interior-Point Methods . . . . .                           | 19        |
| 2.5       | The Central Path Limit . . . . .  | 20        |
| 2.6       | Predictor-Corrector Methods . . . . .                                       | 21        |
| <b>3</b>  | <b>Polarization of the Central Path</b>                                     | <b>24</b> |
| <b>4</b>  | <b>The Max Central Path</b>   | <b>27</b> |
| 4.1       | The Shadow Vertex Simplex Rule . . . . .                                    | 28        |
| 4.2       | Decomposition of Straight Line Segments into Polarized Segments . . . . .   | 29        |
| <b>5</b>  | <b>The Trust Region Step</b>  | <b>32</b> |
| 5.1       | Step-Length Estimates for Trust Region Directions . . . . .                 | 32        |
| 5.2       | The Associated Partition . . . . .  | 34        |
| <b>6</b>  | <b>The Subspace LLS Direction and Cheap Lift Subspaces</b>                  | <b>35</b> |
| 6.1       | The Subspace Layered Least Squares Direction . . . . .                      | 35        |
| 6.2       | Lifting Maps and Operators . . . . .  | 36        |
| 6.3       | Cheap Lift Subspaces . . . . .  | 39        |
| 6.4       | Computing Cheap Lift Subspaces . . . . .                                    | 41        |
| 6.4.1     | Approximating the Singular Subspaces . . . . .                              | 41        |
| 6.4.2     | The Cheap Lift Subspace Algorithm . . . . .                                 | 41        |
| <b>7</b>  | <b>The Subspace Layered Least Squares Algorithm</b>                         | <b>44</b> |
| 7.1       | Description of the Algorithm . . . . .                                      | 44        |
| 7.2       | Correctness . . . . .   | 45        |
| 7.3       | Bounding the Number of Iterations to Traverse a Polarized Segment . . . . . | 46        |
| 7.3.1     | Stability of Singular Values on Polarized Segments . . . . .                | 51        |
| 7.3.2     | Properties of the Ideal Direction . . . . .                                 | 52        |
| 7.3.3     | Dimension and Lifting Cost of Cheap Lift Subspaces . . . . .                | 53        |
| <b>8</b>  | <b>Amortized Iteration Bound for SLLS-IPM</b>                               | <b>54</b> |
| 8.1       | Stability of Singular Values under a Partition Change . . . . .             | 55        |
| 8.2       | Proof of the Amortized Bound . . . . .                                      | 56        |
| <b>9</b>  | <b>Computing Approximate Singular Value Decompositions</b>                  | <b>58</b> |
| 9.1       | Boosting via the Power Method . . . . .                                     | 64        |
| <b>10</b> | <b>Self-Concordant Barrier Central Paths</b>                                | <b>65</b> |
| <b>A</b>  | <b>Missing Proofs in Section 2.3</b>  | <b>70</b> |
| A.1       | Approximate Singular Subspaces . . . . .                                    | 71        |

# 1 Introduction

In this paper, we develop a new interior point method for finding exact solutions to linear programming. Our method is ‘universal’ in the sense that it matches the iteration complexity of any other interior point method up to a small polynomial factor. Our analysis also reveals connections between interior point methods and the simplex method. We consider linear programming (LP) in the following primal-dual form:

$$\begin{aligned} \min \langle c, x \rangle & & \max \langle b, y \rangle \\ \mathbf{A}x = b & & \mathbf{A}^\top y + s = c \\ x \geq \mathbf{0}, & & s \geq \mathbf{0}, \end{aligned} \tag{LP}$$

where  $\mathbf{A} \in \mathbb{R}^{m \times n}$ ,  $b \in \mathbb{R}^m$  and  $c \in \mathbb{R}^n$ , and  $\text{rk}(\mathbf{A}) = m \leq n$ . We let

$$\mathcal{P} := \{x \in \mathbb{R}^n : \mathbf{A}x = b, x \geq \mathbf{0}\}, \quad \mathcal{D} := \{s \in \mathbb{R}^n : \exists y \text{ s.t. } \mathbf{A}^\top y + s = c, s \geq \mathbf{0}\}$$

denote the primal and dual feasible regions and furthermore

$$\mathcal{P}_{++} := \{x \in \mathcal{P} : x > \mathbf{0}\}, \quad \mathcal{D}_{++} := \{s \in \mathcal{D} : s > \mathbf{0}\}$$

denote the primal and dual strictly feasible regions. We summarize the notation used in the paper in Section 2. Throughout, we assume that (LP) is feasible and bounded; consequently,  $\mathcal{P}, \mathcal{D} \neq \emptyset$ . We let  $v^*$  denote the optimum value of (LP). Our focus is on LP algorithms that find exact primal and dual optimal solutions.

The *simplex method* traverses a path formed by vertices and edges of  $\mathcal{P}$  according to a certain *pivot rule*. Albeit efficient in practice, there is no polynomial-time variant known, and there are exponential worst case examples for several pivot rules. The first such construction was given by Klee and Minty [37] for Dantzig’s pivot rule.

Breakthrough developments in the seventies and eighties led to the first polynomial-time algorithms for linear programming: the *ellipsoid method* by Khachiyan [36], and *interior point methods* introduced by Karmarkar [34]. These algorithms run in time  $\text{poly}(n, L)$ , where  $L$  denotes the encoding-length  $L$  of the rational input  $(\mathbf{A}, b, c)$  of (LP).

While the simplex method may require exponentially many iterations, it is never worse: for any non-cycling pivot rule, the number of pivot steps can be bounded by the number of bases, at most  $\binom{n}{m} < 2^n$ . Whereas the bound  $\text{poly}(n, L)$  is typically much better, the encoding length  $L$  may be arbitrarily large. To the extent of our knowledge, no variant of the ellipsoid or interior point methods have been shown to admit a bound  $f(n)$  on the number of iterations for any function  $f: \mathbb{N} \rightarrow \mathbb{N}$  prior to our work.

Even though LPs with exponential encoding length do not frequently appear in practice, there are examples when the binary encoding is exponential yet one could efficiently implement arithmetic operations using a different encoding, see Megiddo [43]. The net present value problem in project scheduling is a particular example of a natural optimization problem that can be reformulated as an LP of exponential encoding length, see Grinold [29]. From a theoretical perspective, finding an interior point method with an absolute bound  $f(n)$  on the number of iterations connects to the fundamental open question on finding a strongly polynomial algorithm for linear programming. In such an algorithm, the number of arithmetic operations is bounded as  $\text{poly}(n)$ , and the algorithm uses polynomial space. This question takes its roots in the development of the simplex method, and appears in Smale’s list of open problems for the 21<sup>st</sup> century [55].

**Interior point methods and the central path** Whereas the simplex method moves on the boundary of the feasible region  $\mathcal{P}$ , interior point methods (IPM) reach an optimal solution by iterating through the strict interior of  $\mathcal{P}$ . Path-following interior point methods are driven to an optimal point by following a smooth trajectory called the *central path*. In the most standard setting [51], the latter is defined as the parametric curve  $\mu \in (0, \infty) \mapsto z^{\text{cp}}(\mu) := (x^{\text{cp}}(\mu), s^{\text{cp}}(\mu))$ , where  $x^{\text{cp}}(\mu)$  and  $(y^{\text{cp}}(\mu), s^{\text{cp}}(\mu))$  are the unique solutions to the system

$$\begin{aligned} \mathbf{A}x^{\text{cp}}(\mu) &= b, \quad x^{\text{cp}}(\mu) > \mathbf{0} \\ \mathbf{A}^\top y^{\text{cp}}(\mu) + s^{\text{cp}}(\mu) &= c, \quad s^{\text{cp}}(\mu) > \mathbf{0} \\ x^{\text{cp}}(\mu)_i s^{\text{cp}}(\mu)_i &= \mu \quad \text{for all } i \in [n]. \end{aligned} \tag{1}$$

This system arises from the optimality conditions of convex problems obtained by penalizing the original linear programs with the logarithmic barrier, i.e., respectively adding terms of the form  $-\mu \sum_{i=1}^n \log x_i$

and  $\mu \sum_{i=1}^n \log s_i$  to the objective functions of the primal and dual (LP). The weight of the penalty is given by the parameter  $\mu > 0$ . When  $\mu \searrow 0$ , the central path  $z^{\text{cp}}(\mu)$  converges to a pair of optimal solutions  $(x^*, s^*)$  of (LP), which can be easily deduced from the fact that the duality gap of  $z^{\text{cp}}(\mu)$  is given by  $\langle c, x^{\text{cp}}(\mu) \rangle - \langle b, y^{\text{cp}}(\mu) \rangle = \langle x^{\text{cp}}(\mu), s^{\text{cp}}(\mu) \rangle = n\mu$ . Accordingly, we define the quantity  $\bar{\mu}(z) := \langle x, s \rangle / n$  for any feasible point  $z = (x, s) \in \mathcal{P} \times \mathcal{D}$ , which we refer to as the *normalized duality gap* of  $z$ .

Interior point methods iteratively compute approximations of the points on the central path associated with successive values of  $\mu$  that decrease geometrically; at most  $O(\sqrt{n} \log(\mu/\mu'))$  iterations are needed to decrease the normalized duality gap from  $\mu$  to  $\mu'$ . The iterations follow an improvement direction, e.g., a Newton step, while remaining in a certain neighborhood of the central path, and can be implemented in polynomial time. The classical analysis yields a bound  $O(\sqrt{n}L)$  on the number of iterations for solving (LP) for a rational input  $(\mathbf{A}, b, c)$  of total encoding length  $L$ . There have been significant improvements in recent years both for general LP as well as for special classes, see Section 1.3.

A running time bound dependent on  $L$  requires a rational input; in contrast, the simplex method can be implemented in  $2^n \text{poly}(n)$  even in the *real model of computation*. Whereas standard IPMs use bit-complexity arguments to terminate, they have also been extended to the real model of computation, e.g., by Vavasis and Ye [65]. The running time of such algorithms is parametrized by *condition numbers* that capture geometric properties of the input. In a remarkable paper, Vavasis and Ye [66] introduced a *layered least squares (LLS)* interior point method that runs in  $O(n^{3.5} \log(\bar{\chi}_{\mathbf{A}} + n))$  iterations, where  $\bar{\chi}_{\mathbf{A}}$  is the Dikin–Stuart–Todd condition number associated with the kernel of  $\mathbf{A}$  (but independent of  $b$  and  $c$ ). As a consequence, they also derive a structural characterization of the central path: there are at most  $\binom{n}{2}$  ‘short and curved’ segments, possibly separated by ‘long and straight’ segments. The LLS directions are refined Newton steps that can traverse the latter segments.

Lan, Monteiro and Tsuchiya [38] gave a scaling invariant *trust region IPM* taking  $O(n^{3.5} \log(\bar{\chi}_{\mathbf{A}}^* + n))$  iterations. Here,  $\bar{\chi}_{\mathbf{A}}^*$  is the minimum value of  $\bar{\chi}_{\mathbf{A}}$  that can be achieved by any column rescaling. However, computing the step directions in this algorithm has a weakly polynomial dependence on  $b$  and  $c$ . In recent work, Dadush, Huiberts, Natura, and Vég h [18] gave a scaling invariant LLS algorithm with iteration bound  $O(n^{2.5} \log(n) \log(\bar{\chi}_{\mathbf{A}}^* + n))$ , where the step directions can be computed by solving linear systems. We discuss the literature on these IPMs in more detail in Section 1.3.

**Lower bounds on interior point methods** LLS methods provide strongly polynomial LP algorithms whenever  $\bar{\chi}_{\mathbf{A}}^* \in 2^{\text{poly}(n)}$ ; this is always the case if the encoding-length of  $\mathbf{A}$  is polynomially bounded. One may wonder if some variant of IPM could be strongly polynomial for all LPs. A negative answer to this question was given in recent work by Allamigeon, Benchimol, Gaubert, and Joswig: they used tropical geometry to build pathological linear programs on which the number of iterations of IPM has to be exponential (in  $m, n$ ) [1, 2]. Their construction shows that, when the entries of  $\mathbf{A}$ ,  $b$ , and  $c$  are of very different orders of magnitude, the central path can be significantly deformed to the boundary of the feasible set. Allamigeon, Gaubert and Vandame later extended this result to the broad class of path-following IPMs using any self-concordant barrier function [3]; concurrently, Zong, Lee, and Yue [69] obtained impossibility for short step methods. The paper [3] exhibits a counterexample where the feasible set is an  $n$ -dimensional combinatorial cube and the shape of the central path is analogous to the simplex paths on pathological instances of LP for the simplex method, akin to the Klee–Minty cube [37].

## 1.1 Contributions

**A ‘near-optimal’ interior point method** The papers [1, 2] implicitly rely on the following lower bound: the trajectory of an IPM performing  $T$  iterations in the wide neighborhood of the central path defines a piecewise linear curve. Hence, the minimum number of pieces of *any* piecewise linear curve in the same neighborhood provides a lower bound on the number of iterations. We introduce a new interior point method based on *subspace layered least squared steps* (see **SLLS-IPM**) and show that the number of iterations of our IPM can be upper bounded in such terms.

The algorithm navigates through the  $\ell_2$ -neighborhood of the central path:

$$\mathcal{N}^2(\beta) := \left\{ z = (x, s) \in \mathcal{P}_{++} \times \mathcal{D}_{++} : \left\| \frac{xs}{\bar{\mu}(z)} - \mathbf{1} \right\| \leq \beta \right\}, \quad (0 < \beta < 1/4), \quad (2)$$

where  $xs \in \mathbb{R}^n$  denotes the coordinate-wise product and  $\mathbf{1} \in \mathbb{R}^n$  is the  $n$ -dimensional all-ones vector.

We also define the *wide neighborhood* as follows:

$$\mathcal{N}^{-\infty}(\theta) := \{ z = (x, s) \in \mathcal{P}_{++} \times \mathcal{D}_{++} : xs \geq (1 - \theta)\bar{\mu}(z)\mathbf{1} \}, \quad (0 < \theta < 1). \quad (3)$$

We let  $\overline{\mathcal{N}}^2(\beta) := \text{cl}(\mathcal{N}^2(\beta))$  and  $\overline{\mathcal{N}}^{-\infty}(\beta) := \text{cl}(\mathcal{N}^{-\infty}(\beta))$  denote the closures of these neighborhoods. These also include points  $z = (x, s) \in \mathcal{P}_+ \times \mathcal{D}_+$  with  $\overline{\mu}(z) = 0$ , i.e., optimal solutions. Our algorithm will terminate with an optimal solution in  $\overline{\mathcal{N}}^2(\beta)$ . We show the following iteration bound.

**Theorem 1.1.** *Let  $\beta \in (0, 1/6]$ ,  $\theta \in (0, 1)$  and  $\mu_0 > \mu_1 \geq 0$ . Let  $\Gamma : (\mu_1, \mu_0) \rightarrow \overline{\mathcal{N}}^{-\infty}(\theta)$  be any piecewise linear curve satisfying  $\overline{\mu}(\Gamma(\mu)) = \mu$ ,  $\forall \mu \in (\mu_1, \mu_0)$  with  $T$  linear segments.*

*Starting from any point  $z^0 \in \mathcal{N}^2(\beta)$  such that  $\overline{\mu}(z^0) \leq \mu_0$ , the algorithm SLLS-IPM finds a solution  $z^1 \in \overline{\mathcal{N}}^2(\beta)$  with  $\overline{\mu}(z^1) \leq \mu_1$  within*

$$O\left(\frac{n^{1.5}}{\beta} \log\left(\frac{n}{\beta(1-\theta)}\right) T\right)$$

*iterations.*

At a high level, our strategy is to show that any ‘somewhat straight’ segment of the central path, corresponding to a single straight segment in the wide neighborhood  $\mathcal{N}^{-\infty}(\theta)$ , can be decomposed into at most  $n$  short segments of length  $\text{poly}(n/(1-\theta))$  (as measured by the ratio of the start and end parameter), where consecutive short segments are possibly separated by ‘long and straight’ segments. To traverse the long and straight segments we develop a novel *subspace* LLS step, which generalizes prior LLS steps from coordinate subspaces to general ones. Before describing this in more detail, we present a stronger form of Theorem 1.1, and two applications.

**The max central path and the straight-line complexity** We next formulate a slightly stronger form of Theorem 1.1. The piecewise linear curve  $\Gamma$  in the statement above lives in a  $2n$ -dimensional space; our next statement argues in terms of  $2n$  separate objects in 2-dimensional space.

Recall that  $v^*$  denotes the optimum value of (LP). The *max central path* is defined as the parametric curve  $g \mapsto z^m(g) := (x^m(g), s^m(g)) \in \mathbb{R}_+^{2n}$ , where  $x_i^m(g)$  and  $s_i^m(g)$  are the optimal values of the following parametric LPs, respectively:

$$\begin{array}{ll} \max x_i & \max s_i \\ \mathbf{A}x = b, x \geq \mathbf{0} & \mathbf{A}^\top y + s = c, s \geq \mathbf{0} \\ \langle c, x \rangle \leq v^* + g, & \langle b, y \rangle \geq v^* - g. \end{array} \quad (4)$$

As we show in Section 4, the maps  $x_i^m(g)$  and  $s_i^m(g)$  are piecewise linear concave, and the number of pieces can be related to the complexity of the simplex method with the shadow vertex rule.

The max central path can be seen as a combinatorial proxy to the central path. In Section 4, we show the following relationship. The upper bounds are immediate by noting that the duality gap for  $(x^{\text{cp}}(\mu), s^{\text{cp}}(\mu))$  is  $n\mu$ .

**Lemma 1.2.** *For every  $\mu > 0$  and the central path point  $z^{\text{cp}}(\mu) = (x^{\text{cp}}(\mu), s^{\text{cp}}(\mu))$ ,*

$$\frac{z^m(n\mu)}{2n} \leq z^{\text{cp}}(\mu) \leq z^m(n\mu).$$

For each  $i \in [n]$  and  $\theta \in [0, 1)$ , we define the *primal and dual multiplicative neighborhoods of the max central path* as

$$\begin{aligned} \mathcal{N}_i^{\text{mp}}(\theta) &:= \{(g, z) \in \mathbb{R}_+^2 : (1-\theta)x_i^m(g) \leq z \leq x_i^m(g)\}, \\ \mathcal{N}_i^{\text{md}}(\theta) &:= \{(g, z) \in \mathbb{R}_+^2 : (1-\theta)s_i^m(g) \leq z \leq s_i^m(g)\}. \end{aligned}$$

**Definition 1.3** (Straight-line complexity). For  $i \in [n]$ ,  $\theta \in [0, 1)$ , and  $0 \leq \underline{g} \leq \overline{g}$ , we define the *primal straight-line complexity w.r.t. coordinate  $i$*  as the minimum number of linear segments of any piecewise linear curve traversing the neighborhood  $\mathcal{N}_i^{\text{mp}}(\theta)$  between parameter values  $\overline{g}$  and  $\underline{g}$ , that is,

$$\begin{aligned} \text{SLC}_{i,\theta}^{\text{p}}(\underline{g}, \overline{g}) &:= \min \{p \geq 1 : \exists (g_k, z_k)_{k \in [p+1]} \in \mathbb{R}_+^2, \overline{g} = g_1, \underline{g} = g_{p+1}, \\ &\quad \forall k \in [p], [(g_k, z_k), (g_{k+1}, z_{k+1})] \subset \mathcal{N}_i^{\text{mp}}(\theta)\}. \end{aligned} \quad (5)$$

The *dual straight-line complexity w.r.t. coordinate  $i$*  is defined analogously for the dual neighborhood  $\mathcal{N}_i^{\text{md}}(\theta)$ .

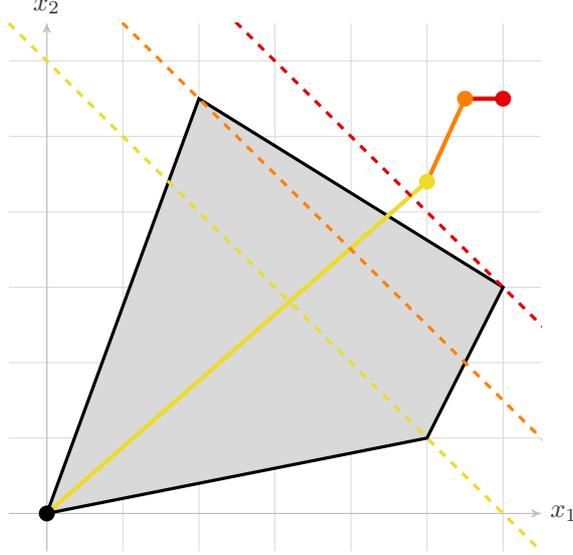


Figure 1: The projection of the max central path on the primal coordinates  $(x_1, x_2)$  for the cost function  $x_1 + x_2$ . Dashed lines correspond to level sets at breakpoints. Note that the max central path does not lie in the feasible region  $\mathcal{P}$ .

We note that Lemma 4.4 shows that one would get an equivalent definition by allowing only breakpoints of the form  $(g_k, z_k) = (g_k, x_i^m(g_k))$ , and requiring  $g_1 > g_2 > \dots > g_{p+1}$ ; this is simple consequence of the concavity of  $g \mapsto x_i^m(g)$  shown in Lemma 1.5.

The stronger form of Theorem 1.1 is as follows:

**Theorem 1.4.** *Let  $\beta \in (0, 1/6]$ ,  $\theta \in [0, 1)$  and  $\mu_0 > \mu_1 \geq 0$ . Given a starting point  $z^0 \in \mathcal{N}^2(\beta)$  such that  $\bar{\mu}(z^0) \leq \mu_0$ , the algorithm SLLS-IPM finds a solution  $z^1 \in \bar{\mathcal{N}}^2(\beta)$  with  $\bar{\mu}(z^1) \leq \mu_1$  in*

$$O\left(\frac{\sqrt{n}}{\beta} \log\left(\frac{n}{\beta(1-\theta)}\right) \min\left\{\sum_{i=1}^n \text{SLC}_{\theta,i}^p(n\mu_1, n\mu_0), \sum_{i=1}^n \text{SLC}_{\theta,i}^d(n\mu_1, n\mu_0)\right\}\right)$$

many iterations. If  $\mu_1 = 0$ , letting  $z^1 = (x^1, s^1) \in \mathcal{P} \times \mathcal{D}$ , the supports  $B := \text{supp}(x^1)$  and  $N := \text{supp}(s^1)$  form a partition of  $[n]$  and the algorithm additionally outputs  $(v^1, w^1) \in \mathbb{R}^{2n}$  satisfying

1.  $v^1 \in \text{im}(A^\top)$ ,  $v_B^1 > 0$  and  $Aw^1 = 0$ ,  $w_N^1 > 0$ .
2.  $\|(x_B^1 v_B^1, s_N^1 w_N^1) - \mathbf{1}_n\| \leq \beta$ .

The algorithm can be implemented in the real RAM model, moreover, each iteration runs in strongly polynomial time in the Turing model.

Some remarks are in order. The computational models and the meaning of strong polynomiality in this context are explained in Section 2.2. The condition for  $\mu_1 = 0$  means that the final output is near the analytic centers of the primal and dual optimal faces, along with a certificate of this fact. This is discussed in Section 2.5. With respect to the iteration bound, the minimum of the primal and dual straight line complexities is just to make the statement symmetric; however, it can be shown that the two terms in the minimum are within a constant factor of each other (see Section 4.2).

From the above statement, Theorem 1.1 follows directly; the proof is in Section 7.3 but we already give the intuition: according to Lemma 1.2, if  $\Gamma : (\mu_1, \mu_0) \rightarrow \bar{\mathcal{N}}^{-\infty}(\theta)$ ,  $\theta \in (0, 1)$  is a piecewise linear curve satisfying  $\bar{\mu}(\Gamma(\mu)) = \mu$ ,  $\forall \mu \in (\mu_1, \mu_0)$  with  $T$  linear segments, then for each  $i \in [n]$ , the projection of  $\Gamma$  to  $x_i$  gives a piecewise linear curve in  $\mathcal{N}_i^{\text{mp}}(1 - (1 - \theta)/(2n))$  for the interval  $[n\mu_1, n\mu_0]$ , and analogously for the  $s_i$ 's.

**An exponential upper bound on the number of iterations** The number of piecewise linear segments of the curves  $x_i^m(g)$  and  $s_i^m(g)$  yield trivial upper bounds on the straight-line complexities in Theorem 1.4.

These can be naturally interpreted in the context of the *shadow vertex simplex rule*. Originally dubbed ‘parametric simplex’ by Gass and Saaty [25], this is one of the most extensively analyzed simplex rules from a theoretical perspective. The shadow vertex rule was used in Borgwardt’s average case analysis [8] and in Spielman and Teng’s smoothed analysis [58]. The interested reader may refer to the recent survey for a detailed exposition by Dadush and Huiberts [17].

Given a pointed polyhedron  $\mathcal{P} \subseteq \mathbb{R}^n$  and two objectives  $c^{(1)}, c^{(2)} \in \mathbb{R}^n$ , the shadow vertex rule consists in iterating over the vertices of  $\mathcal{P}$  successively maximizing the objectives  $(1 - \lambda)c^{(1)} + \lambda c^{(2)}$  as  $\lambda$  goes from 0 to 1. Under non-degeneracy assumptions, the vertices of the path correspond to those vertices of the two-dimensional projection  $\{(\langle c^{(1)}, x \rangle, \langle c^{(2)}, x \rangle) : x \in \mathcal{P}\}$  that maximize some open interval of objectives  $(1 - \lambda)e^1 + \lambda e^2$ ,  $\lambda \in [0, 1]$  (where  $e^1$  and  $e^2$  are the standard basis for  $\mathbb{R}^2$ ). We denote by  $S_{\mathcal{P}}(c^{(1)}, c^{(2)})$  the number of vertices of the projection of the simplex path in this two-dimensional projection; this corresponds to the number of non-degenerate pivots.

Recall that  $(x^*, s^*)$  is the optimal solution of (LP) at the central path limit point. In Section 4, we show that

**Lemma 1.5.** *The following hold:*

(i)  $\forall i \in [n]$ ,  $g \mapsto x_i^m(g)$  is a piecewise linear concave non-decreasing function with  $S_{\mathcal{P}}(-s^*, e^i)$  pieces. That is,  $\text{SLC}_{0,i}^p(0, \infty) = S_{\mathcal{P}}(-s^*, e^i)$ .

(ii)  $\forall i \in [n]$ ,  $g \mapsto s_i^m(g)$  is a piecewise linear concave non-decreasing function with  $S_{\mathcal{D}}(-x^*, e^i)$  pieces. That is,  $\text{SLC}_{0,i}^d(0, \infty) = S_{\mathcal{D}}(-x^*, e^i)$ .

As a consequence, we obtain the following bound:

**Theorem 1.6.** *Given a starting point  $z^0 \in \mathcal{N}^2(1/6)$ , algorithm SLLS-IPM finds an optimal solution of (LP) in a number of iterations bounded by*

$$O\left(\sqrt{n} \log(n) \min\left\{\sum_{i=1}^n S_{\mathcal{P}}(-s^*, e^i), \sum_{i=1}^n S_{\mathcal{D}}(-x^*, e^i)\right\}\right) \leq O\left(\binom{n}{m} n^{1.5} \log(n)\right).$$

Theorem 1.6 thus complements the results of [1, 2] by giving a singly exponential upper bound. We note that the max central path also plays an important if implicit role in the papers [1, 2, 3], as it can be directly related to the tropical central path by the log-limit, see discussion in Section 1.3.

Theorem 1.6 assumes that a feasible starting point  $z^0 \in \mathcal{N}^2(\beta)$  is given. This assumption can be removed e.g. by using the standard homogeneous self-dual embedding [67, Section 5.3.1]. Then, the bounds in the theorem will refer to the shadow vertex paths and the number of vertices in the self-dual program.

**Matching the complexity of any path-following method** The second implication of Theorems 1.1 and 1.4 shows that the number of iterations of SLLS-IPM lower bounds, up to a factor  $n^{1.5} \log n$ , the running time of essentially *any* interior point method.

Let  $f$  be a self-concordant barrier over the polyhedron  $\mathcal{P}$  with complexity value  $\vartheta_f$ ; we introduce these concepts in Section 10. This defines a corresponding central path, with  $x^{\text{cp}(f)}(g)$  denoting the unique point of the central path with gap  $g > 0$ . Recall that  $v^*$  denotes the optimum value of (LP). We define the *wide neighborhood w.r.t. the barrier  $f$  for a parameter  $\theta \in (0, 1)$*  as

$$\mathcal{N}^{\text{cp}(f)}(\theta) := \{x \in \mathcal{P} : x \geq (1 - \theta)x^{\text{cp}(f)}(g) \text{ where } \langle c, x \rangle = v^* + g\}. \quad (6)$$

Note that the  $\ell_2$ -neighborhood  $\mathcal{N}^2(\theta)$  and the wide neighborhood  $\mathcal{N}^{-\infty}(\theta)$  are defined in the primal-dual space  $\mathbb{R}^{2n}$ , whereas  $\mathcal{N}^{\text{cp}(f)}(\theta) \subseteq \mathbb{R}^n$  is in the primal space. Projecting the neighborhoods  $\mathcal{N}^{-\infty}(\theta)$  to primal variables can be shown to be equivalent to the neighborhoods of the form  $\mathcal{N}^{\text{cp}(f)}(\theta)$  for the logarithmic barrier  $f = -\sum_{i \in [n]} \log(x_i)$ . We refer to Section 10 for a discussion on the generality of the neighborhoods  $\mathcal{N}^{\text{cp}(f)}(\theta)$ .

**Theorem 1.7.** *Let  $f$  be a self-concordant barrier over the polyhedron  $\mathcal{P}$  with complexity value  $\vartheta_f$ . Let  $\beta \in (0, 1/6]$ ,  $\theta \in (0, 1)$ , and  $g_0 > g_1 \geq 0$ . Assume an interior point method proceeds through  $T$  straight-line steps from  $x^0$  to  $x^1$  inside the wide neighborhood  $\mathcal{N}^{\text{cp}(f)}(\theta)$  with  $g_0 = \langle c, x^0 \rangle - v^*$  and  $g_1 = \langle c, x^1 \rangle - v^*$ .*

*Given any  $z^0 \in \mathcal{N}^2(\beta)$  with  $n\bar{\mu}(z^0) \leq g_0$ , let  $T'$  be the number of iterations of SLLS-IPM to reach the first iterate  $z^1 \in \overline{\mathcal{N}^2}(\beta)$  with  $n\bar{\mu}(z^1) \leq g_1$ . Then,*

$$T' = O\left(Tn^{1.5} \log\left(\frac{n\vartheta_f}{1 - \theta}\right)\right).$$

Thus—up to a polynomial factor—our algorithm matches the running time of any IPM for any self-concordant barrier function staying in the extremely wide neighborhood  $\mathcal{N}^{\text{cp}(f)}(1 - 1/2^{\text{poly}(n)})$ . In particular, we obtain polynomial-time bounds in case the bit-complexity or a condition number such as  $\bar{\chi}_{\mathbf{A}}^*$  is bounded.

**Comparison to the Trust Region IPM** SLLS-IPM also has an interesting relation to the Trust Region IPM algorithm by Lan, Monteiro, and Tsuchiya [38]. The trust region steps are obtained as optimal solutions to primal and dual quadratic programs (see (8) below). These programs in essence capture the longest possible step achievable at the current point (up to a certain factor). However, it is currently not known how to solve these programs to sufficient accuracy in strongly polynomial time (though this can be done in weakly polynomial time). Lan, Monteiro, and Tsuchiya show in [38] that the number of iterations of the trust region algorithm can be bounded as  $O(n^{3.5} \log(\bar{\chi}_{\mathbf{A}}^* + n))$ , by adapting the analysis of the LLS methods [47, 66].

The step directions used by our algorithm are feasible solutions to (8) for a suitable parameter. This implies that the steps of the Trust Region algorithm are always at least as long as the steps in our algorithm; as a consequence, the iteration bounds of our algorithm are also applicable to the Trust Region algorithm. Whereas any individual step of our algorithm could be arbitrarily worse than the one using the trust region step, Theorem 1.7 implies that overall we may only take  $O(n^{1.5} \log n)$  more iterations. We emphasize that [38] only provides a  $\bar{\chi}_{\mathbf{A}}^*$  dependent iteration bound, and we do not see a way to obtain even an  $f(n)$  bound for their algorithm without using the majority of the analysis of SLLS-IPM.

## 1.2 Techniques

We now explain the key ideas of the algorithm SLLS-IPM and the analysis.

### 1.2.1 Polarization of the Central Path

The first key idea behind the proof of Theorems 1.1 and 1.4 is the following: every linear segment in the wide neighborhood gives rise to a *polarized segment* of the central path. A segment of the central path  $\text{CP}[\mu_1, \mu_0] := \{z^{\text{cp}}(\mu) : \mu \in [\mu_1, \mu_0]\}$ ,  $0 \leq \mu_1 < \mu_0$ , is polarized, if it admits a partition  $B \cup N = [n]$  such that the primal variables in  $B$  are barely changing while those in  $N$  are scaling down linearly with the parameter  $\mu$  (vice versa for the dual variables). More precisely,  $\forall \mu \in [\mu_1, \mu_0]$ , we require

$$\begin{aligned} \gamma x_i^{\text{cp}}(\mu_0) &\leq x_i^{\text{cp}}(\mu) \leq n x_i^{\text{cp}}(\mu_0), \quad \forall i \in B, \\ \frac{\mu}{n\mu_0} x_i^{\text{cp}}(\mu_0) &\leq x_i^{\text{cp}}(\mu) \leq \frac{\mu}{\gamma\mu_0} x_i^{\text{cp}}(\mu_0), \quad \forall i \in N, \end{aligned} \quad (7)$$

where  $\gamma \in (0, 1]$  is a *polarization parameter* (see Definition 3.1 and Corollary 3.4). By definition of the central path, the same relation holds for dual variables  $s^{\text{cp}}(\mu)$ ,  $\mu \in [\mu_1, \mu_0]$ , with the roles of  $N$  and  $B$  swapped. We note that the upper bounds on  $x_i^{\text{cp}}(\mu)$  for  $i \in B$  and the lower bounds on  $x_i^{\text{cp}}(\mu)$  for  $i \in N$  hold by the near-monotonicity property of the central path (see Lemma 2.30); the important parts of the definition are the other two bounds.

For simplicity of notation, let us restrict to line segments between two points on the central path. To relate polarization to the wide neighborhood, we show that if the line segment  $[z^{\text{cp}}(\mu_1), z^{\text{cp}}(\mu_0)]$  between central path points is contained in the wide neighborhood  $\mathcal{N}^{-\infty}(\theta)$ , then the corresponding segment of the central path is polarized with  $\gamma = \frac{(1-\theta)^2}{16n^3}$  with respect to some partition  $B \cup N = [n]$  (see Lemma 3.6 for the general statement).

One should read this last statement as saying that a segment of the central path is ‘approximately linear’ if and only if it is polarized (in fact, one can show that segment is 1-polarized if and only if it is linear). The link between polarization and linearity is surprisingly elementary; it follows from the analysis of the inequalities of the wide neighborhood (3):

$$\begin{aligned} ((1-\alpha)x_i^{\text{cp}}(\mu_0) + \alpha x_i^{\text{cp}}(\mu_1)) &((1-\alpha)s_i^{\text{cp}}(\mu_0) + \alpha s_i^{\text{cp}}(\mu_1)) \\ &\geq (1-\theta)((1-\alpha)\mu_0 + \alpha\mu_1), \quad \forall \alpha \in [0, 1], i \in [n], \end{aligned}$$

where we recall that  $z^{\text{cp}}(\mu_0) = (x^{\text{cp}}(\mu_0), s^{\text{cp}}(\mu_0))$ ,  $z^{\text{cp}}(\mu_1) = (x^{\text{cp}}(\mu_1), s^{\text{cp}}(\mu_1))$ . For example, if  $\theta = 0$ , it is not hard to check that for each  $i \in [n]$ , one must have either  $x_i^{\text{cp}}(\mu_0) = x_i^{\text{cp}}(\mu_1)$  and  $s_i^{\text{cp}}(\mu_1) = \frac{\mu_1}{\mu_0} s_i^{\text{cp}}(\mu_0)$  (i.e.,  $i \in B$ ) or  $x_i^{\text{cp}}(\mu_0) = \frac{\mu_1}{\mu_0} x_i^{\text{cp}}(\mu_1)$  and  $s_i^{\text{cp}}(\mu_1) = s_i^{\text{cp}}(\mu_0)$  (i.e.,  $i \in N$ ).

Given the above, the main task in proving Theorem 1.1, namely traversing linear segments in the wide-neighborhood, can be reduced to traversing  $\gamma$ -polarized segments of the central path. The main guarantee of our algorithm **SLLS-IPM** is in fact that it can traverse any  $\gamma$ -polarized segment of the path in  $O(n^{1.5} \log(n/\gamma))$  iterations (see Theorem 7.4)

To derive the stronger bound in Theorem 1.4, a key step is to use the max central path to guide the decomposition of the central path into polarized segments. We will show that one can decompose the central path into polarized segments where the polarization partitions do not change “too quickly” from segment to segment. Specifically, the sum of partition changes will be bounded by the sum of straight line complexities of either the primal or dual. The formal statement is given below:

**Theorem 1.8.** *Let  $\theta \in [0, 1)$  and  $\mu_0 > \mu_1 \geq 0$ . The segment  $\text{CP}[\mu_1, \mu_0]$  can be decomposed into a sequence of  $\frac{1-\theta}{4n}$ -polarized segments with partitions  $(B^{(k)}, N^{(k)})$ ,  $k \in [T]$ , such that*

$$\max \left\{ \sum_{k=1}^T |N^{(k)} \Delta N^{(k-1)}|, T \right\} \leq 2 \min \left\{ \sum_{i=1}^n \text{SLC}_{\theta, i}^{\text{p}}(n\mu_1, n\mu_0), \sum_{i=1}^n \text{SLC}_{\theta, i}^{\text{d}}(n\mu_1, n\mu_0) \right\}$$

where  $N^{(0)} := \emptyset$ .

For the sake of symmetry, we state the upper bound above terms of the minimum of either primal or dual straight line complexities. However, it can be shown that both are equivalent up to a constant factor (see Lemma 4.5).

We note that polarization plays an important if implicit role in prior layered least squares analyses [18, 48, 66]. In particular, the ‘long and straight’ segments in these works are all polarized. What was unclear in these works, however, is whether polarization *by itself* was sufficient to make a segment easy to traverse. Indeed, these works all crucially rely upon numerical condition numbers of the instance which can be effectively unbounded in the present context. Beyond the LLS context, we are further unaware of central path analyses exploiting the tight connection between approximate linearity and polarization, and we hope this will encourage future study.

As is clear from the definition, polarization provides us extremely useful ‘long-range’ control over the evolution of variables on a segment. Note that  $\gamma$ -polarization is mostly interesting when the segment itself is *long*, namely, when  $\mu_0/\mu_1 \gg 1/\gamma$ . We now explain how to leverage this control to traverse any  $\gamma$ -polarized segment using *subspace LLS steps*.

### 1.2.2 Traversing a Polarized Segment

Let  $\text{CP}[\mu_1, \mu_0]$ ,  $0 \leq \mu_1 < \mu_0$ , be a  $\gamma$ -polarized segment with partition  $B \cup N = [n]$ .

For simplicity of presentation, let us assume that given any iterate  $(x, s)$  in the narrow neighborhood  $\mathcal{N}^2(1/6)$  used in our algorithm, we can jump to the exact central path point  $z^{\text{cp}}(\mu) \in \text{CP} = \mathcal{N}^2(0)$  with  $\mu = \bar{\mu}(x, s)$  for free. Let us further assume that the algorithm knows the partition  $B, N$  (we discuss how to effectively compute it at the end) and that we are given the starting point  $z^{(0)} := z^{\text{cp}}(\mu_0)$ .

Our abstract algorithm will thus compute iterates  $z^{(0)}, z^{(1)}, \dots$  on the central path  $\text{CP}$  with  $\bar{\mu}(z^{(0)}) > \bar{\mu}(z^{(1)}) > \dots$ . To move from  $z^{(t)}$  to  $z^{(t+1)}$ , we first compute a *movement direction*

$$\Delta z^{(t)} = (\Delta x^{(t)}, \Delta s^{(t)}) \in \ker(\mathbf{A}) \times \text{im}(\mathbf{A}^\top) =: W \times W^\perp,$$

together with a step-length  $\alpha^{(t)} \in [0, 1]$ , chosen such that  $z^{(t)} + \alpha \Delta z^{(t)} \in \bar{\mathcal{N}}^2(1/6)$ ,  $0 \leq \alpha \leq \alpha^{(t)}$ . Lastly, assume we can jump for free to  $z^{(t+1)} \in \text{CP}$  satisfying  $\bar{\mu}(z^{(t+1)}) = \bar{\mu}(z^{(t)} + \alpha^{(t)} \Delta z^{(t)})$ .

Given this setup, our goal is to compute movement directions and step-lengths, such that after  $k = O(n^{1.5} \log(n/\gamma))$  iterations, we have  $\bar{\mu}(z^{(k)}) \leq \mu_1$ , i.e., that we have traversed the segment. We would like to emphasize that our algorithm will in fact compute the movement direction  $\Delta z^{(t)}$  using only *local information* at  $z^{(t)}$ , without any explicit knowledge of the polarized segment.

A natural movement direction is *affine scaling* used in predictor-corrector methods, see Section 2.6. This direction guarantees multiplicative  $1 - \Omega(1/\sqrt{n})$  decrease in normalized gap per step. Hence, if  $\mu_0/\mu_1 \leq \text{poly}(n, 1/\gamma)$ , then simply using  $\sqrt{n} \log(\mu_0/\mu_1)$  affine scaling iterations is sufficient for our purposes.

Thus, we may assume that  $\mu_0/\mu_1 \gg \text{poly}(n, 1/\gamma)$ . In this case, we will show that the affine scaling direction  $(\Delta x^{\text{a}}, \Delta s^{\text{a}})$  at the current iterate  $(x^{(t)}, s^{(t)})$  reveals the correct partition  $B \cup N = [n]$  whenever a sufficiently long step exists. This is because the standard affine scaling step itself exhibits a polarized behaviour: we can simply select  $B$  as the set of coordinates  $i$  where  $|\Delta x_i^{\text{a}}/x_i^{(t)}| < |\Delta s_i^{\text{a}}/s_i^{(t)}|$ , i.e., the relative primal movement is smaller than the relative dual movement (see Definition 5.2).

**Trust Region Programs and Subspace LLS** The trust region programs introduced by Lan, Monteiro, Tsuchiya [38] provide a good starting point for defining our movement direction  $\Delta z^{(t)} = (\Delta x^{(t)}, \Delta s^{(t)}) \in W \times W^\perp$  from an iterate  $z^{(t)} = (x^{(t)}, s^{(t)}) \in \text{CP}[\mu_1, \mu_0]$  and a given a partition  $[n] = B \cup N$ :

$$\begin{aligned} & \min_{\Delta x \in W} \left\{ \left\| (x_N^{(t)} + \Delta x_N)/x_N^{(t)} \right\| : \left\| \Delta x_B/x_B^{(t)} \right\| \leq \nu \right\}, \\ & \min_{\Delta s \in W^\perp} \left\{ \left\| (s_B^{(t)} + \Delta s_B)/s_B^{(t)} \right\| : \left\| \Delta s_N/s_N^{(t)} \right\| \leq \nu \right\}, \end{aligned} \quad (8)$$

where  $\nu = O(\beta)$  is sufficient for the induced step to stay inside the  $\mathcal{N}^2(\beta)$  neighborhood. We use the notation  $\Delta x/x^{(t)} := (\Delta x_1/x_1^{(t)}, \dots, \Delta x_n/x_n^{(t)})$  and similarly for  $\Delta s/s^{(t)}$ . The norms  $\|x/x^{(t)}\|$  and  $\|s/s^{(t)}\|$  are the so-called primal and dual *local norms* at  $x^{(t)}$  and  $s^{(t)}$ .<sup>1</sup> By definition, the optimal primal trust region direction  $\Delta x^*$  achieves a maximal multiplicative decrease on the coordinates in  $N$  while ‘barely moving’ the coordinates in  $B$  as measured in the local norm. The optimal dual direction  $\Delta s^*$  achieves the same on the dual side with the role of  $N$  and  $B$  swapped.

Note that these directions mesh well with polarization of the segment  $\text{CP}[\mu_1, \mu_0]$ . In particular, they reflect the idea that the coordinates of  $x^{\text{CP}}(\mu)$  in  $N$  should be linearly scaling down while those in  $B$  are staying mostly fixed, and vice versa for  $s^{\text{CP}}(\mu)$ . As shown in [38] (see also Proposition 5.1), moving in any direction  $\Delta z^{(t)} = (\Delta x^{(t)}, \Delta s^{(t)})$  corresponding to feasible solutions to (8), the normalized gap can be reduced as

$$\bar{\mu}(z^{(t+1)})/\bar{\mu}(z^{(t)}) = O\left(\left\| (x_N^{(t)} + \Delta x_N^{(t)})/x_N^{(t)} \right\| + \left\| (s_B^{(t)} + \Delta s_B^{(t)})/s_B^{(t)} \right\|\right). \quad (9)$$

That is, we can achieve a drop that corresponds to the sum of primal and dual objective values.

In many ways, the trust region direction can be seen as the ‘optimal’ movement direction. However, [38] solves the quadratic convex programs in (8) in weakly polynomial time with dependence on the vectors  $b$  and  $c$  in (LP). It is not known whether a strongly polynomial algorithm (with dependence only on  $n$ ) exists. Further, the analysis in [38] relies on combinatorial progress measures adapted from the LLS analyses, which are too coarse to directly measure progress on a polarized segment (in these analyses, combinatorial progress is only guaranteed every  $\Omega(\sqrt{n} \log(n + \bar{\chi}_A^*))$  iterations).

Instead of optimally solving (8), we introduce what we call *subspace LLS steps* that yield ‘good enough’ approximate solutions to make rapid progress along a polarized segment. We restrict the set of primal and dual directions  $\Delta x \in V^{(t)}$  and  $\Delta s \in U^{(t)}$  for subspaces  $V^{(t)} \subseteq W$  and  $U^{(t)} \subseteq W^\perp$  satisfying:

$$\begin{aligned} \forall \Delta x \in V^{(t)}, \quad & \left\| \Delta x_B/x_B^{(t)} \right\| \leq \tau \left\| \Delta x_N/x_N^{(t)} \right\|, \\ \forall \Delta s \in U^{(t)}, \quad & \left\| \Delta s_N/s_N^{(t)} \right\| \leq \tau \left\| \Delta s_B/s_B^{(t)} \right\|, \end{aligned}$$

where we set  $\tau = O(\nu/\sqrt{n})$ . We call any such subspaces  $V^{(t)}$  and  $U^{(t)}$  *cheap lift subspaces* with lifting cost  $\tau$ . Note that for any partial vector  $\Delta x_N \in \pi_N(V^{(t)}) \subseteq \pi_N(W)$ , the subspace  $V^{(t)}$  provides a way to ‘lift’  $\Delta x_N$  to a full vector  $(\Delta x_B, \Delta x_N) \in W$  at a ‘cost’ of  $\|\Delta x_B/x_B\| \leq \tau \|\Delta x_N/x_N\|$ . Similarly,  $U^{(t)}$  provides a way to lift partial vectors in  $\pi_B(U^{(t)}) \subseteq \pi_B(W^\perp)$  cheaply into  $W^\perp$ . We formally define the associated ‘lifting operator’ in Section 6.2. With a cheap lift subspace at hand, it is not hard to show that any optimal solution to the subspace constrained trust-region program automatically satisfies the norm constraints in program (8). Therefore, by restricting ourselves to search directions  $\Delta x \in V^{(t)}$  and  $\Delta s \in U^{(t)}$  as above, we can solve simple unconstrained minimum-norm point problems in the local norms while still guaranteeing that the computed search directions are feasible for the trust-region program.

There is a lot of flexibility to choose these subspaces. A canonical choice is a lifting of the subspace spanned by the singular vectors of the ‘lifting operator’ (see Definition 6.4) whose corresponding singular values are at most  $\tau$ . While singular values and their corresponding singular subspaces cannot be computed exactly, they can be closely approximated in strongly polynomial time (see Section 9). Furthermore, as we show in Section 7, even very coarse approximations of the singular values and subspaces will suffice.

**Analyzing Subspace LLS** At each iteration, our algorithm computes the affine scaling steps and the subspace LLS steps as above, and uses the one that enables more progress along the central path. For simplicity of exposition, we use the canonical cheap lift subspaces  $U^{(t)}, V^{(t)}$  to compute the subspace LLS direction  $\Delta z^{(t)}$  as above.

<sup>1</sup>Recall the assumption that  $(x^{(t)}, s^{(t)})$  is on the central path.

Let us now explain the key idea in showing that subspace LLS steps can reach the end of the current  $\gamma$ -polarized segment  $\text{CP}[\mu_1, \mu_0]$  in  $O(n^{1.5} \log(n/\gamma))$  iterations. Let  $k = \Omega(\sqrt{n} \log(n/\gamma))$ . Given any iterate  $z^{(t)} \in \text{CP}[\mu_1, \mu_0]$ , if  $\bar{\mu}(z^{(t+k)}) > \mu_1$ —i.e., we have not reached the end of the segment—then we show that both  $\dim(U^{(t+k)}) > \dim(U^{(t)})$  and  $\dim(V^{(t+k)}) > \dim(V^{(t)})$ . The overall bound follows since this can occur at most  $n$  times.

To get this result, we analyze the evolution of what we call the ‘ideal direction’ at  $z^{(t)}$ , which we define to be  $z^{\text{cp}}(\mu_1) - z^{(t)}$ , i.e., the difference between the current iterate and the end of the segment. A crucial observation is that if  $z^{\text{cp}}(\mu_1) - z^{(t)}$  were a feasible solution to (8), then following this direction would get to within a  $\text{poly}(n/\gamma)$  factor for the end of the segment in one step (though we do not know how to compute it). Furthermore, the ideal direction is never far from being feasible, in particular, it is feasible if the bound of  $\nu$  is replaced by  $O(n)$ . Subspace LLS steps will allow us to leverage the ideal direction via the following dichotomy. Given an iterate  $z^{(t)}$ , either the ideal direction  $z^{\text{cp}}(\mu_1) - z^{(t)}$  is mostly “aligned” with the LLS subspaces  $U^{(t)} \times V^{(t)}$ , in which case the LLS step brings us close to the end of the segment, or if not, it brings us close to the time where the cheap lift subspaces increase in dimension. In the latter case, we crucially use the polarization property to analyze the evolution of the singular values of lifting operators. In both cases, the notion of close will mean that decreasing the gap by an additional  $\text{poly}(n/\gamma)$  will be sufficient to enter a new part of the segment. In particular,  $O(\sqrt{n} \log(n/\gamma))$  additional iterations will suffice to make the desired progress.

This concludes our overview of the proof of Theorem 1.1; the detailed argument is presented in Section 7.3. In Section 8, we present an amortized analysis that yields the stronger bound in Theorem 1.4 in terms of the straight-line complexities. In particular, given a piecewise linear curve in the central path neighborhood where the subsequent pieces are polarized with partitions  $(B^{(1)}, N^{(1)})$ ,  $(B^{(2)}, N^{(2)})$ ,  $\dots$ ,  $(B^{(T)}, N^{(T)})$ , we show that the number of iterations can be bounded by

$$O\left(\sqrt{n} \log(n/\gamma) \sum_{t=1}^T (|N^{(t)} \Delta N^{(t-1)}| + 1)\right), \quad (10)$$

with the convention that  $N^{(0)} = \emptyset$  (see Theorem 8.1). This can be better than the previous bound  $O(n^{1.5} \log(n/\gamma)T)$  if the number of indices changing between polarizing partitions in subsequent polarized segments is small compared to  $n$ . The proof of Theorem 1.4 now follows by combining (10) together with Theorem 1.8.

### 1.3 Related Work

Interior points methods have been a tremendously active and fruitful research area since the seminal works of Karmarkar [34] and Renegar [51] in the 80’s. Remarkable advances have been made both in speed as well as applicability of IPMs. We first briefly review works that—unlike the present paper—aim for  $\varepsilon$ -approximate solutions. A key ingredient has been the use of different, self-concordant barrier functions. Like the logarithmic barrier, every such function gives rise to a notion of central path. In the general setting, the iteration complexity to get an  $\varepsilon$ -approximation of the optimal value is bounded by  $O(\vartheta_f^{1/2} \log \varepsilon^{-1})$  for the complexity parameter  $\vartheta_f$ . General bounds on self-concordant barriers were given by Nesterov and Nemirovski [49], improved recently by Lee and Yue [41] and Chewi [15]. Specific barrier functions include Vaidya’s volumetric barrier [60], the entropic barrier by Bubeck and Eldan [9], and the weighted log-barrier by Lee and Sidford [39, 40].

Recent improvements make use of efficient data structures to amortize the cost of the iterative updates, and work with approximate computations, see Cohen, Lee and Song [16], van den Brand [61], and van den Brand, Lee, Sidford, and Song [64]. For special classes of LP such as network flow and matching problems, even faster algorithms have been obtained using, among other techniques, fast Laplacian solvers [57], see e.g. [4, 20, 24, 42, 63, 62], culminating in the very recent almost-linear time minimum-cost flow algorithm [14].

Layered least squares IPMs, initiated by Vavasis and Ye [66] find exact optimal solutions and their running time bound is independent of  $b$  and  $c$ . Improved LLS algorithms were given by Megiddo, Mizuno, and Tsuchiya [44] and Monteiro and Tsuchiya [47, 48]. As discussed previously, scaling invariant algorithms with a  $\bar{\chi}_{\mathbf{A}}^*$  dependence are the Trust Region algorithm by Lan, Monteiro, and Tsuchiya [38], and the LLS algorithm [18] that relies on approximating circuit imbalances.

There is an interesting connection between IPMs and differential geometry. Sonnevend, Stoer, and Zhao [56] introduced a primal-dual curvature concept for the central path, and related the curvature

integral to the iteration complexity of IPMs. Monteiro and Tsuchiya [46] showed that a curvature integral is bounded by  $O(n^{3.5} \log(\bar{\chi}_{\mathbf{A}}^* + n))$ . This has been extended to SDP and symmetric cone programming [33], and it was also studied in the context of information geometry [32].

Relating central paths and simplex paths has been mainly been explored in the context of building LPs with pathological properties. On top of the construction of [3] that we already discussed, Deza, Nematollahi and Terlaky [21] built a Klee–Minty cube with exponentially many redundant inequalities where the central path is distorted into the neighborhood of a simplex path that visits all  $2^n$  vertices.

The max central path studied in this paper is related to the tropical central path in [1, 2, 3]. The latter arises when studying parametric families of LPs where the input  $(\mathbf{A}, b, c)$  depends on a parameter  $t > 1$ . The tropical central path is defined as the log-limit, i.e., the limit as  $t \rightarrow \infty$  of the image under the map  $z \mapsto \log_t z = \frac{\log z}{\log t}$ , of the central path of these LPs. In [1, 2, 3], it was shown that the tropical central path corresponds to the greatest point (entrywise) of the log-limit of the feasible sets of (4). This turns out to be precisely the log-limit of the max central path.

As stated earlier in the introduction, there is no known polynomial time variant of the simplex method which traverses the edges of a base polyhedron. Deviating from this model, Kelner and Spielman [35] gave a weakly-polynomial time LP algorithm which uses the simplex method on random perturbations of a polyhedron. Specifically, their algorithm determines unboundedness for  $\mathcal{P} := \{x \in \mathbb{R}^n : \mathbf{A}x \leq \mathbf{1}\}$ , which is strongly polynomially equivalent to general LP. They apply a shadow vertex simplex method on  $\mathcal{P}$  with randomly perturbed right hand sides as part of a subroutine which either computes a point of large norm – used to make  $\mathcal{P}$  “round” – or computes a suitable certificate of boundedness. While based on the shadow vertex simplex method, their algorithm is inherently weakly polynomial and does not admit a running time that only depends on  $n$  and  $m$ .

## 1.4 Organization of the Paper

In Section 2, we introduce our notation, the basic tools in linear algebra we require (Section 2.3), as well as the important properties of the central path and its neighborhoods (Section 2.4). We also discuss the affine scaling steps used in predictor-correct methods (Section 2.6). Section 3 deals with the polarized segments of the central path and their connection with linear segments in the wide neighborhood. Section 4 studies the max central path and shows how to use it to decompose the central into polarized segments, proving Theorem 1.8. Section 5, proves important properties of trust-region steps, and in particular, how to identify optimal trust-region partitions. Section 6 develops the theory of subspace layered least squares directions and cheap lift subspaces, and also gives algorithms for computing them. From the theory side, the fundamental concept of a lifting operator is introduced in Section 6.2, and the relationship between cheap lift subspaces and approximate singular subspaces of the lifting operator is given in Section 6.3. The algorithm **CHEAP-LIFT-SUBSPACES**, used to compute these subspaces, is presented analyzed in Section 6.4. The algorithm **SLLS-IPM** is introduced in Section 7, where the analysis for polarized segments is presented in Section 7.3, proving Theorem 1.1. Section 8 presents an amortized analysis, leading to the proof of Theorem 1.4. In Section 9, we present a deterministic strongly polynomial algorithm for computing approximate singular value decompositions, which is needed to compute the cheap lift subspaces used by the IPM. Finally, Section 10 discusses interior point methods with self-concordant barrier functions, and proves Theorem 1.7. Omitted proofs are deferred to the Appendix.

## 2 Preliminaries

### 2.1 Notation

We let  $\mathbb{R}_{++}$  denote the set of positive reals,  $\mathbb{R}_+$  the set of nonnegative reals, and  $\mathbb{N} = \{1, 2, \dots\}$  denote the natural numbers. For  $n \in \mathbb{N}$ , we let  $[n] := \{1, 2, \dots, n\}$ .  $B, N \subseteq [n]$  form a partition of  $[n]$  if  $B \cup N = [n]$  and  $B \cap N = \emptyset$ . We say that  $(B, N)$  is a *non-trivial* partition of  $[n]$  if additionally  $B \neq \emptyset$  and  $N \neq \emptyset$ , and is a *trivial* partition otherwise. For  $a \in \mathbb{R}$ ,  $\lceil a \rceil \in \mathbb{Z}$  is the smallest integer greater than or equal to  $a$ , and  $\lfloor a \rfloor \in \mathbb{Z}$  is the largest integer less than or equal to  $a$ . For  $a \in \mathbb{R}_{++}$ , we let  $\log(a)$  denote the natural logarithm of  $a$ . For  $a \in \mathbb{R}$  and  $S \subseteq \mathbb{R}$ , we define  $1[a \in S] = 1$  if  $a \in S$  and 0 otherwise. Throughout, we consider inequalities of vectors coordinate-wise. Let  $e^i \in \mathbb{R}^n$  denote the  $i$ th standard basis vector, and  $\mathbf{1}_n, \mathbf{0}_n \in \mathbb{R}^n$  denote the vector of all ones and all zeros respectively. For  $x \in \mathbb{R}^n$ , we let  $\text{supp}(x) := \{i \in [n] : x_i \neq 0\}$  denote the support of  $x$ . For two points  $x, y \in \mathbb{R}^n$ , we let  $[x, y] := \{\lambda x + (1 - \lambda)y : 0 \leq \lambda \leq 1\}$  denote the line-segment connecting  $x$  and  $y$ . For two sets  $S, T$  we

let  $S\Delta T := (S \setminus T) \cup (T \setminus S)$  be the symmetric difference between  $S$  and  $T$ . If  $S, T \subseteq \mathbb{R}^n$ , we define their Minkowski sum  $S + T := \{s + t : s \in S, t \in T\}$ . For a vector  $t \in \mathbb{R}^n$ , we let  $S + t := S + \{t\}$  for notational simplicity. For a function  $f : S \rightarrow \mathbb{R}$  and  $T \subseteq S$ , we let  $\arg \min_{x \in T} f(x) := \{y \in T : f(y) = \min_{x \in T} f(x)\}$  denote the set of minimizers of  $f$  with respect to  $T$ . By convention,  $\arg \min_{x \in T} f(x) = \emptyset$  if the minimum value of  $f$  inside  $T$  is not attained.

The standard inner product between two vectors is denoted by  $\langle x, y \rangle = x^\top y$ , for  $x, y \in \mathbb{R}^n$ , and the Euclidean norm by  $\|x\| := \sqrt{\sum_{i=1}^n x_i^2}$ . We further let  $\|x\|_\infty := \max_{i \in [n]} |x_i|$  denote the  $\ell_\infty$  norm, and  $\|x\|_1 := \sum_{i=1}^n |x_i|$  denote the  $\ell_1$  norm.

For a vector  $x \in \mathbb{R}^n$ , we let  $\text{diag}(x) \in \mathbb{R}^{n \times n}$  denote the diagonal matrix with  $x$  on the diagonal. For  $x, y \in \mathbb{R}^n$ , we use the notation  $xy \in \mathbb{R}^n$  for the coordinate-wise (Hadamard) product  $xy = \text{diag}(x)y = (x_i y_i)_{i \in [n]}$ . For  $\xi \in \mathbb{R}_{++}^n$  and a linear subspace  $W \subseteq \mathbb{R}^n$ , we use the notation  $\xi W := \{\xi w : w \in W\}$ . For  $p \in \mathbb{Q}$  and  $x \in \mathbb{R}^n$ , we also use the notation  $x^p \in \mathbb{R}^n$  to denote the vector  $(x_i^p)_{i \in [n]}$ , where we will always ensure that the corresponding coordinates are well-defined. Similarly, for  $x \in \mathbb{R}^n$  and  $y \in \mathbb{R}_{++}^n$ , we let  $x/y \in \mathbb{R}^n$  denote the vector  $(x_i/y_i)_{i \in [n]}$ . For vectors  $x^{(1)}, \dots, x^{(k)} \subseteq \mathbb{R}^n$ , we define their linear span as  $\text{span}(x^{(1)}, \dots, x^{(k)}) := \{\sum_{i=1}^k \lambda_i x^{(i)} : \lambda \in \mathbb{R}^k\}$ . For non-empty set  $I \subseteq [n]$ , we define  $\pi_I : \mathbb{R}^n \rightarrow \mathbb{R}^I$  to be the coordinate projection onto  $I$ , that is  $\pi_I(x) := x_I, \forall x \in \mathbb{R}^n$ . By convention, we define  $\mathbb{R}^\emptyset := \{0\}$  to be the trivial subspace, and we let  $\pi_\emptyset(x) := x_\emptyset := 0, \forall x \in \mathbb{R}^n$ . We also define  $\mathbb{R}_I^n := \{x \in \mathbb{R}^n : \text{supp}(x) \subseteq I\}$  to be the set of vectors in  $\mathbb{R}^n$  with support contained in  $I$ .

For a matrix  $\mathbf{M} \in \mathbb{R}^{m \times n}$ , we let  $\mathbf{M}^\top \in \mathbb{R}^{n \times m}$  denote the matrix transpose satisfying  $(\mathbf{M}^\top)_{ji} := \mathbf{M}_{ij}$ , for  $i \in [m], j \in [n]$ . For subsets  $S \subseteq [m], T \subseteq [n]$ , we define  $\mathbf{M}_{S,T}$  to be the matrix with input space  $\mathbb{R}^T$  and output space  $\mathbb{R}^S$  induced by the columns in  $T$  and rows in  $S$  of  $\mathbf{M}$ . We further use the notation  $\mathbf{M}_T := \mathbf{M}_{\bullet, T} := \mathbf{M}_{[m], T}$  to index the corresponding columns of  $\mathbf{M}$ , and  $\mathbf{M}_{S, \bullet} := \mathbf{M}_{S, [n]}$ . By convention, we let  $\mathbf{M}_{S, \emptyset} := [\mathbf{0}_S], \mathbf{M}_{\emptyset, T} := [\mathbf{0}_T]^\top$  and  $\mathbf{M}_{\emptyset, \emptyset} := [0]$  (recall that  $\mathbb{R}^\emptyset = \{0\}$ ).

**Subspace Formulation of Linear Programming** It will be more convenient for our algorithm and analysis to represent (LP) in an equivalent subspace language. Throughout the paper, we let  $W = \ker(\mathbf{A}) \subseteq \mathbb{R}^n$  denote the kernel of  $\mathbf{A}$  and  $W^\perp = \text{im}(\mathbf{A}^\top)$  denote the image of the transpose of  $\mathbf{A}$  (see Section 2.3 for formal definitions). Using this notation, (LP) can be written in the form

$$\begin{aligned} \min \langle c, x \rangle & & \max \langle d, c - s \rangle \\ x \in W + d & & s \in W^\perp + c \\ x \geq \mathbf{0}_n, & & s \geq \mathbf{0}_n, \end{aligned} \tag{11}$$

where  $d \in \mathbb{R}^n$  is any solution  $\mathbf{A}d = b$ . A natural choice of  $d$  is the minimum norm solution, namely,  $d = \arg \min\{\|x\| : \mathbf{A}x = b\}$ .

Note that  $s \in W^\perp + c$  is equivalent to  $\exists y \in \mathbb{R}^m$  such that  $\mathbf{A}^\top y + c = s$ . Hence, the original variable  $y$  is implicit. The feasible regions can be written as

$$\mathcal{P} = \{x \in \mathbb{R}^n : x \in W + d, x \geq \mathbf{0}\}, \quad \mathcal{D} = \{s \in \mathbb{R}^n : s \in W^\perp + c, s \geq \mathbf{0}\}.$$

## 2.2 Models of Computation

As is standard in the interior point literature, our algorithms use the *real RAM model of computation*. In this model, the input is given by  $K$  real numbers. The algorithm may perform a sequence of basic arithmetic operations ( $+$ ,  $-$ ,  $\times$ ,  $/$ ), and comparisons on real numbers. The algorithm is *polynomial in the real RAM model* if the total number of such operations is polynomially bounded in  $K$ . In the case of (LP), the input numbers are the entries of  $(\mathbf{A}, b, c)$  and  $K = n \times m + n + m$ . We note that while square roots appear in the paper, they are only used in the analysis and not in the actual algorithm.

Consider now a problem in the Turing model where the input is given by  $K$  integers. An algorithm is *strongly polynomial in the Turing model* if it performs  $\text{poly}(K)$  arithmetic operations and comparisons as in the real RAM model, and additionally, the algorithm is in PSPACE: the bit-complexity of all numbers during the computations remains bounded polynomially in the input bit-complexity.

The algorithm in Theorem 1.4 is polynomial in the real RAM model as long as the straight-line complexity of the underlying linear program is polynomial. Moreover, every single iteration is strongly polynomial in the Turing model in the following sense. The input of an iteration is given by the current iterate  $z = (x, s)$  and the matrix  $\mathbf{A}$ . If  $(x, s)$  is a rational vector and  $\mathbf{A}$  is a rational matrix, then all computations during a single iteration are polynomially bounded in the bit-length of  $(\mathbf{A}, x, s)$ .

However, even if the total number of iterations is strongly polynomial, this does not suffice for the entire algorithm to be strongly polynomial. This is because the input of each iterate is the output of the previous one, and the bit complexity of the current iterate can increase polynomially in every iteration. Obtaining stronger guarantees in the Turing model requires additional rounding that goes beyond the scope of this paper.

**Strong Polynomial Linear Algebra** Throughout the paper, we will strongly make use of the fact that many basic linear algebraic operations on matrices, such as computing matrix inverses, powers, products and determinants can be performed in strongly polynomial time. The foundational result from this perspective is that of Edmonds [23], who showed that Gaussian elimination can be implemented in strongly polynomial time, where special care is required to maintain reduced representations of rational numbers without performing greatest common divisor computations (as these are not strongly polynomial). Further results by Strassen [59] and then Berkowitz [6] showed that one can compute the determinant of an integer matrix in strongly polynomial time without division, which in particular allows one to implement Gaussian elimination on a matrix without the need to maintain reduced representations (the entries of each iterate can be expressed directly as a ratio of integer matrix determinants). For an overview of strongly polynomial linear algebra, we refer the interested reader to [30, Section 1.4].

### 2.3 Linear Algebra Preliminaries

In this section, we review fundamental concepts in linear algebra from an operator theoretic perspective, including the notion of adjoint and pseudoinverse operators, orthogonal projections, and the singular value decomposition. Throughout the exposition, we restrict to operators between linear subspaces of  $\mathbb{R}^n$  where we fix the standard inner product. The concepts developed here will be needed for the definition of lifting operators and their duality properties (covered in Section 6.2), for the computation of pseudoinverses, and to make the relationship between operators and their associated matrices precise. For a more thorough background, the interested reader may consult the following reference textbooks [68, 11, 5, 7].

For a linear subspace  $U \subseteq \mathbb{R}^n$ , a basis  $u_1, \dots, u_d$  of  $U$  is any maximal subset of linearly independent vectors in  $U$ . The dimension  $\dim(U) := d$  is the number of vectors in any basis. For any two linear subspaces  $V, W \subseteq \mathbb{R}^n$ , we have that  $\dim(V) + \dim(W) = \dim(V \cap W) + \dim(V + W)$ , that is, dimension is a modular function over subspaces.

**Definition 2.1** (Linear Operator).  $T : U \rightarrow V$  is a linear operator between linear subspaces  $U \subseteq \mathbb{R}^n, V \subseteq \mathbb{R}^m$  if  $T(ax + by) = aT(x) + bT(y)$ , for all  $x, y \in U$  and  $a, b \in \mathbb{R}$ . We will only consider linear operators defined on linear subspaces of some  $\mathbb{R}^n$ , on which the standard inner product is always defined. For  $x \in U$ , we will often write  $Tx := T(x)$  for simplicity of notation.

For a matrix  $\mathbf{M} \in \mathbb{R}^{m \times n}$ , the linear operator  $\mathcal{T}(\mathbf{M}) : \mathbb{R}^n \rightarrow \mathbb{R}^m$  induced by  $\mathbf{M}$  is defined by  $\mathcal{T}(\mathbf{M})(x) = \mathbf{M}x, \forall x \in \mathbb{R}^n$ . Similarly, for an operator  $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ , there is a unique matrix  $\mathcal{M}(T) \in \mathbb{R}^{m \times n}$ , defined by  $\mathcal{M}(T)_{ij} = (e^i)^\top T(e^j), \forall i \in [m], j \in [n]$ , satisfying  $\mathcal{T}(\mathcal{M}(T)) = T$ . More generally, for  $T : U \rightarrow V$ , we define  $\mathcal{M}(T) \in \mathbb{R}^{m \times n}$  to be the unique matrix satisfying  $T(u) = \mathcal{M}(T)u, \forall u \in U$ , and  $\mathcal{M}(T)^\top e^i \in U, \forall i \in [m]$ .

**Remark 2.2.** Given the correspondence above, in the remainder of the paper, we will identify an  $m \times n$  matrix  $\mathbf{M}$  with its associated operator  $\mathcal{T}(\mathbf{M}) : \mathbb{R}^n \rightarrow \mathbb{R}^m$  (noting that  $\mathcal{M}(\mathcal{T}(\mathbf{M})) = \mathbf{M}$ ). All operator theoretic definitions will transfer directly to matrices via this identification. Specifically, we will treat the matrix  $\mathbf{M}$  as the operator  $\mathcal{T}(\mathbf{M})$  whenever we need to apply an operator theoretic concept to  $\mathbf{M}$ .

It is useful to note that the input space  $U = \mathbb{R}^n$  and output space  $V = \mathbb{R}^m$  in the definition of the  $\mathcal{T}(\mathbf{M})$  are chosen maximally, that is, they consist of the entire ambient input and output spaces. For a general operator  $T : U \rightarrow V, U \subseteq \mathbb{R}^n, V \subseteq \mathbb{R}^m$ , note that we may have  $\mathcal{T}(\mathcal{M}(T)) \neq T$  (e.g.,  $T$  may be invertible while  $\mathcal{T}(\mathcal{M}(T)) : \mathbb{R}^n \rightarrow \mathbb{R}^m$  need not be).

We let  $\mathbf{I}_U : U \rightarrow U$  denote the identity operator on  $U$ , where we use the shorthand  $\mathbf{I}_n$  for the identity on  $\mathbb{R}^n$ . We let  $\ker(T) = \{x \in U : Tx = \mathbf{0}_m\}$  and  $\text{im}(T) = \{Tx : x \in U\}$  denote the kernel and image of  $T$  respectively, and let  $\text{rank}(T) = \dim(\text{im}(T))$ . A fundamental identity is  $\dim(U) = \text{rank}(T) + \dim(\ker(T))$ .  $T$  is invertible if there exists a  $T^{-1} : V \rightarrow U$  satisfying  $T^{-1} \circ T = \mathbf{I}_U, T \circ T^{-1} = \mathbf{I}_V$ .  $T$  is invertible if and only if  $\ker(T) = \{\mathbf{0}\}$  and  $\text{im}(T) = V$ . For a linear subspace  $S \subseteq U$ , we define the restricted operator  $T|_S : S \rightarrow V$  to be the operator  $T$  restricted to the subspace  $S$ . For  $X \supseteq \text{im}(T|_S)$ , we also define  $T|_S^X : S \rightarrow X$ , which modifies both the input and output space of  $T$  (note that the condition on  $X$  ensures that  $T|_S^X$  is well-defined).

For two linear subspaces  $V, W \subseteq \mathbb{R}^n$ , we write  $V \perp W$  to indicate that  $V$  and  $W$  are orthogonal, that is,  $\langle v, w \rangle = 0, \forall v \in V, w \in W$ . We define  $W^\perp := \{x \in \mathbb{R}^n : \langle x, y \rangle = 0, \forall y \in W\}$  as the orthogonal complement of  $W$ . The orthogonal complement satisfies  $W + W^\perp = \mathbb{R}^n$ ,  $W \cap W^\perp = \{\mathbf{0}_n\}$ , and  $(W^\perp)^\perp = W$ . We say that  $w_1, \dots, w_d$  are orthonormal vectors in  $W$ , if  $\langle w_i, w_j \rangle = 1$  whenever  $i = j$  and 0 otherwise, for  $i, j \in [d]$ . Additionally,  $w_1, \dots, w_d$  form an orthonormal basis of  $W$  if  $d = \dim(W)$ .

The following identities for orthogonal complements will be used throughout the paper. We state them without proof.

**Proposition 2.3.** *For linear subspaces  $V, W \subseteq \mathbb{R}^n$ ,  $(V + W)^\perp = V^\perp \cap W^\perp$ .*

**Proposition 2.4.** *For a partition  $I \cup J = [n]$  and a linear subspace  $W \subseteq \mathbb{R}^n$ ,  $\pi_I(W)^\perp = \pi_I((W + \mathbb{R}^n)^\perp) = \pi_I(W^\perp \cap \mathbb{R}^n)$  holds.*

For every linear operator, there is a corresponding adjoint operator, whose properties we will use heavily.

**Definition 2.5** (Adjoint Linear Operator). For a linear operator  $T: U \rightarrow V$  between linear subspaces  $U \subseteq \mathbb{R}^n, V \subseteq \mathbb{R}^m$ , we define the adjoint operator  $\text{ad}(T): V \rightarrow U$ , to be the unique linear map satisfying  $\langle v, T(u) \rangle = \langle \text{ad}(T)(v), u \rangle, \forall u \in U, v \in V$ . By uniqueness of the adjoint, note  $T = \text{ad}(\text{ad}(T))$ . The inner products we use on  $U$  and  $V$  are the standard inner product on  $\mathbb{R}^n$  and  $\mathbb{R}^m$  respectively.

**Remark 2.6.** Letting  $\mathcal{M}(T) \in \mathbb{R}^{m \times n}$  be the associated matrix as in Definition 2.1 for  $T: U \rightarrow V$ , one has  $\mathcal{M}(\text{ad}(T)) = \mathcal{M}(T)^\top$ . That is, the adjoint operator corresponds to the transpose of the associated matrix. This equality follows from  $\langle v, T(u) \rangle = \langle v, \mathcal{M}(T)u \rangle = \langle \mathcal{M}(T)^\top v, u \rangle$  and  $\mathcal{M}(T)^\top v \in U$  by our assumptions on  $\mathcal{M}(T)$  (in particular,  $\text{im}(\mathcal{M}(T)^\top) \subseteq U$ ), and the uniqueness of the adjoint.

The following proposition collects relevant properties of the adjoint that we will need. The proof is given in Appendix A.

**Proposition 2.7.** *Let  $T: U \rightarrow V$  be a linear operator between  $U \subseteq \mathbb{R}^n, V \subseteq \mathbb{R}^m$ . Then, the following holds:*

1.  $\ker(\text{ad}(T)) \perp \text{im}(T)$  and  $\ker(\text{ad}(T)) + \text{im}(T) = V$ .
2.  $\ker(T) \perp \text{im}(\text{ad}(T))$  and  $\ker(T) + \text{im}(\text{ad}(T)) = U$ .
3.  $\text{im}(T) = T(\text{im}(\text{ad}(T)))$  and  $\text{im}(\text{ad}(T)) = \text{ad}(T)(\text{im}(T))$ .  
In particular,  $\text{rank}(T) = \text{rank}(\text{ad}(T))$ .

**Definition 2.8** (Orthogonal Projection). For a linear subspace  $W \subseteq \mathbb{R}^n$ , we define  $\Pi_W: \mathbb{R}^n \rightarrow \mathbb{R}^n$ , the orthogonal projection onto  $W$ , to be unique linear operator satisfying that  $\Pi_W(x) \in W$  and  $x - \Pi_W(x) \in W^\perp, \forall x \in \mathbb{R}^n$ .

**Remark 2.9.** For  $x \in \mathbb{R}^n$  and  $W \subseteq \mathbb{R}^n$ , since  $\langle \Pi_W(x), x - \Pi_W(x) \rangle = 0$ , we have that  $\|x\|^2 = \langle x, x \rangle = \|\Pi_W(x)\|^2 + \|x - \Pi_W(x)\|^2$ . In particular,  $\|x\| \geq \|\Pi_W(x)\|$ .

The following proposition states the important properties of orthogonal projections that will be used many times throughout the paper. We state it without proof. For a reference proving some of these properties, see for example [54, Theorem 6.57 and Theorem 6.61].

**Proposition 2.10.** *Let  $W \subseteq \mathbb{R}^n$  be a linear subspace. Then, the orthogonal projection  $\Pi_W: \mathbb{R}^n \rightarrow \mathbb{R}^n$  satisfies the following:*

1.  $\Pi_W$  is self-adjoint, that is,  $\text{ad}(\Pi_W) = \Pi_W$ . In particular, when interpreted as an  $n \times n$  matrix,  $\Pi_W$  is symmetric, that is  $\Pi_W^\top = \Pi_W$ .
2.  $\mathbf{I}_n - \Pi_W = \Pi_{W^\perp}$ .
3.  $\Pi_W(x) = \arg \min_{z \in W} \|x - z\| = \arg \min_{y \in x + W^\perp} \|y\|, \forall x \in \mathbb{R}^n$ . In particular,  $\Pi_W(x)$  is the unique point in  $W \cap (W^\perp + x)$ .

The following proposition gives an explicit formula for associated matrices of general operators in terms of projections, and provides some of their basic properties. The proof is given in Appendix A.

**Proposition 2.11.** *Let  $T: U \rightarrow V$ ,  $U \subseteq \mathbb{R}^n$ ,  $V \subseteq \mathbb{R}^m$  be a linear operator. Then,  $\mathcal{M}(T)_{ij} = (e^i)^\top T(\Pi_U e^j)$ ,  $\forall i \in [m], j \in [n]$ , and  $\mathcal{T}(\mathcal{M}(T)) = T|_U^{\mathbb{R}^m} \circ \Pi_U|_{\mathbb{R}^n}^U$ , where  $\Pi_U: \mathbb{R}^n \rightarrow \mathbb{R}^n$  is the orthogonal projection onto  $U$ . Furthermore,  $\text{im}(\mathcal{M}(T)) = \text{im}(T)$  and  $\text{im}(\mathcal{M}(T)^\top) = \text{im}(\text{ad}(T))$ .*

The next proposition shows that restricting the input to any subspace containing the image of the adjoint does not change the associated matrix.

**Proposition 2.12.** *Let  $T: U \rightarrow V$ ,  $U \subseteq \mathbb{R}^n$ ,  $V \subseteq \mathbb{R}^m$  be a linear operator. Let  $\widehat{U}$  be a linear subspace satisfying  $\text{im}(\text{ad}(T)) \subseteq \widehat{U} \subseteq U$  and let  $\widehat{T} = T|_{\widehat{U}}$ . Then,  $\mathcal{M}(T) = \mathcal{M}(\widehat{T})$ .*

*Proof.* By definition,  $\mathcal{M}(T)(u) = T(u) = \widehat{T}(u)$ ,  $\forall u \in \widehat{U}$ . Then by Proposition 2.11,  $\text{im}(\mathcal{M}(T)^\top) = \text{im}(\text{ad}(T)) \subseteq \widehat{U}$ , and thus  $\mathcal{M}(T) = \mathcal{M}(\widehat{T})$  by uniqueness.  $\square$

For the task of computing projection matrices, we will require the pseudoinverse operator.

**Definition 2.13** (Moore-Penrose Pseudoinverse). *Let  $T: U \rightarrow V$ ,  $U \subseteq \mathbb{R}^n$ ,  $V \subseteq \mathbb{R}^m$  be a linear operator. Then, the Moore-Penrose pseudoinverse  $T^+: V \rightarrow U$  of  $T$  is the unique linear operator satisfying  $T^+T(u) = u$ ,  $\forall u \in \text{im}(\text{ad}(T))$  and  $\ker(T^+) = \ker(\text{ad}(T))$ .*

The following proposition gives the relation between pseudoinverses and projection operators and the connection to least-squares problems. For a proof, see [5, Statements 6.69, 6.70].

**Proposition 2.14.** *Let  $\mathbf{M} \in \mathbb{R}^{m \times n}$ . Then, the following hold:*

1.  $\mathbf{M}^+\mathbf{M} = \Pi_{\text{im}(\text{ad}(\mathbf{M}))}$  and  $\mathbf{M}\mathbf{M}^+ = \Pi_{\text{im}(\mathbf{M})}$ .
2.  $\mathbf{M}^+v = \arg \min_{u \in U: \mathbf{M}u = \Pi_{\text{im}(\mathbf{M})}v} \|u\|_2$ .

The next proposition shows that pseudoinverses and orthogonal projections can be computed in strongly polynomial time. For a proof, see for example [11, Theorem 1.3.2 and Algorithm 1.3.1].

**Proposition 2.15.** *Let  $\mathbf{M} \in \mathbb{R}^{m \times n}$  be a matrix and let  $r = \text{rank}(\mathbf{M})$ . Then the following holds:*

1. If  $r = m$ , we have  $\mathbf{M}^+ = \mathbf{M}^\top (\mathbf{M}\mathbf{M}^\top)^{-1}$ .
2. If  $r = n$ , we have  $\mathbf{M}^+ = (\mathbf{M}^\top \mathbf{M})^{-1} \mathbf{M}^\top$ .
3. If  $r < \min\{m, n\}$ , then for any rank factorization  $\mathbf{M} = \mathbf{A}\mathbf{B}$  satisfying  $\mathbf{A} \in \mathbb{R}^{m \times r}$ ,  $\mathbf{B} \in \mathbb{R}^{r \times n}$  with  $r = \text{rank}(\mathbf{A}) = \text{rank}(\mathbf{B})$ , we have

$$\mathbf{M}^+ = \mathbf{B}^+ \mathbf{A}^+ = \mathbf{B}^\top (\mathbf{B}\mathbf{B}^\top)^{-1} (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top.$$

*In particular, on input  $\mathbf{M}$ , the linear operators  $\mathbf{M}^+$ ,  $\Pi_{\text{im}(\mathbf{M})} = \mathbf{M}\mathbf{M}^+$  and  $\Pi_{\text{im}(\mathbf{M}^\top)} = \mathbf{M}^+\mathbf{M}$  can be computed in strongly polynomial time.*

We will need the notions of singular values and singular value decompositions defined below.

**Definition 2.16** (Singular Value Decomposition). *A linear operator  $T: U \rightarrow V$ , where  $U \subseteq \mathbb{R}^n$  and  $V \subseteq \mathbb{R}^m$  are linear subspaces, admits a singular value decomposition (SVD)*

$$\mathcal{M}(T) = \sum_{i=1}^{\text{rank}(T)} \sigma_i(T) v_i u_i^\top \quad (12)$$

where  $v_1, \dots, v_{\text{rank}(T)} \in V$  and  $u_1, \dots, u_{\text{rank}(T)} \in U$  are orthonormal vectors in their respective subspaces and  $\sigma_1(T) \geq \dots \geq \sigma_{\text{rank}(T)}(T) > 0$ . We define

$$\sigma(T) := (\sigma_1(T), \dots, \sigma_{\dim(U)}(T)), \quad (13)$$

where  $\sigma_i(T) := 0$  for  $i \in [\dim(U)] \setminus [\text{rank}(T)]$ , the complete vector of singular values listed in non-increasing order, and

$$\sigma^+(T) := (\sigma_1(T), \dots, \sigma_{\text{rank}(T)}(T)), \quad (14)$$

the subvector of positive singular values. We use the shorthand  $\sigma_{\max}(T) := \sigma_1(T)$  and  $\sigma_{\min}(T) := \sigma_{\dim(U)}(T)$ . By convention, we let  $\sigma_i(T) = 0$  for  $i > \dim(U)$ , though we do not include this as an entry

of  $\sigma(T)$ . We will often need to count the number of singular values inside an interval  $I \subseteq \mathbb{R}_+$ . For this purpose, we use the following notation:

$$\mathcal{C}_\sigma(T, I) := |\{i \in [\dim(U)] : \sigma_i(T) \in I\}|.$$

We will also use the shorthand  $\mathcal{C}_\sigma(T, t) := \mathcal{C}_\sigma(T, [0, t])$  for  $t \geq 0$ . Note that if  $U = \{\mathbf{0}\}$ , then  $\text{rank}(T) = 0$  and  $\mathcal{C}_\sigma(T, \mathbb{R}_+) = 0$ .

**Remark 2.17.** We remark that while the singular value decomposition  $\mathcal{M}(T) = \sum_{i=1}^{\text{rank}(T)} \sigma_i(T) v_i u_i^\top$  above need not be unique (i.e., if some of the non-zero singular values are equal), the vector of singular values  $\sigma(T)$  does not depend on the choice of singular value decomposition.

We will require the standard relation between the adjoint and the singular value decomposition, which we state without proof.

**Proposition 2.18.** *Let  $T: U \rightarrow V$  be a linear operator with singular value decomposition  $\mathcal{M}(T) = \sum_{i=1}^{\text{rank}(T)} \sigma_i v_i u_i^\top$  as in Definition 2.16. Then,  $\text{ad}(T): V \rightarrow U$  has singular value decomposition  $\mathcal{M}(\text{ad}(T)) = \sum_{i=1}^{\text{rank}(T)} \sigma_i u_i v_i^\top$  and  $\sigma^+(T) = \sigma^+(\text{ad}(T))$ .*

We will often need to relate the singular values of an operator with the singular values of its restrictions and its associated matrix. The following direct corollary of Proposition 2.12 gives the precise relations.

**Proposition 2.19.** *Let  $T: U \rightarrow V$ ,  $U \subseteq \mathbb{R}^n$ ,  $V \subseteq \mathbb{R}^m$  a linear operator, and let  $\widehat{U}$  be a linear subspace satisfying  $\text{im}(\text{ad}(T)) \subseteq \widehat{U} \subseteq U$ . Then  $\sigma^+(\mathcal{M}(T)) = \sigma^+(T) = \sigma^+(T|_{\widehat{U}})$ .*

*Proof.* By Proposition 2.12, we have that  $\mathcal{M}(T) = \mathcal{M}(\widehat{T})$ , and hence both operators have the same SVD in Definition 2.16. Therefore, the positive singular value vector is identical. The same holds for  $\sigma^+(\mathcal{M}(T)) := \sigma^+(\mathcal{T}(\mathcal{M}(T)))$ .  $\square$

One of the most useful ways to characterize singular values is via the Courant-Fischer variational characterization. See for example [7, Corollary III.1.2], which gives the variational characterization for the eigenvalues of  $\text{ad}(T)T$  (equal to squared singular values of  $T$ ).

**Proposition 2.20** (Max-Min Principle for Singular Values). *Let  $T: U \rightarrow V$  be a linear operator. Then, for  $1 \leq i \leq \dim(U)$ , we have that*

$$\sigma_i(T) = \min_{\substack{\dim(S) \geq \dim(U) - i + 1 \\ S \subseteq U}} \max_{x \in S \setminus \{\mathbf{0}\}} \frac{\|Tx\|}{\|x\|} \quad (15)$$

$$= \max_{\substack{\dim(S) \geq i \\ S \subseteq U}} \min_{x \in S \setminus \{\mathbf{0}\}} \frac{\|Tx\|}{\|x\|}. \quad (16)$$

where  $S$  ranges over the linear subspaces of  $U$ .

**Remark 2.21.** In Proposition 2.20, one can replace the conditions  $\dim(S) \geq \dim(U) - i + 1$  and  $\dim(S) \geq i$  with  $\dim(S) = \dim(U) - i + 1$  and  $\dim(S) = i$  respectively. In (15), any subspace  $S'$  of  $S$  of dimension exactly  $\dim(U) - i + 1$  satisfies

$$\max_{x' \in S' \setminus \{\mathbf{0}\}} \frac{\|T(x')\|}{\|x'\|} \leq \max_{x \in S \setminus \{\mathbf{0}\}} \frac{\|T(x)\|}{\|x\|},$$

and in (16), any subspace  $S'$  of  $S$  of dimension exactly  $i$  satisfies

$$\min_{x' \in S' \setminus \{\mathbf{0}\}} \frac{\|T(x')\|}{\|x'\|} \geq \min_{x \in S \setminus \{\mathbf{0}\}} \frac{\|T(x)\|}{\|x\|}.$$

The flexibility to use different dimensions will be useful in the sequel, however.

**Remark 2.22.** By Proposition 2.20 above, for any linear subspace  $S \subseteq U$ ,  $1 \leq \dim(S) \leq \dim(U)$ , we have the useful identities

$$\begin{aligned} \sigma_{\max}(T|_S) &:= \sigma_1(T|_S) = \max_{x \in S \setminus \{\mathbf{0}\}} \frac{\|Tx\|}{\|x\|}, \\ \sigma_{\min}(T|_S) &:= \sigma_{\dim(S)}(T|_S) = \min_{x \in S \setminus \{\mathbf{0}\}} \frac{\|Tx\|}{\|x\|}, \end{aligned} \quad (17)$$

where  $T|_S$  is the operator  $T$  restricted to  $S$ . If  $\dim(S) = 0$ , recall that  $\sigma_1(T|_S) = 0$  by convention.

**Remark 2.23.** Let  $\mathcal{M}(T) = \sum_{i=1}^r \sigma_i v_i u_i^\top$ ,  $r = \text{rank}(T)$ , be the SVD of  $T$  and let  $\mathbf{U} = [u_1, \dots, u_r]$ . For  $i \in [\dim(U)]$ , a canonical choice for the subspace attaining the minimum in (15) is  $\text{im}(\mathbf{U}_{\geq i}) + \ker(T)$  for  $i \in [\text{rank}(T)]$ , and  $\ker(T)$  for  $i > \text{rank}(T)$ . For  $i \in [\text{rank}(T)]$ , using that  $\text{im}(\mathbf{U}_{\geq i}) \subseteq \text{im}(\text{ad}(T))$  and  $\text{im}(\text{ad}(T)) \perp \ker(T)$ , this can be verified as follows:

$$\begin{aligned} \sigma_1(T|_{\text{im}(\mathbf{U}_{\geq i}) + \ker(T)}) &= \max_{\substack{w \in \text{im}(\mathbf{U}_{\geq i}), z \in \ker(T) \\ w+z \neq \mathbf{0}_n}} \frac{\|T(w+z)\|}{\|w+z\|} \\ &= \max_{\substack{w \in \text{im}(\mathbf{U}_{\geq i}), z \in \ker(T) \\ w+z \neq \mathbf{0}_n}} \frac{\|Tw\|}{\sqrt{\|w\|^2 + \|z\|^2}} \\ &= \max_{\substack{w \in \text{im}(\mathbf{U}_{\geq i}) \\ w \neq \mathbf{0}_n}} \frac{\|Tw\|}{\|w\|} = \max_{\substack{x \in \mathbb{R}^{r-i+1} \\ x \neq \mathbf{0}_{r-i+1}}} \frac{\|T(\mathbf{U}_{\geq i}x)\|}{\|\mathbf{U}_{\geq i}x\|} \\ &= \max_{\substack{x \in \mathbb{R}^{r-i+1} \\ x \neq \mathbf{0}_{r-i+1}}} \frac{\sqrt{\sum_{j=r-i+1}^r \sigma_j^2 x_j^2}}{\|x\|} = \sigma_{r-i+1}, \end{aligned}$$

where the second to last equality uses the orthonormality of  $v_1, \dots, v_r \in V$  and  $u_1, \dots, u_r \in U$ , and the last equality uses the non-increasing order of  $\sigma$ . Note further that  $\dim(\text{im}(\mathbf{U}_{\geq i}) + \ker(T)) = \dim(\text{im}(\mathbf{U}_{\geq i})) + \dim(\ker(T)) = (\text{rank}(T) - i + 1) + \dim(U) - \text{rank}(T) = \dim(U) - i + 1$ , and hence the subspace has the correct dimension. For  $i \in [\dim(U)] \setminus [\text{rank}(T)]$ , one trivially has  $\sigma_1(T|_{\ker(T)}) = 0$  and  $\dim(\ker(T)) = \dim(U) - \text{rank}(T) \geq \dim(U) - i + 1$ , as needed.

Similarly, a canonical choice for the subspace attaining the maximum in (16) is  $\text{im}(\mathbf{U}_{\leq i})$  if  $i \in [\text{rank}(T)]$ , and  $U$  if  $i > \text{rank}(T)$  (recalling that  $\sigma_i(T) = 0$  for  $i > \text{rank}(T)$ ).

### 2.3.1 Approximate Singular Subspaces

In this subsection, we define approximate singular subspaces for general linear operators and collect their main properties, which will be crucial for the analysis and implementation of our IPM.

**Definition 2.24** (Approximate Singular Subspace). Let  $T: X \rightarrow Y$  be a linear operator. A linear subspace  $S \subseteq X$  is a  $\varrho$ -approximate singular subspace for  $T$ , for  $\varrho \geq 1$ , if  $\sigma_1(T|_S) \leq \varrho \sigma_{\dim(X) - \dim(S) + 1}(T)$ .

**Remark 2.25** (Trivial Subspace). The trivial subspace  $S := \{\mathbf{0}\} \subseteq X$  is always 1-approximate singular subspace for  $T$  as  $\sigma_1(T|_S) = 0 = \sigma_{\dim(X)+1}(T) = \sigma_{\dim(X) - \dim(S) + 1}(T)$  by convention.

When computing approximate singular subspaces, we will use the following lemma to relate approximate singular subspaces of an operator to those of its associated matrix. The proof is given in Appendix A.1.

**Lemma 2.26.** Let  $T: X \rightarrow Y$ ,  $X \subseteq \mathbb{R}^n$ ,  $Y \subseteq \mathbb{R}^m$  be a linear operator and let  $\bar{T} = \mathcal{M}(T)$  be its associated matrix. Then, for a linear subspace  $S$  satisfying  $\ker(\bar{T}) \subseteq S \subseteq \mathbb{R}^n$ , the following holds:

1.  $\ker(T) \subseteq S \cap X = \Pi_X(S)$  and  $\sigma_1(T|_{S \cap X}) = \sigma_1(\bar{T}|_S)$ .
2.  $S$  is a  $\varrho$ -approximate singular subspace for  $\bar{T}$  of dimension  $d \geq 0$  if and only if  $S \cap X$  is a  $\varrho$ -approximate singular subspace for  $T$  of dimension  $d - \dim(X^\perp)$ .

Moreover, for any  $\tau \geq 0$ ,  $\mathcal{C}_\sigma(T, \tau) + \dim(X^\perp) = \mathcal{C}_\sigma(\bar{T}, \tau)$ .

Our definition of an approximate singular subspace  $S$  for  $T$  of dimension  $d$  corresponds to  $S$  be an approximate minimizer of (15), that is

$$\sigma_1(T|_S) \leq \varrho \min_{C \subseteq X, \dim(C)=d} \sigma_1(T|_C) = \varrho \sigma_{\dim(X)-d+1}(T).$$

Given this, one may ask whether there is a relationship between  $S$  and maximizers for the complementary program (16), that is,

$$\max_{D \subseteq \dim(D)=\dim(X)-d} \sigma_{\min}(T|_D) = \sigma_{\dim(X)-d}.$$

The following lemma shows that in fact the orthogonal complement of  $S$  inside  $X$  is indeed an approximate maximizer of (16), under the condition that there is a gap between the singular values  $\sigma_{\dim(X)-d}(T)$  and  $\sigma_{\dim(X)-d+1}(T)$ . This will be important for the error analysis of subspace LLS steps defined in Section 6.1. The proof is given in Appendix A.1.

**Lemma 2.27.** *Let  $T: X \rightarrow Y$  be a linear operator. Let  $S \subseteq X$  be a  $\varrho$ -approximate subspace with  $0 \leq \dim(S) < \dim(X)$  for  $T$ . Then, for  $\bar{S} = X \cap S^\perp$ , we have that*

$$\sigma_{\min}(T|_{\bar{S}})^2 \geq \sigma_{\dim(X)-\dim(S)}(T)^2 - \varrho^2 \sigma_{\dim(X)-\dim(S)+1}(T)^2.$$

## 2.4 Preliminaries on Interior-Point Methods

In this section, we recall standard properties of the central path and IPM that will be required for our algorithm. To ensure that the central path is well-defined, we assume that  $\mathcal{P}$  and  $\mathcal{D}$  admit strictly feasible solutions, i.e., the sets  $\mathcal{P}_{++}$  and  $\mathcal{D}_{++}$  are both nonempty. We recall the notation  $z^{\text{CP}}(\mu) = (x^{\text{CP}}(\mu), s^{\text{CP}}(\mu))$  to denote the central path point at  $\mu$  as in definition (1).

Given  $z = (x, s) \in \mathcal{P} \times \mathcal{D}$ , we recall that the normalized duality gap is defined as  $\bar{\mu}(z) := \frac{\langle x, s \rangle}{n}$ . The following identity is useful in comparing duality gaps.

**Proposition 2.28.** *Given  $x, x' \in W + d$ ,  $s, s' \in W^\perp + c$ , we have that*

$$\langle x, s \rangle + \langle x', s' \rangle = \langle x, s' \rangle + \langle x', s \rangle.$$

*In particular, if  $\langle x', s' \rangle = 0$ , then*

$$\langle x, s \rangle = \langle x, s' \rangle + \langle x', s \rangle.$$

*Proof.* Since  $x - x' \in W$  and  $s - s' \in W^\perp$ , we have that

$$0 = \langle x - x', s - s' \rangle \Leftrightarrow \langle x, s \rangle + \langle x', s' \rangle = \langle x, s' \rangle + \langle x', s \rangle.$$

□

The next proposition shows that the normalized duality gap is a linear function for convex combinations of points.

**Proposition 2.29** (Linearity duality gap). *For  $x^{(1)}, \dots, x^{(k)} \in W + d$ ,  $s^{(1)}, \dots, s^{(k)} \in W^\perp + c$  forming the sequence  $z^{(1)} = (x^{(1)}, s^{(1)}), \dots, z^{(k)} = (x^{(k)}, s^{(k)})$  and  $\lambda \in \mathbb{R}^k$  such that  $\sum_{i=1}^k \lambda_i = 1$ , we have that*

$$\bar{\mu}\left(\sum_{i=1}^k \lambda_i z^{(i)}\right) = \sum_{i=1}^k \lambda_i \bar{\mu}(z^{(i)}).$$

*Proof.* Using that  $\sum_{i=1}^k \lambda_i = 1$  and the orthogonality of  $x^{(i)} - d \in W$  and  $s^{(i)} - c \in W^\perp$  for all  $i \in [k]$  we first get

$$\begin{aligned} \left\langle \sum_{i=1}^k \lambda_i x^{(i)}, \sum_{j=1}^k \lambda_j s^{(j)} \right\rangle &= \left\langle \sum_{i=1}^k \lambda_i (x^{(i)} - d) + d, \sum_{j=1}^k \lambda_j (s^{(j)} - c) + c \right\rangle \\ &= \langle d, c \rangle + \sum_{i=1}^k \lambda_i (\langle x^{(i)} - d, c \rangle + \langle d, s^{(i)} - c \rangle) \\ &= \sum_{i=1}^k \lambda_i (\langle d, c \rangle + \langle x^{(i)} - d, c \rangle + \langle d, s^{(i)} - c \rangle) \\ &= \sum_{i=1}^k \lambda_i \langle x^{(i)}, s^{(i)} \rangle. \end{aligned}$$

Division by  $n$  yields the respective normalized duality gap. □

A key property of the central path is ‘near monotonicity’, formulated in the following lemma, see [66, Lemma 16].

**Lemma 2.30.** *For the central path points at  $0 < \mu' \leq \mu$ , we have*

$$\left\| \frac{x^{\text{CP}}(\mu')}{x^{\text{CP}}(\mu)} + \frac{s^{\text{CP}}(\mu')}{s^{\text{CP}}(\mu)} \right\|_\infty \leq n.$$

In Lemma 2.43, we also present an  $\ell_1$ -variant of this bound.

**Central Path Neighborhoods** The neighborhoods  $\mathcal{N}^2(\beta)$  and  $\mathcal{N}^{-\infty}(\theta)$  introduced in (2) and (3) comprise the points  $z = (x, s) \in \mathcal{P} \times \mathcal{D}$  such that the *centrality error*, i.e., the norm of the vector  $\frac{xs}{\bar{\mu}(z)} - \mathbf{1}$ , is bounded. They use of the  $\ell_2$ -norm and the  $\ell_\infty$ -seminorm  $\|u\|_\infty := \max_{1 \leq i \leq n} \max(0, -u_i)$ , respectively.

We will often use the following proposition which is immediate from the definition of  $\mathcal{N}^2(\beta)$ .

**Proposition 2.31.** *Let  $z = (x, s) \in \mathcal{N}^2(\beta)$  for  $\beta \in (0, 1)$ , and  $\mu = \bar{\mu}(z)$ . Then for each  $i \in [n]$*

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

*Proof.* By definition of  $\mathcal{N}(\beta)$  we have for all  $i \in [n]$  that  $|\frac{x_i s_i}{\mu} - 1| \leq \left\| \frac{xs}{\mu} - \mathbf{1} \right\| \leq \beta$  and so  $(1 - \beta)\mu \leq x_i s_i \leq (1 + \beta)\mu$ .  $\square$

The following proposition gives a bound on the distance between a point  $z \in \mathcal{N}^2(\beta)$  in the  $\beta$ -neighborhood and the corresponding central path point with the same normalized duality gap  $z(\mu)$  for  $\mu = \bar{\mu}(z)$ . See e.g., [28, Lemma 5.4] and [47, Proposition 2.1].

**Proposition 2.32.** *Let  $z = (x, s) \in \mathcal{N}^2(\beta)$  for  $\beta \in (0, 1/4]$  and  $\mu = \bar{\mu}(z)$ , and consider the central path point  $z^{\text{cp}}(\mu) = (x^{\text{cp}}(\mu), s^{\text{cp}}(\mu))$ . For each  $i \in [n]$ ,*

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

We will need the following lemma regarding the near-optimality of the choice  $\bar{\mu}(z)$  as  $\langle x, s \rangle / n$  for a point  $z = (x, s)$  with respect to minimizing centrality error.

**Lemma 2.33** ([47, Lemma 4.4]). *For  $\beta \in (0, 1/4]$ , let  $z = (x, s) \in \mathcal{P}_{++} \times \mathcal{D}_{++}$  and  $\mu' > 0$  satisfy  $\|xs/\mu' - \mathbf{1}\| \leq \beta$ . Then,*

$$(1 - \beta/\sqrt{n})\mu' \leq \bar{\mu}(z) \leq (1 + \beta/\sqrt{n})\mu' \quad \text{and} \quad z \in \mathcal{N}^2(\beta/(1 - \beta)).$$

The next lemma relates a point in the wide neighborhood to the corresponding central path point.

**Lemma 2.34.** *Let  $z = (x', s') \in \mathcal{N}^{-\infty}(\theta)$ ,  $\theta \in [0, 1)$ . Then for  $\mu = \bar{\mu}(z)$ , and the corresponding central path point  $z^{\text{cp}}(\mu) = (x^{\text{cp}}(\mu), s^{\text{cp}}(\mu))$ , we have that*

$$\frac{1}{2n}x' \leq x^{\text{cp}}(\mu) \leq \frac{2n}{1 - \theta}x' \quad \text{and} \quad \frac{1}{2n}s' \leq s^{\text{cp}}(\mu) \leq \frac{2n}{1 - \theta}s'.$$

*Proof.* We only prove the inequalities on  $x'$ ; the proof of the inequalities on  $s'$  is symmetric. Let  $(x, s) := (x^{\text{cp}}(\mu), s^{\text{cp}}(\mu))$ . Using Proposition 2.28, for  $i \in [n]$  we have that

$$\frac{x'_i}{x_i} = \frac{x'_i s_i}{\mu} \leq \frac{1}{\mu}(\langle x', s \rangle + \langle x, s' \rangle) = \frac{1}{\mu}(\langle x', s' \rangle + \langle x, s \rangle) = 2n.$$

This proves the first inequality; note that this part does not use  $z \in \mathcal{N}^{-\infty}(\theta)$ , but only that  $z \in \mathcal{P}_{++} \times \mathcal{D}_{++}$ . For the second inequality,  $z \in \mathcal{N}^{-\infty}(\theta)$  by definition implies

$$\frac{x_i}{x'_i} \leq \frac{x_i s'_i}{\mu(1 - \theta)} \leq \frac{1}{\mu(1 - \theta)}(\langle x', s' \rangle + \langle x, s \rangle) = \frac{2n}{1 - \theta},$$

as needed.  $\square$

## 2.5 The Central Path Limit

Assume that the polytope  $\mathcal{P} = \{x \in \mathbb{R}^n : x \in W + d, x \geq \mathbf{0}\}$  has a strictly positive solution and is bounded. Then, the *analytic center* of  $\mathcal{P}$  is the point

$$\tilde{x} := \arg \max_{x \in \mathcal{P}} \sum_{i=1}^n \log x_i.$$

By Lagrangian duality, there exists a vector  $\tilde{v} \in W^\perp$  such that  $\tilde{x}\tilde{v} = \mathbf{1}_n$ . The limit point  $z^* = (x^*, s^*)$  of the central path corresponds to the analytic centers of the primal and dual optimal faces. Namely, assume  $\mathcal{P}_{++}, \mathcal{D}_{++} \neq \emptyset$ , and let  $(B^*, N^*)$  denote the optimal partition, i.e.,  $x \in \mathcal{P}$  is optimal if and only if  $\text{supp}(x) \subseteq B^*$  and  $s \in \mathcal{D}$  is optimal if and only if  $\text{supp}(s) \subseteq N^*$ . Let

$$\mathcal{P}^* := \mathcal{P} \cap \mathbb{R}_{B^*}^n \quad \text{and} \quad \mathcal{D}^* := \mathcal{D} \cap \mathbb{R}_{N^*}^n$$

denote the set of primal and dual optimal solutions. Then,  $\pi_{B^*}(\mathcal{P}^*) \subseteq \mathbb{R}^{B^*}$  and  $\pi_{N^*}(\mathcal{D}^*) \subseteq \mathbb{R}^{N^*}$  are the projections of the optimal sets to the respective coordinate sets, and the following holds; see e.g., [53, Theorem I.30]

**Theorem 2.35.** *Assume  $\mathcal{P}_{++}, \mathcal{D}_{++} \neq \emptyset$ . Then, the optimal partition  $(B^*, N^*)$  and the limit  $(x^*, s^*) = \lim_{\mu \searrow 0} (x(\mu), s(\mu))$  exist, and  $x_{B^*}^*$  is the analytic center of  $\pi_{B^*}(\mathcal{P}^*)$ , and  $s_{N^*}^*$  is the analytic center of  $\pi_{N^*}(\mathcal{D}^*)$ .*

Consider now the output in Theorem 1.4 in the case  $\mu_1 = 0$ . Since  $B = \text{supp}(x^1)$  and  $N = \text{supp}(s^1)$  form a partition of  $[n]$ , it follows that the solutions  $(x^1, s^1)$  are primal and dual optimal, and  $B = B^*$ ,  $N = N^*$ . The additional output  $(v^1, w^1)$  in Theorem 1.4 satisfying  $\|(x_B^1 v_B^1, s_N^1 w_N^1) - \mathbf{1}\| \leq \beta$  provides Lagrange certificates that the solution  $(x^1, s^1)$  is close to  $(x^*, s^*)$ . Namely,  $v^1 \subseteq \mathbb{R}^B$  certifies that  $x_B^1$  is multiplicatively near the analytic center of  $\pi_B(\mathcal{P}^*)$ , and  $w^1 \subseteq \mathbb{R}^N$  is analogously a certificate for  $s_N^1$  and  $\pi_N(\mathcal{D}^*) \subseteq \mathbb{R}^N$  (see Proposition 2.32).

## 2.6 Predictor-Corrector Methods

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

$$\Delta x \in W \tag{18}$$

$$\Delta s \in W^\perp \tag{19}$$

$$s\Delta x + x\Delta s = \nu\mu\mathbf{1} - xs \tag{20}$$

Predictor-corrector methods, such as the Mizuno–Todd–Ye Predictor-Corrector algorithm [45], alternate between two types of steps. In *corrector steps*, we use  $\nu = 1$ . This gives the *centrality direction*, denoted as  $\Delta z^c = (\Delta x^c, \Delta s^c)$ . In *predictor steps*, we use  $\nu = 0$ . This direction is also called the *affine scaling direction*, and will be denoted as  $\Delta z^a = (\Delta x^a, \Delta s^a)$  throughout.

Let  $z := (x, s) \in \mathcal{N}^2(\beta)$  be our current iterate. In our algorithm, we will first apply a corrector step to get  $z^c := z + \Delta z^c$ , which will reduce our centrality error by a factor 2, that is,  $z^c \in \mathcal{N}^2(\beta/2)$ , without changing the gap  $\bar{\mu}(z)$ . Following this, we apply a predictor step to get  $z^+ := z^c + \alpha^a \Delta z^a$ , for  $\alpha^a \in (0, 1]$ , which will make progress along the central path while maintaining that  $z^+ \in \overline{\mathcal{N}^2}(\beta)$ . Here we slightly abuse notation, by letting  $\Delta z^a := \Delta(z^c)^a$ , that is the predictor direction computed from the recentered iterate  $z^c$ . The step-length  $\alpha^a > 0$  will be chosen such that

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

Thus, we conclude  $z^+ = z^c + \alpha^a \Delta z^a \in \overline{\mathcal{N}^2}(\beta)$ . We remark that the closure allows us to take a step that goes all the way to an optimal solution. If  $z^+ \in \mathcal{N}^2(\beta)$ , i.e., if we have not arrived at an optimal solution, then  $z^+$  remains a valid iterate for the next step.

**Remark 2.36.** In contrast to predictor-corrector methods such as [45], ours is a ‘corrector-predictor’ method, first performing corrector steps followed by predictor steps in each iteration. While the two descriptions are equivalent, it is more convenient for the description of the final iterate, i.e., achieve the termination guarantee in Theorem 1.4 for the same  $\beta$  neighbourhood.

The next proposition summarizes well-known properties of predictor and corrector steps, see e.g. [67, Section 4.5.1].

**Proposition 2.37.** *Let  $z = (x, s) \in \mathcal{N}^2(\beta)$  for  $\beta \in (0, 1/6]$ .*

(i) *For  $z \in \mathcal{N}^2(\beta)$ , let  $\Delta z^c$  be the corrector direction at  $z$ . Then for  $z^c = z + \Delta z^c$ , we have  $\bar{\mu}(z^c) = \bar{\mu}(z)$  and  $z^c \in \mathcal{N}^2(\beta/2)$ .*

(ii) *For the affine scaling step, we have  $\bar{\mu}(z^+) = (1 - \alpha^a)\bar{\mu}(z)$  and  $z^+ \in \overline{\mathcal{N}^2}(\beta)$ .*

(iii) The affine scaling step-length  $\alpha^a$  can be chosen in the range

$$0 \leq \alpha^a \leq \max \left\{ \frac{\beta}{2\sqrt{n}}, 1 - \frac{2 \|\Delta x^a \Delta s^a\|}{\beta \bar{\mu}(z)} \right\}.$$

(iv) After a sequence of  $\left\lceil \frac{3\sqrt{n}}{\beta} \log(1/\varepsilon) \right\rceil$ ,  $\varepsilon \in (0, 1]$ , corrector and predictor steps from  $z$ , assuming the affine scaling step-lengths are all at least  $\frac{\beta}{3\sqrt{n}}$ , we obtain an iterate  $z' = (x', s') \in \mathcal{N}^2(\beta)$  such that  $\bar{\mu}(z') \leq \varepsilon \bar{\mu}(z)$ .

**Remark 2.38.** By (iii), an affine scaling step-length of  $\frac{\beta}{2\sqrt{n}}$  is always valid. Therefore, the assumption in (iv) that the step-lengths are at least  $\frac{\beta}{3\sqrt{n}}$  is conservative. We use this conservative estimate for purely computational reasons as  $\frac{\beta}{2\sqrt{n}}$  may be irrational. In particular, this choice is designed to combine with the step-length computation given by Proposition 2.39 below.

The following proposition explicitly states that the predictor and corrector steps can be computed in strongly polynomial time, and that we can select appropriate step-lengths. The reason is simply that computing the steps amounts to solving linear systems based on the data  $x, s, \mu$  and  $\nu$ .

**Proposition 2.39** (Step Formulas). *Let  $\mathbf{A} \in \mathbb{R}^{m \times n}$ ,  $\text{rank}(\mathbf{A}) = m$ ,  $W = \ker(\mathbf{A})$  and  $W^\perp = \text{im}(\mathbf{A}^\top)$ . Let  $z := (x, s)$ , where  $x \in W$ ,  $x > \mathbf{0}$ ,  $s \in W^\perp$ ,  $s > \mathbf{0}$  and  $t \in \mathbb{R}^n$ . Then, the solution to  $s\Delta x + x\Delta s = t$ ,  $\Delta x \in W$ ,  $\Delta s \in W^\perp$  can be expressed as*

$$\begin{aligned} \Delta s &= \mathbf{A}^\top (\mathbf{A} \text{diag}(x/s) \mathbf{A}^\top)^{-1} \mathbf{A} \frac{t}{s} \\ \Delta x &= \frac{t}{s} - \text{diag}(x/s) \mathbf{A}^\top (\mathbf{A} \text{diag}(x/s) \mathbf{A}^\top)^{-1} \mathbf{A} \frac{t}{s}. \end{aligned}$$

Moreover, both the affine scaling step  $(\Delta x^a, \Delta s^a)$  and corrector step  $(\Delta x^c, \Delta s^c)$ , corresponding to  $t = \nu \bar{\mu}(z) \mathbf{1}_n - xs$  for  $\nu \in \{0, 1\}$  respectively, can be computed in strongly polynomial time. Furthermore, given  $\Delta x, \Delta s \in \mathbb{R}^n$  and  $\beta \in (0, 1/6]$ , one can in strongly polynomial time compute an affine scaling step-length  $\alpha^a \in (0, 1]$  satisfying

$$\max \left\{ \frac{\beta}{3\sqrt{n}}, 1 - \frac{3 \|\Delta x^a \Delta s^a\|}{\beta \bar{\mu}(z)} \right\} \leq \alpha^a \leq \max \left\{ \frac{\beta}{2\sqrt{n}}, 1 - \frac{2 \|\Delta x^a \Delta s^a\|}{\beta \bar{\mu}(z)} \right\}.$$

*Proof.* It is directly verified by inspection that  $s\Delta x + x\Delta s = t$  and that  $\mathbf{A}\Delta x = \mathbf{0}$  and  $\Delta s \in \text{im}(\mathbf{A}^\top)$ . Strongly polynomial computability follows since matrix inversion can be done in strongly polynomial time.

For the affine scaling step-length, note that  $\lceil a \rceil$  can be computed strongly polynomially, using  $O(\log n)$  comparisons as long as  $a = \text{poly}(n)$ . Thus, we can compute  $a = \frac{\beta}{\lceil 2\sqrt{n} \rceil}$ ,  $r = \|\Delta x^a \Delta s^a\|_\infty$  and  $b = 1 - \lceil 2 \|\Delta x^a \Delta s^a\| / r \rceil r / (\beta \bar{\mu}(z))$ . We then return  $\alpha^a := \max\{a, b\}$ . Since  $\lceil 2\sqrt{n} \rceil \in [2\sqrt{n}, 3\sqrt{n}]$  and  $\lceil 2 \|\Delta x^a \Delta s^a\| / r \rceil r \in [2 \|\Delta x^a \Delta s^a\|, 3 \|\Delta x^a \Delta s^a\|]$ , we have that  $\frac{\beta}{3\sqrt{n}} \leq a \leq \frac{\beta}{2\sqrt{n}}$  and  $1 - \frac{3 \|\Delta x^a \Delta s^a\|}{\beta \bar{\mu}(z)} \leq b \leq 1 - \frac{2 \|\Delta x^a \Delta s^a\|}{\beta \bar{\mu}(z)}$ . Thus,  $\alpha^a$  satisfies the requirement.  $\square$

**Minimum-Norm Viewpoint** We introduce some useful notation for the algorithm, and derive the minimum-norm interpretation of the affine scaling steps.

**Definition 2.40** (Normalized Iterates, Gap Vector and Subspaces). For  $z = (x, s) \in \mathcal{P}_{++} \times \mathcal{D}_{++}$ , we let

$$\begin{aligned} \hat{\xi}(z) &:= \sqrt{\frac{xs}{\bar{\mu}(z)}} \in \mathbb{R}^n, \\ \hat{x} &:= x \hat{\xi}(z)^{-1} = \sqrt{\frac{x \bar{\mu}(z)}{s}} \in \mathbb{R}^n, \\ \hat{s} &:= s \hat{\xi}(z)^{-1} = \sqrt{\frac{s \bar{\mu}(z)}{x}} \in \mathbb{R}^n. \end{aligned} \tag{21}$$

We call  $\hat{\xi}(z)$  the *normalized gap vector* and simply use  $\hat{\xi}$  when clear from the context. We call  $\hat{x}$  and  $\hat{s}$  the *normalized primal and dual iterates*, respectively. We also define the *normalized subspaces*

$$\hat{W} := \hat{x}^{-1} W \quad \text{and} \quad \hat{W}^\perp := \hat{s}^{-1} W^\perp.$$

If  $z = (x, s)$  falls on the central path, that is,  $xs = \bar{\mu}(z)\mathbf{1}$ , then  $\hat{\xi}(z) = \mathbf{1}$ ,  $\hat{x} = x$  and  $\hat{s} = s$ . The variables  $\hat{x}$  and  $\hat{s}$  represent natural adjustments for points off the central path. The next statement is immediate from the definitions, using  $\hat{x}\hat{s} = \bar{\mu}(z)\mathbf{1}$ .

**Proposition 2.41.** *The subspaces  $\hat{x}^{-1}W$  and  $\hat{s}^{-1}W^\perp$  are orthogonal.*

The following is a simple corollary of Proposition 2.31.

**Proposition 2.42.** *For  $z = (x, s) \in \mathcal{N}^2(\beta)$  for  $\beta \in (0, 1)$ , we have  $\|\xi\| = \sqrt{n}$ . Moreover,*

$$\begin{aligned} \sqrt{1-\beta}\mathbf{1} \leq \hat{\xi} \leq \sqrt{1+\beta}\mathbf{1}, \\ \sqrt{1-\beta}\hat{x} \leq x \leq \sqrt{1+\beta}\hat{x} \quad \text{and} \quad \sqrt{1-\beta}\hat{s} \leq s \leq \sqrt{1+\beta}\hat{s}. \end{aligned}$$

We will frequently use the rescaled subspaces  $\hat{x}^{-1}W$  and  $\hat{s}^{-1}W^\perp$  that correspond to using the local geometry at the point  $z = (x, s)$ . Throughout, we will refer to  $\|\hat{x}^{-1}w\|$  and  $\|\hat{s}^{-1}w\|$  as the *primal and dual local norms* of the vector  $w \in \mathbb{R}^n$  at the point  $z = (x, s) \in \mathcal{P}_{++} \times \mathcal{D}_{++}$ .

Equation (20) for the predictor step ( $\nu = 0$ ) with update direction  $(\Delta x^a, \Delta s^a)$  can be written as

$$x^{-1}\Delta x^a + x^{-1}\Delta s^a = -\mathbf{1}, \quad (22)$$

or equivalently,

$$\hat{x}^{-1}\Delta x^a + \hat{s}^{-1}\Delta s^a = -\hat{\xi}, \quad (23)$$

which serves the purpose that now  $\hat{x}^{-1}\Delta x^a \in \hat{W}$  and  $\hat{s}^{-1}\Delta s^a \in \hat{W}^\perp$  are orthogonal vectors (Proposition 2.41). Thus,  $\hat{x}^{-1}\Delta x^a$  and  $\hat{s}^{-1}\Delta s^a$  give an orthogonal decomposition of  $-\hat{\xi}$ . This leads to the following formulas:

$$\begin{aligned} \Delta x^a &= -\hat{x}\Pi_{\hat{W}}(\hat{\xi}), \\ \Delta s^a &= -\hat{s}\Pi_{\hat{W}^\perp}(\hat{\xi}). \end{aligned} \quad (24)$$

Equivalently, we can see  $\Delta z^a = (\Delta x^a, \Delta s^a)$  as the optimal solutions of the following minimum-norm problems:

$$\begin{aligned} \Delta x^a &= \hat{x} \arg \min_{\delta \in \hat{W}} \|\hat{\xi} + \delta\|, \\ \Delta s^a &= \hat{s} \arg \min_{\delta \in \hat{W}^\perp} \|\hat{\xi} + \delta\|. \end{aligned} \quad (25)$$

We can rewrite these equivalently as projections in  $W$  and  $W^\perp$ , noting that  $\hat{x}^{-1}x = \hat{s}^{-1}s = \hat{\xi}$ .

$$\begin{aligned} \Delta x^a &= \arg \min_{\Delta x \in W} \|\hat{x}^{-1}(x + \Delta x)\|, \\ \Delta s^a &= \arg \min_{\Delta s \in W^\perp} \|\hat{s}^{-1}(s + \Delta s)\|. \end{aligned} \quad (26)$$

We will require the following generic monotonicity estimate in terms of local norms. The next lemma is an  $\ell_1$ -variant of Lemma 2.30, and the proof is implicit in the proof of [66, Lemma 16].

**Lemma 2.43.** *Let  $z = (x, s) \in \mathcal{N}(\beta)$ ,  $\beta \in (0, 1)$ , and  $z' = (x', s') \in \mathcal{P} \times \mathcal{D}$ . Then, we have that*

$$\|(\hat{x}^{-1}x', \hat{s}^{-1}s')\|_1 \leq \frac{n}{\sqrt{1-\beta}} \left(1 + \frac{\bar{\mu}(z')}{\bar{\mu}(z)}\right). \quad (27)$$

*Proof.*

$$\begin{aligned} \|(\hat{x}^{-1}x', \hat{s}^{-1}s')\|_1 &= \frac{\|(\hat{s}x', \hat{x}s')\|_1}{\bar{\mu}(z)} \leq \frac{1}{\sqrt{1-\beta}} \frac{\|(sx', xs')\|_1}{\bar{\mu}(z)} = \frac{1}{\sqrt{1-\beta}} \frac{\langle s, x' \rangle + \langle x, s' \rangle}{\bar{\mu}(z)} \\ &= \frac{1}{\sqrt{1-\beta}} \frac{\langle s, x \rangle + \langle x', s' \rangle}{\bar{\mu}(z)} = \frac{1}{\sqrt{1-\beta}} \frac{n\bar{\mu}(z) + n\bar{\mu}(z')}{\bar{\mu}(z)} \\ &= \frac{n}{\sqrt{1-\beta}} \left(1 + \frac{\bar{\mu}(z')}{\bar{\mu}(z)}\right), \end{aligned}$$

where the first inequality follows from Proposition 2.42.  $\square$

**Step-Length Estimates** We will also need good estimates on the size on predictor steps beyond affine scaling. Our main estimate in this regard is given below. We use the latter part to compute the step-length associated with layered least squares steps. Note that the algorithm outputs a step-length of 0 in case the requirements of the step are not satisfied.

**Proposition 2.44** (Step-length Estimate for General Directions). *Let  $z = (x, s) \in \mathcal{N}^2(\beta/2)$ ,  $\mu := \bar{\mu}(z)$ ,  $\beta \in (0, 1/6]$ . Consider directions  $\Delta x \in W$ ,  $\Delta s \in W^\perp$  that satisfy*

$$\delta := \left\| \frac{\Delta x \Delta s}{\mu} \right\| \leq \frac{\beta}{9}, \quad \varepsilon := \left\| \frac{(x + \Delta x)(s + \Delta s)}{\mu} \right\| \leq \frac{\beta}{9}. \quad (28)$$

Then  $(x + \alpha \Delta x, s + \alpha \Delta s) \in \bar{\mathcal{N}}^2(\beta)$  and  $\bar{\mu}(x + \alpha \Delta x, s + \alpha \Delta s) \in [1 \pm 1/8](1 - \alpha)\mu$ , for all  $0 \leq \alpha \leq 1 - \frac{8\varepsilon}{\beta}$ . Furthermore, given  $\Delta x \in \mathbb{R}^n$ ,  $\Delta s \in \mathbb{R}^n$ ,  $\mu > 0$ ,  $\beta > 0$ , one can in strongly polynomial time output a step-length  $\alpha^\ell \in [0, 1]$  satisfying  $1 - \frac{9\varepsilon}{\beta} \leq \alpha^\ell \leq 1 - \frac{8\varepsilon}{\beta}$  if  $\max\left\{\left\|\frac{\Delta x \Delta s}{\mu}\right\|, \left\|\frac{(x + \Delta x)(s + \Delta s)}{\mu}\right\|\right\} \leq \beta/9$  and  $\alpha^\ell = 0$  otherwise.

*Proof.* Let  $z_\alpha := (x + \alpha \Delta x, s + \alpha \Delta s)$  for  $0 \leq \alpha \leq 1 - \frac{8\varepsilon}{\beta}$ ,  $\alpha < 1$ . (Note that  $\alpha < 1$  only becomes relevant in case  $\varepsilon = 0$ ; we will discuss this case later.) We first bound the centrality error using the estimate  $(1 - \alpha)\mu$  for  $\mu(z_\alpha)$  as follows:

$$\begin{aligned} & \left\| \frac{(x + \alpha \Delta x)(s + \alpha \Delta s)}{(1 - \alpha)\mu} - \mathbf{1} \right\| \\ &= \left\| \frac{(1 - \alpha)xs + \alpha(x + \Delta x)(s + \Delta s) - \alpha(1 - \alpha)\Delta x \Delta s}{(1 - \alpha)\mu} - \mathbf{1} \right\| \\ &\leq \left\| \frac{xs}{\mu} - \mathbf{1} \right\| + \frac{\alpha}{1 - \alpha} \left\| \frac{(x + \Delta x)(s + \Delta s)}{\mu} \right\| + \alpha \left\| \frac{\Delta x \Delta s}{\mu} \right\| \\ &\leq \beta/2 + \frac{\alpha}{1 - \alpha} \varepsilon + \alpha \delta \leq \beta/2 + \beta/8 + \beta/9 < \frac{3}{4}\beta, \end{aligned}$$

where the last inequality follows since  $\frac{\alpha}{1 - \alpha} \varepsilon \leq \beta/8$  for  $0 \leq \alpha \leq 1 - 8\varepsilon/\beta$  and  $\alpha < 1$  (needed to ensure the denominator  $(1 - \alpha)$  is positive).

By Lemma 2.33 and the above bound, we get that  $\bar{\mu}(z_\alpha) \in [1 \pm \frac{3}{4}\beta/\sqrt{n}](1 - \alpha)\mu \subseteq [1 \pm \frac{1}{8}](1 - \alpha)\mu$ , and  $z_\alpha \in \mathcal{N}^2\left(\frac{\frac{3}{4}\beta}{1 - \frac{3}{4}\beta}\right) \subseteq \mathcal{N}^2(\beta)$ , for  $\beta \in (0, 1/6]$ . If  $\varepsilon = 0$ , letting  $\alpha \rightarrow 1^-$ , we conclude by continuity that  $z_1 := (x + \Delta x, s + \Delta s) \in \bar{\mathcal{N}}^2(\beta)$  and  $\bar{\mu}(z_1) = 0$ , as needed.

For the last part, note that the condition  $\max\left\{\left\|\frac{\Delta x \Delta s}{\mu}\right\|, \left\|\frac{(x + \Delta x)(s + \Delta s)}{\mu}\right\|\right\} \leq \beta/9$  can be checked in strongly polynomial time by squaring both sides. If this check fails, output  $\alpha^\ell = 0$ . Otherwise, compute  $r = \|(x + \Delta x)(s + \Delta s)\|_\infty$  and compute  $\nu = \lceil 8 \|(x + \Delta x)(s + \Delta s)\| / r \rceil \in [8, \lceil 8\sqrt{n} \rceil]$  via binary search in  $O(\log n)$  time, and return  $\alpha^\ell = 1 - \frac{\nu r}{\mu \beta}$ . For correctness, note that  $\|(x + \Delta x)(s + \Delta s)\| \leq \nu(r/8) \leq \frac{9}{8}\|(x + \Delta x)(s + \Delta s)\|$ , and thus the desired inequalities follow recalling that  $\varepsilon := \|(x + \Delta x)(s + \Delta s)\| / \mu$ .  $\square$

### 3 Polarization of the Central Path

We now introduce the notion of polarized segments of the central path. For  $0 \leq \mu_1 \leq \mu_0$ , the central path segment between these values is denoted by

$$\text{CP}[\mu_1, \mu_0] := \{z^{\text{CP}}(\mu) : \mu_1 \leq \mu \leq \mu_0\}. \quad (29)$$

**Definition 3.1** (Polarization). For  $\gamma \in (0, 1]$  and  $\mu_0 > \mu_1 \geq 0$ , we say that the segment  $\text{CP}[\mu_1, \mu_0]$  is  $\gamma$ -polarized if there exists a partition  $B \cup N = [n]$  such that for all  $\mu \in [\mu_1, \mu_0]$ :

$$\begin{aligned} x_i^{\text{CP}}(\mu) &\geq \gamma x_i^{\text{CP}}(\mu_0), \quad \forall i \in B, \\ s_i^{\text{CP}}(\mu) &\geq \gamma s_i^{\text{CP}}(\mu_0), \quad \forall i \in N. \end{aligned}$$

**Remark 3.2.** By continuity of the central path and the condition  $\mu_0 > \mu_1 \geq 0$ , we may restrict the polarization check above to  $\mu \in [\mu_1, \mu_0]$ ,  $\mu > 0$ . For  $i \in N$ , since  $s_i^{\text{CP}}(\mu) = \mu/x_i^{\text{CP}}(\mu)$  for  $\mu > 0$ , the condition  $s_i^{\text{CP}}(\mu) \geq \gamma s_i^{\text{CP}}(\mu_0)$  is equivalent to

$$\mu/x_i^{\text{CP}}(\mu) \geq \gamma \mu_0/x_i^{\text{CP}}(\mu_0) \quad \Leftrightarrow \quad x_i(\mu) \leq \frac{\mu}{\gamma \mu_0} x_i^{\text{CP}}(\mu_0). \quad (30)$$

Thus, the polarization condition can be stated only in terms of the primal central path. Similarly, switching the roles of  $B$  and  $N$ , the polarization condition can also be stated only in terms of the dual central path.

**Remark 3.3.** As stated, the notion of polarization requires an inequality to hold for all  $\mu \in [\mu_1, \mu_0]$ . At the cost of losing a factor  $n$  however, it is in fact sufficient to check the polarization condition only at  $\mu = \mu_1$ . This follows by the near-monotonicity of the central path (Lemma 2.30):

$$\frac{x_i^{\text{CP}}(\mu)}{x_i^{\text{CP}}(\mu_0)} = \frac{x_i^{\text{CP}}(\mu)}{x_i^{\text{CP}}(\mu_1)} \cdot \frac{x_i^{\text{CP}}(\mu_1)}{x_i^{\text{CP}}(\mu_0)} \geq \frac{1}{n} \cdot \frac{x_i^{\text{CP}}(\mu_1)}{x_i^{\text{CP}}(\mu_0)}, \quad \forall i \in [n],$$

The same is true for  $s(\mu)$  by a symmetric argument.

As a direct consequence of the definition together with near-monotonicity, we deduce the following crucial corollary:

**Corollary 3.4.** *Let  $\text{CP}[\mu_1, \mu_0]$ ,  $0 \leq \mu_1 \leq \mu_0$ , be  $\gamma$ -polarized with respect to the partition  $B \cup N = [n]$ . Then, for all  $\mu \in [\mu_1, \mu_0]$ , the following holds:*

- (1)  $\gamma x_i(\mu_0) \leq x_i(\mu) \leq n x_i(\mu_0)$ ,  $i \in B$ .
- (2)  $\gamma s_i(\mu_0) \leq s_i(\mu) \leq n s_i(\mu_0)$ ,  $i \in N$ .
- (3)  $\frac{\mu}{n\mu_0} x_i(\mu_0) \leq x_i(\mu) \leq \frac{\mu}{\gamma\mu_0} x_i(\mu_0)$ ,  $i \in N$ .
- (4)  $\frac{\mu}{n\mu_0} s_i(\mu_0) \leq s_i(\mu) \leq \frac{\mu}{\gamma\mu_0} s_i(\mu_0)$ ,  $i \in B$ .

*Proof.* The first inequalities in (1) and (2) are the definition of  $\gamma$ -polarization and the second inequalities are from Lemma 2.30. (3) and (4) are equivalent to (1) and (2) using the central path relations  $x(\mu_0)s(\mu_0) = \mu_0 \mathbf{1}$  and  $x(\mu)s(\mu) = \mu \mathbf{1}$ .  $\square$

Section 7 introduces the algorithm **SLLS-IPM** that can traverse  $\gamma$ -polarized segments in  $O(n^{1.5} \log(n/\gamma))$  iterations. Theorem 1.1 follows by combining this algorithm with the following decomposition result that is the main result of this section; the proof can be found in Section 7.3. The stronger variant Theorem 1.4 is proved in Section 8 using an additional amortization argument.

**Theorem 3.5.** *Let  $\Gamma : (\mu_1, \mu_0) \rightarrow \overline{\mathcal{N}}^{-\infty}(\theta)$ ,  $\theta \in (0, 1)$ ,  $0 \leq \mu_1 < \mu_0 \leq \infty$ , be a piecewise linear curve satisfying  $\bar{\mu}(\Gamma(\mu)) = \mu$ ,  $\forall \mu \in (\mu_1, \mu_0)$  consisting of  $T$  linear segments. Then,  $\text{CP}[\mu_1, \mu_0]$  can be decomposed into  $T$  segments that are  $\frac{(1-\theta)^2}{16n^3}$ -polarized.*

Theorem 3.5 is a direct consequence of the following key lemma.

**Lemma 3.6.** *For  $\theta \in (0, 1)$ , let  $[z^{(0)}, z^{(1)}] \subseteq \overline{\mathcal{N}}^{-\infty}(\theta)$ ,  $\bar{\mu}(z^{(0)}) > \bar{\mu}(z^{(1)})$ . Then,  $\text{CP}[\bar{\mu}(z^{(1)}), \bar{\mu}(z^{(0)})]$  is  $\frac{(1-\theta)^2}{16n^3}$ -polarized.*

*Proof of Theorem 3.5.* By assumption, the curve  $\Gamma([\mu_1, \mu_0]) = \cup_{i=1}^T [z^{(i-1)}, z^{(i)}] \subseteq \overline{\mathcal{N}}^{-\infty}(\theta)$ , where  $\mu_0 = \bar{\mu}(z^{(0)}) > \bar{\mu}(z^{(1)}) > \dots > \bar{\mu}(z^{(T)}) = \mu_1$ . By Lemma 3.6, each segment  $\text{CP}[\bar{\mu}(z^{(i)}), \bar{\mu}(z^{(i-1)})]$ ,  $i \in [T]$ , is therefore  $\frac{(1-\theta)^2}{16n^3}$ -polarized with respect to some polarization partition  $(B^{(i)}, N^{(i)})$ . This proves the theorem.  $\square$

It remains to prove Lemma 3.6. The proof requires the following simple technical lemma that allows us to relate approximate centrality along lines to polarization.

**Lemma 3.7.** *For any  $u, v > 0$ ,*

$$\min_{\alpha \in [0, 1]} \frac{(1 - \alpha + \alpha u)(1 - \alpha + \alpha v)}{1 - \alpha + \alpha uv} = \min \left\{ 1, \left( \frac{\sqrt{u} + \sqrt{v}}{1 + \sqrt{uv}} \right)^2 \right\} \leq 2(u + v). \quad (31)$$

*Proof.* To show the equality, let  $\mu := uv$ . Note that

$$\begin{aligned} \min_{\alpha \in [0, 1]} \frac{(1 - \alpha + \alpha u)(1 - \alpha + \alpha v)}{1 - \alpha + \alpha uv} &= \min_{\alpha \in [0, 1]} \frac{(1 - \alpha)^2 + \alpha^2 \mu + \alpha(1 - \alpha)(u + v)}{1 - \alpha + \alpha \mu} \\ &= 1 + \min_{\alpha \in [0, 1]} (u + v - (1 + \mu)) \frac{\alpha(1 - \alpha)}{1 - \alpha + \alpha \mu}. \end{aligned} \quad (32)$$

**Case I:**  $\left(\frac{\sqrt{u}+\sqrt{v}}{1+\sqrt{uv}}\right)^2 \geq 1$ . In this case, we need to show that the minimum of the expression is 1. It is easy to see that the condition equivalent to  $u+v \geq 1+uv = 1+\mu$ . Thus, the minimum value of (32) is clearly 1, attained at  $\alpha \in \{0, 1\}$ .

**Case II:**  $\left(\frac{\sqrt{u}+\sqrt{v}}{1+\sqrt{uv}}\right)^2 < 1$ , or equivalently,  $u+v < 1+\mu$ . In this case, the minimizer of (32) corresponds to the maximizer of  $\frac{\alpha(1-\alpha)}{1-\alpha+\alpha\mu}$ . This function takes value 0 at  $\alpha \in \{0, 1\}$  and is strictly positive for  $0 < \alpha < 1$ . Furthermore, the unique critical point in the interval  $[0, 1]$  occurs at  $\alpha^* = \frac{1}{1+\sqrt{\mu}}$ , which is thus the maximizer. The minimum value of (32) is therefore

$$1 + (u+v - (1+\mu)) \frac{\alpha^*(1-\alpha^*)}{1-\alpha^*+\alpha^*\mu} = \frac{u+v+2\sqrt{\mu}}{(1+\sqrt{\mu})^2},$$

as required. The inequality in the statement follows easily as

$$\left(\frac{\sqrt{u}+\sqrt{v}}{1+\sqrt{uv}}\right)^2 \leq (\sqrt{u}+\sqrt{v})^2 = 2(u+v) - (\sqrt{u}-\sqrt{v})^2.$$

□

*Proof of Lemma 3.6.* For  $\alpha \in [0, 1]$ , let  $z^{(\alpha)} := (x^{(\alpha)}, s^{(\alpha)}) := (1-\alpha)z^{(0)} + \alpha z^{(1)}$ . By Proposition 2.29, we first note that the normalized gap function  $\bar{\mu}(z)$  is in fact linear on  $[z^{(0)}, z^{(1)}]$ . That is,

$$\mu_\alpha := \bar{\mu}(z^{(\alpha)}) = (1-\alpha)\bar{\mu}(z^{(0)}) + \alpha\bar{\mu}(z^{(1)}).$$

For any  $i \in [n]$ ,  $z^{(\alpha)} \in \mathcal{N}^{-\infty}(\theta)$  implies

$$\begin{aligned} \frac{x_i^{(\alpha)} s_i^{(\alpha)}}{(1-\alpha)x_i^{(0)} s_i^{(0)} + \alpha x_i^{(1)} s_i^{(1)}} &\geq \frac{(1-\theta)\mu_\alpha}{(1-\alpha)x_i^{(0)} s_i^{(0)} + \alpha x_i^{(1)} s_i^{(1)}} \\ &\geq \frac{(1-\theta)\mu_\alpha}{n((1-\alpha)\bar{\mu}(z^{(0)}) + \alpha\bar{\mu}(z^{(1)}))} = \frac{1-\theta}{n}. \end{aligned}$$

Note that the above expression is the same as in Lemma 3.7 for  $u = x_i^{(1)}/x_i^{(0)}$ ,  $v = s_i^{(1)}/s_i^{(0)}$ . Since the bound is true for any  $\alpha \in [0, 1]$ , the Lemma implies

$$\frac{1-\theta}{n} \leq 2 \left( \frac{x_i^{(1)}}{x_i^{(0)}} + \frac{s_i^{(1)}}{s_i^{(0)}} \right).$$

Let

$$B := \left\{ i \in [n] : \frac{x_i^{(1)}}{x_i^{(0)}} \geq \frac{s_i^{(1)}}{s_i^{(0)}} \right\}, \quad N := [n] \setminus B.$$

Then,  $x_i^{(1)}/x_i^{(0)} \geq \frac{1-\theta}{4n}$ , for all  $i \in B$ , and  $s_i^{(1)}/s_i^{(0)} \geq \frac{1-\theta}{4n}$  for all  $i \in N$ .

For any  $\alpha \in [0, 1]$  and  $i \in B$ ,

$$\frac{x_i^{(\alpha)}}{x_i^{(0)}} = (1-\alpha) + \alpha \frac{x_i^{(1)}}{x_i^{(0)}} \geq \min \left\{ 1, \frac{1-\theta}{4n} \right\} = \frac{1-\theta}{4n}.$$

Similarly, for  $i \in N$ ,  $s_i^{(\alpha)}/s_i^{(0)} \geq \frac{1-\theta}{4n}$ .

For the central path point  $z^{\text{CP}}(\mu_\alpha) = (x^{\text{CP}}(\mu_\alpha), s^{\text{CP}}(\mu_\alpha))$  at  $\mu_\alpha$ , the bounds in Lemma 2.34 relating points in a neighborhood with central path points give

$$\frac{x_i^{\text{CP}}(\mu_\alpha)}{x_i^{\text{CP}}(\mu_0)} \geq \frac{x_i^{(\alpha)}/(2n)}{\frac{2n}{1-\theta} x_i^{(0)}} \geq \frac{(1-\theta)^2}{16n^3}, \quad \forall i \in B.$$

By a similar argument, we also have  $s_i^{\text{CP}}(\mu_\alpha)/s_i^{\text{CP}}(\mu_0) \geq \frac{(1-\theta)^2}{16n^3}$ ,  $\forall i \in N$ . Thus,  $\text{CP}[\mu_1, \mu_0]$  is  $\frac{(1-\theta)^2}{16n^3}$ -polarized. □

## 4 The Max Central Path

In this section, we derive key properties of the max central path and how to use the max central path to decompose the central path into polarized segments. In particular, we prove Lemma 1.2, Lemma 1.5 and Theorem 1.8. Given  $g \geq 0$ , we denote by

$$\begin{aligned}\mathcal{P}_g &:= \{x \in \mathbb{R}^n : \mathbf{A}x = b, x \geq \mathbf{0}, \langle c, x \rangle \leq v^* + g\}, \\ \mathcal{D}_g &:= \{s \in \mathbb{R}^n : \exists y \in \mathbb{R}^m \mathbf{A}^\top y + s = c, s \geq \mathbf{0}, \langle b, y \rangle \geq v^* - g\}.\end{aligned}$$

We will also use the subspace formulation (11) with  $d \in \mathbb{R}^n$  such  $\mathbf{A}d = b$ ; in these terms, we can write  $\mathcal{P}_g = \{x \in \mathbb{R}^n : x \in W + d, x \geq \mathbf{0}, \langle c, x \rangle \leq v^* + g\}$  and  $\mathcal{D}_g = \{s \in \mathbb{R}^n : s \in W^\top + c, s \geq \mathbf{0}, \langle d, c - s \rangle \geq v^* - g\}$ . the feasible sets of the linear programs in (4). They correspond to the sets of the primal and dual feasible points  $(x, s) \in \mathcal{P} \times \mathcal{D}$  with objective value within  $g$  from the optimum  $v^*$ , respectively. By the assumption that (LP) is feasible and bounded, it follows that  $\mathcal{P}_g$  and  $\mathcal{D}_g$  are both non-empty for any  $g \geq 0$ .

We recall that the duality gap of any pair  $(x, (y, s))$  of primal-dual feasible points of (LP) fulfills  $\langle c, x \rangle - \langle b, y \rangle = \langle x, s \rangle$ . In particular, we have  $\langle x, s^* \rangle = \langle c, x \rangle - v^*$  and  $\langle x^*, s \rangle = v^* - \langle b, y \rangle$ . Thus, the two sets  $\mathcal{P}_g$  and  $\mathcal{D}_g$  are equivalently given by

$$\mathcal{P}_g = \{x \in \mathcal{P} : \langle x, s^* \rangle \leq g\}, \quad \mathcal{D}_g = \{s \in \mathcal{D} : \langle x^*, s \rangle \leq g\}.$$

These expressions are in fact independent of the choice of optimal solutions  $(x^*, s^*)$ . The following claim is immediate by our assumption that  $\mathcal{P}_{++}$  and  $\mathcal{D}_{++}$  are non-empty.

**Proposition 4.1.** *For all  $g \geq 0$ , the sets  $\mathcal{P}_g$  and  $\mathcal{D}_g$  are bounded.*

*Proof.* We restrict to the proof of the boundedness of  $\mathcal{P}_g$ , since the proof is analogous for  $\mathcal{D}_g$ . Let  $s^\circ \in \mathcal{D}_{++}$  be a strictly feasible point of the dual, and  $x \in \mathcal{P}_g$ . By Proposition 2.28, we have

$$\langle x, s^* \rangle + \langle x^*, s^\circ \rangle = \langle x, s^\circ \rangle + \langle x^*, s^* \rangle.$$

Since  $\langle x^*, s^* \rangle = 0$ , we deduce that  $\langle x, s^\circ \rangle \leq g + \langle x^*, s^\circ \rangle$ . As  $s^\circ > \mathbf{0}$ , this implies that  $x_i \leq (g + \langle x^*, s^\circ \rangle) / s_i^\circ$  for all  $i \in [n]$ .  $\square$

We denote by  $\text{MCP} = \{z^m(g) : g \geq 0\}$  the whole max central path. The max central path point  $z^m(g) = (x^m(g), s^m(g))$  is the entry-wise maximum of the set  $\mathcal{P}_g \times \mathcal{D}_g$ .

While the points of the max central path are not feasible in general, the following theorem shows that the max central path shares important similarities with the central path:

**Theorem 4.2** (Centrality of the max central path). *For all  $g \geq 0$ , we have that*

$$g \leq x_i^m(g) s_i^m(g) \leq 2g \quad \forall i \in [n].$$

*Proof.* We first prove the upper bound. For  $i \in [n]$ , let  $x^{(i)} \in \arg \max\{x_i : x \in \mathcal{P}_g\}$  and  $s^{(i)} \in \arg \max\{s_i : s \in \mathcal{D}_g\}$ . Note that  $x^{(i)}, s^{(i)}$  exist by Proposition 4.1. Then,

$$x_i^m(g) s_i^m(g) = x_i^{(i)} s_i^{(i)} \leq \langle x^{(i)}, s^{(i)} \rangle = \langle x^{(i)}, s^* \rangle + \langle x^*, s^{(i)} \rangle \leq 2g,$$

where the last equality follows from Proposition 2.28. We now prove the lower bound. We assume  $g > 0$ , since the statement is trivial otherwise.

Note that the dual program of  $\max\{x_i : x \in W + d, x \geq \mathbf{0}, \langle x, s^* \rangle \leq g\}$  can be expressed as

$$\min \{\alpha g + \langle u, x^* \rangle : \alpha s^* + u \geq e^i, u \in W^\perp, \alpha \geq 0\},$$

using that  $\langle u, x^* \rangle = \langle u, d \rangle$  since  $d - x^* \in W$ ,  $u \in W^\perp$ . Similarly, the dual program of  $\max\{s_i : s \in W^\perp + c, s \geq \mathbf{0}, \langle s, x^* \rangle \leq g\}$  can be expressed as

$$\min \{\beta g + \langle v, s^* \rangle : \beta x^* + v \geq e^i, v \in W, \beta \geq 0\}.$$

Let us pick optimal  $(\alpha, u)$  and  $(\beta, v)$  to these two programs. The product of the objective values is thus equal to  $x_i^m(g) s_i^m(g)$ ; the proof is complete by showing a lower bound  $g$ .

We first claim that

$$\langle u, x^* \rangle \geq 0 \quad \text{and} \quad \langle v, s^* \rangle \geq 0. \quad (33)$$

By symmetry, it suffices to prove the first claim. Recall  $x^* \geq 0$ ; we show that whenever  $x_j^* > 0$  then we must have  $u_j \geq 0$ . This follows since by complementary,  $s_j^* = 0$ , and that we have the constraint  $\alpha s_j^* + u_j \geq e_j^i$ .

Next, note that the constraints in the two programs imply

$$1 = \langle e^i, e^i \rangle \leq \langle \alpha s^* + u, \beta x^* + v \rangle = \alpha \langle v, s^* \rangle + \beta \langle u, x^* \rangle. \quad (34)$$

Now, the product of the objective values can be written as

$$\begin{aligned} x_i^m(g) s_i^m(g) &= (\alpha g + \langle u, x^* \rangle)(\beta g + \langle v, s^* \rangle) \\ &= \alpha \beta g^2 + g(\alpha \langle v, s^* \rangle + \beta \langle u, x^* \rangle) + \langle u, x^* \rangle \cdot \langle v, s^* \rangle \geq g, \end{aligned}$$

using (33) and (34). This concludes the proof.  $\square$

Given the above, we are ready to show

**Lemma 1.2.** *For every  $\mu > 0$  and the central path point  $z^{\text{CP}}(\mu) = (x^{\text{CP}}(\mu), s^{\text{CP}}(\mu))$ ,*

$$\frac{z^m(n\mu)}{2n} \leq z^{\text{CP}}(\mu) \leq z^m(n\mu).$$

*Proof.* Recall that  $z^{\text{CP}}(\mu) = (x^{\text{CP}}(\mu), s^{\text{CP}}(\mu))$  with

$$\langle x^{\text{CP}}(\mu), s^{\text{CP}}(\mu) \rangle = \langle x^{\text{CP}}(\mu), s^* \rangle + \langle x^*, s^{\text{CP}}(\mu) \rangle = n\mu$$

using Proposition 2.28. Therefore,  $x^{\text{CP}}(\mu) \in \mathcal{P}_{n\mu}$  and  $s^{\text{CP}}(\mu) \in \mathcal{D}_{n\mu}$ . By definition of the max central path,  $z^{\text{CP}}(\mu) = (x^{\text{CP}}(\mu), s^{\text{CP}}(\mu)) \leq (x^m(n\mu), s^m(n\mu)) = z^m(n\mu)$ . For the second inequality, note that

$$\begin{aligned} x^{\text{CP}}(\mu) &= \frac{\mu}{s^{\text{CP}}(\mu)} \geq \frac{\mu}{s^m(n\mu)} = \frac{\mu}{x^m(n\mu)s^m(n\mu)} x^m(n\mu) \\ &\stackrel{\text{Thm. 4.2}}{\geq} \frac{\mu}{2n\mu} \mathbf{1} x^m(n\mu) = \frac{x^m(n\mu)}{2n}. \end{aligned}$$

By a symmetric argument,  $s^{\text{CP}}(\mu) \geq s^m(n\mu)/2n$ .  $\square$

## 4.1 The Shadow Vertex Simplex Rule

Given a pointed polyhedron  $P \subseteq \mathbb{R}^n$  (that means that it has at least one vertex) and two objectives  $c^{(1)}, c^{(2)} \in \mathbb{R}^n$  such that  $\max_{x \in P} \langle c^{(1)}, x \rangle, \max_{x \in P} \langle c^{(2)}, x \rangle < \infty$ , we recall that the shadow vertex simplex rule consists in pivoting over vertices of  $P$  maximizing the objectives  $(1 - \lambda)c^{(1)} + \lambda c^{(2)}$  as  $\lambda$  goes from 0 to 1. More formally, a sequence of vertices  $v^{(1)}, \dots, v^{(k)} \in P$  is a  $(c^{(1)}, c^{(2)})$ -shadow vertex path on  $P$  if

- $[v^{(i)}, v^{(i+1)}]$  is an edge of  $P$ ,  $\forall i \in [k - 1]$ ,
- $\langle c^{(2)}, v^{(i)} \rangle < \langle c^{(2)}, v^{(i+1)} \rangle$ ,  $\forall i \in [k - 1]$ , and
- there exists  $0 = \lambda_0 < \lambda_1 \leq \dots \leq \lambda_{k-1} < \lambda_k = 1$  such that  $\forall i \in [k]$ ,  $\langle v^{(i)}, (1 - \alpha)c^{(1)} + \alpha c^{(2)} \rangle = \max_{x \in P} \langle x, (1 - \alpha)c^{(1)} + \alpha c^{(2)} \rangle$ ,  $\forall \alpha \in [\lambda_{i-1}, \lambda_i]$ .

To analyze shadow vertex paths further, we define the two-dimensional projection

$$P[c^{(1)}, c^{(2)}] := \left\{ \left( \langle c^{(1)}, x \rangle, \langle c^{(2)}, x \rangle \right) : x \in P \right\} = \left( c^{(1)}, c^{(2)} \right)^\top \cdot P.$$

The vertices of  $P[c^{(1)}, c^{(2)}]$  maximizing an open interval of objectives in  $(1 - \lambda)e^1 + \lambda e^2$ ,  $\lambda \in [0, 1]$  are precisely the projections of vertices  $v^{(i)}$ ,  $i \in [k]$ , on the shadow path such that  $\lambda_{i-1} < \lambda_i$ .

We define  $S_P(c^{(1)}, c^{(2)})$  as the number of vertices of  $P[c^{(1)}, c^{(2)}]$  maximizing an open interval of objectives in  $(1 - \lambda)e^1 + \lambda e^2$ ,  $\lambda \in [0, 1]$ . By the preceding observations, we have that  $S_P(c^{(1)}, c^{(2)})$  is a lower bound on the number of vertices of any  $(c^{(1)}, c^{(2)})$ -shadow vertex path.

In the above, we restricted both starting and ending objectives  $c^{(1)}, c^{(2)}$  to have finite objective value on  $P$ . It will be useful in the sequel to extend to the case where  $P$  might be unbounded in direction  $c^{(2)}$ . In this case, we define the shadow vertex path as above, with the only modification being that we let  $\lambda_k := \max\{\lambda \in [0, 1] : \max_{x \in P} \langle x, (1 - \lambda)c^{(1)} + \lambda c^{(2)} \rangle < \infty\}$ , that is, the simplex path stops just before reaching an unbounded ray for  $c^{(2)}$ . In this setting, note that  $S_P(c^{(1)}, c^{(2)})$  is still well-defined and continues to be a lower bound on the number of vertices on any  $c^{(1)}, c^{(2)}$  shadow vertex path.

We are ready to prove Lemma 1.5.

**Lemma 1.5.** *The following hold:*

(i)  $\forall i \in [n]$ ,  $g \mapsto x_i^m(g)$  is a piecewise linear concave non-decreasing function with  $S_{\mathcal{P}}(-s^*, e^i)$  pieces. That is,  $\text{SLC}_{0,i}^{\mathcal{P}}(0, \infty) = S_{\mathcal{P}}(-s^*, e^i)$ .

(ii)  $\forall i \in [n]$ ,  $g \mapsto s_i^m(g)$  is a piecewise linear concave non-decreasing function with  $S_{\mathcal{D}}(-x^*, e^i)$  pieces. That is,  $\text{SLC}_{0,i}^{\mathcal{D}}(0, \infty) = S_{\mathcal{D}}(-x^*, e^i)$ .

*Proof.* We only prove part (i); part (ii) follows analogously. For  $i \in [n]$ , let  $Q_i = P[s^*, e^i]$ . We note that  $x_i^m(g) = \max\{v_2 : (v_1, v_2) \in Q_i, v_1 \leq g\}$ . In particular, the map  $x_i^m(g)$  is a non-decreasing function of  $g$ . Moreover, it is easy to verify that  $x_i^m(g)$  is concave.

Again by definition,  $S_{\mathcal{P}}(-s^*, e^i)$  equals the number of vertices of  $Q_i$  maximizing an open interval of objectives in  $\mathcal{O} := \{-(1-\lambda)e^1 + \lambda e^2 : \lambda \in [0, 1]\} \subset \mathbb{R}^2$ .

Define  $\bar{u}_i(g) := \sup\{v_2 : (g, v_2) \in Q_i\}$ , which is defined to equal  $-\infty$  if  $\{(g, v_2) \in Q_i\} = \emptyset$ . By Proposition 4.1, note that  $\bar{u}_i(g) < \infty$  for all  $g \geq 0$ . By convexity of  $Q_i$ ,  $\bar{u}_i$  is concave function on  $\mathbb{R}_+$ . Let  $h = \sup\{\bar{u}_i(g) : g \geq 0\}$ .

Assume  $h = \infty$ . By concavity,  $\bar{u}_i$  must be a strictly increasing function on  $\mathbb{R}_+$ . In particular,  $\bar{u}_i(g) = x_i^m(g)$ . Given this, we see that the linear pieces of  $x_i^m(g)$  are in one to one correspondence with the edges of  $Q_i$  on the upper convex hull whose projection onto the  $e_1$ -axis have positive length (i.e., excluding the potential edge  $\{(0, v_2) \in Q_i\}$ ). Since  $Q_i \subseteq \mathbb{R}_+^2$ , every such edge can be uniquely associated with its left endpoint (which is always a vertex of  $Q_i$ ). It is now easy to check geometrically that the set of such endpoints exactly corresponds to the set of vertices that are maximizers of the objectives in an open interval of  $\mathcal{O}$ .

Assume  $h < \infty$ . Let  $g_h := \min\{g \geq 0 : x_i^m(g) = h\}$ . It is direct to see that  $x_i^m(g) = \bar{u}_i(g)$  if  $g \leq g_h$  and that  $x_i^m(g) = h$  for  $g \geq g_h$ . Furthermore,  $x_i^m(g)$  is strictly increasing on  $[0, g_h]$ . From this, it is easy to see geometrically that the number of linear pieces of  $x_i^m$  is one plus the number of edges of  $Q_i$  on the upper convex hull lying in the band  $\{(v_1, v_2) : 0 \leq v_1 \leq g\}$ , where the extra linear segment corresponds to constant segment between  $g_h$  and  $\infty$ . As in the previous case, these linear segments can be uniquely identified with their left endpoints, which correspond to vertices of  $Q_i$ . Furthermore, it is easy to check that these correspond to vertices of  $Q_i$  maximizing an open interval of objectives in  $\mathcal{O}$ .  $\square$

## 4.2 Decomposition of Straight Line Segments into Polarized Segments

In this section we prove Theorem 1.8. To begin, we provide some simple consequences of the concavity of the max central path coordinates.

**Lemma 4.3.** *Let  $g \geq 0$  and  $\alpha \geq 1$ . Then for all  $i \in [n]$  we have  $x_i^m(\alpha g) \leq \alpha x_i^m(g)$  and  $x_i^m(\frac{1}{\alpha}g) \geq \frac{1}{\alpha}x_i^m(g)$ .*

*Proof.* We use the concavity of  $g \mapsto x_i^m(g)$  (Lemma 1.5) and deduce that  $(1 - \frac{1}{\alpha})x_i^m(0) + \frac{1}{\alpha}x_i^m(\alpha g) \leq x_i^m(g)$ . The first inequality of the lemma follows from  $x_i^m(0) \geq 0$ . The second inequality follows from the first applied to  $\tilde{g} = \frac{g}{\alpha}$ .  $\square$

The next lemma shows that the definition of the primal straight-line complexity (Definition 1.3) can be restricted to breakpoints  $(g_k, z_k) = (g_k, x_i^m(g_k))$ . By symmetry, the analogous statement for the dual straight-line complexity also holds.

**Lemma 4.4.** *Let  $0 \leq \underline{g} < \bar{g}$ . We have*

$$\begin{aligned} \text{SLC}_{\theta,i}^{\mathcal{P}}(\underline{g}, \bar{g}) = \min \{p \geq 1 : \exists (g_k)_{k \in [p+1]} \in \mathbb{R}, \bar{g} = g_1 > g_2 > \dots > g_{p+1} = \underline{g}, \\ \forall k \in [p], \forall \lambda \in [0, 1], (1-\lambda)x_i^m(g_k) + \lambda x_i^m(g_{k+1}) \\ \geq (1-\theta)x_i^m((1-\lambda)g_k + \lambda g_{k+1})\} \end{aligned} \quad (35)$$

*Proof.* Let  $\bar{g} = g_1 > g_2 > \dots > g_{p+1} = \underline{g}$ . By the concavity of  $g \mapsto x_i^m(g)$ , the condition  $(1-\lambda)x_i^m(g_k) + \lambda x_i^m(g_{k+1}) \leq x_i^m((1-\lambda)g_k + \lambda g_{k+1})$  is satisfied for every  $k \in [p]$  and  $\lambda \in [0, 1]$ . We deduce that  $\text{SLC}_{\theta,i}(\underline{g}, \bar{g})$  is less than or equal to the right-hand side of (35).

We now prove the converse inequality. Let  $(g_1, z_1), \dots, (g_{p+1}, z_{p+1})$  as in the right-hand side of (5). Up to removing or shortening segments in the sequence  $\left( [(g_k, z_k), (g_{k+1}, z_{k+1})] \right)_{k \in [p]}$ , we can assume that  $\bar{g} = g_1 \geq g_k > g_{p+1} = \underline{g}$  for every  $k \in [p]$ . As  $z_k \leq x_i^m(g_k)$  for all  $k \in [p+1]$ , we have  $(1-\lambda)x_i^m(g_k) + \lambda x_i^m(g_{k+1}) \geq (1-\lambda)z_k + \lambda z_{k+1} \geq (1-\theta)x_i^m((1-\lambda)g_k + \lambda g_{k+1})$  for all  $k \in [p]$ . It remains us to show that, up to removing some  $g_k$ , we can ensure  $g_1 > g_2 > \dots > g_{p+1}$ . Suppose that there is  $k \in [p]$  such

that  $g_1 > \dots > g_k$  and  $g_{k+1} \geq g_k$ . We introduce the smallest integer  $l \geq k$  such that  $g_{l+1} < g_k$  ( $l$  is well-defined as  $g_{p+1} = \underline{g} < g_k$ ). By concavity of  $x_i^m$ , we have  $x_i^m(g_k) \geq (1 - \bar{\lambda})x_i^m(g_l) + \bar{\lambda}x_i^m(g_{l+1})$  for  $\bar{\lambda} = \frac{g_k - g_l}{g_{l+1} - g_l} \in [0, 1]$ . Then, for all  $\lambda \in [0, 1]$ ,

$$\begin{aligned} (1 - \lambda)x_i^m(g_k) + \lambda x_i^m(g_{l+1}) &\geq (1 - \lambda')x_i^m(g_l) + \lambda'x_i^m(g_{l+1}) \\ &\geq (1 - \theta)x_i^m((1 - \lambda')g_l + \lambda'g_{l+1}) \\ &= (1 - \theta)x_i^m((1 - \lambda)g_k + \lambda g_{l+1}) \end{aligned}$$

where  $\lambda' = \lambda + \bar{\lambda}(1 - \lambda) \in [0, 1]$ . By removing the  $g_{k+1}, \dots, g_l$  and recursively applying the same argument on the remaining sequence, we end up with a sequence  $\bar{g} = g_1 > \dots > g_{q+1} = \underline{g}$  as in the right-hand side of (35), with  $q \leq p$ . This shows that  $\text{SLC}_{\theta, i}^p(\underline{g}, \bar{g})$  is greater than or equal to the right-hand side of (35).  $\square$

In the next lemma, we show that the primal and dual straight line complexity for a single component of the maximum central path are related up to a factor 2, modulo a slight widening of the neighborhood. This shows that that primal and dual straight line complexities are essentially equivalent. The proof proceeds by decomposing straight line segments into polarized segments.

**Lemma 4.5.** *Let  $L := \text{SLC}_{1-\alpha, i}^p(\underline{g}, \bar{g})$ ,  $\alpha \in (0, 1]$ ,  $i \in [n]$ , and let  $\bar{g} = g_1 > g_2 > \dots > g_{L+1} = \underline{g}$  be the minimizer of (35). Define  $\phi_i : [\underline{g}, \bar{g}] \rightarrow \mathbb{R}_+$  so that  $\phi_i(g) = \max\{x_i^m(g_{j+1}), \frac{g}{g_j}x_i^m(g_j)\}$  for  $g \in [g_j, g_{j+1}]$ ,  $j \in [L]$ . Then,  $\phi_i(g)$  and  $\frac{g}{\phi_i(g)}$  are continuous piecewise linear functions on  $(\underline{g}, \bar{g}]$  with at most  $2L$  pieces, and*

$$x_i^m(g) \geq \phi_i(g) \geq \frac{\alpha}{2}x_i^m(g), \quad (36)$$

$$s_i^m(g) \geq \frac{\alpha g}{2\phi_i(g)} \geq \frac{\alpha}{4}s_i^m(g), \quad \forall g \in (\underline{g}, \bar{g}]. \quad (37)$$

Moreover,  $\text{SLC}_{1-\alpha/4, i}^d(\underline{g}, \bar{g}) \leq 2\text{SLC}_{1-\alpha, i}^p(\underline{g}, \bar{g})$ .

**Remark 4.6.** For  $\alpha \in (0, 1/2]$ , it can be shown that straight line complexity satisfies  $\text{SLC}_{1-\alpha, i}^d(\underline{g}, \bar{g}) \leq O(1)\text{SLC}_{1-\alpha/4, i}^d(\underline{g}, \bar{g})$ . That is, shrinking the neighborhood size by a constant factor can only change the straight line complexity of a single coordinate of the max central path by a constant factor. In particular, this implies that  $\text{SLC}_{1-\alpha, i}^d(\underline{g}, \bar{g}) \leq O(1)\text{SLC}_{1-\alpha, i}^p(\underline{g}, \bar{g})$  for  $\alpha \in (0, 1/2]$ . As this does not have a significant effect on our iteration bounds, we do not prove this here.

*Proof of Lemma 4.5.* Noting that  $x_i^m(g_j) \geq \frac{g_j}{g_{j-1}}x_i^m(g_{j-1})$ ,  $j \in \{2, \dots, L+1\}$ , by Lemma 4.3 and  $x_i^m(g_j) \geq x_i^m(g_{j+1})$ ,  $j \in [L]$ , by monotonicity, we have that  $\phi(g_j) = x_i^m(g_j)$ ,  $\forall j \in [L+1]$ . In particular,  $\phi_i$  is well-defined. Since each function  $\max\{x_i^m(g_{j+1}), \frac{g}{g_j}x_i^m(g_j)\}$ ,  $i \in [L]$ , is continuous piecewise linear with at most 2 pieces,  $\phi_i$  is also continuous with at most  $2L$  pieces. By our assumption that  $\mathcal{P}$  is strictly feasible, we have that  $x_i^m(g_j) > 0$  for  $j \in [L]$  since  $g_j > g_{L+1} \geq 0$ . In particular, for  $g \in (g_{j+1}, g_j]$ ,  $i \in [L]$ , we have that  $\frac{g}{\phi_i(g)} = \min\{\frac{g}{x_i^m(g_{j+1})}, \frac{g_j}{x_i^m(g_j)}\}$  if  $x_i^m(g_{j+1}) > 0$  and  $\frac{g}{\phi_i(g)} = \frac{g_j}{x_i^m(g_j)}$  otherwise. Thus,  $\frac{g}{\phi_i(g)}$  is also continuous piecewise linear with at most  $2L$  pieces on  $(\underline{g}, \bar{g}]$ .

We now prove (36), (37). For  $j \in [L]$ ,  $g \in [g, \bar{g}]$ , let

$$l_i^j(g) := \left( x_i^m(g_{j+1}) - g_{j+1} \frac{x_i^m(g_j) - x_i^m(g_{j+1})}{g_j - g_{j+1}} \right) + \frac{x_i^m(g_j) - x_i^m(g_{j+1})}{g_j - g_{j+1}} g := a_j + b_j g,$$

denote the linear interpolation of  $x_i^m$  through  $g_j$  and  $g_{j+1}$ . By concavity, monotonicity of  $x_i^m$  and  $0 \leq g_{j+1} < g_j$ , the intercept  $a_j$  and slope  $b_j$  of  $l_i^j$  satisfy

$$0 \leq x_i^m(0) \leq l_i^j(0) = a_j = x_i^m(g_{j+1}) - g_{j+1} \frac{x_i^m(g_j) - x_i^m(g_{j+1})}{g_j - g_{j+1}} \leq x_i^m(g_{j+1}), \quad (38)$$

$$0 \leq \frac{x_i^m(g_j) - x_i^m(g_{j+1})}{g_j - g_{j+1}} = b_j \leq \frac{x_i^m(g_j) - x_i^m(0)}{g_j - 0} \leq \frac{x_i^m(g_j)}{g_j}. \quad (39)$$

For  $g \in (g_{j+1}, g_j]$ , then by (38) and (39) we get that

$$\phi_i(g) = \max \left\{ l_i^j(g_{j+1}), g \frac{l_i^j(g_j)}{g_j} \right\} \quad (40)$$

$$\begin{aligned} &\leq \max \left\{ l_i^j(g_{j+1}) + b_j(g_j - g_{j+1}), \left(1 - \frac{g}{g_j}\right) l_i^j(0) + \frac{g}{g_j} l_i^j(g_j) \right\} = l_i^j(g) \\ &= a_j + gb_j \leq x_i^m(g_{j+1}) + g \frac{x_i^m(g_j)}{g_j} \leq 2\phi_i(g). \end{aligned} \quad (41)$$

Since  $[(g_j, x_i^m(g_j)), (g_{j+1}, x_i^m(g_{j+1}))] \subset \mathcal{N}_i^{\text{mp}}(1 - \alpha)$  by construction, for  $g \in (g_{j+1}, g_j]$  we have that

$$x_i^m(g) \geq l_i^j(g) \geq \phi_i(g) \geq \frac{1}{2} l_i^j(g) \geq \frac{\alpha}{2} x_i^m(g).$$

Since the above holds for all  $j \in [L]$ , this proves (36). To prove (37), we combine (36) with Theorem 4.2 as follows: for  $g \in (\underline{g}, \bar{g}]$ ,

$$s_i^m(g) = \frac{x_i^m(g) s_i^m(g)}{x_i^m(g)} \geq \frac{g}{x_i^m(g)} \geq \frac{\alpha g}{2\phi_i(g)} \geq \frac{\alpha g}{2x_i^m(g)} = \frac{\alpha s_i^m(g)}{2} \frac{g}{x_i^m(g) s_i^m(g)} \geq \frac{\alpha s_i^m(g)}{4}.$$

From here, the moreover  $\text{SLC}_{1-\alpha/4, i}^d(\underline{g}, \bar{g}) \leq 2 \text{SLC}_{1-\alpha, i}^p(\underline{g}, \bar{g})$  follows immediately from (37) and the fact that  $\frac{\alpha g}{2\phi_i(g)}$  is continuous piecewise linear with a number of pieces bounded by  $2 \text{SLC}_{1-\alpha, i}^p(\underline{g}, \bar{g})$  on  $(\underline{g}, \bar{g}]$  (note that we can extend  $\frac{\alpha g}{2\phi_i(g)}$  to  $[\underline{g}, \bar{g}]$  by continuity).  $\square$

We are now ready to prove Theorem 1.8; we restate it here for convenience.

**Theorem 1.8.** *Let  $\theta \in [0, 1)$  and  $\mu_0 > \mu_1 \geq 0$ . The segment  $\text{CP}[\mu_1, \mu_0]$  can be decomposed into a sequence of  $\frac{1-\theta}{4n}$ -polarized segments with partitions  $(B^{(k)}, N^{(k)})$ ,  $k \in [T]$ , such that*

$$\max \left\{ \sum_{k=1}^T |N^{(k)} \Delta N^{(k-1)}|, T \right\} \leq 2 \min \left\{ \sum_{i=1}^n \text{SLC}_{\theta, i}^p(n\mu_1, n\mu_0), \sum_{i=1}^n \text{SLC}_{\theta, i}^d(n\mu_1, n\mu_0) \right\}$$

where  $N^{(0)} := \emptyset$ .

*Proof.* We only prove the statement for the primal straight-line complexity bound; the proof for the dual bound follows analogously.

Let  $\underline{g} := n\mu_1$  and  $\bar{g} := n\mu_0$  denote the gaps at  $\mu_1$  and  $\mu_0$ . For  $i \in [n]$ , apply Lemma 4.5 on  $x_i^m$  on  $[\bar{g}, \underline{g}]$  to get  $\bar{g} = g_{i,1} > \dots > g_{i,L_i+1} = \underline{g}$ ,  $L_i = \text{SLC}_{\theta, i}^p(\underline{g}, \bar{g})$ , and the corresponding piecewise linear function  $\phi_i(\cdot)$  satisfying  $x_i^m(g) \geq \phi_i(g) \geq \frac{1-\theta}{2} x_i^m(g)$ ,  $\forall g \in [\underline{g}, \bar{g}]$ . For  $j \in [L_i]$ , define  $g_{i,j}^+ := \frac{g_{i,j} x_i^m(g_{i,j+1})}{x_i^m(g_{i,j})} \in [g_{i,j+1}, g_{i,j}]$ ,

which we note satisfies  $\phi_i(g_{i,j}^+) = \max \left\{ x_i^m(g_{i,j+1}), \frac{g_{i,j}^+ x_i^m(g_{i,j})}{g_{i,j}} \right\} = \frac{g_{i,j}^+ x_i^m(g_{i,j})}{g_{i,j}} = x_i^m(g_{i,j+1})$ . From here,

let  $G_i := \{g_{i,j} : i \in [L_i + 1]\} \cup \{g_{i,j}^+ : i \in [L_i]\}$  denote the breakpoints of the function  $\phi_i$ . For a closed interval  $[a, b] \subseteq [\underline{g}, \bar{g}]$  with  $a < b$ , we say that  $[a, b]$  is type (B) for  $\phi_i$  if  $\phi_i$  is constant on  $[a, b]$  and is type (N) for  $\phi_i$  if  $\phi_i(g') = \frac{g'}{g} \phi_i(g)$ ,  $\forall g', g \in [a, b]$ ,  $g' \leq g$  and  $g > 0$ . By construction of  $\phi_i$  and  $G_i$ , if  $G_i \cap (a, b) = \emptyset$ , then  $[a, b]$  must be either type (B) or type (N) for  $\phi_i$  (note that these options are mutually exclusive).

Define  $G = \cup_{i=1}^n G_i$  to be the union of the breakpoints of the functions  $\phi_i$ ,  $i \in [n]$ . Let  $\underline{g} = g_T < \dots < g_0 = \bar{g}$  satisfy  $G = \{g_i : i \in \{0, \dots, T\}\}$ . Since  $\underline{g} \in G_i$ ,  $\forall i \in [n]$ , we have that

$$T = |G| - 1 \leq \sum_{i=1}^n (|G_i| - 1) \leq 2 \sum_{i=1}^n L_i := 2 \sum_{i=1}^n \text{SLC}_{\theta, i}^p(\underline{g}, \bar{g}). \quad (42)$$

Define  $(B^{(0)}, N^{(0)}) := ([n], \emptyset)$ , and for  $k \in [T]$ , define  $(B^{(k)}, N^{(k)})$  by

$$\begin{aligned} B^{(k)} &= \{i \in [n] : [g_k, g_{k-1}] \text{ is type (B) for } \phi_i\}, \\ N^{(k)} &= \{i \in [n] : [g_k, g_{k-1}] \text{ is type (N) for } \phi_i\}. \end{aligned}$$

By construction of  $G$ , we have  $(g_k, g_{k-1}) \cap G_i = \emptyset$ ,  $i \in [n]$ , and hence  $(B^{(k)}, N^{(k)})$  partitions  $[n]$ . The sum of partition differences then satisfies the bound

$$\begin{aligned} \sum_{k=1}^T |N^{(k)} \Delta N^{(k-1)}| &= \sum_{i=1}^n \sum_{k=1}^T |1[i \in N^{(k)}] - 1[i \in N^{(k-1)}]| \leq \sum_{i=1}^n \sum_{k=1}^T 1[g_{k-1} \in G_i] \\ &= \sum_{i=1}^n (|G_i| - 1) \leq 2 \sum_{i=1}^n \text{SLC}_{\theta, i}^p(\underline{g}, \bar{g}). \end{aligned} \quad (43)$$

Define  $h_i := g_i/n$ ,  $i \in \{0, \dots, T\}$ , where we have  $h_0 = g_0/n = \bar{g}/n = \mu_0$  and  $h_T = g_T/n = \underline{g}/n = \mu_1$ . For  $k \in [T]$ , we claim that the central path segment  $\text{CP}[h_k, h_{k-1}]$  is  $\gamma = \frac{1-\theta}{4n}$ -polarized with polarization partition  $(B^{(k)}, N^{(k)})$ . Together with the claim, the bounds (42) and (43) directly imply theorem.

We now prove the claim for  $k \in [T]$ . For  $\mu \in [h_k, h_{k-1}] = [g_k/n, g_{k-1}/n]$  and  $i \in B^{(k)}$ , by Lemma 1.2 and the guarantees on  $\phi_i$ , we have that

$$x_i^{\text{CP}}(\mu) \geq \frac{1}{2n} x_i^{\text{m}}(n\mu) \geq \frac{1}{2n} \phi_i(n\mu) = \frac{1}{2n} \phi_i(g_{k-1}) \geq \frac{1-\theta}{4n} x_i^{\text{m}}(g_{k-1}) \geq \frac{1-\theta}{4n} x_i^{\text{CP}}(h_{k-1}), \quad (44)$$

where in the first equality we used that  $[g_k, g_{k-1}]$  is type (B) for  $\phi_i$ . Similarly, for  $i \in N^{(k)}$ , we have that

$$\begin{aligned} x_i^{\text{CP}}(\mu) &\leq x_i^{\text{m}}(n\mu) \leq \frac{2}{1-\theta} \phi_i(n\mu) = \frac{2}{1-\theta} \frac{\mu}{h_{k-1}} \phi_i(nh_{k-1}) \\ &\leq \frac{2}{1-\theta} \frac{\mu}{h_{k-1}} x_i^{\text{m}}(nh_{k-1}) \leq \frac{4n}{1-\theta} \frac{\mu}{h_{k-1}} x_i^{\text{CP}}(h_{k-1}), \end{aligned} \quad (45)$$

where in the first equality we used that  $[g_k, g_{k-1}]$  is type (N) for  $\phi_i$ . By Remark 3.2, the inequalities (44) and (45) imply that  $\text{CP}[h_k, h_{k-1}]$  is  $\frac{1-\theta}{4n}$ -polarized with polarization partition  $(B^{(k)}, N^{(k)})$  as desired.  $\square$

## 5 The Trust Region Step

In this section, we prove important properties for the trust-region step of Lan, Monteiro and Tsuchiya [38], as defined in the program (8) in the introduction. The estimates proved here will play a crucial role in the analysis of the IPM, and correspond to refinements of existing estimates in [38] and [48] that are more adapted to our IPM.

Our first estimate, given in Section 5.1 is a general relationship between the parameters of trust-region directions and the achievable step-length guarantees one can derive from them. The second main result, given in Section 5.2, regards the computation of the optimal partition  $(B, N)$  to use when computing trust-region directions. We show that if a suitably good trust-region direction exists with respect to some partition  $(B, N)$ , then this partition can be read off from the coordinates of the affine scaling direction (which implies that it is unique). We prove this by showing that the affine scaling direction must be close to any good enough trust region direction. This is analogous to a result of [48], who showed that the same statement holds if we replace the trust-region direction by the layered least squares direction.

We dub the partition induced by the affine scaling direction the *associated partition* (in [38] this was called the AS bipartition). With respect to our IPM, the associated partition will be used to identify the current polarization partition.

### 5.1 Step-Length Estimates for Trust Region Directions

We now formally link the properties of trust-region program solutions to the achievable step-lengths in Proposition 2.44. A similar estimate is proved in [38, Lemma 2.7].

**Proposition 5.1.** *Let  $x, s \in \mathbb{R}_{++}^n$  satisfy  $\|xs/\mu - \mathbf{1}_n\| \leq 1/6$ , where  $\mu := \bar{\mu}(x, s)$ , and let  $B \cup N = [n]$  be a partition. For  $\Delta x, \Delta s \in \mathbb{R}^n$ , assume that*

$$\|(\Delta \hat{x}_B, \Delta \hat{s}_N)\| \leq \delta \leq 1/6, \quad \left\| (\hat{\xi}_B + \Delta \hat{s}_B, \hat{\xi}_N + \Delta \hat{x}_N) \right\| \leq \varepsilon \leq 1/6, \quad (46)$$

where  $\hat{\xi} := \sqrt{xs/\mu}$ ,  $(\Delta \hat{x}, \Delta \hat{s}) := (\hat{x}^{-1} \Delta x, \hat{s}^{-1} \Delta s)$  and  $(\hat{x}, \hat{s}) := (x \hat{\xi}^{-1}, s \hat{\xi}^{-1})$ . Then,

$$\left\| \frac{\Delta x \Delta s}{\mu} \right\| \leq \sqrt{2} \delta, \quad (47)$$

$$\left\| \frac{(x + \Delta x)(s + \Delta s)}{\mu} \right\| \leq \sqrt{2} \varepsilon. \quad (48)$$

*Proof.* To derive (47) and (48), we prove the following more general statements:

$$\left\| \frac{\Delta x \Delta s}{\mu} \right\|^2 \leq \|\Delta \hat{x}_B\|^2 \left(1 + \frac{1}{12} + \|\hat{\xi}_B + \Delta \hat{s}_B\|\right)^2 + \|\Delta \hat{s}_N\|^2 \left(1 + \frac{1}{12} + \|\hat{\xi}_N + \Delta \hat{x}_N\|\right)^2, \quad (49)$$

$$\left\| \frac{(x + \Delta x)(s + \Delta s)}{\mu} \right\|^2 \leq \|\hat{\xi}_B + \Delta \hat{s}_B\|^2 \left(1 + \frac{1}{12} + \|\Delta \hat{x}_B\|\right)^2 + \|\hat{\xi}_N + \Delta \hat{x}_N\|^2 \left(1 + \frac{1}{12} + \|\Delta \hat{s}_N\|\right)^2. \quad (50)$$

To derive (47), we combine (49) and the assumed bounds (46) as follows:

$$\begin{aligned} \left\| \frac{\Delta x \Delta s}{\mu} \right\|^2 &\leq \|\Delta \hat{x}_B\|^2 \left(1 + \frac{1}{12} + \|\hat{\xi}_B + \Delta \hat{s}_B\|\right)^2 + \|\Delta \hat{s}_N\|^2 \left(1 + \frac{1}{12} + \|\hat{\xi}_N + \Delta \hat{x}_N\|\right)^2 \\ &\leq \|(\Delta \hat{x}_B, \Delta \hat{s}_N)\|^2 (1 + \frac{1}{12} + \varepsilon)^2 \leq \delta^2 (1 + \frac{1}{12} + \frac{1}{6})^2 \leq \delta^2 (\frac{5}{4})^2 \leq 2\delta^2. \end{aligned}$$

Similarly, (48) is derived by combining (50) and (46) as

$$\begin{aligned} \left\| \frac{(x + \Delta x)(s + \Delta s)}{\mu} \right\|^2 &\leq \|\hat{\xi}_B + \Delta \hat{s}_B\|^2 \left(1 + \frac{1}{12} + \|\Delta \hat{x}_B\|\right)^2 + \|\hat{\xi}_N + \Delta \hat{x}_N\|^2 \left(1 + \frac{1}{12} + \|\Delta \hat{s}_N\|\right)^2 \\ &\leq \varepsilon^2 (1 + \frac{1}{12} + \delta)^2 \leq 2\varepsilon^2. \end{aligned}$$

We now focus on the proofs of (49) and (50).

**Proof of inequality (49)** To begin, note that

$$\left\| \frac{\Delta x \Delta s}{\mu} \right\|^2 = \|\Delta \hat{x} \Delta \hat{s}\|^2 = \|\Delta \hat{x}_B \Delta \hat{s}_B\|^2 + \|\Delta \hat{x}_N \Delta \hat{s}_N\|^2.$$

For the  $B$  term on the right hand side, we have that

$$\begin{aligned} \|\Delta \hat{x}_B \Delta \hat{s}_B\|^2 &\leq \|\Delta \hat{x}_B\|^2 \|\Delta \hat{s}_B\|_\infty^2 \\ &\leq \|\Delta \hat{x}_B\|^2 \left( \|\hat{\xi}_B\|_\infty + \|\hat{\xi}_B + \Delta \hat{s}_B\|_\infty \right)^2 \\ &\leq \|\Delta \hat{x}_B\|^2 \left( 1 + \frac{1}{12} + \|\hat{\xi}_B + \Delta \hat{s}_B\| \right)^2, \end{aligned}$$

where the last inequality follows since  $\|\hat{\xi}_B + \Delta \hat{s}_B\|_\infty \leq \|\hat{\xi}_B + \Delta \hat{s}_B\|$  and

$$\|\hat{\xi}_B\|_\infty \leq \|\hat{\xi}\|_\infty = \max_{i \in [n]} \sqrt{\frac{x_i s_i}{\mu}} \leq \sqrt{1 + \frac{1}{6}} \leq 1 + \frac{1}{12}.$$

By a symmetric argument, swapping the roles of  $(x, s)$  and  $(B, N)$ , we also get that

$$\|\Delta \hat{x}_N \Delta \hat{s}_N\|^2 \leq \|\Delta \hat{s}_N\|^2 \left( 1 + \frac{1}{12} + \|\hat{\xi}_N + \Delta \hat{x}_N\| \right)^2.$$

Inequality (49) now follows by combining for the  $B$  and  $N$  bounds above.

**Proof of inequality (50)** Similar to the above, we have that

$$\begin{aligned} \left\| \frac{(x + \Delta x)(s + \Delta s)}{\mu} \right\|^2 &= \|(\hat{\xi} + \Delta \hat{x})(\hat{\xi} + \Delta \hat{s})\|^2 \\ &= \|(\hat{\xi}_B + \Delta \hat{x}_B)(\hat{\xi}_B + \Delta \hat{s}_B)\|^2 + \|(\hat{\xi}_N + \Delta \hat{x}_N)(\hat{\xi}_N + \Delta \hat{s}_N)\|^2. \end{aligned}$$

As before, we bound the  $B$  and  $N$  parts separately. For the  $B$  part, we have that

$$\begin{aligned} \|(\hat{\xi}_B + \Delta \hat{x}_B)(\hat{\xi}_B + \Delta \hat{s}_B)\|^2 &\leq \|\hat{\xi}_B + \Delta \hat{s}_B\|^2 \|\hat{\xi}_B + \Delta \hat{x}_B\|_\infty^2 \\ &\leq \|\hat{\xi}_B + \Delta \hat{s}_B\|^2 \left( \|\hat{\xi}_B\|_\infty + \|\Delta \hat{x}_B\|_\infty \right)^2 \\ &\leq \|\hat{\xi}_B + \Delta \hat{s}_B\|^2 \left( 1 + \frac{1}{12} + \|\Delta \hat{x}_B\| \right)^2, \end{aligned}$$

where the last inequality from identical to the first part. Again, by a symmetric argument,

$$\|(\hat{\xi}_N + \Delta \hat{x}_N)(\hat{\xi}_N + \Delta \hat{s}_N)\|^2 \leq \|\hat{\xi}_N + \Delta \hat{x}_N\|^2 \left( 1 + \frac{1}{12} + \|\Delta \hat{s}_N\| \right)^2.$$

Inequality (50) now follows by combining for the  $B$  and  $N$  bounds above.  $\square$

## 5.2 The Associated Partition

The trust-region step is applicable for any non-trivial partition  $B \cup N = [n]$  and  $z \in \mathcal{N}^2(\beta)$ . Following [38], we choose a natural partition derived from the size of normalized coordinates of the affine scaling direction:

**Definition 5.2** (Associated partition). For  $z = (x, s) \in \mathcal{N}^2(\beta)$ , let  $(\Delta x^a, \Delta s^a)$  be the affine scaling step as in (24). Let us define the associated partition  $\tilde{B}_z \cup \tilde{N}_z = [n]$  as

$$\tilde{B}_z := \left\{ i : \left| \frac{\Delta x_i^a}{x_i} \right| < \left| \frac{\Delta s_i^a}{s_i} \right| \right\}, \quad \tilde{N}_z := [n] \setminus \tilde{B}_z.$$

The affine scaling step is the canonical candidate for an improving direction. Namely, for each  $i \in \tilde{B}_z$  the variable  $s_i$  decreases at a faster rate than  $x_i$ , and vice versa for  $i \in \tilde{N}_z$ .

As we show below, as long as the trust-region program admits a sufficiently good solution with respect to the polarization partition  $(B, N)$ , then the affine scaling step is close to this trust-region direction and has associated partition  $(B, N)$ . This implies that the optimal choice of partition is in fact unique under the assumption that a long steps exists.

**Lemma 5.3.** Let  $z = (x, s) \in \mathcal{N}^2(\beta)$ ,  $\beta \in (0, 1/6]$ ,  $\mu := \bar{\mu}(z)$ ,  $B \cup N = [n]$  be a partition. Assume that there exists  $(\Delta x, \Delta s) \in W \times W^\perp$  satisfying

$$\|(\Delta \hat{x}_B, \Delta \hat{s}_N)\| \leq \delta \leq 1/30, \quad \left\| (\hat{\xi}_N + \Delta \hat{x}_N, \hat{\xi}_B + \Delta \hat{s}_B) \right\| \leq \epsilon \leq 1/30, \quad (51)$$

where  $\hat{\xi} := \sqrt{xs}/\mu$ ,  $(\Delta \hat{x}, \Delta \hat{s}) := (\hat{x}^{-1} \Delta x, \hat{s}^{-1} \Delta s)$ ,  $(\hat{x}, \hat{s}) := (x \hat{\xi}^{-1}, s \hat{\xi}^{-1})$ . Then, the affine scale direction  $(\Delta x^a, \Delta s^a)$  at  $z$  satisfies

$$\left\| \frac{\Delta x^a \Delta s^a}{\mu} \right\| = \left\| \frac{(x + \Delta x^a)(s + \Delta s^a)}{\mu} \right\| \leq 3.5(\delta + \epsilon), \quad (52)$$

and the associated partition at  $z$  satisfies  $(\tilde{B}_z, \tilde{N}_z) = (B, N)$ .

*Proof.* Recall that the affine scaling step  $(\Delta x^a, \Delta s^a)$  is defined by

$$\Delta \hat{x}^a + \Delta \hat{s}^a = -\hat{\xi},$$

where  $(\Delta \hat{x}^a, \Delta \hat{s}^a) := (\hat{x}^{-1} \Delta x^a, \hat{s}^{-1} \Delta s^a) \in \hat{x}^{-1} W \times \hat{s}^{-1} W^\perp$  form an orthogonal decomposition of  $-\hat{\xi}$ . By orthogonality, we therefore have that

$$\|\Delta \hat{x}^a - \Delta \hat{x}\|^2 + \|\Delta \hat{s}^a - \Delta \hat{s}\|^2 = \|\hat{\xi} + \Delta \hat{x} + \Delta \hat{s}\|^2.$$

By the triangle inequality,

$$\begin{aligned} \left\| \hat{\xi}_B + \Delta \hat{x}_B + \Delta \hat{s}_B \right\|^2 &\leq 2 \|\Delta \hat{x}_B\|^2 + 2 \left\| \hat{\xi}_B + \Delta \hat{s}_B \right\|^2, \\ \left\| \hat{\xi}_N + \Delta \hat{x}_N + \Delta \hat{s}_N \right\|^2 &\leq 2 \|\Delta \hat{s}_N\|^2 + 2 \left\| \hat{\xi}_N + \Delta \hat{x}_N \right\|^2. \end{aligned}$$

Therefore, by (51),

$$\begin{aligned} \left\| \hat{\xi} + \Delta \hat{x} + \Delta \hat{s} \right\|^2 &= \left\| \hat{\xi}_B + \Delta \hat{x}_B + \Delta \hat{s}_B \right\|^2 + \left\| \hat{\xi}_N + \Delta \hat{x}_N + \Delta \hat{s}_N \right\|^2 \\ &\leq 2(\|\Delta \hat{x}_B\|^2 + \|\Delta \hat{s}_N\|^2) + \left\| \hat{\xi}_N + \Delta \hat{x}_N \right\|^2 + \left\| \hat{\xi}_B + \Delta \hat{s}_B \right\|^2 \\ &\leq 2(\delta^2 + \epsilon^2). \end{aligned}$$

In particular,

$$\begin{aligned} \|(\Delta \hat{x}_B^a, \Delta \hat{s}_N^a)\| &\leq \sqrt{2(\delta^2 + \epsilon^2)} + \|(\Delta \hat{x}_B, \Delta \hat{s}_N)\| \leq \sqrt{2}(\delta + \epsilon) + \delta \\ &\leq (\sqrt{2} + 1)(\delta + \epsilon) \leq \frac{1}{6}, \\ \left\| (\hat{\xi}_N + \Delta \hat{x}_N^a, \hat{\xi}_B + \Delta \hat{s}_B^a) \right\| &\leq \sqrt{2(\delta^2 + \epsilon^2)} + \left\| (\hat{\xi}_N + \Delta \hat{x}_N, \hat{\xi}_B + \Delta \hat{s}_B) \right\| \\ &\leq (\sqrt{2} + 1)(\delta + \epsilon) \leq \frac{1}{6}. \end{aligned}$$

Therefore by Proposition 5.1, since affine scaling satisfies  $\Delta x^a \Delta s^a = (x + \Delta x^a)(s + \Delta s^a)$ , we get that

$$\left\| \frac{\Delta x^a \Delta s^a}{\mu} \right\| = \left\| \frac{(x + \Delta x^a)(s + \Delta s^a)}{\mu} \right\| \leq \sqrt{2}(\sqrt{2} + 1)(\delta + \varepsilon) \leq 3.5(\delta + \varepsilon).$$

For the last part, notice that  $|\Delta x_i^a/x_i| < |\Delta s_i^a/s_i| \Leftrightarrow |\Delta \hat{x}_i^a| < |\Delta \hat{s}_i^a|, \forall i \in [n]$ . By Proposition 2.42, since  $\beta \in (0, 1/6]$ , we have that  $\hat{\xi} \geq \sqrt{1 - \beta} \mathbf{1}_n \geq (1 - 1/6) \mathbf{1}_n$ . For  $i \in B$ , using that  $(\sqrt{2} + 1)(\delta + \varepsilon) \leq 3(\delta + \varepsilon) \leq 1/3$  together with (53) we see that

$$|\Delta \hat{x}_i^a| \leq 3(\delta + \varepsilon) < (1 - 1/6) - 3(\delta + \varepsilon) \leq \hat{\xi}_i - |\Delta \hat{x}_i^a| \leq |\hat{\xi}_i + \Delta \hat{x}_i^a| = |\Delta \hat{s}_i^a|.$$

By a symmetric argument, we also have  $|\Delta \hat{s}_i^a| < |\Delta \hat{x}_i^a|$ , for  $i \in N$ . Therefore,  $(\tilde{B}_z, \tilde{N}_z) = (B, N)$ .  $\square$

**Remark 5.4.** Given that the affine scaling step is always close to a trust-region direction, as long as the direction yields a sufficiently long step, one may wonder why the trust-region direction, or its approximate version the subspace LLS direction (defined in the next section), is even needed. At a quantitative level, this comes from the fact that the trust-region direction assumed to exist in Lemma 5.3 would in fact allow us to decrease the gap by a factor  $\Theta(\varepsilon)$  (see Proposition 2.44), whereas the “nearby” affine scaling would only achieve  $\Theta(\delta + \varepsilon)$ , which can be arbitrarily worse. Indeed, recall that in (8) we only enforce  $\delta = O(\beta)$ , that is, proportional to the neighborhood size, while trying to minimize  $\varepsilon$ .

## 6 The Subspace LLS Direction and Cheap Lift Subspaces

In this section, we formally define the subspace layered least squares (SLLS) steps and show how to compute the cheap lift subspaces as described in the Introduction. In Section 6.1, we formally define the SLLS step and overview its basic computational aspects. Then, in Section 6.2, we introduce the lifting map and operator and overview their basic properties and give an algorithm to compute them. In Section 6.3, we define cheap lift subspaces, that will be used for taking subspace LLS steps, as well as the lifting operators, and show how to construct cheap lift subspaces from approximate singular subspaces (as defined in Definition 2.24). In Section 6.4, we show how to compute cheap lift subspaces of approximately maximum dimension using an approximate singular value decomposition (see the  $\varrho$ -SVD problem).

### 6.1 The Subspace Layered Least Squares Direction

We now introduce a new update direction, called the *subspace layered least squares* (SLLS) update direction, which will allow us to accelerate our IPM over long straight parts of the central path. As discussed in the previous sections, such straight segments of the central path are in fact *polarized* according to a partition  $B \cup N = [n]$  (Definition 3.1). Within this segment, the primal variables indexed by  $N$  scale down linearly with respect to the parameter  $\mu$ , while the primal variables in  $B$  will remain roughly constant. For the dual, the situation is reversed, the variables in  $B$  scale down while the variables in  $N$  remain roughly constant.

As discussed, in the Introduction, the “optimal” step direction that mimics the polarization pattern is the *trust region step* of Lan, Monteiro and Tsuchiya (8). To be able to suitably approximate this step in strongly polynomial time, we require the SLLS update direction, which we formally define below.

**Definition 6.1** (Subspace Layered Least Squares direction). Let  $z := (x, s) \in \mathcal{N}^2(\beta)$ ,  $\mu = \bar{\mu}(z)$ ,  $B \cup N = [n]$  be a non-trivial partition. Let  $V \subseteq W$ ,  $U \subseteq W^\perp$  be linear subspaces satisfying  $\dim(\pi_N(V)) = \dim(V)$ , and  $\dim(\pi_B(U)) = \dim(U)$ . The *Subspace LLS (SLLS)* update direction  $(\Delta x^\ell, \Delta s^\ell) \in W \times W^\perp$  at  $z$  with respect to  $(B, N, V, U)$  is defined as

$$\Delta x^\ell := \arg \min_{\delta \in V} \left\| \frac{x_N + \delta_N}{\hat{x}_N} \right\|^2, \quad (53)$$

$$\Delta s^\ell := \arg \min_{\delta \in U} \left\| \frac{s_B + \delta_B}{\hat{s}_B} \right\|^2, \quad (54)$$

where  $\hat{x} := \sqrt{x\mu/s}$  and  $\hat{s} := \sqrt{s\mu/x}$ .

**Remark 6.2.** Note that  $\Delta x^\ell, \Delta s^\ell$  are indeed well-defined, as our assumption that  $\dim(\pi_N(V)) = \dim(V)$  and  $\dim(\pi_B(U)) = \dim(U)$  allows to uniquely determine  $\Delta x^\ell, \Delta s^\ell$  from their coordinates in  $N$  and  $B$  respectively. In particular, if  $N = \emptyset$ , then  $V = \{\mathbf{0}_n\}$  and hence  $\Delta x^\ell = \mathbf{0}_n$ . Similarly, if  $B = \emptyset$ , then  $U = \{\mathbf{0}_n\}$  and  $\Delta s^\ell = \mathbf{0}_n$ .

As with affine scaling, one can interpret the SLLS step directions in terms of orthogonal projections. Using that  $\hat{\xi} := \sqrt{x s / \mu} = x / \hat{x} = s / \hat{s}$ , in direct analogy with (24) in Section 2.6,  $(\hat{x}^{-1} \Delta x)_N$  and  $(\hat{s}^{-1} \Delta s)_B$  are respectively the orthogonal projections of  $-\hat{\xi}_N$  and  $-\hat{\xi}_B$  onto  $\pi_N(\hat{x}^{-1} V)$  and  $\pi_B(\hat{s}^{-1} U)$ . In contrast, whereas the affine scaling direction minimizes the norm of the primal-dual residual vector  $(\hat{\xi} + \Delta \hat{x}, \hat{\xi} + \Delta \hat{s})$ , the subspace LLS direction instead tries to minimize the norm of  $(\hat{\xi}_N + \Delta \hat{x}_N, \hat{\xi}_B + \Delta \hat{s}_B)$ .

Similarly to the affine scaling and corrector direction, the SLLS direction can be computed in strongly polynomial time, given an appropriate representation of the subspaces. The formulas for the subspace LLS step directions are given in the next proposition. These are computed by solving the linear system which sets the gradient of the corresponding quadratic optimization problems to zero.

**Proposition 6.3** (Subspace LLS Step Formulas). *Let  $z = (x, s) \in \mathbb{R}_{++}^{2n}$  be an iterate,  $B \cup N = [n]$  be a non-trivial partition, and  $\mathbf{M}^P \in \mathbb{R}^{n \times m_p}, \mathbf{M}^D \in \mathbb{R}^{n \times m_d}$  be matrices where  $V := \text{im}(\mathbf{M}^P), U := \text{im}(\mathbf{M}^D)$  satisfy  $\dim(V) = \dim(\pi_N(V))$  and  $\dim(U) = \dim(\pi_B(U))$ . Then, the SLLS direction  $(\Delta x^\ell, \Delta s^\ell)$  at  $z$  with respect to  $(B, N, U, V)$  can be computed in strongly polynomial time as follows:*

$$\begin{aligned} \Delta x^\ell &= -\mathbf{M}^P \left( (\mathbf{M}_{N, \bullet}^P)^\top \text{diag}(s_N/x_N) \mathbf{M}_{N, \bullet}^P \right)^+ (\mathbf{M}_{N, \bullet}^P)^\top s_N, \\ \Delta s^\ell &= -\mathbf{M}^D \left( (\mathbf{M}_{B, \bullet}^D)^\top \text{diag}(x_B/s_B) \mathbf{M}_{B, \bullet}^D \right)^+ (\mathbf{M}_{B, \bullet}^D)^\top x_B. \end{aligned}$$

In the above, by convention,  $\Delta x^\ell = \mathbf{0}_n$  if  $N = \emptyset$  and  $\Delta s^\ell = \mathbf{0}_n$  if  $B = \emptyset$ .

*Proof.* The strong polynomiality follows directly from the fact that computing pseudoinverses is strongly polynomial. We thus focus on correctness. We prove correctness of the formula for  $\Delta x^\ell$  as the analysis for  $\Delta s^\ell$  is symmetric. The formula is clearly correct if  $N = \emptyset$  by convention, so assume  $N \neq \emptyset$ . Let

$$z_x := - \left( (\mathbf{M}_{N, \bullet}^P)^\top \text{diag}(s_N/x_N) \mathbf{M}_{N, \bullet}^P \right)^+ (\mathbf{M}_{N, \bullet}^P)^\top s_N \in \mathbb{R}^{m_p}.$$

Recalling that  $V := \text{im}(\mathbf{M}^P)$ , we have that  $\Delta x^\ell = \mathbf{M}^P z_x \in V$ . Thus, letting  $g_x(z) = \left\| \frac{x_N + \mathbf{M}_{N, \bullet}^P z}{\hat{x}_N} \right\|^2$ , it suffices to check that

$$\begin{aligned} z_x \in \arg \min_{z \in \mathbb{R}^{m_p}} g_x(z) &\Leftrightarrow \nabla g_x(z_x) = \mathbf{0}_{m_p} \Leftrightarrow 2 (\mathbf{M}_{N, \bullet}^P)^\top \left( \frac{x_N + \mathbf{M}_{N, \bullet}^P z_x}{\hat{x}_N^2} \right) = \mathbf{0}_{m_p} \\ &\Leftrightarrow (\mathbf{M}_{N, \bullet}^P)^\top (s_N + \text{diag}(s_N/x_N) \mathbf{M}_{N, \bullet}^P z_x) = \mathbf{0}_{m_p}, \end{aligned} \quad (55)$$

where we have used convexity of  $g_x$  and  $\hat{x} := \sqrt{x \mu / s}$  where  $\mu := \bar{\mu}(x, s) > 0$ . Note that while  $g_x$  need not have a unique minimizer (i.e., the columns of  $\mathbf{M}^P$  may be linearly dependent), the condition  $\dim(V) = \dim(\pi_N(V)) = \text{rank}(\mathbf{M}_{N, \bullet}^P)$  indeed ensures that  $\Delta x^\ell = \mathbf{M}^P z_x$  is the unique minimizer to (53) assuming  $z_x$  minimizes  $g_x$ . From here, it suffices to check that  $z_x$  satisfies (55), and hence minimizes  $g_x$ . For this purpose, letting  $\mathbf{P} = (\mathbf{M}_{N, \bullet}^P)^\top \text{diag}(s_N/x_N) \mathbf{M}_{N, \bullet}^P$ , by Proposition 2.14 part (1) and  $x, s > \mathbf{0}_n$

$$\begin{aligned} (\mathbf{M}_{N, \bullet}^P)^\top s_N &= \Pi_{\text{im}((\mathbf{M}_{N, \bullet}^P)^\top)} (\mathbf{M}_{N, \bullet}^P)^\top s_N = \Pi_{\text{im}(\mathbf{P})} (\mathbf{M}_{N, \bullet}^P)^\top s_N \\ &= \mathbf{P} \mathbf{P}^+ (\mathbf{M}_{N, \bullet}^P)^\top s_N, \end{aligned}$$

where we have used that  $\text{im}((\mathbf{M}_{N, \bullet}^P)^\top) = \text{im}((\mathbf{M}_{N, \bullet}^P)^\top \text{diag}(s_N/x_N) \mathbf{M}_{N, \bullet}^P)$ .  $\square$

## 6.2 Lifting Maps and Operators

The algorithm in Section 7 and the analysis in Section 7.3 will rely crucially on the properties of *lifting maps and operators*. These will be used to compute cheap lift subspaces for the purpose of computing subspace layered least squares directions. These maps have appeared in many prior works on layered least squares algorithms [66, 47, 19]. In this section, we give a self-contained overview of all their main properties. Importantly, we show their duality properties and that they can be computed in strongly polynomial time.

**Definition 6.4.** Given a non-trivial partition  $I \cup J = [n]$  and a subspace  $W \subseteq \mathbb{R}^n$ , we define the lifting map  $L_I^W : \mathbb{R}^I \rightarrow \mathbb{R}^n$  as follows:

$$L_I^W(x) := \arg \min\{\|w\| : w \in W, w_I = \Pi_{\pi_I(W)}(x)\}. \quad (56)$$

We further define the lifting operator  $\ell_I^W : \pi_I(W) \rightarrow \pi_J(W^\perp)$  by

$$\ell_I^W(x) := (L_I^W(x))_J = \pi_J(L_I^W(x)), \forall x \in \pi_I(W). \quad (57)$$

Note that if  $x \in \pi_I(W)$ , then  $w = L_I^W(x) = (\ell_I^W(x), x)$  is the minimum-norm point in  $W$  with  $w_I = x$ .

By convention, if  $(I, J)$  is a trivial partition, we define  $L_I^W := \Pi_W$  if  $I = [n]$ , and define  $L_I^W$  to be the linear operator from  $\mathbb{R}^\emptyset := \{0\}$  to  $\mathbb{R}^n$  if  $I = \emptyset$ . We furthermore define  $\ell_I^W$  to be the linear operator from  $W$  to  $\mathbb{R}^\emptyset$  if  $I = [n]$  and from  $\mathbb{R}^\emptyset$  to  $W^\perp$  if  $I = \emptyset$ .

The computation of lifting operators will be crucial to our IPM. The following proposition, whose proof is deferred to later in the section, gives an explicit formula for the associated matrices of the lifting map and operator and shows how to compute them in strongly polynomial time.

**Proposition 6.5** (Computing Lifting Maps). *Let  $\mathbf{A} \in \mathbb{R}^{m \times n}$ ,  $\text{rank}(\mathbf{A}) = m$ , and  $I \cup J = [n]$  be a partition. Then, for  $W = \ker(\mathbf{A})$ , one can in strongly polynomial time compute the associated matrices (as in Definition 2.1) for the lifting map and operator as follows:*

$$\mathcal{M}(L_I^W) = \mathbf{P}_{\bullet, I}(\mathbf{P}_{I, I})^+, \quad \mathcal{M}(\ell_I^W) = \mathbf{P}_{J, I}(\mathbf{P}_{I, I})^+,$$

where  $\mathbf{P} := \mathbf{I}_n - \mathbf{A}^\top(\mathbf{A}\mathbf{A}^\top)^{-1}\mathbf{A}$  is the orthogonal projection onto  $W$ , i.e., the matrix associated with  $\Pi_W$ .

Towards proving the above, we first show that  $L_I^W$  and  $\ell_I^W$  are both well-defined linear operators, which is not directly obvious from the definitions. Linearity of these operators, as well as other key properties, is proven in Lemma 6.6 and Lemma 6.7 below.

**Lemma 6.6.** *For a linear subspace  $W \subseteq \mathbb{R}^n$ , and a partition  $I \cup J = [n]$ , then the function  $L_I^W : \mathbb{R}^I \rightarrow \mathbb{R}^n$  as defined in Definition 6.4 is a linear operator. In particular, for  $x \in \mathbb{R}^I$ ,  $w = L_I^W(x)$  is the unique solution to the following linear system:*

$$\begin{aligned} w &\in \Pi_W(\mathbb{R}_I^n), \\ \pi_I(w) &= \Pi_{\pi_I(W)}(x). \end{aligned} \quad (58)$$

Furthermore,  $\pi_I(W) = \pi_I(\Pi_W(\mathbb{R}_I^n))$ .

*Proof.* If  $(I, J)$  is a trivial partition, the characterization follows directly by our convention, so we may assume that  $(I, J)$  is non-trivial.

By construction, the solution set of  $L_I^W(x)$  is non-empty. Furthermore, the minimum norm solution exists and is unique by strict convexity of the squared Euclidean norm. Thus,  $w = L_I^W(x)$  is well-defined. The definition requires  $w_I = \pi_I(w) = \Pi_{\pi_I(W)}(x)$ ; we show that subject to this,  $w = L_I^W(x)$  if and only if  $w \in \Pi_W(\mathbb{R}_I^n)$ . By Lagrangian duality, using that the gradient of the squared norm objective is  $\nabla(\|w\|^2) = 2w$ , it follows that  $w$  is optimal to the program in the definition (56) if and only if

$$\begin{aligned} w_I &= \Pi_{\pi_I(W)}(x), \\ w &\in W, \quad \text{and} \\ w &\perp (W \cap \mathbb{R}_I^n), \end{aligned} \quad (59)$$

where we have used that the difference of any two feasible solutions to (56) lives in  $W \cap \mathbb{R}_I^n$ . The last requirement can be written as  $w \in (W \cap \mathbb{R}_I^n)^\perp = W^\perp + \mathbb{R}_I^n$ , where the equality follows by Proposition 2.3. By orthogonal decomposition, we note that  $\Pi_W(\mathbb{R}_I^n) = W \cap (W^\perp + \mathbb{R}_I^n)$ .

Thus, the last two requirements are equivalent to  $w \in \Pi_W(\mathbb{R}_I^n)$ , completing the proof. Further, since (59) can be equivalently written as a linear system of equations, it follows that  $L_I^W$  is a linear map.

We now prove the furthermore:  $\pi_I(W) = \pi_I(\Pi_W(\mathbb{R}_I^n))$ . Since  $\Pi_W(\mathbb{R}_I^n) \subseteq W$ , we have  $\pi_I(\Pi_W(\mathbb{R}_I^n)) \subseteq \pi_I(W)$ . To show the reverse containment  $\pi_I(W) \subseteq \pi_I(\Pi_W(\mathbb{R}_I^n))$ , take any  $x \in \pi_I(W)$ . We have that  $L_I^W(x) \in \Pi_W(\mathbb{R}_I^n)$  by the first part, and  $\pi_I(L_I^W(x)) = x$ . Therefore,  $\pi_I(W) \subseteq \pi_I(\Pi_W(\mathbb{R}_I^n))$ , as needed.  $\square$

**Lemma 6.7.** *For a linear subspace  $W \subseteq \mathbb{R}^n$ , a partition  $I \cup J = [n]$ ,  $\ell_I^W : \pi_I(W) \rightarrow \pi_J(W^\perp)$  as in Definition 6.4 is a well-defined linear operator. Moreover,  $\text{im}(\ell_I^W) = \pi_J(W) \cap \pi_J(W^\perp)$ ,  $\ker(\ell_I^W) = \pi_I(W \cap \mathbb{R}_I^n)$  and  $\text{im}(\text{ad}(\ell_I^W)) = \pi_I(W) \cap \pi_I(W^\perp)$ .*

*Proof.* If  $(I, J)$  is a trivial partition, the lemma follows trivially by the definition of  $L_I^W$ , so we may assume that  $(I, J)$  is non-trivial. We first prove the moreover statement  $\text{im}(\ell_I^W) = \pi_J(W) \cap \pi_J(W^\perp)$ . Assuming this, the well-definedness of  $\ell_I^W : \pi_I(W) \rightarrow \pi_J(W^\perp)$  follows immediately from the inclusion  $\text{im}(\ell_I^W) \subseteq \pi_J(W^\perp)$ . Furthermore, linearity of  $\ell_I^W$  follows from linearity of  $L_I^W$ , proved in Lemma 6.6, and the definition  $\ell_I^W(x) = \pi_J(L_I^W(x))$ ,  $\forall x \in \pi_I(W)$ .

We start by showing the inclusion  $\text{im}(\ell_I^W) \subseteq \pi_J(W) \cap \pi_J(W^\perp)$ . For  $x \in \pi_I(W)$ , by Lemma 6.6 we have that  $L_I^W(x) \in \Pi_W(\mathbb{R}_I^n)$  and  $\ell_I^W(x) = \pi_J(L_I^W(x))$ . Since  $\Pi_W(\mathbb{R}_I^n) = W \cap (W^\perp + \mathbb{R}_I^n)$ , we see that  $\ell_I^W(x) \in \pi_J(\Pi_W(\mathbb{R}_I^n)) \subseteq \pi_J(W) \cap \pi_J(W^\perp + \mathbb{R}_I^n) = \pi_J(W) \cap \pi_J(W^\perp)$ , as needed. We now show the reverse inclusion  $\pi_J(W) \cap \pi_J(W^\perp) \subseteq \text{im}(\ell_I^W)$ . Take  $y \in \pi_J(W) \cap \pi_J(W^\perp)$ . By definition, there exists  $w \in W$  such that  $w_J = y$ . It now suffices to show that  $L_I^W(w_I) = w$ , since then  $\ell_I^W(w_I) = y$ . Clearly  $\pi_I(L_I^W(w_I)) = w_I$  since  $w \in W$ . By Lemma 6.6, it suffices to show that  $w \in \Pi_W(\mathbb{R}_I^n) = W \cap (W^\perp + \mathbb{R}_I^n)$ . From here, we see that  $w \in W^\perp + \mathbb{R}_I^n \Leftrightarrow \pi_J(w) \subseteq \pi_J(W^\perp + \mathbb{R}_I^n) = \pi_J(W^\perp)$ . The inclusion now follows since  $w_J = y \in \pi_J(W^\perp)$  by assumption.

We now characterize the kernel of  $\ell_I^W$  by

$$\ker(\ell_I^W) = \{x \in \pi_I(W) : \ell_I^W(x) = \mathbf{0}_J\} = \{x \in \pi_I(W) : (x, \mathbf{0}_J) \in W\} = \pi_I(W \cap \mathbb{R}_I^n),$$

as needed. Finally, we characterize  $\text{im}(\text{ad}(\ell_I^W))$ . By Proposition 2.7, we have that

$$\begin{aligned} \text{im}(\text{ad}(\ell_I^W)) &= \pi_I(W) \cap \ker(\ell_I^W)^\perp = \pi_I(W) \cap \pi_I(W \cap \mathbb{R}_I^n)^\perp \\ &= \pi_I(W) \cap \pi_I((W \cap \mathbb{R}_I^n)^\perp \cap \mathbb{R}_I^n) \\ &= \pi_I(W) \cap \pi_I((W^\perp + \mathbb{R}_I^n) \cap \mathbb{R}_I^n) = \pi_I(W) \cap \pi_I(W^\perp), \end{aligned}$$

as needed.  $\square$

We now have the tools to prove Proposition 6.5.

*Proof of Proposition 6.5.* To begin, recall that  $\mathbf{A}^\top(\mathbf{A}\mathbf{A}^\top)^{-1}\mathbf{A}$  is the orthogonal projection onto  $\text{im}(\mathbf{A}^\top) = \ker(\mathbf{A})^\perp = W^\perp$ , and hence  $\mathbf{P} = \mathbf{I}_n - \mathbf{A}^\top(\mathbf{A}\mathbf{A}^\top)^{-1}\mathbf{A}$  is the orthogonal projection onto  $W$ . Furthermore, the inverse  $(\mathbf{A}\mathbf{A}^\top)^{-1}$  is well-defined since  $\mathbf{A}$  has full row rank.

To justify the formula for  $\mathcal{M}(L_I^W)$ , it suffices to show that  $L_I^W(x) = \mathbf{P}_{\bullet, I}(\mathbf{P}_{I, I})^+x$  for any  $x \in \mathbb{R}^I$  since the input space of  $L_I^W$  is  $\mathbb{R}^I$ . By Lemma 6.6, we have that  $w = L_I^W(x)$  is the unique solution to  $w_I = \Pi_{\pi_I(W)}(x) = x$  and  $w \in \Pi_W(\mathbb{R}_I^n) = \mathbf{P}(\mathbb{R}_I^n)$ . By Proposition 2.14 part (1), we have  $\mathbf{P}_{I, I}(\mathbf{P}_{I, I})^+x = \Pi_{\text{im}(\mathbf{P}_{I, I})}x = \Pi_{\pi_I(W)}x$ , where the last equality is  $\pi_I(W) = \pi_I(\mathbf{P}(\mathbb{R}_I^n)) = \text{im}(\mathbf{P}_{I, I})$  in Lemma 6.6. Since by construction  $\mathbf{P}_{\bullet, I}(\mathbf{P}_{I, I})^+x \in \mathbf{P}(\mathbb{R}_I^n)$ , by uniqueness we have that  $w = \mathbf{P}_{\bullet, I}(\mathbf{P}_{I, I})^+x$  as needed.

We now justify the formula for  $\mathcal{M}(\ell_I^W)$ . The identity  $\ell_I^W(x) = \mathbf{P}_{J, I}(\mathbf{P}_{I, I})^+x$ , for  $x \in \pi_I(W)$ , follows directly from  $\ell_I^W(x) := \pi_J(L_I^W(x))$ . To conclude the proof, it suffices to show that  $\text{im}((\mathbf{P}_{J, I}(\mathbf{P}_{I, I})^+)^\top) \subseteq \text{im}(((\mathbf{P}_{I, I})^+)^\top) \subseteq \pi_I(W)$ , recalling that  $\ell_I^W : \pi_I(W) \rightarrow \pi_J(W^\perp)$ . Since  $\ker((\mathbf{P}_{I, I})^+) = \ker((\mathbf{P}_{I, I})^\top)$  by definition of the pseudoinverse, we have that  $\text{im}(((\mathbf{P}_{I, I})^+)^\top) = \text{im}(\mathbf{P}_{I, I}) = \pi_I(W)$ , where the first equality is by Proposition 2.7 part (2). This proves the statement.

For the strongly polynomial computability, this follows directly from the fact that computing matrix products, inverses and pseudoinverses (Proposition 2.15) is strongly polynomial.  $\square$

To conclude this section, we give the fundamental duality relation between lifting operators, which will be crucial to the analysis of the IPM.

**Lemma 6.8.** *For a linear subspace  $W \subseteq \mathbb{R}^n$ , and a partition  $I \cup J = [n]$ ,  $\ell_I^W = -\text{ad}(\ell_J^{W^\perp})$ . In particular,  $\sigma^+(\ell_I^W) = \sigma^+(\ell_J^{W^\perp})$ .*

*Proof.* If  $(I, J)$  is a trivial partition, the the statement follows by our definition of  $\ell_I^W$ , so may assume that  $(I, J)$  is non-trivial. To prove the statement, it suffices to show that for all  $x \in \pi_I(W)$  that

$$L_I^W(x) = (x, \ell_I^W(x)) = \left(x, -\text{ad}(\ell_J^{W^\perp})(x)\right).$$

Letting  $z := \left(x, -\text{ad}\left(\ell_J^{W^\perp}\right)(x)\right)$ , by Lemma 6.6, it suffices to show that  $z \in \Pi_W(\mathbb{R}^n)$   
 $= W \cap (\mathbb{R}_I^n + W^\perp)$ . To show that  $z \in W$ , we must show that  $\langle v, z \rangle = 0, \forall v \in W^\perp$ . For  $v \in W^\perp$ , we see that

$$\begin{aligned} \langle z, v \rangle &= \langle x, v_I \rangle - \left\langle \text{ad}\left(\ell_J^{W^\perp}\right)(x), v_J \right\rangle \\ &= \langle x, v_I \rangle - \left\langle x, \ell_J^{W^\perp}(v_J) \right\rangle \quad (\text{since } v_J \in \pi_J(W^\perp)) \\ &= \left\langle x, v_I - \ell_J^{W^\perp}(v_J) \right\rangle = \left\langle L_I^W(x), (v_I - \ell_J^{W^\perp}(v_J), 0_J) \right\rangle \\ &= \left\langle L_I^W(x), v - L_J^{W^\perp}(v_J) \right\rangle = 0, \end{aligned}$$

where the last equality follows since  $L_I^W(x) \in W$  and  $v - L_J^{W^\perp}(v_J) \in W^\perp$ .

The inclusion  $z \in \mathbb{R}_I^n + W^\perp$  follows directly from  $z_J = -\text{ad}\left(\ell_J^{W^\perp}\right)(x) \in \pi_J(W^\perp)$  by definition of the adjoint. Thus,  $\ell_I^W = -\text{ad}\left(\ell_J^{W^\perp}\right)$  as needed.

The equality of the positive singular values now follows from Proposition 2.18, noting that  $\sigma^+(\ell_I^W) = \sigma^+(-\text{ad}(\ell_J^W)) = \sigma^+(\ell_J^{W^\perp})$ .  $\square$

### 6.3 Cheap Lift Subspaces

To argue for the usefulness of the SLLS step direction, and to select suitable subspaces  $V$  and  $U$  for a given partition  $(B, N)$ , we recall the discussion of the trust region step from the Introduction (Section 1). As long as the step primal and dual directions  $(\Delta x, \Delta s)$  are feasible to the systems (8) for a suitably small threshold, we are guaranteed to make progress as measured by the primal and dual objective values as in (9).

Simply selecting  $V = W$  and  $U = W^\perp$  would attain the smallest possible objective values; however, the constraints bounding the local norms of  $\Delta x_B^\ell$  and  $\Delta s_N^\ell$  in (8) could be arbitrarily violated. We will select the subspaces  $V$  and  $U$  so that  $\|\hat{x}^{-1}\Delta x_B^\ell\|$  and  $\|\hat{s}^{-1}\Delta s_N^\ell\|$  are guaranteed to be small.

To ensure that the constraints are satisfied, we will restrict the primal and dual movement directions to *cheap lift subspaces*  $V$  and  $U$ . These are formally defined below:

**Definition 6.9** (Cheap Lift Subspace). Let  $W \subseteq \mathbb{R}^n$  be a subspace,  $\Xi \in \mathbb{R}_{++}^n$ ,  $(B, N)$  be a partition of  $[n]$ . Then,  $V \subseteq W$  is a cheap lift subspace for  $(W, \Xi, B, N)$  with lifting cost  $\tau \geq 0$  if

$$\|\Xi_B x_B\| \leq \tau \|\Xi_N x_N\|, \forall x \in V.$$

**Remark 6.10.** Note that for the above inequality to hold, we must have that  $\dim(\pi_N(V)) = \dim(V)$ , since otherwise there exists a vector  $x \in V$  with  $x_B \neq \mathbf{0}_B$  and  $x_N = \mathbf{0}_N$ . In particular, if  $N = \emptyset$ , we must have  $V = \{\mathbf{0}_n\}$ , which has lifting cost 0. Furthermore, if  $N = [n]$ , then  $V = W$  is a cheap lift subspace for  $(W, \Xi, B, N)$  with lifting cost 0 of maximum dimension (recall that  $(\Xi x)_\emptyset := \pi_\emptyset(\Xi x) = 0$  by convention).

In the context of solving the primal trust-region program, if  $V \subseteq W$  is a cheap lift subspace with respect to  $(W, \hat{x}^{-1}, B, N)$  with lifting cost  $\tau = \frac{\nu}{\sqrt{n}}$ , the primal subspace LLS direction with respect to  $(W, V, N, B)$

$$\Delta x^\ell := \arg \min_{\Delta x \in V} \|\hat{x}_N^{-1}(x_N + \Delta x_N)\|$$

automatically satisfies the trust-region constraint  $\|\hat{x}_B^{-1}\Delta x_B^\ell\| \leq \nu$ . This is because  $\hat{x}_N^{-1}\Delta x_N^\ell$  is the orthogonal projection of  $-\hat{x}_N^{-1}x_N = -\hat{\xi}_N$  onto the subspace  $\hat{x}_N^{-1}\pi_N(V)$ , and hence  $\|\hat{x}_N^{-1}\Delta x_N^\ell\| \leq \|\hat{x}_N^{-1}x_N\| = \sqrt{\sum_{i \in N} x_i s_i / \mu} \leq \sqrt{n}$ . By the lifting cost condition on  $V$ , we then have  $\|\hat{x}_B^{-1}\Delta x_B^\ell\| \leq \tau \|\hat{x}_N^{-1}\Delta x_N^\ell\| \leq \sqrt{n}\tau = \nu$ , as needed. Analogously, any cheap lift subspace  $U \subseteq W^\perp$  for  $(W^\perp, \hat{s}^{-1}, N, B)$  with lifting cost at most  $\tau$  will also suffice for the dual subspace LLS direction  $\Delta s^\ell$  as in Definition 6.1 to be feasible.

While low lifting cost subspaces are sufficient to get feasible solutions to the trust-region program, they do not necessarily yield useful approximations. In particular, one can always choose the subspace  $V = \{\mathbf{0}\}$ , which is trivially cheap. To make significant progress along a polarized segment, we will require that the use of cheap lift subspaces of *maximum dimension* subject to a lifting cost bound  $\tau$ . These will allow us to either quickly increase the dimension of the cheap lift subspace, or to get past the end of the of the polarized segment.

We now explain how to find these cheap lift subspaces and what their achievable dimensions are. For this purpose, we will rely on approximate singular subspaces of lifting operator, as defined in Definition 6.4. The following lemma gives the precise relation between cheap lift subspaces and approximate singular subspaces of the lifting operator, and will be the main tool underlying the computation of cheap lift subspaces.

**Lemma 6.11.** *Let  $W \subseteq \mathbb{R}^n$  be a linear subspace,  $\Xi \in \mathbb{R}_{++}^n$ ,  $B \cup N = [n]$  be a partition. Let  $L_N^{\Xi W} : \mathbb{R}^N \rightarrow \mathbb{R}^n$  and  $\ell_N^{\Xi W} : \pi_N(\Xi W) \rightarrow \pi_B(\Xi^{-1}W^\perp)$  be the lifting operator and map as in Definition 6.4.*

*Then, the maximum dimension of a cheap lift subspace  $V$  for  $(W, \Xi, B, N)$  with lifting cost  $\tau \geq 0$  is*

$$\mathcal{C}_\sigma(\ell_N^{\Xi W}, \tau) := |\{i \in [\dim(\pi_N(\Xi W))] : \sigma_i(\ell_N^{\Xi W}) \leq \tau\}|,$$

*the number of singular values of  $\ell_N^{\Xi W}$  of value at most  $\tau$  as in Definition 2.16. Furthermore, if  $S \subseteq \pi_N(\Xi W)$  is a  $\varrho$ -approximate singular subspace for  $\ell_N^{\Xi W}$  of dimension  $\mathcal{C}_\sigma(\ell_N^{\Xi W}, \frac{\tau}{\varrho})$ , then*

$$V := \Xi^{-1}L_N^{\Xi W}(S)$$

*is a cheap lift subspace for  $(W, \Xi, B, N)$  of dimension  $\dim(S) = \mathcal{C}_\sigma(\ell_N^{\Xi W}, \frac{\tau}{\varrho})$  with lifting cost*

$$\sigma_1(\ell_N^{\Xi W}|_S) \leq \varrho \sigma_{\dim(\pi_N(\Xi W)) - \dim(S) + 1}(\ell_N^{\Xi W}) \leq \tau.$$

*Proof.* We first show the upper bound of  $\mathcal{C}_\sigma(\ell_N^{\Xi W}, \tau)$  on the dimension of any cheap lift subspace. Let  $V \subseteq W$  be a cheap lift subspace as above with  $\dim(V) = d$ , and let  $\widehat{V} := \Xi V$  and  $\widehat{W} := \Xi W$ . For any  $y \in \widehat{V} \subseteq \widehat{W}$ , by definition of the lifting map and operator, we have that

$$L_N^{\widehat{W}}(y) \in \widehat{W}, (L_N^{\widehat{W}}(y))_N = y_N, \text{ and } \|y_B\| \geq \|(L_N^{\widehat{W}}(y))_B\| = \|\ell_N^{\widehat{W}}(y_N)\|,$$

where we note that this still holds if  $(B, N)$  is trivial as  $y_\emptyset := \pi_\emptyset(y) = 0$  by convention. Therefore, without loss of generality, we may assume that

$$\widehat{V} = L_N^{\widehat{W}}(\pi_N(\widehat{V})),$$

since this can only decrease lifting cost while maintaining the dimension, recalling that  $\dim(\pi_N(\widehat{V})) = \dim(\widehat{V})$  for any cheap lift subspace. If  $d = 0$  then trivially  $d \leq \mathcal{C}_\sigma(\ell_N^{\widehat{W}}, \tau)$ , and there is nothing left to prove. So assume  $d \geq 1$ . Then, since  $\pi_N(\widehat{V}) \subseteq \pi_N(\widehat{W})$  and  $\dim(\pi_N(\widehat{V})) = d$ , the lifting cost of  $V$  satisfies

$$\begin{aligned} \tau &\geq \max_{x \in V, x_N \neq \mathbf{0}_N} \frac{\|\Xi_B x_B\|}{\|\Xi_N x_N\|} = \max_{y \in \widehat{V}, y_N \neq \mathbf{0}_N} \frac{\|y_B\|}{\|y_N\|} = \max_{z \in \pi_N(\widehat{V}) \setminus \{\mathbf{0}_N\}} \frac{\|\ell_N^{\widehat{W}}(z)\|}{\|z\|} \\ &= \sigma_1(\ell_N^{\widehat{W}}|_{\pi_N(\widehat{V})}) \geq \sigma_{\dim(\pi_N(\widehat{W})) - d + 1}(\ell_N^{\widehat{W}}), \end{aligned} \quad (60)$$

where the last inequality follows from (15). Since  $\tau \geq \sigma_{\dim(\pi_N(\widehat{W})) - d + 1}(\ell_N^{\widehat{W}})$  and the singular values are in non-increasing order, we have that

$$\begin{aligned} d &= |\{i \in [\dim(\pi_N(\widehat{W}))] : \dim(\pi_N(\widehat{W})) - d + 1 \leq i\}| \\ &\leq |\{i \in [\dim(\pi_N(\widehat{W}))] : \sigma_i(\ell_N^{\widehat{W}}) \leq \tau\}| = \mathcal{C}_\sigma(\ell_N^{\widehat{W}}, \tau), \end{aligned}$$

as needed.

For the second part, if  $S \subseteq \pi_N(\widehat{W})$ , then tracing the reduction of the first part backwards, we immediately get from (60) that  $\Xi^{-1}L_N^{\widehat{W}}(S) \subseteq W$  is a cheap lift subspace for  $(W, \Xi, B, N)$  of dimension  $\dim(S)$  and lifting cost  $\sigma_1(\ell_N^{\widehat{W}}|_S)$ . If  $S$  is a  $\varrho$ -approximate singular subspace of  $\ell_N^{\widehat{W}}$  of dimension  $\mathcal{C}_\sigma(\ell_N^{\widehat{W}}, \frac{\tau}{\varrho})$ , then by definition  $\sigma_1(\ell_N^{\widehat{W}}|_S) \leq \varrho \sigma_{\dim(\pi_N(\widehat{W})) - \dim(S) + 1}(\ell_N^{\widehat{W}}) \leq \varrho(\tau/\varrho) = \tau$ , as needed.  $\square$

We remark that so far we have been treating the tasks of finding primal and dual cheap lift subspaces separately. Fortunately, it turns out the singular values of the corresponding primal and dual lifting operators are identical. Thus, the corresponding problems of finding cheap lift subspace on both sides are intimately linked. In particular, we derive the following relation as a direct corollary of Lemma 6.8.

**Corollary 6.12.** Let  $\hat{x}, \hat{s} \in \mathbb{R}_{++}^n$ , satisfy  $\hat{x}\hat{s} = \mu \mathbf{1}_n$  where  $\mu = \bar{\mu}(\hat{x}, \hat{s})$ . Then, for a subspace  $W \subseteq \mathbb{R}^n$  and non-trivial partition  $B \cup N = [n]$ ,  $\sigma^+(\ell_N^{\hat{x}^{-1}W}) = \sigma^+(\ell_B^{\hat{s}^{-1}W^\perp})$ . Furthermore, if  $\sum_{i=1}^r \sigma_i u_i v_i^\top$  is the SVD of  $\ell_N^{\hat{x}^{-1}W}$ , then  $-\sum_{i=1}^r \sigma_i v_i u_i^\top$  is the SVD of  $\ell_B^{\hat{s}^{-1}W^\perp}$ .

Given the above, computing cheap lift subspaces for either the primal or dual side is closely tied to the problem of computing an *approximate* SVD of the lifting operator.

## 6.4 Computing Cheap Lift Subspaces

In this section, we explain how to compute cheap lift subspaces.

### 6.4.1 Approximating the Singular Subspaces

To compute cheap lift subspaces, we will rely on a solver for the  $\varrho$ -SVD problem defined below, which provides the requisite approximate singular value decomposition.

**Definition 6.13** ( $\varrho$ -SVD problem). For  $\varrho \geq 1$ , the input to the  $\varrho$ -SVD problem is a matrix  $\mathbf{M} \in \mathbb{R}^{m \times n}$ , and the output is an *orthogonal* basis  $\mathbf{B} \in \mathbb{R}^{n \times n}$  and  $s \in \mathbb{R}_+^n$ ,  $s$  sorted in non-increasing order, such that

$$\|\mathbf{M}x\| \leq s_i \|x\|, \forall x \in \text{im}(\mathbf{B}_{\geq i}), \forall i \in [n], \quad \text{and} \quad (61)$$

$$\varrho^{-1}s \leq \sigma(\mathbf{M}) \leq s. \quad (62)$$

**Remark 6.14.** The requirement that  $s \geq \sigma(\mathbf{M})$  in (62) is automatically satisfied given (61) and (15). Furthermore, the requirement that  $s$  be sorted in non-increasing order is without loss of generality: if  $s$  satisfies the requirements (61) and (62), then  $\bar{s}$  defined by  $\bar{s}_i = \min_{j \in [i]} s_j$ ,  $\forall i \in [n]$ , also satisfies the requirements and is non-increasing.

**Remark 6.15.** The guarantees of  $\varrho$ -SVD imply that for each  $i \in [m]$ , the subspace  $\text{im}(\mathbf{B}_{\geq i})$  is a  $\varrho$ -approximate singular subspace of  $\mathbf{M}$  (i.e., of the operator  $\mathcal{T}(\mathbf{M})$ ) of dimension  $n - i + 1$ . In particular, given a threshold  $\tau \geq 0$ , letting  $c_\tau = |\{i \in [n] : s_i \leq \tau\}|$ , the subspace  $S_\tau = \text{im}(\mathbf{B}_{\geq n - c_\tau + 1})$  is the subspace of largest dimension in this collection that satisfies  $\sigma_1(\mathbf{M}|_{S_\tau}) \leq \tau$  under the guarantee of (61).

For the sake of being able to execute each iteration of our IPM in strongly polynomial time, we will need a  $\varrho$ -SVD solver that uses a strongly polynomial number of basic arithmetic operations (i.e., depending only on the dimension of the underlying matrix), and that is numerically stable, i.e., that ensures iterates with polynomial bit complexity. In particular, the algorithm cannot rely on computing square root computations.

In Section 9, we show that for any fixed  $\varrho > 1$ ,  $\varrho$ -SVD can be solved in deterministic strongly polynomial time. Our main theorem is as follows.

**Theorem 6.16.** For  $\varrho > 1$ , there is a deterministic algorithm for solving  $\varrho$ -SVD that on an input  $m \times n$  matrix requires  $O(n^2 \max(m, n)^3 \log(n + \frac{1}{\varrho-1}))$  operations and space polynomial in  $n, m, \max\{1, \frac{1}{\varrho-1}\}$  and the bit-encoding length of the input matrix.

We defer the discussion of how this theorem is proved and how it relates to earlier work on approximate SVDs to Section 9.

Our IPM can be implemented using any  $\varrho$ -SVD solver (modulo strongly polynomial considerations). While the iteration complexity of the IPM does depend on the approximation factor  $\varrho$ , this dependency is very mild: as we show in Section 7.3, the iteration bound of our IPM will depend only logarithmically on the approximation factor. The guarantees of our IPM in Theorem 1.4 will be derived by using the above for  $\varrho = 2$ .

### 6.4.2 The Cheap Lift Subspace Algorithm

We now overview the **CHEAP-LIFT-SUBSPACES** algorithm. At a high level, the algorithm is a direct algorithmic implementation of Lemma 6.11 applied to both the primal and dual separately. Specifically, it computes the primal and dual lifting operators, followed by approximate singular subspaces associated with the singular values smaller than 1, and then lifts them to build the cheap lift subspaces. One complication however is that the approximate  $\varrho$ -SVD solver directly computes approximate singular subspaces for the associated matrices of the lifting operators and not the lifting operators themselves. The difference here being that the computed subspaces do not necessarily reside in the input spaces of

---

**Algorithm 1: CHEAP-LIFT-SUBSPACES**


---

**Input** : Matrix  $\mathbf{A} \in \mathbb{R}^{m \times n}$ ,  $\text{rank}(\mathbf{A}) = m$ ,  $\Xi \in \mathbb{R}_{++}^n$ , partition  $B \cup N = [n]$  and SVD approximation factor  $\varrho \geq 1$ .

**Output**: Matrices  $\mathbf{M}^P \in \mathbb{R}^{n \times m_p}$ ,  $\mathbf{M}^D \in \mathbb{R}^{n \times m_d}$ ,  $m_p, m_d \in [n]$ . Letting  $W = \ker(\mathbf{A})$  and  $\ell = \ell_N^{\Xi W}$ ,  $\ell^\perp = \ell_B^{\Xi^{-1} W^\perp}$ , the subspaces  $V = \text{im}(\mathbf{M}^P)$ ,  $U = \text{im}(\mathbf{M}^D)$  satisfy:

1.  $W \cap \mathbb{R}_N^n \subseteq V$  and  $W^\perp \cap \mathbb{R}_B^n \subseteq U$ .
2.  $V$  is a cheap lift subspace for  $(W, \Xi, B, N)$  with lifting cost  $\varrho \sigma_{\dim(\pi_N(W)) - \dim(V) + 1}(\ell)$ , and  $U$  is a cheap lift subspace for  $(W^\perp, \Xi^{-1}, N, B)$  with lifting cost  $\varrho \sigma_{\dim(\pi_B(W^\perp)) - \dim(U) + 1}(\ell^\perp)$ .
3.  $\mathcal{C}_\sigma\left(\ell, \frac{1}{\varrho}\right) \leq \dim(V) \leq \mathcal{C}_\sigma(\ell, 1)$  and  $\mathcal{C}_\sigma\left(\ell^\perp, \frac{1}{\varrho}\right) \leq \dim(U) \leq \mathcal{C}_\sigma(\ell^\perp, 1)$ .

```

1 // Compute Primal and Dual Projection Maps
2  $\mathbf{P}^\perp \leftarrow \text{diag}(\Xi^{-1}) \mathbf{A}^\top (\mathbf{A} \text{diag}(\Xi^{-2}) \mathbf{A}^\top)^{-1} \mathbf{A} \text{diag}(\Xi^{-1})$ ;  $\mathbf{P} \leftarrow \mathbf{I}_n - \mathbf{P}^\perp$ ;
3 if  $B = \emptyset$  then
4    $(\mathbf{M}^P, \mathbf{M}^D) \leftarrow (\mathbf{P}, [\mathbf{0}_n])$ ;
5 else if  $N = \emptyset$  then
6    $(\mathbf{M}^P, \mathbf{M}^D) \leftarrow ([\mathbf{0}_n], \mathbf{P}^\perp)$ ;
7 else
8   // Compute Primal and Dual Lifting Maps and Operators
9    $(\bar{L}, \bar{L}^\perp) \leftarrow (\mathbf{P}_{\bullet, N}(\mathbf{P}_{N, N})^+, \mathbf{P}_{\bullet, B}^\perp(\mathbf{P}_{B, B}^\perp)^+)$ , associated matrices of lifting maps  $L_N^{\Xi W}, L_B^{\Xi^{-1} W^\perp}$ ;
10   $(\bar{\ell}, \bar{\ell}^\perp) \leftarrow (\bar{L}_{B, \bullet}, \bar{L}_{N, \bullet}^\perp)$ , associated matrices of lifting operators  $\ell_N^{\Xi W}, \ell_B^{\Xi^{-1} W^\perp}$ ;
11  // Approximate SVDs of  $\bar{\ell}$  and  $\bar{\ell}^\perp$ 
12   $(\bar{\mathbf{C}}, \bar{s}) \leftarrow \varrho$ -SVD on matrix  $\bar{\ell} \in \mathbb{R}^{B \times N}$ ;  $(\bar{\mathbf{C}}^\perp, \bar{s}^\perp) \leftarrow \varrho$ -SVD on matrix  $\bar{\ell}^\perp \in \mathbb{R}^{N \times B}$ ;
13  // Number of Small Primal and Dual Singular Values
14   $\bar{c}_p \leftarrow |\{i \in [|N|] : \bar{s}_i \leq 1\}|$ ;  $\bar{c}_d \leftarrow |\{i \in [|B|] : \bar{s}_i^\perp \leq 1\}|$ ;
15  // Generators of Primal and Dual Cheap Lift Subspaces
16   $(\mathbf{M}^P, \mathbf{M}^D) \leftarrow (\Xi^{-1} \bar{L}(\bar{\mathbf{C}}_{\geq |N| - \bar{c}_p + 1}^\perp), \Xi \bar{L}^\perp(\bar{\mathbf{C}}_{\geq |B| - \bar{c}_d + 1}^\perp))$ ;
17 return  $(\mathbf{M}^P, \mathbf{M}^D)$ ;

```

---

the lifting operators. Fortunately, the lifting maps automatically correct for this discrepancy as they first project onto the input space. In the proof of correctness, we justify that these projections yield the desired approximate singular subspaces using Lemma 2.26.

**Remark 6.17.** On line 16 of **CHEAP-LIFT-SUBSPACES**, recall that by convention, if  $\bar{c}_p = 0$  then  $\bar{\mathbf{C}}_{\geq |N| - \bar{c}_p + 1}^\perp = \bar{\mathbf{C}}_\emptyset = [\mathbf{0}_B]$ . In particular,  $\mathbf{M}^P = \Xi^{-1} \bar{L}(\bar{\mathbf{C}}_{\geq |N| - \bar{c}_p + 1}^\perp) = [\mathbf{0}_n]$ . Similarly, if  $\bar{c}_d = 0$ , then  $\mathbf{M}^D = \Xi \bar{L}^\perp(\bar{\mathbf{C}}_{\geq |B| - \bar{c}_d + 1}^\perp) = [\mathbf{0}_n]$ . In a similar vein, the checks on whether  $B$  or  $N$  are empty are not strictly speaking necessary. It can be verified that the output of the algorithm is the same without these checks, however we have decided to include them for the sake of clarity.

**Remark 6.18.** One may wonder why **CHEAP-LIFT-SUBSPACES** puts the singular value cutoff at 1 instead of  $\nu/\sqrt{n}$ ,  $\nu = O(\beta)$ , as described in Section 6.1. As the analysis will show, we will only need the cheap lift subspaces when there is a large multiplicative gap between the singular values less than 1 and greater than 1. In particular, all the singular values less than 1 in this case will be significantly smaller than the threshold  $\nu/\sqrt{n}$ .

**Remark 6.19.** Given the duality lifting operators it is natural to wonder whether the cheap lift subspace  $V$  for  $(W, \Xi, B, N)$  outputted by **CHEAP-LIFT-SUBSPACES** can be transformed into a cheap lift subspace  $U$  for  $(W^\perp, \Xi^{-1}, N, B)$  satisfying the output conditions (and vice versa). Assuming  $\Xi = \mathbf{1}_n$  for simplicity, one can in fact show that setting  $U = L_B^{W^\perp}((\ell_B^{W^\perp})^+(\pi_N(V))) + (W^\perp \cap \mathbb{R}_B^n)$  indeed satisfies the desired conditions. For the sake of keeping the analysis and description of the algorithm as simple as possible however, we have opted to avoid using and proving this relation here.

We now prove correctness of **CHEAP-LIFT-SUBSPACES** and give its running time guarantee.

**Lemma 6.20.** *Algorithm **CHEAP-LIFT-SUBSPACES** is correct. Furthermore, assuming a strongly polynomial solver for  $\varrho$ -SVD, **CHEAP-LIFT-SUBSPACES** runs in time strongly polynomial in  $n$ .*

*Proof.* The strong polynomiality claim is immediate from the assumption on the  $\varrho$ -SVD solver and the strong polynomiality of computing pseudoinverses (Proposition 2.15). We thus focus on correctness.

For correctness, we must show that the outputted matrices  $\mathbf{M}^P, \mathbf{M}^D$  and correspondings subspaces  $V = \text{im}(\mathbf{M}^P)$ ,  $U = \text{im}(\mathbf{M}^D)$  satisfy the output conditions (1), (2), (3) and that  $m_p, m_d \in [n]$ .

By Proposition 2.15, on line 2 the algorithm computes the orthogonal projection matrices  $\mathbf{P}^\perp, \mathbf{P} = \mathbf{I}_n - \mathbf{P}^\perp$  onto  $\Xi^{-1} \text{im}(\mathbf{A}^\top) = \Xi^{-1}W^\perp$  and  $(\Xi^{-1}W^\perp)^\perp = \Xi W = \Xi \ker(\mathbf{A})$ , respectively.

After this line, the algorithm is completely symmetric for the primal and dual, so we restrict to proving (1), (2), (3) for the primal subspace  $V$  and proving  $m_p \in [n]$ , recalling the shorthand  $\ell := \ell_N^{\Xi W}$ .

Let us first assume that  $(B, N)$  is a trivial partition. If  $B = \emptyset$ , then  $N = [n]$ ,  $V = \text{im}(\mathbf{P}) = W = W \cap \mathbb{R}_N^n$  and  $m_p = \dim(W) \in [n]$ , which verifies (1) and the requirement on  $m_p$ . Furthermore, since  $\ell := \ell_{[n]}^W : W \rightarrow \mathbb{R}^\emptyset$ , we have  $\dim(V) = \mathcal{C}_\sigma(\ell, 0) = \mathcal{C}_\sigma(\ell, 1/\varrho) = \mathcal{C}_\sigma(\ell, 1)$ , which verifies (3). The lifting cost condition (2) also holds trivially. If  $N = \emptyset$ , then  $V = \text{im}([\mathbf{0}_n]) = \{\mathbf{0}_n\} = W \cap \mathbb{R}_\emptyset^n$  and  $m_p = 1$ , which verifies (1) and the requirement on  $m_p$ . From here, since  $\ell := \ell_\emptyset^W : \mathbb{R}^\emptyset \rightarrow W^\perp$ , we have  $\dim(V) = 0 = \mathcal{C}_\sigma(\ell, \mathbb{R}^+) = \mathcal{C}_\sigma(\ell, 1/\varrho) = \mathcal{C}_\sigma(\ell, 1)$ , which verifies (3). Again, the lifting cost condition (2) is trivial.

Now assume that  $(B, N)$  is non-trivial. Then by Proposition 6.5,  $\bar{L}$  computed on line 9 equals  $\mathcal{M}(L_N^{\Xi W})$  and  $\bar{\ell}$  computed on line 10 equals  $\mathcal{M}(\ell_N^{\Xi W})$ , the associated matrix of the primal lifting map and operator.

Let  $(\bar{\mathbf{C}}, \bar{s})$  be the  $\varrho$ -SVD of  $\bar{\ell} \in \mathbb{R}^{B \times N}$  as computed on line 12 and let  $\bar{c}_p = |\{i \in [N] : \bar{s}_i \leq 1\}|$  be as computed on line 14. By the guarantees of  $\varrho$ -SVD, we have that  $\sigma(\bar{\ell}) \leq \sqrt{\bar{s}} \leq \varrho \sigma(\bar{\ell})$ , and therefore

$$\mathcal{C}_\sigma(\bar{\ell}, 1/\varrho) \leq \bar{c}_p \leq \mathcal{C}_\sigma(\bar{\ell}, 1). \quad (63)$$

Similarly, letting  $\bar{k}_p := \dim(\ker(\bar{\ell})) = \mathcal{C}_\sigma(\bar{\ell}, 0)$ , we have that

$$\sigma(\bar{\ell})_{[[N] \setminus [N - \bar{k}_p]]} = \bar{s}_{[[N] \setminus [N - \bar{k}_p]]} = \mathbf{0}_{[[N] \setminus [N - \bar{k}_p]]}.$$

By the guarantees of  $\varrho$ -SVD, this implies that

$$\text{im}(\bar{\mathbf{C}}_{\geq |N| - \bar{k}_p + 1}) = \ker(\bar{\ell}), \quad (64)$$

since  $\sigma_1(\bar{\ell})_{\text{im}(\bar{\mathbf{C}}_{\geq |N| - \bar{k}_p + 1})} = 0$  and  $\dim(\text{im}(\bar{\mathbf{C}}_{\geq |N| - \bar{k}_p + 1})) = \bar{k}_p = \dim(\ker(\bar{\ell}))$ .

Since  $\bar{c}_p \geq \bar{k}_p$  we get that  $\bar{V}_N := \text{im}(\bar{\mathbf{C}}_{\geq |N| - \bar{c}_p + 1}) \supseteq \text{im}(\bar{\mathbf{C}}_{\geq |N| - \bar{k}_p + 1}) = \ker(\bar{\ell})$  is a  $\varrho$ -approximate singular subspace for  $\bar{\ell}$  of dimension  $\bar{c}_p$  containing  $\ker(\bar{\ell})$ . Let us use the shorthand  $X := \pi_N(\Xi W)$  and  $Y := \pi_B(\Xi^{-1}W^\perp)$ , recalling that  $\ell : X \rightarrow Y$ . Defining  $\hat{V}_N := \Pi_X(\bar{V}_N)$ , by Lemma 2.26 we have that  $\hat{V}_N$  is a  $\varrho$ -approximate singular subspace for  $\ell$  containing  $\ker(\ell)$  of dimension  $\bar{c}_p - \dim(X^\perp)$ . Again by Lemma 2.26, for  $\tau \geq 0$  we have  $\mathcal{C}_\sigma(\ell, \tau) = \mathcal{C}_\sigma(\bar{\ell}, \tau) - \dim(X^\perp)$ , and hence we conclude that

$$\begin{aligned} \mathcal{C}_\sigma(\ell, 1/\varrho) &= \mathcal{C}_\sigma(\bar{\ell}, 1/\varrho) - \dim(X^\perp) \leq \bar{c}_p - \dim(X^\perp) = \dim(\hat{V}_N) \\ &\leq \mathcal{C}_\sigma(\bar{\ell}, 1) - \dim(X^\perp) = \mathcal{C}_\sigma(\ell, 1). \end{aligned} \quad (65)$$

By line 16, recall that  $\mathbf{M}^P = \Xi^{-1} \bar{L}(\bar{\mathbf{C}}_{\geq n - \bar{c}_p + 1})$ . The requirement  $m_p \in [n]$ , then follows from  $m_p = \max\{1, \bar{c}_p\} \in [N] \subseteq [n]$ . Recalling that  $L_N^{\Xi W}(\Pi_X(\cdot)) = L_N^{\Xi W}(\cdot)$  by definition and that  $\bar{L} = \mathcal{M}(L_N^{\Xi W})$ , we have that

$$\begin{aligned} V &:= \text{im}(\mathbf{M}^P) = \Xi^{-1} L_N^{\Xi W}(\text{im}(\bar{\mathbf{C}}_{\geq n - \bar{c}_p + 1})) = \Xi^{-1} L_N^{\Xi W}(\Pi_X \text{im}(\bar{\mathbf{C}}_{\geq n - \bar{c}_p + 1})) \\ &= \Xi^{-1} L_N^{\Xi W}(\hat{V}_N). \end{aligned} \quad (66)$$

By Lemma 6.11, we then get that  $V$  is cheap lift subspace for  $(W, \Xi, B, N)$  of dimension  $\dim(V) = \dim(\hat{V}_N)$  with lifting cost at most

$$\varrho \sigma_{\dim(X) - \dim(V) + 1}(\ell) = \varrho \sigma_{\dim(\pi_N(W)) - \dim(V) + 1}(\ell).$$

Together with (65) this establishes (2), (3) for  $V$ . By Lemma 6.7, we have that  $\pi_N(\Xi W \cap \mathbb{R}_N^n) = \ker(\ell) \subseteq \hat{V}_N$  and therefore

$$W \cap \mathbb{R}_N^n = \Xi^{-1} L_N^{\Xi W}(\pi_N(\Xi W \cap \mathbb{R}_N^n)) \subseteq \Xi^{-1} L_N^{\Xi W}(\hat{V}_N) = V,$$

which verifies (1).  $\square$

## 7 The Subspace Layered Least Squares Algorithm

In this section, we present our SLLS based IPM (**SLLS-IPM**) and give the proof of Theorem 1.1 from the introduction. The pseudocode for our IPM is provided in Section 7.1, and its correctness is proved in Section 7.2. In Section 7.3, we bound the number of iterations needed to traverse a polarized segment of the central path from which Theorem 1.1 follows directly. We note that the iteration bound depends only logarithmically on the SVD approximation factor  $\varrho$ . A more fine-grained amortized iteration bound is deferred to Section 8.

---

### Algorithm 2: SLLS-IPM

---

**Input** : Instance of (LP) with primal constraint matrix  $\mathbf{A} \in \mathbb{R}^{m \times n}$ ,  $\text{rank}(\mathbf{A}) = m$ , and initial iterate  $(x^0, s^0) \in \mathcal{N}^2(\beta)$ ,  $\beta \in (0, 1/6]$ , and SVD approximation factor  $\varrho \geq 1$ .

**Output**:  $(x^*, s^*, v^*, w^*)$  satisfying:

1.  $x^* \in \mathcal{P}, s^* \in \mathcal{D}, \langle x^*, s^* \rangle = 0$ .
2.  $B := \{i \in [n] : x_i^* > 0\}, N := \{i \in [n] : s_i^* > 0\}$  satisfy  $B \cup N = [n]$ .
3.  $v^* \in W^\perp := \text{im}(\mathbf{A}^\top), v_B^* > \mathbf{0}_B, w^* \in W := \ker(\mathbf{A}), w_N^* > \mathbf{0}_N$ .
4.  $\|(x_B^* v_B^*, s_N^* w_N^*) - \mathbf{1}_n\| \leq \beta$ .

```

1  $(x, s) \leftarrow (x^0, s^0)$ ;
2 while  $\bar{\mu}(x, s) > 0$  do
3   Compute corrector direction  $(\Delta x^c, \Delta s^c)$  at  $(x, s)$ ;
4    $(x, s) \leftarrow (x + \Delta x^c, s + \Delta s^c)$ ;
5   Compute affine scaling direction  $(\Delta x^a, \Delta s^a)$  at  $(x, s)$ ;
6   Set  $\alpha^a$  for  $(\Delta x^a, \Delta s^a)$  according to Proposition 2.39 with parameter  $\beta$ ;
7    $\tilde{B} \leftarrow \left\{ i \in [n] : \left| \frac{\Delta x_i^a}{x_i} \right| < \left| \frac{\Delta s_i^a}{s_i} \right| \right\}, \tilde{N} \leftarrow [n] \setminus \tilde{B}$ ;
8    $(\tilde{\mathbf{M}}^P, \tilde{\mathbf{M}}^D) \leftarrow \text{Cheap-Lift-Subspaces}(\mathbf{A}, \frac{1}{x}, \tilde{B}, \tilde{N}, \varrho)$ ;
9    $(\tilde{V}, \tilde{U}) \leftarrow (\text{im}(\tilde{\mathbf{M}}^P), \text{im}(\tilde{\mathbf{M}}^D))$ ;
10  Compute the subspace LLS direction  $(\Delta x^\ell, \Delta s^\ell)$  at  $(x, s)$  with respect to  $(\tilde{B}, \tilde{N}, \tilde{V}, \tilde{U})$  using
    Proposition 6.3 on input  $(x, s, \tilde{B}, \tilde{N}, \tilde{\mathbf{M}}^P, \tilde{\mathbf{M}}^D)$ ;
11  Set  $\alpha^\ell$  for  $(\Delta x^\ell, \Delta s^\ell)$  according to Proposition 2.44 with parameters  $\mu = \bar{\mu}(x, s), \beta$ ;
12  if  $\bar{\mu}(x + \alpha^a \Delta x^a, s + \alpha^a \Delta s^a) < \bar{\mu}(x + \alpha^\ell \Delta x^\ell, s + \alpha^\ell \Delta s^\ell)$  then
13     $\lfloor (x, s) \leftarrow (x + \alpha^a \Delta x^a, s + \alpha^a \Delta s^a)$ ;
14  else
15     $\lfloor (x, s) \leftarrow (x + \alpha^\ell \Delta x^\ell, s + \alpha^\ell \Delta s^\ell)$ ;
16  $(x^*, s^*) \leftarrow (x, s)$ ;
17  $\mu \leftarrow \langle x^* - \Delta x^\ell, s^* - \Delta s^\ell \rangle / n$ ;
18  $(v^*, w^*) \leftarrow (-\Delta s^\ell / \mu, -\Delta x^\ell / \mu)$ ;
19 return  $(x^*, s^*, v^*, w^*)$ ;

```

---

### 7.1 Description of the Algorithm

We are ready to describe the predictor-corrector algorithm **SLLS-IPM**, shown in Algorithm 2. We are given a starting iterate  $z = (x, s) \in \mathcal{N}^2(\beta)$ . In each iteration, we first compute a corrector step to move  $z$  into the  $\mathcal{N}^2(\beta/2)$  neighborhood. We then compute the affine scaling direction  $(\Delta x^a, \Delta s^a)$  at  $z$  and identify the associated partition  $(\tilde{B}_z, \tilde{N}_z)$ . Using this partition, we compute cheap lift subspaces  $\tilde{V}$  and  $\tilde{U}$  using **CHEAP-LIFT-SUBSPACES** (Algorithm 1) from Section 6.4.

We then compute the subspace LLS direction  $(\Delta x^\ell, \Delta s^\ell)$  for  $(\tilde{B}, \tilde{N}, \tilde{V}, \tilde{U})$ . For both directions, we then compute the feasible step-lengths according to the bounds in Proposition 2.37 and Proposition 2.44, and use the better of these two possible steps to obtain the next iterate.

Once the algorithm has found optimal solutions  $(x^*, s^*) \in \bar{\mathcal{N}}^2(\beta)$ , that is, satisfying  $\bar{\mu}(x^*, s^*) = 0$ , it uses the direction of the last segment of the central path to compute certificates  $(v^*, w^*)$  that certify that the optimal solutions are close to the analytic centers of the respective optimal faces.

**Remark 7.1** (Solving Linear Programs to Target Accuracy  $\varepsilon \geq 0$ ). One can easily modify **SLLS-IPM** to terminate once it has found a primal and dual solution  $(x, s)$  with gap  $\langle x, s \rangle \leq \varepsilon$ , where  $\varepsilon \geq 0$  is the target accuracy. This is achieved by changing the while loop check on line 1 from  $\bar{\mu}(x, s) > 0$  to  $\langle x, s \rangle > \varepsilon$ . Once the while loop terminates, if  $\varepsilon > 0$ , we immediately return  $(x, s) \in \bar{\mathcal{N}}^2(\beta)$  and skip the computation of the additional certificates  $(v^*, w^*)$ . If  $\varepsilon = 0$ , the IPM remains unchanged. We prove approximately optimal iteration bounds for any target gap  $\varepsilon \geq 0$  in Section 7.3, and give more refined amortized bounds in Section 8.

**Remark 7.2** (Choice of Cheap Lift Subspace Scaling). One may wonder why we use the scaling  $1/x$  instead of the normalized scaling  $1/\hat{x} := \sqrt{\frac{s}{x\bar{\mu}(x,s)}}$  in the call **Cheap-Lift-Subspaces**( $\mathbf{A}, \frac{1}{x}, \tilde{B}, \tilde{N}, \varrho$ ) on line 8, which would have been more adapted to the definition of the SLLS direction computed on line 10. The main reason is to avoid the need to compute square roots within the algorithm, which is not a strongly polynomial operation. Relying on the fact that  $x \approx \hat{x}$  (see Proposition 2.42), this choice of rescaling is safe and has little additional impact on the analysis or the iteration bound.

## 7.2 Correctness

In this subsection, we prove that upon termination, the **SLLS-IPM** satisfies its output requirements, namely it outputs optimal primal and dual solutions that are close to the analytic centers of the corresponding optimal faces. This will mainly depend on the guarantees on the computed step lengths for subspace LLS and affine scaling, which are given in Proposition 2.44 and Proposition 2.39.

**Lemma 7.3.** *The output of **SLLS-IPM** is correct. Furthermore, each iteration of the algorithm runs in strongly polynomial time.*

*Proof.* We first show that the output  $(x^*, s^*, v^*, w^*)$  satisfies the requirements (1), (2),(3),(4) listed in the output description.

To argue this, we first claim that during the last iteration of the while loop, either the affine scaling step length  $\alpha^a$  or the subspace LLS step length  $\alpha^\ell$  equals 1. Firstly, by the assumption that  $\bar{\mu}(x_0, s_0) > 0$ , the algorithm enters the while loop. Secondly, for any iterate  $(x, s)$ , by Proposition 2.37 part (i) and (ii), the corrector step leaves  $\bar{\mu}(x, s)$  unchanged, whereas the affine scaling step satisfies  $\bar{\mu}(x + \alpha^a \Delta x^a, s + \alpha^a \Delta s^a) = (1 - \alpha^a) \bar{\mu}(x, s)$ . Similarly, by Proposition 2.44,  $9/8(1 - \alpha^\ell) \bar{\mu}(x, s) \geq \bar{\mu}(x + \alpha^\ell \Delta x^\ell, s + \alpha^\ell \Delta s^\ell) \geq 7/8(1 - \alpha^\ell) \bar{\mu}(x, s)$ . Thus, the only way to exit the loop, corresponding to the condition  $\bar{\mu}(x, s) = 0$ , is if  $\max\{\alpha^a, \alpha^\ell\} = 1$ .

We now further claim that upon exiting the while loop we have  $\max\{\alpha^a, \alpha^\ell\} = \alpha^\ell = 1$ . For this purpose, let  $(x, s) \in \mathcal{N}^2(\beta/2)$  be the iterate computed on line 4 during the last iteration of the while loop. From here, let  $(\Delta x^a, \Delta s^a), (\Delta x^\ell, \Delta s^\ell)$ , be the affine scaling and subspace LLS directions computed at  $(x, s)$ , and let  $(\tilde{B}, \tilde{N}), (\tilde{V}, \tilde{U})$  be the corresponding associated partition and cheap lift subspaces at  $(x, s)$ . We now claim that if  $\alpha^a = 1$ , then the affine scaling and subspace LLS directions are equal. Assuming  $\alpha^a = 1$ , we have that  $(x + \Delta x^a, s + \Delta s^a) \in \mathcal{P} \times \mathcal{D}$  is an optimal primal-dual pair. In particular,  $\mathbf{0}_n = (x + \Delta x^a)(s + \Delta s^a) = \Delta x^a \Delta s^a$ , where the first equality is by complementary slackness and the second equality is by the defining equation for the affine scaling direction (20). Recalling that  $\tilde{B} = \{i \in [n] : |\Delta x_i^a|/x_i < |\Delta s_i^a|/s_i\}$ ,  $\tilde{N} = [n] \setminus \tilde{B}$  and that  $x, s > \mathbf{0}_n$ , we conclude that  $\Delta x^a = (\mathbf{0}_{\tilde{B}}, -x_{\tilde{N}}) \in W \cap \mathbb{R}_{\tilde{N}}^n$  and  $\Delta s = (-s_{\tilde{B}}, \mathbf{0}_{\tilde{N}}) \in W^\perp \cap \mathbb{R}_{\tilde{B}}^n$ . Therefore by the output guarantee (1) of **CHEAP-LIFT-SUBSPACES**, we have that  $(\Delta x^a, \Delta s^a) \in \tilde{V} \times \tilde{U}$ . Since  $\|(x_{\tilde{N}} + x_{\tilde{N}}^a, s_{\tilde{B}} + s_{\tilde{B}}^a)\| = 0$ , we must have that  $(x^a, s^a) \in \tilde{V} \times \tilde{U}$  are the (unique) optimal solutions to the subspace LLS programs (53) and (54) at  $(x, s)$  with respect to  $(\tilde{B}, \tilde{N}, \tilde{V}, \tilde{U})$ . In particular,  $(\Delta x^a, \Delta s^a) = (\Delta x^\ell, \Delta s^\ell)$ , as claimed. By Proposition 2.44, using that  $0 = \|(x + \Delta x^\ell)(s + \Delta s^\ell)\| = \|\Delta x^\ell \Delta s^\ell\|$ , we get  $\alpha^\ell = 1$ , as needed.

Let  $(x^*, s^*)$  be the iterate defined right after the while loop on line 16, and let  $B = \text{supp}(x^*)$  and  $N = \text{supp}(s^*)$ . Since  $\alpha^\ell = 1$  by the above, we have that  $(x^*, s^*) = (x + \Delta x^\ell, s + \Delta s^\ell) \in \mathcal{P} \times \mathcal{D}$  and  $\bar{\mu}(x^*, s^*) = 0$ . Therefore,  $(x^*, s^*)$  is an optimal primal-dual pair, which proves (1).

Let  $\mu$  be as defined on line 17. By the above,  $\mu = \bar{\mu}(x, s)$ . By the guarantees of Proposition 2.44 used on line 11, we must have that  $\|\Delta x^\ell \Delta s^\ell\| \leq \beta\mu/9$  and  $\|(x + \Delta x^\ell)(s + \Delta s^\ell)\| = 0$ . For  $i \in [n]$ , using that  $|\Delta x_i^\ell \Delta s_i^\ell| \leq \beta\mu/9 < (1 - \beta/2)\mu \leq x_i s_i$  (Proposition 2.31) and  $(x_i + \Delta x_i^\ell)(s_i + \Delta s_i^\ell) = 0$ , we conclude that either  $x_i^* = x_i + \Delta x_i^\ell > 0$  and  $s_i^* = s_i + \Delta s_i^\ell = 0$ , and thus  $i \in B$ , or that  $x_i^* = x_i + \Delta x_i^\ell = 0$  and  $s_i^* = s_i + \Delta s_i^\ell > 0$ , and thus  $i \in N$ . In particular, we see that  $(B, N)$  partitions  $[n]$ . This verifies the output guarantee (2).

Given the above, for  $(v^*, w^*) := (-\Delta s^\ell/\mu, -\Delta x^\ell/\mu) \in W^\perp \times W$  as set on line 18, we have that  $v_B^* = s_B/\mu > \mathbf{0}_B$  and  $u_N^* = x_N/\mu > \mathbf{0}_N$  which verifies (3). We now establish (4) as follows:

$$\begin{aligned} \|(x_B^* v_B^*, s_N^* w_N^*) - \mathbf{1}_n\| &= \left\| \frac{((x_B + \Delta x_B^\ell) s_B, (s_N + \Delta s_N^\ell) x_N)}{\mu} - \mathbf{1}_n \right\| \\ &= \left\| \frac{x s}{\mu} - \mathbf{1}_n - \frac{\Delta x^\ell \Delta s^\ell}{\mu} \right\| \leq \left\| \frac{x s}{\mu} - \mathbf{1}_n \right\| + \left\| \frac{\Delta x^\ell \Delta s^\ell}{\mu} \right\| \\ &\leq \beta/2 + \beta/9 \leq \beta. \end{aligned} \tag{67}$$

For the strongly polynomial bound on each iteration, we must simply show that all the computations performed within the while loop are strongly polynomial. In particular, we must check the corrector, affine scaling and subspace LLS steps can all be computed in strongly polynomial, and that the calls to **CHEAP-LIFT-SUBSPACES** run in strongly polynomial time (under the assumption that the  $\rho$ -SVD solver is strongly polynomial). These claims are verified in Proposition 2.39 (corrector and affine scaling), Proposition 6.3 (subspace LLS direction), Proposition 2.44 (subspace LLS step length) and Lemma 6.20 (**CHEAP-LIFT-SUBSPACES**). Thus, each iteration runs in strongly polynomial time as needed.  $\square$

### 7.3 Bounding the Number of Iterations to Traverse a Polarized Segment

In this section, we prove the following bound.

**Theorem 7.4.** *For  $\mu_0 > \mu_1 \geq 0$ , let  $\text{CP}[\mu_1, \mu_0]$  be  $\gamma$ -polarized for  $\gamma \in (0, 1]$ . Then, given an iterate  $z \in \mathcal{N}^2(\beta)$ ,  $\beta \in (0, 1/6]$ , with parameter  $\bar{\mu}(z) \in (\mu_1, \mu_0)$ , the algorithm **SLLS-IPM** (Algorithm 2) takes  $O\left(\frac{n^{1.5}}{\beta} \log\left(\frac{n\mu}{\beta\gamma}\right)\right)$  many iterations to find  $z' \in \overline{\mathcal{N}}^2(\beta)$  such that  $\bar{\mu}(z') \leq \mu_1$ .*

Together with Theorem 3.5, we obtain the proof of Theorem 1.1.

*Proof of Theorem 1.1.* As in the statement of the theorem, let  $\Gamma: (\mu_1, \mu_0) \rightarrow \overline{\mathcal{N}}^{-\infty}(\theta)$  be any piecewise linear curve satisfying  $\bar{\mu}(\Gamma(\mu)) = \mu$ ,  $\forall \mu \in (\mu_1, \mu_0)$  with  $T$  linear segments. According to Theorem 3.5, the segment  $\text{CP}[\mu_1, \mu_0]$  of the central path can be decomposed into  $T$  segments that are  $\frac{(1-\theta)^2}{16n^3}$ -polarized. By Theorem 7.4, **SLLS-IPM** instantiated with the 2-SVD solver (i.e.,  $\rho = 2$ ) from Theorem 6.16 traverses these segments in  $O\left(\frac{n^{1.5}}{\beta} \log\left(\frac{n}{\beta(1-\theta)}\right) T\right)$  iterations.  $\square$

The stronger form Theorem 1.4 will be shown in Section 8 by amortizing the running time estimates over subsequent polarized segments. We now introduce the main potential for the analysis.

Our main focus will be on the analyzing the evolution of the singular values of the lifting operator in the normalized subspace at the current iterate. We define  $z := (x, s)$  to be a *basic iterate* if it either corresponds to an iterate computed just after the corrector step on line 4 during the course of the while loop of **SLLS-IPM**, in which case  $z \in \mathcal{N}^2(\beta/2)$ , or if it corresponds to the final iterate  $z \in \overline{\mathcal{N}}^2(\beta)$  after the end of the while loop.

For a basic iterate  $z \in \mathcal{N}^2(\beta/2)$  with  $\bar{\mu}(z) \in [\mu_1, \mu_0]$ , we use the shorthands

$$\ell_z := \ell_N^{\hat{x}^{-1}W} \quad \text{and} \quad \ell_z^\perp := \ell_B^{\hat{s}^{-1}W^\perp}$$

to denote the primal and dual lifting operators with respect to  $(B, N)$  in the respective normalized subspaces, as well as  $\sigma(\ell_z)$  to be the vector of singular values of the lifting operator. If  $\mu_1 = 0$  and  $z \in \overline{\mathcal{N}}^2(\beta)$  is the basic iterate with  $\bar{\mu}(z) = 0$ , let us by convention additionally define  $\ell_z$  to be the zero operator  $\pi_N(W)$  to  $\pi_B(W^\perp)$  and  $\ell_z^\perp$  to be the zero operator from  $\pi_B(W^\perp)$  to  $\pi_N(W)$ . Recall that  $\sigma^+(\ell_z) = \sigma^+(\ell_z^\perp)$  by Lemma 6.8, that is, the non-zero singular values of  $\ell_z$  and  $\ell_z^\perp$  are identical. Define the parameter

$$\tau := \frac{\beta}{256n}.$$

The combinatorial potential we use to measure progress on the polarized segment is

$$\zeta(z) := \mathcal{C}_\sigma\left(\ell_z^\perp, \left(\frac{\tau}{\theta}, \infty\right)\right) = \mathcal{C}_\sigma\left(\ell_z, \left(\frac{\tau}{\theta}, \infty\right)\right) := \left| \left\{ i \geq 1 : \sigma_i(\ell_z) > \frac{\tau}{\theta} \right\} \right|, \tag{68}$$

the number of singular values of  $\ell_z$  (and  $\ell_z^\perp$ ) that are larger than the threshold  $\frac{\tau}{\theta}$ . Note that if  $\zeta(z) > 0$ , then

$$\sigma_{\zeta(z)}(\ell_z) = \sigma_{\zeta(z)}(\ell_z^\perp) = \min \left\{ \sigma_i(\ell_z) : \sigma_i(\ell_z) > \frac{\tau}{\theta}, i \geq 1 \right\}.$$

By convention, recall that if  $B = \emptyset$ , then  $\ell_z$  is the unique linear operator from  $W$  to  $\{0\}$ , and if  $N = \emptyset$ , then  $\ell_z$  is the unique linear operator from  $\{0\}$  to  $W^\perp$ . In particular,  $(B, N)$  is a trivial partition, then  $\zeta(z) = 0$ . The key property of **SLLS-IPM** is given by the following lemma.

**Lemma 7.5.** *Let  $z \in \mathcal{N}^2(\beta)$ ,  $\beta \in (0, 1/6]$ ,  $\bar{\mu}(z) \in (\mu_1, \mu_0]$ , be a basic iterate on the polarized segment  $\text{CP}[\mu_1, \mu_0]$  with a polarization partition  $B \cup N = [n]$ . Then, there exists a constant  $C_{\text{it}} \geq 1$ , such that for any basic iterate  $z' \in \bar{\mathcal{N}}^2(\beta)$  computed after at least  $\lceil C_{\text{it}} \frac{\sqrt{n}}{\beta} \log(\frac{n\varrho}{\beta\gamma}) \rceil$  iterations from  $z$ , either  $\bar{\mu}(z') \leq \mu_1$  or  $\mu_1 < \bar{\mu}(z') \leq \bar{\mu}(z)$  and  $\zeta(z') < \zeta(z)$ .*

**Remark 7.6.** In the above lemma, if  $\zeta(z) = 0$ , then  $z'$  must pass the end of segment  $(\mu_1, \mu_0]$ , since the potential is  $\zeta(z')$  a non-negative integer. For convenience, if  $z^*$  denotes the final optimal iterate, we define the next basic iterate after  $z^*$  to be  $z^*$  itself. In this way, Lemma 7.5 still directly applies if we reach  $z^*$  before  $\lceil C_{\text{it}} \frac{\sqrt{n}}{\beta} \log(\frac{n\varrho}{\beta\gamma}) \rceil$  iterations after  $z$ .

*Proof of Theorem 7.4.* The potential  $\zeta(z) \leq \text{rank}(\ell_z) \leq n$  at the start of the segment and decreases by 1 every  $\lceil C_{\text{it}} \frac{\sqrt{n}}{\beta} \log(\frac{n\varrho}{\beta\gamma}) \rceil$  iterations while we remain in the segment by Lemma 7.5.  $\square$

The rest of the section is dedicated to the proof of Lemma 7.5. We will require the following technical lemma which describes the evolution of the singular values of the lifting operator along a polarized segment. The proof is deferred to Section 7.3.1.

**Lemma 7.7** (Stability of singular values on polarized segments). *Let  $\text{CP}[\mu_1, \mu_0]$  be a  $\gamma$ -polarized segment of the central path with partition  $B \cup N = [n]$ . Let  $z \in \mathcal{N}^2(\beta)$ ,  $z' \in \bar{\mathcal{N}}^2(\beta)$ ,  $\beta \in (0, 1/6]$ , such that  $\mu := \bar{\mu}(z)$  and  $\mu' := \bar{\mu}(z')$  satisfy  $\mu_0 \geq \mu \geq \mu' \geq \mu_1$ . Then we have:*

$$\frac{\gamma^2}{4n^2} \cdot \frac{\mu'}{\mu} \sigma(\ell_z) \leq \sigma(\ell_{z'}) \leq \frac{4n^2}{\gamma^2} \cdot \frac{\mu'}{\mu} \sigma(\ell_z). \quad (69)$$

Lemma 7.7 asserts that, up to  $\text{poly}(n/\gamma)$  factors, the singular values *scale down* by a  $\frac{\mu'}{\mu}$  factor. Recalling that the normalized subspaces for  $z, z'$  are  $\widehat{W} = \hat{x}^{-1}W$  and  $\widehat{W}' = \hat{x}'^{-1}W$ , where  $\hat{x} := \sqrt{x\mu/s} \approx x$  and  $\hat{x}' := \sqrt{x'\mu'/s'} \approx x'$ , the above relation follows straightforwardly from the polarization guarantee  $(x_B, x'_B) \approx x_B^{\text{CP}}(\mu_0)$  and  $(x_N, x'_N) \approx (\frac{\mu}{\mu_0} x_N^{\text{CP}}(\mu_0), \frac{\mu'}{\mu_0} x_N^{\text{CP}}(\mu_0))$  (up to  $\text{poly}(n/\gamma)$  multiplicative factors) together with the variational characterization of singular values.

The proof of Lemma 7.5 relies on the following concepts. For a segment  $\text{CP}[\mu', \mu]$ ,  $0 \leq \mu' \leq \mu$  of the central path, we say that **SLLS-IPM** traverses this segment in at most  $T \geq 0$  iterations, if the number of iterations from the first iterate  $z^{\text{start}} \in \mathcal{N}^2(\beta)$  with  $\bar{\mu}(z^{\text{start}}) \leq \mu$  to the first iterate  $z^{\text{end}} \in \mathcal{N}^2(\beta)$  with  $\bar{\mu}(z^{\text{end}}) \leq \mu'$  is at most  $T$ . Important to the analysis is the distinction between *long* and *short* segments. We say that  $\text{CP}[\mu', \mu]$  is *long* if  $\mu/\mu' \geq \text{poly}(\frac{n\varrho}{\beta\gamma})$  and *short* otherwise, where  $\gamma \in (0, 1]$  is the polarization parameter. We recall from (iv) in Proposition 2.37, the standard predictor-corrector IPM traverses  $\text{CP}[\mu', \mu]$  in at most  $\lceil \frac{3\sqrt{n}}{\beta} \log(\frac{\mu}{\mu'}) \rceil$  iterations. In particular, short segments can be traversed in  $O(\frac{\sqrt{n}}{\beta} \log(\frac{n\varrho}{\beta\gamma}))$  iterations. As **SLLS-IPM** always takes predictor steps that are at least as long as the affine scaling steps, the same guarantee holds for our IPM.

*Proof of Lemma 7.5.* The analysis is divided into two cases.

**Case 1:**  $\mu_1 < \bar{\mu}(z) \leq \mu_1 \frac{72n^{1.5}}{\beta\gamma}$  or  $\mathcal{C}_\sigma(\ell_z, (\tau/\varrho, \varrho/\tau]) \neq 0$  If the first condition holds, the segment  $\text{CP}[\mu_1, \bar{\mu}(z)]$  is short, and hence the IPM passes the end of the polarized segment in at most  $O(\frac{\sqrt{n}}{\beta} \log(\frac{n\varrho}{\beta\gamma}))$  iterations, as needed. Let us now assume that  $\mathcal{C}_\sigma(\ell_z, (\tau/\varrho, \varrho/\tau]) \neq 0$  (note that  $(B, N)$  is non-trivial in this case), or equivalently, that  $\zeta(z) \neq 0$  and  $\sigma_{\zeta(z)}(\ell_z) \in (\tau/\varrho, \varrho/\tau]$ .

Using the stability of singular values, we will show that the number of large singular values quickly drops by one as long as we remain in the segment. Let  $z'$  be any basic iterate satisfying  $\bar{\mu}(z') \leq \frac{\tau^2}{\varrho^2} \cdot \frac{\gamma^2}{4n^2} \cdot \bar{\mu}(z)$ . If  $\bar{\mu}(z') \geq \mu_1$  and  $i \geq \zeta(z)$ , then by Lemma 7.7,

$$\sigma_i(\ell_{z'}) \leq \frac{\tau^2}{\varrho^2} \cdot \sigma_i(\ell_z) \leq \frac{\tau^2}{\varrho^2} \cdot \sigma_{\zeta(z)}(\ell_z) \leq \frac{\tau^2}{\varrho^2} \cdot \frac{\varrho}{\tau} = \frac{\tau}{\varrho}. \quad (70)$$

In particular, if  $\bar{\mu}(z') \geq \mu_1$ , then  $\zeta(z') < \zeta(z)$ . Since the IPM traverses  $\text{CP}[\frac{\tau^2}{\varrho^2} \cdot \frac{\gamma^2}{4n^2} \cdot \bar{\mu}(z), \bar{\mu}(z)]$  in at most  $O(\frac{\sqrt{n}}{\beta} \log(\frac{n\varrho}{\beta\gamma}))$  iterations, recalling that  $\tau = \frac{\beta}{256n}$ , this proves the lemma in this case.

**Case 2:**  $\bar{\mu}(z) > \mu_1 \frac{72n^{1.5}}{\beta\gamma}$  and  $\mathcal{C}_\sigma(\ell_z, (\tau/\varrho, \varrho/\tau]) = 0$  We first note that the condition on the singular values can be restated equivalently as  $\zeta(z) = 0$  (that is,  $\mathcal{C}_\sigma(\ell_z, (\frac{\tau}{\varrho}, \infty)) = 0$ ) or  $\zeta(z) > 0$  and  $\sigma_{\zeta(z)}(\ell_z) > \varrho/\tau$ . In this case, there is a large gap in the singular values around the threshold, namely, for  $i > \zeta(z)$ , we have  $\sigma_i(\ell_z) \leq \tau/\varrho$ , and for  $1 \leq i \leq \zeta(z)$ , we have  $\sigma_i(\ell_z) > \varrho/\tau$ .

In this setting, an identical analysis as in (70) applies as long as we can *quickly* compute a basic iterate  $z'$  such that

$$\bar{\mu}(z') \leq \hat{\mu} := \begin{cases} \mu_1, & \text{if } \zeta(z) = 0, \\ \max \left\{ \mu_1, \frac{1}{\sigma_{\zeta(z)}(\ell_z)} \cdot \frac{\tau}{\varrho} \cdot \frac{\gamma^2}{4n^2} \cdot \bar{\mu}(z) \right\}, & \text{if } \zeta(z) \geq 1. \end{cases} \quad (71)$$

That is, for any basic iterate  $z'$  as above, either  $\bar{\mu}(z') \leq \mu_1$  or  $\mu_1 < \bar{\mu}(z') \leq \bar{\mu}(z)$  and  $\zeta(z') < \zeta(z)$ . Note that  $\bar{\mu}(z)/\hat{\mu}$  may be *arbitrarily large* in this setting, and thus, we may need to take a *huge step* down the central path. It is precisely for this purpose that we require subspace LLS steps.

Let  $z^+$  denote the next basic iterate after  $z$ . Then, by Lemma 7.8 below, we have that

$$\begin{aligned} \frac{\bar{\mu}(z^+)}{\hat{\mu}} &\leq \frac{11}{\beta} \begin{cases} \frac{4n^{1.5}\mu_1}{\beta}, & \text{if } \zeta(z) = 0, \\ \frac{4n^{1.5}\mu_1}{\gamma\hat{\mu}} + \frac{12n}{\sigma_{\zeta(z)}(\ell_z)} \frac{\bar{\mu}(z)}{\hat{\mu}}, & \text{if } \zeta(z) \geq 1. \end{cases} \\ &\leq \frac{11}{\beta} \begin{cases} \frac{4n^{1.5}}{\gamma}, & \text{if } \zeta(z) = 0, \\ \frac{4n^{1.5}}{\gamma} + \frac{48n^3\varrho}{\tau\gamma^2}, & \text{if } \zeta(z) \geq 1. \end{cases} \end{aligned} \quad (72)$$

This guarantees that after the next basic iterate  $z^+$  after  $z$ , at most  $O(\frac{\sqrt{n}}{\beta} \log(\frac{n\varrho}{\beta\gamma}))$  additional iterations are needed to obtain  $z'$  as in (71).

Combining the two case analyses, the constant  $C_{\text{it}} \geq 1$  in the lemma statement can be chosen so that  $C_{\text{it}} \frac{\sqrt{n}}{\beta} \log(\frac{n\varrho}{\beta\gamma})$  iterations is an upper bound on 1 plus the number of affine scaling iterations needed to divide the normalized gap by a factor

$$\frac{11}{\beta} \max \left\{ \frac{\varrho^2}{\tau^2} \cdot \frac{4n^2}{\gamma^2}, \frac{4n^{1.5}}{\gamma} + \frac{48n^3\varrho}{\tau\gamma^2} \right\} \leq \left( \frac{n\varrho}{\beta\gamma} \right)^{12}, \quad (73)$$

where we have used  $\tau := \frac{\beta}{256n}$  and  $\beta \in (0, 1/6]$ . Thus, by (iv) in Proposition 2.37, the choice  $C_{\text{it}} = 3 \cdot 12 + 2 = 38$  is sufficient.  $\square$

**Lemma 7.8.** *Let  $\text{CP}[\mu_1, \mu_0]$  be a  $\gamma$ -polarized segment with polarization partition  $B \cup N = [n]$ . Let  $z = (x, s) \in \mathcal{N}(\beta/2)$ ,  $\beta \in (0, 1/6]$  be a basic iterate satisfying  $\mu := \bar{\mu}(z) \in (\frac{72n}{\beta\gamma} \cdot \mu_1, \mu_0]$  and  $\mathcal{C}_\sigma(\ell_z, (\frac{\tau}{\varrho}, \frac{\varrho}{\tau}]) = 0$ . Then, the next basic iterate  $z^+$  computed in SLLS-IPM after  $z$  satisfies*

$$\bar{\mu}(z^+) \leq \frac{11}{\beta} \cdot \begin{cases} \frac{4n^{1.5}\mu_1}{\beta}, & \text{if } \zeta(z) = 0, \\ \frac{4n^{1.5}\mu_1}{\gamma} + \frac{12n}{\sigma_{\zeta(z)}(\ell_z)} \bar{\mu}(z), & \text{if } \zeta(z) \geq 1. \end{cases} \quad (74)$$

It remains to prove Lemma 7.8. We will show the subspace LLS step from  $z = (x, s) \in \mathcal{N}^2(\beta/2)$  achieves (74). We will show this by comparing these steps to the *ideal direction*, which goes straight to the end of the polarized segment, as defined below:

**Definition 7.9** (Ideal Direction). For an  $z \in \mathcal{N}(\beta)$ ,  $\beta \in (0, 1/6]$ , satisfying  $\bar{\mu}(z) > \mu_1$ , we define the *ideal direction* from  $z$  towards  $\mu_1$  to be

$$\Delta z^{\text{ideal}} := z^{\text{CP}}(\mu_1) - z := (x^{\text{CP}}(\mu_1) - x, s^{\text{CP}}(\mu_1) - s) =: (\Delta x^{\text{ideal}}, \Delta s^{\text{ideal}}).$$

That is,  $\Delta z^{\text{ideal}}$  is the direction from the current iterate to the central path point  $z^{\text{CP}}(\mu_1)$ .

We now recall the notation used for cheap lift subspaces in the algorithm. We define

$$\begin{aligned} (\mathbf{M}^{\text{P}}, \mathbf{M}^{\text{D}}) &\leftarrow \text{Cheap-Lift-Subspace} \left( \mathbf{A}, \frac{1}{x}, B, N, \varrho \right), \\ (V, U) &\leftarrow (\text{im}(\mathbf{M}^{\text{P}}), \text{im}(\mathbf{M}^{\text{D}})), \end{aligned} \quad (75)$$

to be the output of the call to algorithm **CHEAP-LIFT-SUBSPACES** with respect to the *true partition*  $(B, N)$ . By the guarantees of **CHEAP-LIFT-SUBSPACES**, recall that if  $(B, N) = (\emptyset, [n])$ , then  $(V, U) = (W, \{\mathbf{0}_n\})$ , and if  $(B, N) = ([n], \emptyset)$ , then  $(V, U) = (\{\mathbf{0}_n\}, W^\perp)$ .

The crux of the argument will be to show that a suitable projection of  $\Delta x^{\text{ideal}}$  onto  $V$  and  $\Delta s^{\text{ideal}}$  onto  $U$  induces a step satisfying (74). We recall the notation  $\hat{\xi} := \sqrt{xs}/\mu$ ,  $(\hat{x}, \hat{s}) := (x\hat{\xi}^{-1}, s\hat{\xi}^{-1})$ ,  $\mu := \bar{\mu}(z)$ ,  $\widehat{W} := \hat{x}^{-1}W$ , and  $\widehat{W}^\perp := \hat{s}^{-1}W^\perp$  from Definition 2.40. Similarly, we let  $\widehat{V} := \hat{x}^{-1}V$ ,  $\widehat{U} := \hat{s}^{-1}U$  and  $(\Delta \hat{x}^{\text{ideal}}, \Delta \hat{s}^{\text{ideal}}) := (\hat{x}^{-1}\Delta x^{\text{ideal}}, \hat{s}^{-1}\Delta s^{\text{ideal}})$ .

We now define the projected ideal directions

$$\Delta x^{\text{proj}} := \arg \min_{v \in V} \|\hat{x}_N^{-1}(v_N - \Delta x_N^{\text{ideal}})\|, \quad \Delta s^{\text{proj}} := \arg \min_{u \in U} \|\hat{s}_B^{-1}(u_B - \Delta s_B^{\text{ideal}})\|,$$

where we recall that the minimizers are unique since  $\dim(\pi_N(V)) = \dim(V)$  and  $\dim(\pi_B(U)) = \dim(U)$ . Define the normalized projections by  $(\Delta \hat{x}^{\text{proj}}, \Delta \hat{s}^{\text{proj}}) := (\hat{x}^{-1}\Delta x^{\text{proj}}, \hat{s}^{-1}\Delta s^{\text{proj}})$ . By construction of the projected ideal directions, note that

$$(\Delta \hat{x}_N^{\text{proj}}, \Delta \hat{s}_B^{\text{proj}}) = (\Pi_{\pi_N(\widehat{V})}(\Delta \hat{x}_N^{\text{ideal}}), \Pi_{\pi_B(\widehat{U})}(\Delta \hat{s}_B^{\text{ideal}})). \quad (76)$$

Let us now define

$$\widehat{V}_N^\perp := \pi_N(\widehat{W}) \cap \pi_N(\widehat{V})^\perp \quad \widehat{U}_B^\perp := \pi_B(\widehat{W}^\perp) \cap \pi_B(\widehat{U})^\perp, \quad (77)$$

to be the corresponding orthogonal complements inside  $\pi_N(\widehat{W})$  and  $\pi_B(\widehat{W}^\perp)$ , where we note that

$$\Delta \hat{x}_N^{\text{ideal}} - \Delta \hat{x}_N^{\text{proj}} \in \widehat{V}_N^\perp \quad \Delta \hat{s}_B^{\text{ideal}} - \Delta \hat{s}_B^{\text{proj}} \in \widehat{U}_B^\perp.$$

Our goal will be to show that  $(\Delta x^{\text{proj}}, \Delta s^{\text{proj}})$  induces a trust-region step achieving the guarantees of (74). This will then imply that the subspace LLS step  $(\Delta x^\ell, \Delta s^\ell)$  also achieves the guarantee in (74).

The analysis relies on two auxiliary lemmas. The first one exposes basic properties of the ideal direction, while the second one gives basic properties of the cheap lift subspaces. The proofs are deferred to subsections 7.3.2 and 7.3.3.

**Proposition 7.10.** *Let  $\text{CP}[\mu_1, \mu_0]$  be a  $\gamma$ -polarized segment with partition  $B \cup N = [n]$ . Let  $z = (x, s) \in \mathcal{N}(\beta)$ ,  $\beta \in (0, 1/6]$ , with  $\mu := \bar{\mu}(z) \in (\frac{72n^{1.5}}{\beta\gamma} \cdot \mu_1, \mu_0]$ . Then, the ideal direction  $\Delta z^{\text{ideal}}$  towards  $\mu_1$  satisfies*

$$\|(\Delta \hat{x}^{\text{ideal}}, \Delta \hat{s}^{\text{ideal}})\| \leq 4n, \quad (78)$$

$$\|(\hat{\xi}_N + \Delta \hat{x}_N^{\text{ideal}}, \hat{\xi}_B + \Delta \hat{s}_B^{\text{ideal}})\| \leq \frac{2n^{1.5}\mu_1}{\gamma\mu} \leq \frac{\beta}{36}. \quad (79)$$

**Lemma 7.11.** *Let  $\text{CP}[\mu_1, \mu_0]$  be a  $\gamma$ -polarized segment with partition  $B \cup N = [n]$ . Let  $z = (x, s) \in \mathcal{N}(\beta)$ ,  $\beta \in (0, 1/6]$ . Let  $V := \text{im}(\mathbf{M}^P)$ ,  $U := \text{im}(\mathbf{M}^D)$  be the subspaces returned by **CHEAP-LIFT-SUBSPACES** on  $(\mathbf{A}, x^{-1}, B, N, \varrho)$  as in (75), and  $\ell_z := \ell_N^{\hat{x}^{-1}W}$ ,  $\ell_z^\perp := \ell_B^{\hat{s}^{-1}W^\perp}$ . Then, if  $\mathcal{C}_\sigma(\ell_z, (\tau/\varrho, \varrho/\tau)) = 0$ , we have*

1.  $\dim(V) = \dim(\pi_N(W)) - \zeta(z)$ ,  $\dim(U) = \dim(\pi_B(W^\perp)) - \zeta(z)$ .
2.  $\|\hat{v}_B\| \leq 2\varrho\sigma_{\zeta(z)+1}(\ell_z) \|\hat{v}_N\| \leq 2\tau \|\hat{v}_N\|$ ,  $\forall \hat{v} \in \widehat{V} := \hat{x}^{-1}V$ , and  
 $\|\hat{u}_N\| \leq 2\varrho\sigma_{\zeta(z)+1}(\ell_z^\perp) \|\hat{u}_B\| \leq 2\tau \|\hat{u}_B\|$ ,  $\forall \hat{u} \in \widehat{U} := \hat{s}^{-1}U$ .

We are ready to prove Lemma 7.8.

*Proof of Lemma 7.8.* We show that the subspace LLS step  $(\Delta x^\ell, \Delta s^\ell)$  with respect to  $(B, N)$  is sufficient to achieve the guarantees of (74). As we show below, this will correspond to a long trust-region step and hence Lemma 5.3 will guarantee that  $(\tilde{B}_z, \tilde{N}_z) = (B, N)$ . In particular, the subspace LLS direction computed by the algorithm will be the same as  $(\Delta x^\ell, \Delta s^\ell)$ .

We now analyze the norms and residuals of the subspace LLS direction  $(\Delta x^\ell, \Delta s^\ell)$  with respect to  $(B, N)$ . We first provide sufficient conditions to ensure that  $z^\ell := (x + \alpha^\ell \Delta x^\ell, s + \alpha^\ell \Delta s^\ell)$ , corresponding to the subspace LLS step applied to  $z$ , satisfies (74). Letting  $\mu := \bar{\mu}(z)$ , we will show below that

$$\|(\Delta \hat{x}_B^\ell, \Delta \hat{s}_N^\ell)\| \leq \frac{\beta}{18}, \quad (80)$$

$$\|(\hat{\xi}_N + \Delta \hat{x}_N^\ell, \hat{\xi}_B + \Delta \hat{s}_B^\ell)\| \leq \begin{cases} \frac{2n^{1.5}\mu_1}{\gamma\mu}, & \text{if } \zeta(z) = 0, \\ \frac{2n^{1.5}\mu_1}{\gamma\mu} + \frac{6n}{\sigma_{\zeta(z)}(\ell_z)}, & \text{if } \zeta(z) \geq 1. \end{cases} \quad (81)$$

By our assumption that  $\mu > \frac{72n^{1.5}}{\beta\gamma} \cdot \mu_1$  and, if  $\zeta(z) \geq 1$ , that  $\sigma_{\zeta(z)}(\ell_z) > \frac{\varrho}{\tau} \geq \frac{256n}{\beta}$ , the upper bound in (81) is always at most  $\beta/18$ . From Proposition 5.1, we now get

$$\delta := \left\| \frac{\Delta x^\ell \Delta s^\ell}{\mu} \right\| \leq \frac{\beta}{9} \quad \text{and}$$

$$\varepsilon := \left\| \frac{(x + \Delta x^\ell)(s + \Delta s^\ell)}{\mu} \right\| \leq \begin{cases} \frac{4n^{1.5}\mu_1}{\gamma}, & \text{if } \zeta(z) = 0, \\ \frac{4n^{1.5}\mu_1}{\gamma\mu} + \frac{12n}{\sigma_{\zeta(z)}(\ell_z)}, & \text{if } \zeta(z) \geq 1. \end{cases}$$

and  $\varepsilon \leq \beta/9$ . Consequently, Proposition 2.44 is applicable. Choosing  $1 - \frac{9\varepsilon}{\beta} \leq \alpha^\ell \leq 1 - \frac{8\varepsilon}{\beta}$ , we have

$$\bar{\mu}(z^\ell) \leq (1 + \frac{1}{8}) \frac{9\varepsilon\mu}{\beta} \leq \frac{11}{\beta} \varepsilon\mu \leq \frac{11}{\beta} \cdot \begin{cases} \frac{4n^{1.5}\mu_1}{\gamma}, & \text{if } \zeta(z) = 0, \\ \frac{4n^{1.5}\mu_1}{\gamma} + \frac{12n}{\sigma_{\zeta(z)}(\ell_z)}\mu, & \text{if } \zeta(z) \geq 1. \end{cases}$$

which is precisely the guarantee required for (74). Since (80) and (81) are at most  $\beta/18 \leq 1/30$ , then assuming these bounds hold, Lemma 5.3 implies that the associated partition  $(\tilde{B}_z, \tilde{N}_z) = (B, N)$ . In particular, if the bounds hold, SLLS-IPM correctly computes the subspace LLS direction  $(\Delta x^\ell, \Delta s^\ell)$ . Since  $\bar{\mu}(z^+) \leq \bar{\mu}(z^\ell)$ , as we take the better of the AS and subspace LLS step, this will yield the claimed bound (74).

We now prove (80). By Lemma 7.11 part (2), we have that the subspace LLS direction  $(\Delta x^\ell, \Delta s^\ell)$  satisfies

$$\|(\Delta \hat{x}_B^\ell, \Delta \hat{s}_N^\ell)\| \leq 2\varrho\sigma_{\zeta(z)+1}(\ell_z) \|(\Delta \hat{x}_N^\ell, \Delta \hat{s}_B^\ell)\| \leq 2\tau \|(\hat{\xi}_N, \hat{\xi}_B)\| = 2\tau\sqrt{n} \leq \frac{\beta}{18},$$

as needed, where the second inequality uses  $(\Delta \hat{x}_N^\ell, \Delta \hat{s}_B^\ell) = (\Pi_{\pi_N(\hat{V})}(\hat{\xi}_N), \Pi_{\pi_B(\hat{U})}(\hat{\xi}_B))$  as per Remark 6.2.

We now prove (81). By definition of subspace LLS (Definition 6.1) and the triangle inequality, we have that

$$\begin{aligned} \|(\hat{\xi}_N + \Delta \hat{x}_N^\ell, \hat{\xi}_B + \Delta \hat{s}_B^\ell)\| &\leq \|(\hat{\xi}_N + \Delta \hat{x}_N^{\text{proj}}, \hat{\xi}_B + \Delta \hat{s}_B^{\text{proj}})\| \\ &\leq \|(\hat{\xi}_N + \Delta \hat{x}_N^{\text{ideal}}, \hat{\xi}_B + \Delta \hat{s}_B^{\text{ideal}})\| + \\ &\quad \|(\Delta \hat{x}_N^{\text{proj}} - \Delta \hat{x}_N^{\text{ideal}}, \Delta \hat{s}_B^{\text{proj}} - \Delta \hat{s}_B^{\text{ideal}})\| \\ &\leq \frac{2n^{1.5}\mu_1}{\gamma\mu} + \|(\Delta \hat{x}_N^{\text{proj}} - \Delta \hat{x}_N^{\text{ideal}}, \Delta \hat{s}_B^{\text{proj}} - \Delta \hat{s}_B^{\text{ideal}})\|, \end{aligned} \quad (82)$$

where the last inequality follows by (79) in Proposition 7.10.

We now bound the second term of (82). Note that if  $\zeta(z) = 0$ , then by Lemma 7.11 part (1), we have that  $\pi_N(\hat{V}) = \pi_N(\hat{W})$  and  $\pi_B(\hat{U}) = \pi_B(\hat{W}^\perp)$  and hence  $\hat{V}_N^\perp = \{\mathbf{0}_n\}$  and  $\hat{U}_B^\perp = \{\mathbf{0}_n\}$ . In particular,

$$\|\Delta \hat{x}_N^{\text{proj}} - \Delta \hat{x}_N^{\text{ideal}}\| = 0 \quad \|\Delta \hat{s}_B^{\text{proj}} - \Delta \hat{s}_B^{\text{ideal}}\| = 0.$$

Thus (81) follows directly from (82). Now assume that  $\zeta(z) \geq 1$ . Note that this implies that  $(B, N)$  is a non-trivial partition. By Lemma 7.11, we have that  $\pi_N(\hat{V})$  is a  $(2\varrho)$ -approximate singular subspace for  $\ell_z$  with dimension  $\dim(\pi_N(\hat{W})) - \zeta(z) < \dim(\pi_N(\hat{W}))$ . By our assumption that  $\mathcal{C}_\sigma(\ell_z, (\tau/\varrho, \varrho/\tau)) = 0$ , we have that

$$\sigma_{\zeta(z)}(\ell_z) > \frac{\varrho}{\tau} = \frac{\varrho^2}{\tau^2} \cdot \frac{\tau}{\varrho} \geq \frac{\varrho^2}{\tau^2} \sigma_{\zeta(z)+1}(\ell_z) \geq 6\varrho\sigma_{\zeta(z)+1}(\ell_z).$$

By Lemma 2.27, recalling that  $\hat{V}_N^\perp := \pi_N(\hat{W}) \cap \pi_N(\hat{V})^\perp$ , we therefore have that

$$\sigma_{\min}(\ell_z|_{\hat{V}_N^\perp})^2 \geq \sigma_{\zeta(z)}(\ell_z)^2 - (2\varrho\sigma_{\zeta(z)+1}(\ell_z))^2 \geq \frac{8}{9}\sigma_{\zeta(z)}(\ell_z)^2. \quad (83)$$

By a symmetric argument, we have that  $\hat{U}_B^\perp := \pi_B(\hat{W}^\perp) \cap \pi_B(\hat{U}_B)^\perp$  is a  $\varrho$ -approximate singular subspace for  $\ell^\perp$  of dimension  $\dim(\pi_B(\hat{W}^\perp)) - \zeta(z)$  and satisfies

$$\sigma_{\min}(\ell_z^\perp|_{\hat{U}_B^\perp})^2 \geq \frac{8}{9}\sigma_{\zeta(z)}(\ell_z^\perp)^2 = \frac{8}{9}\sigma_{\zeta(z)}(\ell_z)^2. \quad (84)$$

We now bound the second term in (82) as follows:

$$\begin{aligned}
\left\| \Delta \hat{x}_N^{\text{proj}} - \Delta \hat{x}_N^{\text{ideal}} \right\|^2 &\leq \frac{\left\| \ell_z (\Delta \hat{x}_N^{\text{proj}} - \Delta \hat{x}_N^{\text{ideal}}) \right\|^2}{\sigma_{\min}(\ell_z |_{\widehat{V}_N^\perp})^2} \leq \frac{\left\| \Delta \hat{x}_B^{\text{proj}} - \Delta \hat{x}_B^{\text{ideal}} \right\|^2}{\sigma_{\min}(\ell_z |_{\widehat{V}_N^\perp})^2} \\
&\leq \frac{2 \left\| \Delta \hat{x}_B^{\text{proj}} \right\|^2 + 2 \left\| \Delta \hat{x}_B^{\text{ideal}} \right\|^2}{\sigma_{\min}(\ell_z |_{\widehat{V}_N^\perp})^2} \\
&\leq \frac{2(2\tau)^2 \left\| \Delta \hat{x}_N^{\text{proj}} \right\|^2 + 2 \left\| \Delta \hat{x}_B^{\text{ideal}} \right\|^2}{\sigma_{\min}(\ell_z |_{\widehat{V}_N^\perp})^2} \\
&\leq \frac{2 \left\| \Delta \hat{x}_N^{\text{ideal}} \right\|^2 + 2 \left\| \Delta \hat{x}_B^{\text{ideal}} \right\|^2}{\sigma_{\min}(\ell_z |_{\widehat{V}_N^\perp})^2} \\
&= \frac{2 \left\| \Delta \hat{x}_N^{\text{ideal}} \right\|^2}{\sigma_{\min}(\ell_z |_{\widehat{V}_N^\perp})^2} \leq \frac{9 \cdot 2 \left\| \Delta \hat{x}_N^{\text{ideal}} \right\|^2}{8\sigma_{\zeta(z)}(\ell_z)^2} = \frac{9 \left\| \Delta \hat{x}_N^{\text{ideal}} \right\|^2}{4\sigma_{\zeta(z)}(\ell_z)^2}. \tag{85}
\end{aligned}$$

By a symmetric argument,

$$\left\| \Delta \hat{s}_B^{\text{proj}} - \Delta \hat{s}_B^{\text{ideal}} \right\|^2 \leq \frac{2 \left\| \Delta \hat{s}_B^{\text{ideal}} \right\|^2}{\sigma_{\min}(\ell_z^\perp |_{\widehat{U}_B^\perp})^2} \leq \frac{9 \left\| \Delta \hat{s}_B^{\text{ideal}} \right\|^2}{4\sigma_{\zeta(z)}(\ell_z)^2}. \tag{86}$$

Combining (85) and (86) together with (78) in Proposition 7.10, we get that

$$\begin{aligned}
\left\| (\Delta \hat{x}_N^{\text{proj}} - \Delta \hat{x}_N^{\text{ideal}}, \Delta \hat{s}_B^{\text{proj}} - \Delta \hat{s}_B^{\text{ideal}}) \right\| &\leq \frac{3 \left\| (\Delta \hat{x}_N^{\text{ideal}}, \Delta \hat{s}_B^{\text{ideal}}) \right\|}{2\sigma_{\zeta(z)}(\ell_z)} \leq \frac{3 \cdot 4n}{2\sigma_{\zeta(z)}(\ell_z)} \\
&= \frac{6n}{\sigma_{\zeta(z)}(\ell_z)}. \tag{87}
\end{aligned}$$

The desired bound (81) now follows by combining (82) and (87).  $\square$

### 7.3.1 Stability of Singular Values on Polarized Segments

We now present the proof of Lemma 7.7 on the evolution of singular values of the map  $\ell_z$  on polarized segments of the central path. This will rely on the next lemma that bounds the change in the singular values under a rescaling of the space.

**Lemma 7.12** (Stability of singular values for multiplicative perturbation). *Let  $W \subseteq \mathbb{R}^n$  be a subspace, and let  $B \cup N = [n]$  be a non-trivial partition. Let  $y \in \mathbb{R}_{++}^n$ , and let  $\ell_N^{y^{-1}W} : \pi_N(y^{-1}W) \rightarrow \pi_B(yW^\perp)$  and  $\ell_N^W : \pi_N(W) \rightarrow \pi_B(W^\perp)$  be defined according to Definition 6.4. Let  $\sigma := \sigma(\ell_N^{y^{-1}W})$  and  $\hat{\sigma} := \sigma(\ell_N^W)$  denote their respective singular values. Then, we have that*

$$\frac{1}{\left\| y_B^{-1} \right\|_\infty \left\| y_N \right\|_\infty} \sigma \leq \hat{\sigma} \leq \left\| y_B \right\|_\infty \left\| y_N^{-1} \right\|_\infty \sigma. \tag{88}$$

*Proof.* We only prove the second inequality. The first inequality then follows by swapping  $y$  with  $y^{-1}$  and  $\sigma$  with  $\hat{\sigma}$ . We denote  $\ell := \ell_N^{y^{-1}W}$  and  $\hat{\ell} := \ell_N^W$ . Since  $y \in \mathbb{R}_{++}^n$ , we clearly have that  $p := \dim(\pi_N(W)) = \dim(\pi_N(y^{-1}W))$ . Therefore  $\sigma$  and  $\hat{\sigma}$  are both vectors in  $\mathbb{R}_+^p$ . By (15), recall that for  $i \in [p]$ , we have that

$$\sigma_k = \min_{\substack{S \subseteq \pi_N(y^{-1}W) \\ \dim(S) \geq p-k+1}} \sigma_1(\ell|_S), \quad \hat{\sigma}_k = \min_{\substack{\hat{S} \subseteq \pi_N(W) \\ \dim(\hat{S}) \geq p-k+1}} \sigma_1(\hat{\ell}|_{\hat{S}}).$$

To prove the inequality, we will show that for any  $S \subseteq \pi_N(y^{-1}W)$ , there exists  $\hat{S} \subseteq \pi_N(W)$  with  $\dim(S) = \dim(\hat{S})$  satisfying  $\sigma_1(\hat{\ell}|_{\hat{S}}) \leq \left\| y_B \right\|_\infty \left\| y_N^{-1} \right\|_\infty \sigma_1(\ell|_S)$ . For this purpose, define  $\hat{S} = y_N S$ , where

clearly  $\dim(\hat{S}) = \dim(S)$  and  $\hat{S} \subseteq y_N \pi_N(y^{-1}W) = \pi_N(W)$ . From here, we have that

$$\begin{aligned} \sigma_1(\hat{\ell}|_{\hat{S}}) &= \max_{\hat{x} \in \hat{S} \setminus \{0\}} \frac{\|\hat{\ell}(\hat{x})\|}{\|\hat{x}\|} = \max_{x \in S \setminus \{0\}} \frac{\|\hat{\ell}(y_N x)\|}{\|y_N x\|} \leq \max_{x \in S \setminus \{0\}} \frac{\|y_B \ell(x)\|}{\|y_N x\|} \\ &\leq \max_{x \in S \setminus \{0\}} \|y_B\|_\infty \|y_N^{-1}\|_\infty \frac{\|\ell(x)\|}{\|x\|} = \|y_B\|_\infty \|y_N^{-1}\|_\infty \sigma_1(\ell|_S), \end{aligned}$$

where  $\|\hat{\ell}(y_N x)\| \leq \|y_B \ell(x)\|$  follows since  $(y_B \ell(x), y_N x) \in W$  and by the optimality of the lift provided by  $\hat{\ell}(y_N x)$ . The statement thus follows.  $\square$

**Lemma 7.7** (Stability of singular values on polarized segments). *Let  $\text{CP}[\mu_1, \mu_0]$  be a  $\gamma$ -polarized segment of the central path with partition  $B \cup N = [n]$ . Let  $z \in \mathcal{N}^2(\beta)$ ,  $z' \in \overline{\mathcal{N}}^2(\beta)$ ,  $\beta \in (0, 1/6]$ , such that  $\mu := \bar{\mu}(z)$  and  $\mu' := \bar{\mu}(z')$  satisfy  $\mu_0 \geq \mu \geq \mu' \geq \mu_1$ . Then we have:*

$$\frac{\gamma^2}{4n^2} \cdot \frac{\mu'}{\mu} \sigma(\ell_z) \leq \sigma(\ell_{z'}) \leq \frac{4n^2}{\gamma^2} \cdot \frac{\mu'}{\mu} \sigma(\ell_z). \quad (69)$$

*Proof.* Let  $z := (x, s)$  and  $z' := (x', s')$ . Since  $z \in \mathcal{N}^2(\beta)$ , note that  $\mu = \bar{\mu}(z) > 0$ . If  $(B, N)$  is a trivial partition, then  $\ell_z, \ell_{z'}$  are both zero operators on input spaces of the same dimension and hence the statement trivially holds. If  $\mu' = 0$ , then  $\ell_{z'}$  is a zero operator on  $\pi_N(W)$ , and hence again the statement holds trivially. Thus, we may assume that  $\mu' > 0$  and that  $(B, N)$  is a non-trivial partition of  $[n]$ . Consequently,  $z' \in \mathcal{N}^2(\beta)$ .

Let  $\hat{x} = \sqrt{\frac{x\mu}{s}}$  and  $\hat{x}' = \sqrt{\frac{x'\mu'}{s'}}$  denote the normalized iterates. To prove the bounds, we apply Lemma 7.12 to the subspace  $\frac{1}{\hat{x}'}W$  and  $y := \frac{\hat{x}}{\hat{x}'}$ , noting that  $y\hat{x}' = \hat{x}$ . Since  $z, z' \in \mathcal{N}^2(\beta)$ , these subspaces are well-defined as  $\hat{x}, \hat{x}' \in \mathbb{R}_{++}^n$ . By Proposition 2.42, we have that

$$\frac{\sqrt{1-\beta}}{\sqrt{1+\beta}} \frac{x}{x'} \leq \frac{\hat{x}}{\hat{x}'} \leq \frac{\sqrt{1+\beta}}{\sqrt{1-\beta}} \frac{x}{x'}. \quad (89)$$

Combining the above with Proposition 2.32 and Corollary 3.4, we get that

$$\begin{aligned} \|y_B\|_\infty &= \left\| \frac{\hat{x}_B}{\hat{x}'_B} \right\|_\infty \leq \frac{\sqrt{1+\beta}}{\sqrt{1-\beta}} \left\| \frac{x_B}{x'_B} \right\|_\infty \leq \frac{\sqrt{1+\beta}(1+2\beta)}{(1-\beta)^{3/2}} \left\| \frac{x^{\text{CP}}(\mu)_B}{x^{\text{CP}}(\mu')_B} \right\|_\infty \leq \frac{2n}{\gamma}, \\ \|y_N^{-1}\|_\infty &= \left\| \frac{\hat{x}'_N}{\hat{x}_N} \right\|_\infty \leq \frac{\sqrt{1+\beta}}{\sqrt{1-\beta}} \left\| \frac{x'_N}{x_N} \right\|_\infty \leq \frac{\sqrt{1+\beta}(1+2\beta)}{(1-\beta)^{3/2}} \left\| \frac{x^{\text{CP}}(\mu')_N}{x^{\text{CP}}(\mu)_N} \right\|_\infty \leq \frac{2n}{\gamma} \cdot \frac{\mu'}{\mu}, \end{aligned} \quad (90)$$

and

$$\begin{aligned} \|y_B^{-1}\|_\infty &= \left\| \frac{\hat{x}'_B}{\hat{x}_B} \right\|_\infty \leq \frac{\sqrt{1+\beta}}{\sqrt{1-\beta}} \left\| \frac{x'_B}{x_B} \right\|_\infty \leq \frac{\sqrt{1+\beta}(1+2\beta)}{(1-\beta)^{3/2}} \left\| \frac{x^{\text{CP}}(\mu')_B}{x^{\text{CP}}(\mu)_B} \right\|_\infty \leq \frac{2n}{\gamma}, \\ \|y_N\|_\infty &= \left\| \frac{\hat{x}_N}{\hat{x}'_N} \right\|_\infty \leq \frac{\sqrt{1+\beta}}{\sqrt{1-\beta}} \left\| \frac{x_N}{x'_N} \right\|_\infty \leq \frac{\sqrt{1+\beta}(1+2\beta)}{(1-\beta)^{3/2}} \left\| \frac{x^{\text{CP}}(\mu)_N}{x^{\text{CP}}(\mu')_N} \right\|_\infty \leq \frac{2n}{\gamma} \cdot \frac{\mu}{\mu'}, \end{aligned} \quad (91)$$

using  $\frac{(1+2\beta)(1+\beta)^{1/2}}{(1-\beta)^{3/2}} \leq 2$  for  $\beta \in (0, 1/6]$ ,  $\left\| \frac{x^{\text{CP}}(\mu)_B}{x^{\text{CP}}(\mu')_B} \right\|_\infty = \max_{i \in B} \frac{x^{\text{CP}}(\mu)_i}{x^{\text{CP}}(\mu')_i} \leq n \cdot \frac{1}{\gamma}$  and  $\left\| \frac{x^{\text{CP}}(\mu')_N}{x^{\text{CP}}(\mu)_N} \right\|_\infty = \max_{i \in N} \frac{x^{\text{CP}}(\mu')_i}{x^{\text{CP}}(\mu)_i} \leq \frac{\mu'}{\mu} \cdot \frac{n\mu_0}{\mu}$ , together with the the analogous bounds for the remaining inequalities.

Plugging these estimates into Lemma 7.12 yields the result.  $\square$

### 7.3.2 Properties of the Ideal Direction

**Proposition 7.10.** *Let  $\text{CP}[\mu_1, \mu_0]$  be a  $\gamma$ -polarized segment with partition  $B \cup N = [n]$ . Let  $z = (x, s) \in \mathcal{N}(\beta)$ ,  $\beta \in (0, 1/6]$ , with  $\mu := \bar{\mu}(z) \in (\frac{72n^{1.5}}{\beta\gamma} \cdot \mu_1, \mu_0]$ . Then, the ideal direction  $\Delta z^{\text{ideal}}$  towards  $\mu_1$  satisfies*

$$\|(\Delta \hat{x}^{\text{ideal}}, \Delta \hat{s}^{\text{ideal}})\| \leq 4n, \quad (78)$$

$$\left\| (\hat{\xi}_N + \Delta \hat{x}_N^{\text{ideal}}, \hat{\xi}_B + \Delta \hat{s}_B^{\text{ideal}}) \right\| \leq \frac{2n^{1.5}\mu_1}{\gamma\mu} \leq \frac{\beta}{36}. \quad (79)$$

*Proof.* We first prove (78)

$$\begin{aligned}
\|(\Delta \hat{x}^{\text{ideal}}, \Delta \hat{s}^{\text{ideal}})\| &= \|(\hat{x}^{-1} x^{\text{cp}}(\mu_1) - \hat{\xi}, \hat{s}^{-1} s_N^{\text{cp}}(\mu_1) - \hat{\xi})\| \\
&\leq \|(\hat{x}^{-1} x^{\text{cp}}(\mu_1), \hat{s}^{-1} s^{\text{cp}}(\mu_1))\| + \|(\hat{\xi}, \hat{\xi})\| \\
&\leq \|(\hat{x}^{-1} x^{\text{cp}}(\mu_1), \hat{s}^{-1} s^{\text{cp}}(\mu_1))\|_1 + \sqrt{2n} \\
&\leq \frac{n}{\sqrt{1-\beta}}(1 + \mu_1/\mu) + \sqrt{2n} \quad (\text{by Lemma 2.43}) \\
&\leq 4n,
\end{aligned}$$

where the last inequality follows from  $\beta \in (0, 1/6]$  and  $\mu \geq 6\mu_1$ .

We now continue with (79),

$$\begin{aligned}
\|(\hat{\xi}_N + \Delta \hat{x}_N^{\text{ideal}}, \hat{\xi}_B + \Delta \hat{s}_B^{\text{ideal}})\| &= \|(\hat{x}_N^{-1} x_N^{\text{cp}}(\mu_1), \hat{s}_B^{-1} s_B^{\text{cp}}(\mu_1))\| \\
&\leq \left\| \left( \frac{x_N^{\text{cp}}(\mu_1)}{x_N}, \frac{s_B^{\text{cp}}(\mu_1)}{s_B} \right) \right\|_{\infty} \|(\hat{x}_N^{-1} x_N, \hat{s}_B^{-1} s_B)\| \\
&= \left\| \left( \frac{x_N^{\text{cp}}(\mu_1)}{x_N}, \frac{s_B^{\text{cp}}(\mu_1)}{s_B} \right) \right\|_{\infty} \|\hat{\xi}\| \\
&\leq \frac{1}{1-\beta} \left\| \left( \frac{x_N^{\text{cp}}(\mu_1)}{x_N^{\text{cp}}(\mu)}, \frac{s_B^{\text{cp}}(\mu_1)}{s_B^{\text{cp}}(\mu)} \right) \right\|_{\infty} \sqrt{n} \quad (\text{by Proposition 2.32}) \\
&\leq \frac{n\mu_1}{(1-\beta)\mu\gamma} \cdot \sqrt{n} \leq \frac{2n^{1.5}\mu_1}{\gamma\mu}, \quad (\text{by Corollary 3.4}) \\
&\leq \frac{\beta}{36}. \quad \left( \text{since } \mu \geq \mu_1 \frac{72n^{1.5}}{\beta\gamma} \right)
\end{aligned}$$

□

### 7.3.3 Dimension and Lifting Cost of Cheap Lift Subspaces

**Lemma 7.11.** *Let  $\text{CP}[\mu_1, \mu_0]$  be a  $\gamma$ -polarized segment with partition  $B \cup N = [n]$ . Let  $z = (x, s) \in \mathcal{N}(\beta)$ ,  $\beta \in (0, 1/6]$ . Let  $V := \text{im}(\mathbf{M}^{\text{P}})$ ,  $U := \text{im}(\mathbf{M}^{\text{D}})$  be the subspaces returned by **CHEAP-LIFT-SUBSPACES** on  $(\mathbf{A}, x^{-1}, B, N, \varrho)$  as in (75), and  $\ell_z := \ell_N^{\hat{x}^{-1}W}$ ,  $\ell_z^{\perp} := \ell_B^{\hat{s}^{-1}W^{\perp}}$ . Then, if  $\mathcal{C}_{\sigma}(\ell_z, (\tau/\varrho, \varrho/\tau)) = 0$ , we have*

1.  $\dim(V) = \dim(\pi_N(W)) - \zeta(z)$ ,  $\dim(U) = \dim(\pi_B(W^{\perp})) - \zeta(z)$ .
2.  $\|\hat{v}_B\| \leq 2\varrho\sigma_{\zeta(z)+1}(\ell_z) \|\hat{v}_N\| \leq 2\tau \|\hat{v}_N\|$ ,  $\forall \hat{v} \in \hat{V} := \hat{x}^{-1}V$ , and  
 $\|\hat{u}_N\| \leq 2\varrho\sigma_{\zeta(z)+1}(\ell_z^{\perp}) \|\hat{u}_B\| \leq 2\tau \|\hat{u}_B\|$ ,  $\forall \hat{u} \in \hat{U} := \hat{s}^{-1}U$ .

*Proof.* Assume first that  $(B, N)$  is a trivial partition. Recall that  $\zeta(\ell_z) = \zeta(\ell_z^{\perp}) = 0$ . By the guarantees of **CHEAP-LIFT-SUBSPACES**, we have  $(V, U) = (W, \{\mathbf{0}_n\})$  if  $B = \emptyset$  and  $(V, U) = (\{\mathbf{0}_n\}, W^{\perp})$  if  $N = \emptyset$ , and thus (1) and (2) follow directly.

Now assume that  $(B, N)$  is a non-trivial partition. We first prove the dimension guarantees for  $V$  and  $U$ . By Lemma 7.12, using that  $x^{-1}W = \hat{\xi}^{-1}\hat{x}^{-1}W$ , we have that

$$\left( \|\hat{\xi}_B^{-1}\|_{\infty} \|\hat{\xi}_N\|_{\infty} \right)^{-1} \sigma(\ell_N^{x^{-1}W}) \leq \sigma(\ell_N^{\hat{x}^{-1}W}) = \sigma(\ell_z) \leq \|\hat{\xi}_B\|_{\infty} \|\hat{\xi}_N^{-1}\|_{\infty} \sigma(\ell_N^{x^{-1}W}).$$

By Proposition 2.42,  $\sqrt{1-\beta}\mathbf{1}_n \leq \hat{\xi} \leq \sqrt{1+\beta}\mathbf{1}_n$ , and since  $\beta \in (0, 1/6]$  we have that

$$\max\left\{ \|\hat{\xi}_B^{-1}\|_{\infty} \|\hat{\xi}_N\|_{\infty}, \|\hat{\xi}_B\|_{\infty} \|\hat{\xi}_N^{-1}\|_{\infty} \right\} \leq \sqrt{\frac{1+\beta}{1-\beta}} \leq \sqrt{2}.$$

Applying the same argument to  $\ell_B^{xW^{\perp}}$  and  $\ell_B^{\hat{s}^{-1}W^{\perp}} = \ell_z^{\perp}$ , using that  $xW^{\perp} = \frac{x}{\mu}W^{\perp} = \hat{\xi}\hat{s}^{-1}W^{\perp}$ , we get that

$$\frac{1}{\sqrt{2}}\sigma(\ell_N^{x^{-1}W}) \leq \sigma(\ell_z) \leq \sqrt{2}\sigma(\ell_N^{x^{-1}W}), \quad \frac{1}{\sqrt{2}}\sigma(\ell_B^{xW^{\perp}}) \leq \sigma(\ell_z^{\perp}) \leq \sqrt{2}\sigma(\ell_B^{xW^{\perp}}), \quad (92)$$

where we recall that  $\sigma^+(\ell_z) = \sigma^+(\ell_z^\perp)$  by Lemma 6.8. By the guarantees of [CHEAP-LIFT-SUBSPACES](#) and (92), we have that

$$\mathcal{C}_\sigma\left(\ell_z, \frac{1}{\sqrt{2}\varrho}\right) \leq \mathcal{C}_\sigma\left(\ell_N^{x^{-1}W}, 1/\varrho\right) \leq \dim(V) \leq \mathcal{C}_\sigma\left(\ell_N^{x^{-1}W}, 1\right) \leq \mathcal{C}_\sigma\left(\ell_z, \sqrt{2}\right).$$

Since  $\mathcal{C}_\sigma\left(\ell_z, \left(\frac{\tau}{\varrho}, \frac{\varrho}{\tau}\right)\right) = 0$  and  $\frac{\tau}{\varrho} \leq \frac{1}{\sqrt{2}\varrho} \leq \sqrt{2} \leq \frac{\varrho}{\tau}$ , we have  $\dim(V) = \mathcal{C}_\sigma\left(\ell_z, \frac{\tau}{\varrho}\right) = \mathcal{C}_\sigma\left(\ell_z, \frac{\varrho}{\tau}\right)$ . In particular,

$$\begin{aligned} \dim(V) &= \mathcal{C}_\sigma\left(\ell_z, \frac{\tau}{\varrho}\right) = \mathcal{C}_\sigma(\ell_z, [0, \infty)) - \mathcal{C}_\sigma\left(\ell_z, \left(\frac{\tau}{\varrho}, \infty\right)\right) := \dim(\pi_N(\hat{x}^{-1}W)) - \zeta(z) \\ &= \dim(\pi_N(W)) - \zeta(z). \end{aligned}$$

By a completely symmetric argument on the dual, we get that

$$\dim(U) = \mathcal{C}_\sigma\left(\ell_z^\perp, \frac{\tau}{\varrho}\right) = \dim(\pi_B(W^\perp)) - \zeta(z).$$

We now prove the lifting cost guarantees. For  $v \in V$  and the guarantees of [CHEAP-LIFT-SUBSPACES](#), we have that

$$\begin{aligned} \|\hat{x}_B^{-1}v_B\| &\leq \left\|\hat{\xi}_B\right\|_\infty \|x_B^{-1}v_B\| \leq \left\|\hat{\xi}_B\right\|_\infty \varrho^{\dim(\pi_N(W)) - \dim(V) + 1} (\ell_N^{x^{-1}W}) \|x_N^{-1}v_N\| \\ &\leq \left\|\hat{\xi}_B\right\|_\infty \left\|\hat{\xi}_N^{-1}\right\|_\infty \sqrt{2}\varrho\sigma_{\zeta(z)+1}(\ell_z) \|\hat{x}_N^{-1}v_N\| \leq 2\varrho\sigma_{\zeta(z)+1}(\ell_z) \|\hat{x}_N^{-1}v_N\| \\ &\leq 2\tau \|\hat{x}_N^{-1}v_N\|. \end{aligned}$$

By a symmetric algorithm for the dual, for  $u \in U$ , we have that

$$\|\hat{s}_N^{-1}u_N\| \leq 2\varrho\sigma_{\zeta(z)+1}(\ell_z^\perp) \|\hat{s}_B^{-1}u_B\| \leq 2\tau \|\hat{s}_B^{-1}u_B\|,$$

as needed.  $\square$

## 8 Amortized Iteration Bound for [SLLS-IPM](#)

In this section we prove Theorem 1.4. For this purpose we will rely on Theorem 1.8, which decomposes the central path into polarized segments where the sum of partition changes is bounded by the straight line complexity. The proof of Theorem 1.4 will then follow from the following theorem, which shows that the number of iterations of our IPM can be upper bounded in terms of a sum of partition changes.

**Theorem 8.1.** *For any  $T \in \mathbb{N}$  let  $0 \leq \mu^{(T)} < \mu^{(T-1)} < \dots < \mu^{(1)} < \mu^{(0)}$  such that for all  $1 \leq i \leq T$  we have that  $\text{CP}[\mu^{(i)}, \mu^{(i-1)}]$  is  $\gamma$ -polarized with polarizing partition  $B^{(i)} \cup N^{(i)}$ , where  $N^{(0)} := \emptyset$ . Then, [SLLS-IPM](#) equipped with a  $\varrho$ -SVD solver, starting from any point  $z^{(0)} \in \mathcal{N}^2(\beta)$  such that  $\bar{\mu}(z^{(0)}) \leq \mu^{(0)}$ , finds a point  $z \in \bar{\mathcal{N}}^2(\beta)$  such that  $\bar{\mu}(z) \leq \mu^{(T)}$  in*

$$O\left(\frac{\sqrt{n}}{\beta} \log\left(\frac{n\varrho}{\beta\gamma}\right) \sum_{i=1}^T \left(|N^{(i)} \Delta N^{(i-1)}| + 1\right)\right) \quad (93)$$

many iterations.

As explained in the introduction, the above theorem strengthens Theorem 1.1 by reducing the  $n^{1.5}$  factor in the iteration bound for traversing the  $i$ th polarized segment down to  $\sqrt{n} |N^{(i)} \Delta N^{(i-1)}|$ . This will yield an overall amortized improvement as long as the polarizing partitions do not change too rapidly.

We are now ready to give the proof of Theorem 1.4.

*Proof of Theorem 1.4.* By Lemma 7.3, the output  $(x^*, s^*, v^*, w^*)$  of [SLLS-IPM](#) is correct when  $\mu_1 = 0$ . The (trivial) modification to [SLLS-IPM](#) for  $\mu_1 > 0$ , where it only outputs  $z^1 = (x^1, s^1)$  with  $\bar{\mu}(z^1) \leq \mu_1$ , is explained in Remark 7.1. The desired upper bound on the number of iterations to traverse the segment  $\text{CP}[\mu_1, \mu_0]$  follows directly by combining Theorem 1.8 and Theorem 8.1 with  $\varrho = 2$ , where we use Theorem 6.16 to instantiate the 2-SVD solver.  $\square$

It remains to prove Theorem 8.1. In Section 8.1, we show that the singular values cannot change too quickly under a change in polarization partition, and finally we prove Theorem 8.1 in Section 8.2.

## 8.1 Stability of Singular Values under a Partition Change

The crucial statement towards proving Theorem 8.1 is the following lemma on lifting operators for different partitions. This will ensure that we do not lose too much progress, in terms gaining large singular values, when moving from one partition to another.

**Lemma 8.2.** *Given a subspace  $W \subseteq \mathbb{R}^n$  and two partitions  $B \cup N = [n]$  and  $\hat{B} \cup \hat{N} = [n]$ , the lifting operators  $\ell := \ell_N^W$  and  $\hat{\ell} := \ell_{\hat{N}}^W$  satisfy*

$$\sigma_i(\ell) \geq \sigma_{i+|N \Delta \hat{N}|}(\hat{\ell}), \quad \forall i \geq 1. \quad (94)$$

*Proof.* If either  $\hat{B}$  or  $\hat{N}$  is empty,  $\sigma_i(\hat{\ell}) = 0, \forall i \geq 1$ , by convention, and hence the statement is trivial. If  $N \cap \hat{N} = \emptyset$ , then  $\sigma_{i+|N \Delta \hat{N}|}(\hat{\ell}) \leq \sigma_{i+|\hat{N}|}(\hat{\ell}) = 0$  since  $\text{rank}(\hat{\ell}) \leq |\hat{N}|$ , and hence again the statement is trivial. Thus, we may assume that  $(\hat{B}, \hat{N})$  is non-trivial and that  $N \cap \hat{N} \neq \emptyset$ .

Let  $\check{N} := N \cap \hat{N} \neq \emptyset, \check{B} := [n] \setminus \check{N} = B \cup \hat{B} \neq \emptyset$ , and define  $\check{\ell} := \ell_{\check{N}}^W$ . By our assumptions, note that  $(\check{B}, \check{N})$  is a non-trivial partition of  $[n]$ .

First, we are going to prove that for all  $i \geq 1$  we have  $\sigma_i(\ell) \geq \sigma_{i+|N \setminus \check{N}|}(\check{\ell})$ . Note that  $\dim(\pi_N(W)) \geq \dim(\pi_{\check{N}}(W))$ . For  $i > \dim(\pi_N(W))$ , we have  $\sigma_i(\ell) = \sigma_{i+|N \setminus \check{N}|}(\check{\ell}) = 0$ . Assume now  $i \leq \dim(\pi_N(W))$ . By (15), there exists a subspace  $S^{(i)} \subseteq \pi_N(W)$ ,  $\dim(S) = \dim(\pi_N(W)) - i + 1$  with

$$\sigma_i(\ell) = \sigma_{\max}(\ell|_{S^{(i)}}). \quad (95)$$

Consider  $T^{(i)} := S^{(i)} \cap \mathbb{R}_{\check{N}}^N$ . Note that  $\dim(\pi_{\check{N}}(T^{(i)})) = \dim(T^{(i)})$  as  $\pi_{N \setminus \check{N}}(T^{(i)}) = \{\mathbf{0}_{N \setminus \check{N}}\}$  by definition. Further,  $\pi_{\check{N}}(T^{(i)}) \subseteq \pi_{\check{N}}(W)$  by construction. Therefore, we have that

$$\begin{aligned} \dim(\pi_{\check{N}}(T^{(i)})) &= \dim(T^{(i)}) \geq \dim(S^{(i)}) - |N \setminus \check{N}| \\ &= \dim(\pi_N(W)) - i + 1 - |N \setminus \check{N}|, \end{aligned} \quad (96)$$

where the inequality follows as  $T^{(i)}$  arises from  $S^{(i)}$  by adding  $N \setminus \check{N}$  homogeneous linear equations to its defining system. Finally, note that for any  $v \in \pi_{\check{N}}(T^{(i)})$  we have that

$$\|\check{\ell}(v)\| = \min_{w \in W, w_{\check{N}} = v} \|w_{[n] \setminus \check{N}}\| \leq \min_{\substack{w \in W, w_{\check{N}} = v, \\ w_{N \setminus \check{N}} = \mathbf{0}_{N \setminus \check{N}}} } \|w_{[n] \setminus \check{N}}\| = \|\ell(\mathbf{0}_{N \setminus \check{N}}, v)\|, \quad (97)$$

and so in particular

$$\sigma_{\max}(\check{\ell}|_{\pi_{\check{N}}(T^{(i)})}) \leq \sigma_{\max}(\ell|_{T^{(i)}}). \quad (98)$$

From here, since  $\dim(\pi_{\check{N}}(W)) \leq \dim(\pi_N(W))$  we get that

$$\begin{aligned} \sigma_{i+|N \setminus \check{N}|}(\check{\ell}) &\stackrel{(96)}{\leq} \sigma_{\dim(\pi_N(W)) - \dim(\pi_{\check{N}}(T^{(i)})) + 1}(\check{\ell}) \stackrel{(15)}{\leq} \sigma_{\max}(\check{\ell}|_{\pi_{\check{N}}(T^{(i)})}) \\ &\stackrel{(98)}{\leq} \sigma_{\max}(\ell|_{T^{(i)}}) \leq \sigma_{\max}(\ell|_{S^{(i)}}) = \sigma_i(\ell). \end{aligned} \quad (99)$$

It remains to show that we can use similar shift argumentation between the operators  $\check{\ell}$  and  $\hat{\ell}$ , that is, that  $\sigma_i(\check{\ell}) \geq \sigma_{i+|\hat{N} \setminus \check{N}|}(\hat{\ell}), \forall i \geq 1$ . Using Lemma 6.8, we may equivalently consider the dual operator  $\ell_{\check{B}}^{W^\perp} = -\text{ad}(\ell_{\check{N}}^W)$ , where  $\check{B} = [n] \setminus \check{N}$ , and show that

$$\sigma_i(\ell_{\check{B}}^{W^\perp}) \geq \sigma_{i+|\check{B} \setminus \hat{B}|}(\ell_{\hat{B}}^{W^\perp}), \quad \forall i \geq 1. \quad (100)$$

Since  $\emptyset \neq \hat{B} \subseteq \check{B}$ , the inequality above follows directly from same arguments we used to obtain (99). Therefore, for all  $i \geq 1$ , by Lemma 6.8 we get that

$$\sigma_i(\check{\ell}) = \sigma_i(\ell_{\check{B}}^{W^\perp}) \geq \sigma_{i+|\check{B} \setminus \hat{B}|}(\ell_{\hat{B}}^{W^\perp}) = \sigma_{i+|\hat{N} \setminus \check{N}|}(\hat{\ell}) \quad (101)$$

holds, where we used that  $|\hat{N} \setminus \check{N}| = |\check{B} \setminus \hat{B}|$ . Combining (99) and (101) gives the result as now

$$\begin{aligned} \sigma_i(\ell) &\geq \sigma_{i+|N \setminus \check{N}|}(\check{\ell}) = \sigma_{i+|N \setminus \check{N}|}(\check{\ell}) \geq \sigma_{i+|N \setminus \check{N}|+|\hat{N} \setminus \check{N}|}(\hat{\ell}) \\ &= \sigma_{i+|N \setminus \check{N}|+|\hat{N} \setminus \check{N}|}(\hat{\ell}) = \sigma_{i+|N \Delta \hat{N}|}(\hat{\ell}), \quad \forall i \geq 1. \end{aligned} \quad (102)$$

□

## 8.2 Proof of the Amortized Bound

We need few more preparations to prove Theorem 8.1. First, consider the basic iterates  $z^{(0)}, \dots, z^{(K)}$  of the algorithm, where basic is as defined in Section 7.3, with  $z^{(K)}$  being the first basic iterate with  $\bar{\mu}(z^{(K)}) \leq \mu^{(T)}$ . Without loss of generality, we may assume that  $\mu^{(1)} < \bar{\mu}(z^{(0)}) \leq \mu^{(0)}$ .

For  $i \in \{0, \dots, T\}$ , define  $t_i$  to be the smallest index of an iterate with  $\mu(z^{(t_i)}) \leq \mu^{(i)}$ . With this definition, note that  $t_0 = 0$ ,  $t_T = K$ , and for  $i \in [T]$ , that  $t_i - t_{i-1} = |\{j \in [K] : \mu^{(i)} < \bar{\mu}(z^{(j)}) \leq \mu^{(i-1)}\}|$ . Further, define  $y^{(i)} := z^{\text{cp}}(\mu^{(i)})$ ,  $i \in \{0, \dots, T\}$ , to be the central path point with parameter  $\mu^{(i)}$ .

For  $z = (x, s) \in \mathcal{N}^2(\beta)$ ,  $\beta \in (0, 1/6)$ , let us extend the notation (68) to  $\ell_{z, N^{(i)}} := \ell_{N^{(i)}}^{\hat{x}^{-1}W}$  and

$$\zeta^{(i)}(z) := \left| \left\{ j \geq 1 : \sigma_j(\ell_{z, N^{(i)}}) > \frac{\tau}{\rho} \right\} \right| = \mathcal{C}_\sigma \left( \ell_{z, N^{(i)}}, \left( \frac{\tau}{\rho}, \infty \right) \right). \quad (103)$$

**Proof Overview** We start by giving an outline of the proof of Theorem 8.1. Assume for simplicity of presentation that  $\bar{\mu}(z^{(t_i)}) = \mu^{(i)}$  (i.e.,  $y^{(i)} = z^{(t_i)}$ ),  $i \in \{0, \dots, T\}$ , which can be achieved by adding artificial iterates in each segment. Consider the behaviour of the algorithm on the polarized segment  $\text{CP}[\mu^{(i)}, \mu^{(i-1)}]$  with partition  $(B^{(i)}, N^{(i)})$ , which starts with  $z^{(t_{i-1})}$  and ends with  $z^{(t_i)}$ . By Lemma 7.5, we have that after every  $D = \Theta\left(\frac{\sqrt{n}}{\beta} \ln\left(\frac{n\sigma}{\beta\gamma}\right)\right)$  iterations, either we decrease the potential  $\zeta^{(i)}(\cdot)$  by one or we pass the end of the segment. Specifically, for  $k \geq 1$ , either  $t_i + kD > t_{i+1}$ , or  $\zeta^{(i)}(z^{(t_i+kD)}) \leq \zeta^{(i)}(z^{(t_i)}) - k$ .

By Lemma 7.7, for  $z^{(t')}$ ,  $z^{(t)}$  with  $t_i \leq t' < t \leq t_{i-1}$  (i.e.,  $\mu^{(i)} \leq \bar{\mu}(z^{(t')}) < \bar{\mu}(z^{(t)}) \leq \mu^{(i-1)}$ ), recall that  $\sigma(\ell_{z^{(t')}, N^{(i)}}) \leq \Gamma \frac{\bar{\mu}(z^{(t')})}{\bar{\mu}(z^{(t)})} \sigma(\ell_{z^{(t)}, N^{(i)}})$  where  $\Gamma = \frac{4n^2}{\gamma^2}$ . Let us assume for now that the inequality holds for  $\Gamma = 1$  instead of  $\frac{4n^2}{\gamma^2}$ . Then, under this assumption, the singular values would be non-increasing over the segment, and hence the potential  $\zeta^{(i)}(z^{(t)})$  would be monotonic in  $t$ . Therefore, we would have that

$$\zeta^{(i)}(z^{(t_i)}) \leq \zeta^{(i)}(z^{(t_{i-1})}) - \lfloor (t_i - t_{i-1})/D \rfloor \Rightarrow t_i - t_{i-1} \leq D(\zeta^{(i)}(z^{(t_{i-1})}) - \zeta^{(i)}(z^{(t_i)}) + 1). \quad (104)$$

That is, the drop in potential from the start to the end of the segment pays for the number of iterations. From here, the handover to the next segment is controlled by Lemma 8.2, which implies that  $\zeta^{(i)}(z^{(t_{i-1})}) \leq |N^{(i)} \Delta N^{(i-1)}| + \zeta^{(i-1)}(z^{(t_{i-1})})$  for  $i \in [T]$ . Putting these bounds together, we would get

$$\begin{aligned} K = t_T - t_0 &= \sum_{i=1}^T t_i - t_{i-1} \leq \sum_{i=1}^T D(\zeta^{(i)}(z^{(t_{i-1})}) - \zeta^{(i)}(z^{(t_i)}) + 1) \\ &\leq \sum_{i=1}^T D(|N^{(i)} \Delta N^{(i-1)}| + 1) + D(\zeta^{(i-1)}(z^{(t_{i-1})}) - \zeta^{(i)}(z^{(t_i)})) \\ &= \sum_{i=1}^T D(|N^{(i)} \Delta N^{(i-1)}| + 1) + D(\zeta^{(0)}(z^{(0)}) - \zeta^{(T)}(z^{(t_T)})) \\ &\leq D \sum_{i=1}^T (|N^{(i)} \Delta N^{(i-1)}| + 1), \end{aligned} \quad (105)$$

where the last inequality follows since  $\zeta^{(0)}(z^{(0)}) = 0$  as  $N^{(0)} = \emptyset$ . This is precisely the desired bound in Theorem 8.1.

Unfortunately, this argument breaks due to the  $(4n^2/\gamma^2)$  factor in Lemma 7.7. The singular values may in fact increase on the short term. This would not be a problem on sufficiently ‘long’ segments. However, on ‘short’ segments, this could lead to an increase in singular values, where there may not be sufficiently many— $D$ —iterations to compensate. Such increments could even aggregate over a sequence of short segments.

The proof below works with a more robust version of the potential in (103). The robust potential is defined as a minimum over parametrized potentials: for  $i \geq 0$  and  $z \in \mathcal{N}^2(\beta)$ ,

$$\bar{\zeta}^{(i)}(z) := \min_{q \geq 0} \zeta^{(i)}(z, q), \quad \zeta^{(i)}(z, q) := \mathcal{C}_\sigma \left( \ell_{z, N^{(i)}}, \left( \frac{\tau}{\rho} \Gamma^q, \infty \right) \right) + q, \text{ for } q \geq 0, \quad (106)$$

where  $\Gamma := \frac{4n^2}{\gamma^2}$  as above.

Each parametrized potential keep track of the singular values above a threshold  $\Gamma^q \frac{\tau}{\rho}$  for some  $q \geq 0$ . The additional factor represents possible increases accumulated on previous short segments. However, the contributions  $|N^{(j)} \Delta N^{(j-1)}| + 1$  of these short segments will ‘pay’ for bringing down large singular values

later, which corresponds to the additive  $q$  factor in these potentials. Note that the original potential in (103) corresponds to the parametrized potential with  $q = 0$ .

The following technical lemma shows that the drop in robust potential  $\bar{\zeta}^{(i)}(\cdot)$  over a polarized segment indeed pays for the number of iterations, in analogy to (104) in the ideal setting (where we assumed monotonicity of the potential). With this lemma in hand, the iteration bound in Theorem 8.1 follows along identical lines to the analysis in (105).

**Lemma 8.3.** *Define  $D_{\text{it}} := 3\lceil C_{\text{it}} \frac{\sqrt{n}}{\beta} \log(\frac{nq}{\beta\gamma}) \rceil$ , where  $C_{\text{it}}$  is as in Lemma 7.5. Let  $t_i, N^{(i)}, y^{(i)}$ ,  $i \in \{0, \dots, T\}$  be as defined above, let  $k_i := \max\{0, \lfloor (t_i - t_{i-1} - 1)/D_{\text{it}} \rfloor\}$ ,  $i \in [T]$ . Then, for  $i \in [T]$ , the following holds*

- (i)  $\bar{\zeta}^{(i)}(y^{(i-1)}) \leq \bar{\zeta}^{(i-1)}(y^{(i-1)}) + |N^{(i)} \Delta N^{(i-1)}|$ .
- (ii)  $\bar{\zeta}^{(i)}(y^{(i)}) + k_i \leq \bar{\zeta}^{(i)}(y^{(i-1)}) + 1$ .

*Proof.* We prove (i) and (ii) separately below.

**Proof of (i)** By Lemma 8.2,  $\sigma_{i+|N^{(i)} \Delta N^{(i-1)}|}(\ell_{y^{(i-1)}, N^{(i)}}) \leq \sigma_i(\ell_{y^{(i-1)}, N^{(i-1)}})$ ,  $\forall i \geq 1$ . This implies

$$\mathcal{C}_\sigma \left( \ell_{y^{(i-1)}, N^{(i-1)}}, \left( \frac{\tau}{\rho} \Gamma^q, \infty \right) \right) + |N^{(i)} \Delta N^{(i-1)}| \geq \mathcal{C}_\sigma \left( \ell_{y^{(i-1)}, N^{(i)}}, \left( \frac{\tau}{\rho} \Gamma^q, \infty \right) \right),$$

and hence  $\zeta^{(i)}(y^{(i-1)}, q) \leq \zeta^{(i-1)}(y^{(i-1)}, q) + |N^{(i)} \Delta N^{(i-1)}|$ , for  $q \geq 0$ . Therefore,  $\bar{\zeta}^{(i)}(y^{(i-1)}) \leq \bar{\zeta}^{(i-1)}(y^{(i-1)}) + |N^{(i)} \Delta N^{(i-1)}|$ .

**Proof of (ii)** For  $q \geq 0$ , we claim that

$$\zeta^{(i)}(y^{(i)}, \max\{0, q+1-k_i\}) + k_i \leq \zeta^{(i)}(y^{(i-1)}, q) + 1. \quad (107)$$

Assuming the claim, (ii) follows noting that

$$\begin{aligned} \bar{\zeta}^{(i)}(y^{(i)}) + k_i &\leq \min_{q \geq 0} \zeta^{(i)}(y^{(i)}, \max\{0, q+1-k_i\}) + k_i \leq \min_{q \geq 0} \zeta^{(i)}(y^{(i-1)}, q) + 1 \\ &= \bar{\zeta}^{(i)}(y^{(i-1)}) + 1. \end{aligned}$$

To prove (107), we need the following intermediate inequality:

$$\sigma(\ell_{y^{(i)}, N^{(i)}}) \leq \Gamma^{1-k_i} \sigma(\ell_{y^{(i-1)}, N^{(i)}}). \quad (108)$$

We prove this inequality first. Recall that  $\text{CP}[\mu^{(i)}, \mu^{(i-1)}]$  is a  $\gamma$ -polarized segment with partition  $(B^{(i)}, N^{(i)})$ . Therefore, for  $z' \in \overline{\mathcal{N}^2}(\beta)$ ,  $z \in \mathcal{N}^2(\beta)$ ,  $\beta \in (0, 1/6)$ , with  $\mu^{(i)} \leq \bar{\mu}(z') \leq \bar{\mu}(z) \leq \mu^{(i-1)}$ , by Lemma 7.7 we have that

$$\sigma(\ell_{z', N^{(i)}}) \leq \Gamma \frac{\bar{\mu}(z')}{\bar{\mu}(z)} \sigma(\ell_{z, N^{(i)}}). \quad (109)$$

Assume that  $k_i = 0$ . Then, (108) follows from (109) with  $z' = y^{(i)}$  and  $z = y^{(i-1)}$ , noting that  $\bar{\mu}(y^{(i)}) = \mu^{(i)} < \mu^{(i-1)} = \bar{\mu}(y^{(i-1)})$  and  $\Gamma = \Gamma^{1-k_i}$ .

Now assume that  $k_i \geq 1$ . Let  $z^{(i)}$ ,  $i \in \{0, \dots, K\}$ , be the basic iterates as defined in the overview, and let  $t = t_{i-1} + D_{\text{it}}/3$  and  $t' = t_i - 1 - D_{\text{it}}/3$ . Since  $k_i \geq 1$ , we have that

$$t' - t \geq k_i D_{\text{it}} - 2D_{\text{it}}/3 \geq k_i D_{\text{it}}/3. \quad (110)$$

By the choice of  $C_{\text{it}}$  in (73),  $D_{\text{it}}/3 = \lfloor C_{\text{it}} \frac{\sqrt{n}}{\beta} \log(\frac{nq}{\beta\gamma}) \rfloor$  iterations are sufficient to divide the normalized gap by  $\Gamma$ . That is,  $\bar{\mu}(z^{(i+D_{\text{it}}/3)}) \leq \bar{\mu}(z^{(i)})/\Gamma$ , for  $i \in \{0, \dots, K - D_{\text{it}}/3\}$ . In particular,

$$\bar{\mu}(z^{(t')}) \leq \bar{\mu}(z^{(t)})/\Gamma^{\lfloor \frac{3(t'-t)}{D_{\text{it}}} \rfloor} \leq \bar{\mu}(z^{(t)})/\Gamma^{k_i}.$$

Similarly,  $\mu^{(i)} \leq \bar{\mu}(z^{(t_i-1)}) \leq \bar{\mu}(z^{(t')})/\Gamma$  and  $\bar{\mu}(z^{(t)}) \leq \bar{\mu}(z^{(t_i-1)})/\Gamma \leq \mu^{(i-1)}/\Gamma$ . Therefore, by (109) we have that

$$\sigma(\ell_{y^{(i)}, N^{(i)}}) \leq \sigma(\ell_{z^{(t')}, N^{(i)}}) \leq \sigma(\ell_{z^{(t)}, N^{(i)}})/\Gamma^{k_i-1} \leq \sigma(\ell_{y^{(i-1)}, N^{(i)}})/\Gamma^{k_i-1}, \quad (111)$$

completing the proof of (108).

We now prove claim (107) by dividing the analysis into two additional cases.

**Case I.  $k_i \leq q + 1$  :**

$$\begin{aligned} \zeta^{(i)}(y^{(i)}, q + 1 - k_i) + k_i &= \mathcal{C}_\sigma \left( \ell_{y^{(i)}, N^{(i)}} \left( \frac{\tau}{\varrho} \Gamma^{q+1-k}, \infty \right) \right) + (q + 1 - k_i) + k_i \\ &\stackrel{(108)}{\leq} \mathcal{C}_\sigma \left( \ell_{y^{(i-1)}, N^{(i)}} \left( \frac{\tau}{\varrho} \Gamma^q, \infty \right) \right) + q + 1 \\ &= \zeta^{(i)}(y^{(i-1)}, q) + 1, \text{ as needed.} \end{aligned}$$

**Case II.  $k_i \geq q + 2$  :** Let  $\hat{t} = t + (q + 1)D_{\text{it}}/3$ . By our assumption that  $k_i \geq q + 2$ , using (110) we have that that  $t' - \hat{t} \geq (k_i - q - 1)D_{\text{it}}/3 \geq D_{\text{it}}/3$ . Via an identical calculation to (111), we have that

$$\sigma(\ell_{z^{(i)}, N^{(i)}}) \leq \sigma(\ell_{z^{(t)}, N^{(i)}})/\Gamma^q \leq \sigma(\ell_{y^{(i-1)}, N^{(i)}})/\Gamma^q.$$

Using the above, we get that

$$\begin{aligned} \zeta^{(i)}(z^{(i)}) &= \mathcal{C}_\sigma \left( \ell_{z^{(i)}, N^{(i)}} \left( \frac{\tau}{\varrho}, \infty \right) \right) \leq \mathcal{C}_\sigma \left( \ell_{y^{(i-1)}, N^{(i)}} \left( \frac{\tau}{\varrho} \Gamma^q, \infty \right) \right) \\ &= \zeta^{(i)}(y^{(i-1)}, q) - q. \end{aligned} \tag{112}$$

By Lemma 7.5, since  $t' - \hat{t} \geq (k_i - q - 1)D_{\text{it}}/3 = (k_i - q - 1)[C_{\text{it}} \frac{\sqrt{n}}{\beta} \log(\frac{n\varrho}{\beta\gamma})]$ , we have that

$$\begin{aligned} \zeta^{(i)}(y^{(i)}, 0) &:= \zeta^{(i)}(y^{(i)}) \stackrel{(111)}{\leq} \zeta^{(i)}(z^{(t')}) \stackrel{7.5}{\leq} \zeta^{(i)}(z^{(\hat{t})}) + q + 1 - k_i \\ &\stackrel{(112)}{\leq} \zeta^{(i)}(y^{(i-1)}, q) + 1 - k_i. \end{aligned} \tag{113}$$

The claimed bound now follows by rearranging. □

*Proof of Theorem 8.1.* Let  $t_i, y^{(i)}, N^{(i)}$ ,  $i \in \{0, \dots, T\}$  be as defined in the overview. Recall that  $N^{(0)} = \emptyset$ ,  $t_0 = 0$  and  $t_T = K$ . Let  $D_{\text{it}}$  and  $k_i := \max\{0, \lfloor (t_i - t_{i-1} - 1)/D_{\text{it}} \rfloor\}$ ,  $i \in [T]$ , be as in Lemma 8.3. Proceeding in a similar manner to (105), we bound the number of iteration as follows:

$$\begin{aligned} K = t_T - t_0 &= \sum_{i=1}^T (t_i - t_{i-1} - 1) + 1 \leq \sum_{i=1}^T D_{\text{it}}(k_i + 1) \quad (b + 1 \leq a(\lfloor \frac{b}{a} \rfloor + 1), a, b \in \mathbb{N}) \\ &\leq \sum_{i=1}^T D_{\text{it}} \left( \bar{\zeta}^{(i)}(y^{(i-1)}) - \bar{\zeta}^{(i)}(y^{(i)}) + 2 \right) \quad (\text{by Lemma 8.3(ii)}) \\ &\leq \sum_{i=1}^T D_{\text{it}} \left( \bar{\zeta}^{(i-1)}(y^{(i-1)}) - \bar{\zeta}^{(i)}(y^{(i)}) + |N^{(i)} \Delta N^{(i-1)}| + 2 \right) \quad (\text{by Lemma 8.3(i)}) \\ &= \left( \sum_{i=1}^T D_{\text{it}} \left( |N^{(i)} \Delta N^{(i-1)}| + 2 \right) \right) + D_{\text{it}}(\bar{\zeta}^{(0)}(y^{(0)}) - \bar{\zeta}^{(T)}(y^{(T)})) \\ &\leq \sum_{i=1}^T D_{\text{it}} \left( |N^{(i)} \Delta N^{(i-1)}| + 2 \right), \end{aligned}$$

where the last inequality uses that  $\bar{\zeta}^{(0)}(y^{(0)}) \leq \zeta^{(0)}(y^{(0)}) = 0$  since  $N^{(0)} = \emptyset$ . □

## 9 Computing Approximate Singular Value Decompositions

In this section, we give our algorithm for computing approximate SVDs and prove Theorem 6.16.

We begin by presenting a deterministic strongly polynomial method for obtaining a  $(n2^n)$ -approximate SVD for matrices  $\mathbf{M} \in \mathbb{R}^{m \times n}$  using QR decomposition with greedy column permutations. In the next subsection, we demonstrate how to enhance the approximation factor to  $(1 + \varepsilon)$  by leveraging the classical power iteration.

Using QR decompositions to approximate the SVD is a well-established technique that dates back to the mid-1960s [10, 26]. Many existing approaches focus on finding rank-revealing QR factorizations

instead of a full  $\varrho$ -approximate SVD. These methods aim to identify a number  $k \leq n$  such that  $\sigma_k(\mathbf{M}) \gg \sigma_{k+1}(\mathbf{M})$ , along with an approximate subspace for the eigenvalues  $\sigma_1(\mathbf{M}), \dots, \sigma_k(\mathbf{M})$ . This is achieved by greedily selecting column permutations, followed by a standard QR factorization. The top  $k \times k$  block of the resulting upper triangular matrix provides an approximation of the subspace corresponding to the top  $k$  singular values.

One of the earliest algorithms that utilizes column pivoting and subsequent QR decomposition was proposed by Chan [12], who attributes the procedure to [27]. He achieves approximation guarantees similar to ours in Lemma 9.3. Algorithms with tighter bounds (even polynomial instead of exponential) for rank-revealing QR factorizations can be found in [13, 31, 50].

However, these algorithms are designed to work only for a fixed  $k$ , not as the  $\varrho$ -approximate SVD requires for all  $k$  simultaneously. Further, the initial algorithms [10, 26] critically rely on Householder reflection, which intrinsically make use of unit vectors in the computations. Unit vectors however, can not in general be used in the strongly polynomial model as they require the computation of square roots. The mentioned approach by Chan [12] requires the computation of the smallest singular vector of the matrix. This again can not be done in strongly polynomial time. More precisely, given a matrix  $\mathbf{M} \in \mathbb{R}^{m \times n}$ , the algorithm requires solving  $\min_{x \in \mathbb{R}^n, \|x\|=1} \|\mathbf{M}x\|$ . The algorithm we present in this section, instead solves the problem  $\min_{x \in \mathbb{R}^n, \|x\|_\infty=1} \|\mathbf{M}x\|$ . As we will see below, this problem can be solved in strongly polynomial time. Also note, that the objective values of the two minimization problems above, differ by at most a factor of  $\sqrt{m}$ . This will explain why our algorithm achieves similar approximation guarantees as the algorithm of Chan [12].

In general, most works in the literature have not focused on strong polynomiality. Beyond the use of square roots, many of these results require the computation of orthonormal matrices, which in general is not achievable in strongly polynomial space. Additionally, it is important to avoid the sequential computation of  $\Omega(n)$  Gram-Schmidt orthogonalizations (GSO). While a single GSO can be performed in strongly polynomial time [30], the size of the numbers may increase by a polynomial factor. It remains unclear whether a sequence of super-logarithmically many GSO computations can be performed in strongly polynomial time.

A good overview of QR-type algorithms with column pivoting can be found in [13]. They also present new algorithms for rank-revealing QR factorizations with improved approximation guarantees. Although the authors do not emphasize this fact, some algorithms in [13], in particular GREEDY-I.1, GREEDY-I.2 and GREEDY-I.3, can be implemented in strongly polynomial time. Furthermore, they give similar approximation guarantees as us in Lemma 9.3 for the algorithm GREEDY-I.3.

Nonetheless, we present a self-contained new algorithm here for the sake of completeness and to focus on achieving strong polynomiality. While the framework of our algorithm fits into the regime of performing column pivoting and subsequent QR decomposition, we believe that the exact rule for the column pivoting is novel.

Very recently, Diakonikolas, Tzamos, and Kane [22] also provided a strongly polynomial  $(1+\varepsilon)$ -singular value decomposition algorithm based on randomized power iteration (the randomness corresponds to a random choice of initial basis of the input space). At a high level, our algorithm removes the need for randomness by using a suitable greedy QR decomposition (see the description below). In contrast to our algorithm however, the approximate singular value decomposition of [22] outputs a decomposition of a matrix  $\widetilde{\mathbf{M}}$  (as in Definition 2.16) that is  $\varepsilon$ -“spectrally close” to  $\mathbf{M}$ . This provides a somewhat stronger guarantee than what is needed for our formalization of the  $(1+\varepsilon)$ -SVD problem, which only requires an orthogonal basis inducing a chain of  $(1+\varepsilon)$ -approximate singular subspaces. While we expect that one can extract such a decomposition in a blackbox manner from any  $(1+\varepsilon')$ -SVD solution, for a suitably chosen  $\varepsilon' \ll \varepsilon$ , for the sake of simplicity we do not pursue this direction here.

We proceed by describing our main algorithm on a high level. For a matrix  $\mathbf{M} \in \mathbb{R}^{m \times n}$ , we will use the following procedure (Algorithm 3): For all columns  $j$  of  $\mathbf{M}$ , consider the projection of the column  $\mathbf{M}_{\bullet,j}$  onto the orthogonal complement of the span of all the other columns  $\text{im}(\mathbf{M}_{\bullet,[n] \setminus \{j\}})$ . Then, remove the column  $j_{\min}$  from the matrix for which the norm of this projection is the smallest and recurse on the remaining matrix  $\mathbf{M}_{\bullet,[n] \setminus \{j_{\min}\}}$ . When this process finishes, we obtain a permutation  $\mathbf{MP}$  of the columns, given by the order in which they were removed from the matrix. It turns out, that the norms of the columns of the orthogonal matrix  $\mathbf{Q}$  obtained from a Gram-Schmidt process on the permuted matrix  $\mathbf{MP}$  (the first column removed from  $\mathbf{M}$  is the last column in its reordering) provide an exponential approximation of the singular values of the original matrix  $\mathbf{M}$ . The main observation for the proof is that the matrix  $\mathbf{R}$  in the Gram-Schmidt process, uniquely defined by  $\mathbf{MPR} = \mathbf{Q}$ , has only exponential condition number.

It is not hard to see that the output of GREEDY-SVD satisfies that  $\mathbf{Q}$  is precisely the result of Gram

---

**Algorithm 3: GREEDY-SVD**


---

**Input** : Matrix  $\mathbf{M} \in \mathbb{R}^{m \times n}$ .  
**Output**: Matrix  $\mathbf{B} \in \mathbb{R}^{n \times n}$  and a vector  $s \in \mathbb{R}^n$  such that  $(\mathbf{B}, s)$  is a  $(n2^n)$ -approximate SVD (Definition 6.13).

```

1  $J_n \leftarrow [n]$ ;
2 for  $t = n$  down to 1 do
3   for  $i \in J_t$  do
4      $w^{(t,i)} \leftarrow \mathbf{0}_n$ ;
5      $w_i^{(t,i)} \leftarrow 1$ ;
6      $w_{J_t \setminus \{i\}}^{(t,i)} \leftarrow -[\mathbf{M}_{\bullet, J_t \setminus \{i\}}]^+ \mathbf{M}_{\bullet, i}$ ;
7    $K_t \leftarrow \arg \min_{j \in J_t} \|\mathbf{M} w^{(t,j)}\|$ ;
8    $\pi(t) \leftarrow \text{any element in } \arg \min_{j \in K_t} \|w^{(t,j)}\|$ ;
9    $v^{(t)} \leftarrow w^{(t, \pi(t))}$ ;
10   $q^{(t)} \leftarrow \mathbf{M} v^{(t)}$ ;
11   $J_{t-1} \leftarrow J_t \setminus \{\pi(t)\}$ ;
12  $\mathbf{P} \leftarrow (1_{i=\pi(j)})_{i,j \in [n]}$ ;
13  $\mathbf{R} \leftarrow \mathbf{P}^\top [v^{(1)} \ \dots \ v^{(n)}]$ ;
14  $\mathbf{Q} \leftarrow [q^{(1)} \ \dots \ q^{(n)}]$ ;
15  $\mathbf{B} \leftarrow \mathbf{P}\mathbf{R}$ ;
16 for  $i = 1, \dots, n-1$  do
17    $\mathbf{B}_{\bullet, i} \leftarrow \mathbf{B}_{\bullet, i} - \Pi_{\text{im}(\mathbf{B}_{\bullet, > i})}(\mathbf{B}_{\bullet, i})$ ;           // Reverse Gram-Schmidt Orthogonalization
18  $s \leftarrow \mathbf{0}_n$ ;
19 for  $i = 1, \dots, n$  do
20    $s_i \leftarrow 2^n \|\mathbf{Q}_{\bullet, i}\|_1$ ;
21 return  $(\mathbf{B}, s)$ 

```

---

Schmidt orthogonalization (GSO) on the matrix  $\mathbf{M}\mathbf{P}$ , i.e., the matrix  $\mathbf{M}$  after its columns have been permuted according to  $\mathbf{P}$ .

Before we begin the analysis, we introduce some additional notation and recall some fundamental matrix inequalities. For a matrix  $\mathbf{M} \in \mathbb{R}^{m \times n}$ , we use  $\|\mathbf{M}\|_2 := \sigma_1(\mathbf{M})$  to denote the operator norm,  $\|\mathbf{M}\|_F := \sqrt{\sum_{i \in [m], j \in [n]} \mathbf{M}_{ij}^2}$  to denote the Frobenius norm, and  $\|\mathbf{M}\|_\infty := \max_{i \in [m], j \in [n]} |\mathbf{M}_{ij}|$  to denote the maximum absolute value of any entry. We will use the inequality  $\|\mathbf{M}\|_2 \leq \|\mathbf{M}\|_F$ , and if  $\mathbf{M}$  is invertible, the relation  $\sigma_{\min}(\mathbf{M}) = \frac{1}{\|\mathbf{M}^{-1}\|_2} \geq \frac{1}{\|\mathbf{M}^{-1}\|_F}$ .

**Lemma 9.1.** *The matrices  $\mathbf{P}$ ,  $\mathbf{R}$ , and  $\mathbf{Q}$  constructed in GREEDY-SVD (Algorithm 3) have the following properties:*

- (i)  $\mathbf{P} \in \mathbb{R}^{n \times n}$  is a permutation matrix,
- (ii)  $\mathbf{R} \in \mathbb{R}^{n \times n}$  is upper triangular,
- (iii)  $\mathbf{Q} \in \mathbb{R}^{n \times n}$  has orthogonal columns,
- (iv)  $\mathbf{M}\mathbf{P}\mathbf{R} = \mathbf{Q}$ ,
- (v)  $\|\mathbf{Q}_{\bullet, s}\| \geq \|\mathbf{Q}_{\bullet, t}\|$  for all  $s, t \in [n]$  such that  $s \leq t$ ,
- (vi)  $\|\mathbf{Q}_{\bullet, t}\| = \min_{j \in [t]} \min_{x \in \mathbb{R}^t, x_j = 1} \|\mathbf{M}\mathbf{P}_{\bullet, [t]} x\|$ ,
- (vii)  $\|\mathbf{R}\|_\infty = 1$  and  $\mathbf{R}_{t,t} = 1, \forall t \in [n]$ .

*Proof.* It is easy to see that  $\pi$  is a bijection from  $[n]$  to  $[n]$ . Therefore, by definition of  $\mathbf{P}$  in line 12 we obtain (i).

Let  $t \in [n]$  and  $i \in J_t$ , Then, noting that  $w_i^{(t,i)} = 1$ , we have that

$$\begin{aligned}
\mathbf{M}w^{(t,i)} &= \mathbf{M}_{\bullet, J_t \setminus \{i\}} w_{J_t \setminus \{i\}}^{(t,i)} + \mathbf{M}_{\bullet, i} \\
&= -\mathbf{M}_{\bullet, J_t \setminus \{i\}} [\mathbf{M}_{\bullet, J_t \setminus \{i\}}]^+ \mathbf{M}_{\bullet, i} + \mathbf{M}_{\bullet, i} \\
&= -\Pi_{\text{im}(\mathbf{M}_{\bullet, J_t \setminus \{i\}})} \mathbf{M}_{\bullet, i} + \mathbf{M}_{\bullet, i} \\
&= \Pi_{\text{im}(\mathbf{M}_{\bullet, J_t \setminus \{i\}})^\perp} \mathbf{M}_{\bullet, i},
\end{aligned} \tag{114}$$

where the third equality follows by Proposition 2.14 part (1). This furthermore implies, that

$$\begin{aligned}
\|\mathbf{M}w^{(t,i)}\| &= \|\Pi_{\text{im}(\mathbf{M}_{\bullet, J_t \setminus \{i\}})^\perp} \mathbf{M}_{\bullet, i}\| = \min_{z \in \text{im}(\mathbf{M}_{\bullet, J_t \setminus \{i\}})} \|\mathbf{M}_{\bullet, i} + z\| \\
&= \min_{x \in \mathbb{R}^{J_t}, x_i=1} \|\mathbf{M}_{\bullet, J_t} x\|.
\end{aligned} \tag{115}$$

Now, (vi) follows by (115) and the definition of  $\mathbf{P}$ , which gives  $\mathbf{M}_{\bullet, J_t} = \mathbf{M}\mathbf{P}_{\bullet, [t]}$  as  $\pi([t]) = J_t$ . From here, we also conclude that for any  $j \in J_t \setminus \{i\}$  with  $w_j^{(t,i)} \neq 0$ , we have that  $[w_j^{(t,i)}]^{-1} w_j^{(t,i)}$  is a feasible vector for the minimization problem on the right most side of (115) for index  $j$ . Therefore, we have that

$$\|\mathbf{M}w^{(t,j)}\| \leq \frac{1}{|w_j^{(t,i)}|} \|\mathbf{M}w^{(t,i)}\|. \tag{116}$$

We distinguish two cases for the norm of  $\mathbf{M}w^{(t,\pi(t))}$ . If  $\|\mathbf{M}w^{(t,\pi(t))}\| > 0$ , then for all  $j \in J_t$ , the inequality (116) gives  $|w_j^{(t,\pi(t))}| \leq \|\mathbf{M}w^{(t,\pi(t))}\| / \|\mathbf{M}w^{(t,i)}\| \leq 1$  by  $\pi(t) \in K_t$  and the definition of  $K_t$ .

In the other case we have that  $\|\mathbf{M}w^{(t,\pi(t))}\| = 0$ . Then, notice that for any  $j \in J_t$  such that  $w_j^{(t,\pi(t))} \neq 0$ , we have by (116) that  $\|\mathbf{M}w^{(t,j)}\| = 0$  and therefore  $j \in K_t$ . In particular, we have that  $\mathbf{M}_{\bullet, j} \in \text{im}(\mathbf{M}_{\bullet, J_t \setminus \{j\}})$ . Therefore, by Proposition 2.14 part (2), for  $j \in \text{supp}(w^{(t,\pi(t))})$  we obtain

$$\begin{aligned}
\|w_{J_t \setminus \{j\}}^{(t,j)}\| &= \|(\mathbf{M}_{\bullet, J_t \setminus \{j\}})^+ \mathbf{M}_{\bullet, j}\| \\
&= \min \{ \|u\| : u \in \mathbb{R}^{|J_t| - 1}, \mathbf{M}_{\bullet, J_t \setminus \{j\}} u = \Pi_{\text{im}(\mathbf{M}_{\bullet, J_t \setminus \{j\}})}(\mathbf{M}_{\bullet, j}) \} \\
&= \min \{ \|u\| : u \in \mathbb{R}^{|J_t| - 1}, \mathbf{M}_{\bullet, J_t \setminus \{j\}} u = \mathbf{M}_{\bullet, j} \} \\
&\leq \frac{\|w_{J_t \setminus \{j\}}^{(t,\pi(t))}\|}{|w_j^{(t,\pi(t))}|},
\end{aligned} \tag{117}$$

where the inequality follows by noting that the vector  $u \in \mathbb{R}^{J_t \setminus \{j\}}$  defined as  $u = -[w_j^{(t,\pi(t))}]^{-1} w_{J_t \setminus \{j\}}^{(t,\pi(t))}$  is a feasible solution to the minimization problem in (117) as

$$\mathbf{M}_{\bullet, J_t \setminus \{j\}} u = -[w_j^{(t,\pi(t))}]^{-1} \mathbf{M}_{\bullet, J_t \setminus \{j\}} w_{J_t \setminus \{j\}}^{(t,\pi(t))} = [w_j^{(t,\pi(t))}]^{-1} \mathbf{M}_{\bullet, j} w_j^{(t,\pi(t))} = \mathbf{M}_{\bullet, j}.$$

This gives for all  $j \in \text{supp}(w^{(t,\pi(t))}) \subseteq K_t$  that

$$|w_j^{(t,\pi(t))}| \leq \frac{\|w_{J_t \setminus \{\pi(t)\}}^{(t,\pi(t))}\|}{\|w_{J_t \setminus \{i\}}^{(t,i)}\|} = \frac{\sqrt{\|w^{(t,\pi(t))}\|^2 - 1}}{\sqrt{\|w^{(t,i)}\|^2 - 1}} \leq 1,$$

where the last inequality follows by definition of  $\pi(t)$ . We can therefore conclude that  $\|v^{(t)}\|_\infty = \|w^{(t,\pi(t))}\|_\infty = 1$  for all  $t \in [n]$ , recalling that  $v_{\pi(t)}^{(t)} = w_{\pi(t)}^{(t,\pi(t))} = 1$ .

Let us now show the statements for  $\mathbf{R} = \mathbf{P}^\top [v^{(1)} \dots v^{(n)}]$ . As  $\mathbf{P}$  is a permutation matrix by (i), we have shown that  $\|\mathbf{R}\|_\infty = 1$  and furthermore  $\mathbf{R}_{t,t} = 1, \forall t \in [n]$ , as  $\mathbf{R}_{t,t} = (\mathbf{P}^\top)_{t,\bullet} v^{(t)} = v_{\pi(t)}^{(t)} = 1$  for all  $t \in [n]$ . Therefore, (vii) holds. To show that  $\mathbf{R}$  is upper triangular, recall that  $\text{supp}(v_t) \subseteq J_t$  and  $\pi([t]) = J_t$ . Therefore, letting  $\pi^{-1}$  denote the inverse permutation, we get  $\text{supp}(\mathbf{P}^\top v^{(t)}) \subseteq \pi^{-1}(J_t) = [t]$  as needed for (ii).

Further, note that by construction  $\mathbf{M}v^{(t)} = q^{(t)}, t \in [n]$ . Therefore,

$$\mathbf{M}\mathbf{P}\mathbf{R} = \mathbf{M}\mathbf{P}\mathbf{P}^\top [v^{(1)} \dots v^{(n)}] = \mathbf{M} [v^{(1)} \dots v^{(n)}] = [q^{(1)} \dots q^{(n)}] = \mathbf{Q},$$

as needed for (iv). To see that  $\mathbf{Q}$  has orthogonal columns, note that for by definition of  $q^{(t)}$  we have with (114) that

$$q^{(t)} = \Pi_{\text{im}(\mathbf{M}_{\bullet, J_t \setminus \{\pi(t)\}})^\perp} \mathbf{M}_{\bullet, \pi(t)}, \quad (118)$$

and so in particular  $q^{(t)} \in \text{im}(\mathbf{M}_{\bullet, J_t \setminus \{\pi(t)\}})^\perp$ . However, since for  $1 \leq s < t$  we have that  $q^{(s)} \in \text{im}(\mathbf{M}_{\bullet, J_s \setminus \{\pi(s)\}})$ . We conclude that  $q^{(t)}$  is orthogonal to  $q^{(s)}$ , proving (iii).

To show (v), note that for  $1 \leq s < t \leq n$ , the inclusion  $\text{im}(\mathbf{M}_{\bullet, J_s \setminus \{\pi(s)\}}) \subseteq \text{im}(\mathbf{M}_{\bullet, J_t \setminus \{\pi(t)\}})$  holds, and hence

$$\begin{aligned} \|\mathbf{Q}_{\bullet, s}\| &= \|q^{(s)}\| \\ &= \left\| \Pi_{\text{im}(\mathbf{M}_{\bullet, J_s \setminus \{\pi(s)\}})^\perp} \mathbf{M}_{\bullet, \pi(s)} \right\| \quad (\text{by (118)}) \\ &\geq \left\| \Pi_{\text{im}(\mathbf{M}_{\bullet, J_t \setminus \{\pi(t)\}})^\perp} \mathbf{M}_{\bullet, \pi(s)} \right\| \quad (\text{im}(\mathbf{M}_{\bullet, J_t \setminus \{\pi(t)\}})^\perp \subseteq \text{im}(\mathbf{M}_{\bullet, J_s \setminus \{\pi(s)\}})^\perp) \\ &= \left\| \mathbf{M}w^{(t, \pi(s))} \right\| \quad (\text{by (114)}) \\ &\geq \left\| \mathbf{M}w^{(t, \pi(t))} \right\| \quad (\text{by choice of } \pi(t)) \\ &= \left\| \Pi_{\text{im}(\mathbf{M}_{\bullet, J_t \setminus \{\pi(t)\}})^\perp} \mathbf{M}_{\bullet, \pi(t)} \right\| \quad (\text{by (114)}) \\ &= \|q^{(t)}\| \quad (\text{by (118)}) \\ &= \|\mathbf{Q}_{\bullet, t}\|. \end{aligned}$$

This proves the lemma.  $\square$

The following proposition shows that the matrix  $\mathbf{R}$  computed by GREEDY-SVD is “well-conditioned”.

**Proposition 9.2.** *Let  $\mathbf{R} \in \mathbb{R}^{n \times n}$  be an upper triangular matrix with diagonal  $\mathbf{1}_n$  and with entries of absolute value at most 1. Then,  $\mathbf{R}^{-1}$  is upper triangular with diagonal  $\mathbf{1}_n$  and  $|\mathbf{R}_{ij}^{-1}| \leq \max\{1, 2^{j-i-1}\}$ ,  $i, j \in [n]$ ,  $i \leq j$ . In particular,  $\|\mathbf{R}^{-1}\|_2 \leq 2^n$ .*

*Proof.* The proof goes by induction on  $n$ . The base case  $n = 1$  is trivial, so assume  $n > 1$ . Then, it is directly verifiable that

$$\mathbf{R}^{-1} = \begin{bmatrix} \mathbf{R}_{[n-1], [n-1]}^{-1} & -\mathbf{R}_{[n-1], [n-1]}^{-1} \mathbf{R}_{[n-1], n} \\ \mathbf{0}_{n-1} & 1 \end{bmatrix}.$$

By the induction hypothesis, it holds that  $|\mathbf{R}_{ij}^{-1}| \leq \max\{1, 2^{j-i-1}\}$ ,  $i, j \in [n-1]$ ,  $i \leq j$ . Note that  $\mathbf{R}^{-1}$  is upper triangular as claimed. We now prove the coefficient bound for  $\mathbf{R}_{i,n}^{-1}$ , for  $i \leq n$ . If  $i = n$ ,  $\mathbf{R}_{i,i}^{-1} = 1$ , as needed. For  $i < n$ , using that  $\mathbf{R}$  has entries at most 1 we get that

$$\begin{aligned} |\mathbf{R}_{in}^{-1}| &= |(\mathbf{R}_{[n-1], [n-1]}^{-1} \mathbf{R}_{[n-1], n})_i| = \left| \sum_{j=i}^{n-1} \mathbf{R}_{i,j}^{-1} \mathbf{R}_{j,n} \right| \leq \sum_{j=i}^{n-1} |\mathbf{R}_{i,j}^{-1}| \\ &\leq \sum_{j=i}^{n-1} \max\{1, 2^{j-i-1}\} = 1 + \sum_{l=0}^{n-i-2} 2^l = \max\{1, 2^{n-i-1}\}, \end{aligned}$$

as needed. For the last statement, by a direct calculation

$$\|\mathbf{R}^{-1}\|_2^2 \leq \|\mathbf{R}^{-1}\|_F^2 \leq \sum_{i=1}^n \sum_{j=i}^n \max\{1, 4^{j-i-1}\} = \frac{2}{3}n - \frac{1}{9} + \frac{4^n}{9} \leq 4^n,$$

as needed.  $\square$

We now show how GREEDY-SVD produces a  $(n2^n)$ -approximation of the singular values and corresponding subspaces of  $\mathbf{M}$ .

**Lemma 9.3.** *Let  $\mathbf{M} \in \mathbb{R}^{m \times n}$  be a matrix and  $\mathbf{P}$ ,  $\mathbf{R}$  and  $\mathbf{Q}$  be the matrices constructed in GREEDY-SVD applied to  $\mathbf{M}$ . Let  $V_{\geq i} := \text{im}(\mathbf{P}\mathbf{R}_{\bullet, \geq i})$  and similarly  $V_{\leq i} := \text{im}(\mathbf{P}\mathbf{R}_{\bullet, \leq i})$ . Then, we have that*

$$2^{-n} \max_{v \in V_{\geq i} \setminus \{\mathbf{0}_n\}} \frac{\|\mathbf{M}v\|}{\|v\|} \leq \|\mathbf{Q}_{\bullet, i}\| \leq n \min_{v \in V_{\leq i} \setminus \{\mathbf{0}_n\}} \frac{\|\mathbf{M}v\|}{\|v\|}.$$

Furthermore,  $\|\mathbf{Q}_{\bullet, i}\| \leq \sqrt{n} \sigma_i(\mathbf{M})$ .

*Proof.* For the lower bound on  $\|\mathbf{Q}_{\bullet,i}\|$ , we have that

$$\begin{aligned}
\max_{v \in V_{\geq i}} \frac{\|\mathbf{M}v\|}{\|v\|} &= \max_{x \in \mathbb{R}^{n-i+1} \setminus \{\mathbf{0}\}} \frac{\|\mathbf{MPR}_{\bullet, \geq i}x\|}{\|\mathbf{PR}_{\bullet, \geq i}x\|} \\
&= \max_{x \in \mathbb{R}^{n-i+1} \setminus \{\mathbf{0}\}} \frac{\|\mathbf{Q}_{\bullet, \geq i}x\|}{\|\mathbf{R}_{\bullet, \geq i}x\|} \\
&\leq \max_{x \in \mathbb{R}^{n-i+1} \setminus \{\mathbf{0}\}} \|\mathbf{Q}_{\bullet,i}\| \frac{\|x\|}{\|\mathbf{R}_{\bullet, \geq i}x\|} \quad (\text{Lemma 9.1 (iii), (v)}) \\
&\leq \|\mathbf{Q}_{\bullet,i}\| \|\mathbf{R}^{-1}\| \\
&\leq 2^n \|\mathbf{Q}_{\bullet,i}\|, \quad (\text{Proposition 9.2})
\end{aligned}$$

proving the inequality.

For the upper bound on  $\|\mathbf{Q}_{\bullet,i}\|$ , note that

$$\begin{aligned}
\min_{v \in V_{\leq i}} \frac{\|\mathbf{M}v\|}{\|v\|} &= \min_{x \in \mathbb{R}^i \setminus \{\mathbf{0}\}} \frac{\|\mathbf{MPR}_{\bullet, \leq i}x\|}{\|\mathbf{PR}_{\bullet, \leq i}x\|} \\
&= \min_{x \in \mathbb{R}^i \setminus \{\mathbf{0}\}} \frac{\|\mathbf{Q}_{\bullet, \leq i}x\|}{\|\mathbf{R}_{\bullet, \leq i}x\|} \\
&\geq \min_{x \in \mathbb{R}^i \setminus \{\mathbf{0}\}} \|\mathbf{Q}_{\bullet,i}\| \frac{\|x\|}{\|\mathbf{R}_{\bullet, \leq i}x\|} \quad (\text{Lemma 9.1 (iii), (v)}) \\
&\geq \|\mathbf{Q}_{\bullet,i}\| \|\mathbf{R}\|_2^{-1} \\
&\geq \frac{\|\mathbf{Q}_{\bullet,i}\|}{n}. \quad (\|\mathbf{R}\| \leq n \text{ by Lemma 9.1 (vii)})
\end{aligned}$$

It remains to prove the furthermore statement. By Proposition 2.20, there exists a subspace  $U_i \subseteq \mathbb{R}^n$  with  $\dim(U_i) = n - i + 1$  such that  $\max_{x \in U_i \setminus \{\mathbf{0}\}} \frac{\|\mathbf{MP}x\|}{\|x\|} = \sigma_i(\mathbf{MP}) = \sigma_i(\mathbf{M})$ . By dimension counting, there exists  $\bar{x} \in U_i \setminus \{\mathbf{0}_n\}$  such that  $\text{supp}(\bar{x}) \subseteq [i]$ . Therefore,

$$\sigma_i(\mathbf{M}) \geq \frac{\|\mathbf{MP}\bar{x}\|}{\|\mathbf{P}\bar{x}\|} \geq \frac{1}{\sqrt{n}} \frac{\|\mathbf{MP}\bar{x}\|}{\|\bar{x}\|_\infty} \geq \frac{1}{\sqrt{n}} \min_{j \in [i]} \min_{x \in \mathbb{R}^i, x_j=1} \|\mathbf{MP}_{\bullet, [i]}x\| = \frac{\|\mathbf{Q}_{\bullet,i}\|}{\sqrt{n}},$$

where the last equality follows from Lemma 9.1 (vi). This concludes the proof.  $\square$

**Theorem 9.4.** *Let  $\mathbf{M} \in \mathbb{R}^{m \times n}$  be a matrix. The output of GREEDY-SVD on  $\mathbf{M}$  is a  $(n2^n)$ -SVD approximation of  $\mathbf{M}$  in the sense of Definition 6.13. Furthermore, the algorithm runs in strongly polynomial time  $O(n^2 \max(m, n)^3)$ .*

*Proof.* First, note that  $\mathbf{B}$  is an orthogonal basis of  $\mathbb{R}^n$ . Furthermore,  $\text{im}(\mathbf{B}_{\geq i}) = V_{\geq i}$ . Using Lemma 9.3, we therefore have that

$$\max_{v \in \text{im}(\mathbf{B}_{\geq i}) \setminus \{\mathbf{0}_n\}} \frac{\|\mathbf{M}v\|}{\|v\|} = \max_{v \in V_{\geq i} \setminus \{\mathbf{0}_n\}} \frac{\|\mathbf{M}v\|}{\|v\|} \leq 2^n \|\mathbf{Q}_{\bullet,i}\| \leq 2^n \sqrt{n} \sigma_i(\mathbf{M}). \quad (119)$$

Using the variational characterization of singular values (15) we obtain that

$$\sigma_i(\mathbf{M}) \leq \max_{v \in V_{\geq i} \setminus \{\mathbf{0}_n\}} \frac{\|\mathbf{M}v\|}{\|v\|} \leq 2^n \|\mathbf{Q}_{\bullet,i}\| \quad (120)$$

Together with Lemma 9.3, this proves that

$$\frac{1}{\sqrt{n}} \|\mathbf{Q}_{\bullet,i}\| \leq \sigma_i(\mathbf{M}) \leq 2^n \|\mathbf{Q}_{\bullet,i}\|. \quad (121)$$

In particular, by choice of  $s_i = 2^n \|\mathbf{Q}_{\bullet,i}\|_1$  for  $i \in [n]$ , we get that

$$\frac{1}{2^n n} s \leq \sigma(\mathbf{M}) \leq s, \quad (122)$$

as needed for Definition 6.13. The choice of  $\|\cdot\|_1$  instead of  $\|\cdot\|_2$  in definition of  $s$  is necessary to ensure that the output of the algorithm has polynomial bit complexity.

For the strongly polynomial guarantees, it remains to check that all the intermediate iterates have polynomial bit complexity. To see this, note first that the columns of  $\mathbf{R}$  correspond to solutions of a well-described linear system in the original matrix  $\mathbf{M}$ , and thus has bit-complexity polynomially related to that of  $\mathbf{M}$ . From here,  $\mathbf{B}$  corresponds to a “reverse order” GSO applied to  $\mathbf{PR}$  (or again, the columns are solutions to well-described linear systems in  $\mathbf{PR}$ ), and hence  $\mathbf{B}$  also has bit-complexity polynomially related to that of  $\mathbf{M}$  (see, e.g., [30, Section 1.4] for a thorough overview of the bit-complexity of GSO).

Let us now prove the statment about the running time. Using the characterization in Proposition 2.15 and further noting that the required rank factorization can be computed in  $O(\max(m, n)^3)$ , we are able to compute the pseudoinverse of an  $m \times n$  matrix as well as the projection onto the image of an  $m \times n$  matrix in time  $O(\max(m, n)^3)$ . Therefore, a single execution of both Lines 6 and 17 of Algorithm 3 takes  $O(\max(m, n)^3)$ . The repeated execution of these lines dominates the overall running time of the algorithm, hence the overall running time of the algorithm can be bounded by  $O(n \cdot n \cdot \max(m, n)^3) = O(n^2 \max(m, n)^3)$ .

This completes the proof of the theorem.  $\square$

## 9.1 Boosting via the Power Method

In this section, we demonstrate a method to enhance the approximation ratio of any algorithm for the  $\varrho$ -SVD problem (as in Definition 6.13) to a  $(1 + \varepsilon)$ -SVD approximation. This is possible provided  $\varrho = \varrho(\dim(\mathbf{M}))$ , i.e., the approximation ratio  $\varrho$  is a function of the dimension of the matrix only and does not depend on the matrix’s conditioning. The strategy involves using a low accuracy  $\varrho$ -SVD approximation to calculate an approximate SVD of the matrix  $\mathbf{S} := (\mathbf{M}^\top \mathbf{M})^p$ . Here, a large power  $p = O(n/\varepsilon \cdot \log(\varrho/\varepsilon))$  is used to achieve the desired error  $\varepsilon$ . We will then show that the approximate SVD of  $\mathbf{S}$  is in fact a  $(1 + \varepsilon)$ -approximation of the SVD of  $\mathbf{M}$ .

**Lemma 9.5.** *Let  $0 < \varepsilon \leq 1/2$  and  $\varrho \geq \varepsilon$ . Then, there is a reduction from  $(1 + \varepsilon)$ -SVD on an  $m \times n$  matrix to  $\varrho$ -SVD on one  $n \times n$  and one  $m \times n$  matrix that runs in time  $O(mn^2 + n^3 \log(\log(2 + \varrho)/\varepsilon))$ , and requires space polynomial in  $m, n, \log(2 + \varrho)/\varepsilon$  and the bit-encoding length of  $\varepsilon, \varrho$  and the input matrix.*

*Proof.* Let  $\mathbf{M} \in \mathbb{R}^{m \times n}$  be a  $m \times n$  matrix. The reduction proceeds as follows. Compute  $\mathbf{S} := (\mathbf{M}^\top \mathbf{M})^p$  for  $p := \lceil \log_{1+\varepsilon/3}(\varrho/2) \rceil$ . Let  $(\mathbf{B}, s)$  be the output of  $\varrho$ -SVD on  $\mathbf{S}$  and  $(\cdot, s')$  be the output of  $\varrho$ -SVD on  $\mathbf{M}$ . For each  $i \in [n]$ , apply binary search to compute  $\hat{s}_i = s'_i(1 + \varepsilon/3)^{-2k_i}$ ,  $-1 \leq k_i \leq 2p$ , satisfying  $s_i^{1/(2p)} \leq \hat{s}_i \leq (1 + \varepsilon/3)s_i^{1/(2p)}$ . Finally, return  $(\mathbf{B}, \hat{s})$ .

We first proof correctness of the reduction. Consider a singular value decomposition  $\mathbf{M} = \sum_{i=1}^r \sigma_i(\mathbf{M}) u_i v_i^\top$  of  $\mathbf{M}$ , where  $r = \text{rank}(\mathbf{M})$ . Fix  $i \in [n]$ . By definition of  $\mathbf{B}$ , we have for all  $x \in \text{im}(\mathbf{B}_{\geq i}) \setminus \{\mathbf{0}_n\}$  that  $\|\mathbf{S}x\|_2 / \|x\|_2 \leq s_i \leq \varrho \sigma_i(\mathbf{S})$ , which implies that

$$\begin{aligned} \sum_{j=1}^r \sigma_j(\mathbf{M})^{4p} \langle v_j, x \rangle^2 &= \left\| \sum_{j=1}^r \sigma_j(\mathbf{M})^{2p} v_j \langle v_j, x \rangle \right\|^2 = \|\mathbf{S}x\|^2 \\ &\leq s_i^2 \|x\|^2 \leq \varrho^2 \sigma_i(\mathbf{S})^2 \|x\|^2 = \varrho^2 \sigma_i(\mathbf{M})^{4p} \|x\|^2. \end{aligned} \quad (123)$$

Therefore,  $\forall x \in \text{im}(\mathbf{B}_{\geq i}) \setminus \{\mathbf{0}_n\}$ , we have that

$$\begin{aligned} \|\mathbf{M}x\|^{4p} &= \left( \sum_{j=1}^r \sigma_j(\mathbf{M})^2 \langle v_j, x \rangle^2 \right)^{2p} \\ &= \left( \|x\|^2 \cdot \sum_{j=1}^r \sigma_j(\mathbf{M})^2 \frac{\langle v_j, x \rangle^2}{\|x\|^2} \right)^{2p} \\ &\leq \|x\|^{4p} \sum_{j=1}^r \sigma_j(\mathbf{M})^{4p} \frac{\langle v_j, x \rangle^2}{\|x\|^2} \quad (\text{Jensen's inequality}) \\ &\leq \|x\|^{4p} s_i^2 \\ &\leq \|x\|^{4p} \cdot \varrho^2 \sigma_i(\mathbf{M})^{4p} \quad (\text{by (123)}), \end{aligned} \quad (124)$$

and so

$$\max_{x \in \text{im}(\mathbf{B}_{\geq i}) \setminus \{\mathbf{0}_n\}} \frac{\|\mathbf{M}x\|}{\|x\|} \leq s_i^{1/(2p)} \leq \varrho^{1/(2p)} \sigma_i(\mathbf{M}) \leq (1 + \varepsilon/3) \sigma_i(\mathbf{M}). \quad (125)$$

From the above, note that  $(\mathbf{B}, s^{1/(2p)})$  is already a solution to  $(1 + \varepsilon)$ -SVD on  $\mathbf{M}$ . Unfortunately, we cannot compute the vector  $s^{1/(2p)}$  in strongly polynomial time, which motivates the binary search used to construct the approximation  $\hat{s}$ . By the guarantees of  $s$  and  $s'$ , we have that

$$s'(1 + \varepsilon/3)^{-2p} \leq \frac{s'}{\varrho} \leq \sigma(\mathbf{M}) \leq s^{1/(2p)} \leq (1 + \varepsilon/3)\sigma(\mathbf{M}) \leq s'(1 + \varepsilon/3),$$

and thus the claimed range for  $k_i$ ,  $i \in [n]$ , in the binary search is correct. From here, by the guarantees on  $\hat{s}$ , since  $\varepsilon \in (0, 1/2]$ , we have that

$$s^{1/(2p)} \leq \hat{s} \leq (1 + \varepsilon/3)s^{1/(2p)} \leq (1 + \varepsilon/3)^2\sigma(\mathbf{M}) \leq (1 + \varepsilon)\sigma(\mathbf{M}).$$

Therefore,  $(\mathbf{B}, \hat{s})$  is a valid solution to  $(1 + \varepsilon)$ -SVD.

We now justify the running time of the reduction. Since  $p = \Theta(\log(2 + \varrho)/\varepsilon)$ , we can compute  $\mathbf{S} = (\mathbf{M}^\top \mathbf{M})^p$  in time  $O(mn^2 + n^3 \log(p)) = O(mn^2 + n^3 \log(\log(2 + \varrho)/\varepsilon))$  by repeated squaring of  $\mathbf{M}^\top \mathbf{M}$ . For each  $i \in [n]$ , the corresponding binary search requires  $O(\log p)$  comparisons, where each comparison are of the form  $s_i \leq (s')^{2p}(1 + \varepsilon/3)^{-2k}$  for  $-1 \leq k \leq 2p$ . Lastly, the reduction makes two  $\varrho$ -SVD calls, one to  $\mathbf{M} \in \mathbb{R}^{m \times n}$  and one to  $\mathbf{S} \in \mathbb{R}^{n \times n}$ .

It remains to show the claim about the space needed by the algorithm. To this end, it suffices to show that the matrix  $\mathbf{S}$  has a bit-encoding length polynomial in  $\log(2 + \varrho)/\varepsilon^{-1}$  and the bit-encoding of  $\mathbf{M}$ . To this end, let the input matrix  $\mathbf{M}$  be given as  $\mathbf{M}_{i,j} = p_{ij}/q_{ij}$  for  $i \in [m], j \in [n]$  with  $p_{ij} \in \mathbb{Z}$  and  $q_{ij} \in \mathbb{Z}_{>0}$ . We can write  $\mathbf{M} = k^{-1}\bar{\mathbf{M}}$ , where  $k = \text{lcm}_{i,j} q_{ij}$  and  $\bar{\mathbf{M}}_{i,j} = |p_{ij}/q_{ij}| \cdot \text{lcm}_{i,j} q_{ij} \in \mathbb{Z}$ , where lcm denotes the least common multiple. In particular,  $\bar{\mathbf{M}} \in \mathbb{Z}^{m \times n}$  and  $\|\bar{\mathbf{M}}\|_{\text{F}} \leq \sum_{i,j} |p_{ij}| \cdot \prod_{i,j} q_{ij}$ . Therefore, we can write  $\mathbf{S} = (\mathbf{M}^\top \mathbf{M})^p = k^{-2p}(\bar{\mathbf{M}}^\top \bar{\mathbf{M}})^p$ , where  $\|(\bar{\mathbf{M}}^\top \bar{\mathbf{M}})^p\|_{\text{F}} \leq \|\bar{\mathbf{M}}\|_{\text{F}}^{2p} \leq (\sum_{i,j} |p_{ij}| \prod_{i,j} q_{ij})^{2p}$ . In particular, the entries of  $\mathbf{S}$  can be written in a number of bits that is polynomial in  $p$  and the number of bits of  $\mathbf{M}$ . Recalling that  $p = O(\log(2 + \varrho)/\varepsilon)$  finishes the proof.  $\square$

Theorem 6.16 is a now a simple consequence of Lemma 9.5 and Theorem 9.4.

**Theorem 6.16.** *For  $\varrho > 1$ , there is a deterministic algorithm for solving  $\varrho$ -SVD that on an input  $m \times n$  matrix requires  $O(n^2 \max(m, n)^3 \log(n + \frac{1}{\varrho-1}))$  operations and space polynomial in  $n, m, \max\{1, \frac{1}{\varrho-1}\}$  and the bit-encoding length of the input matrix.*

*Proof.* Let  $\varepsilon = \min\{\varrho - 1, 1/2\}$ , and apply the reduction in Lemma 9.5 with parameters  $(\varepsilon, n2^n)$ , where we use Theorem 9.4 to implement the  $(n2^n)$ -SVD algorithm. The space guarantees now follow from the guarantees of Theorem 9.4 and Lemma 9.5. For the running time, observe that Theorem 9.4 requires  $O(n^2 \max(m, n)^3)$  operations, while the application of Lemma 9.5 requires  $O(mn^2 + n^3 \log(\log(2 + n2^n)/\min(\varrho - 1, \frac{1}{2}))) = O(mn^2 + n^3 \log(n + \frac{1}{\varrho-1}))$  operations. The sum of these running times is  $O(n^2 \max(m, n)^3 \log(n + \frac{1}{\varrho-1}))$ , which completes the proof.  $\square$

## 10 Self-Concordant Barrier Central Paths

The theory of self-concordant barriers, introduced by Nesterov and Nemirovski [49], is used to describe interior point methods in a general setting. In this section, we show that the central paths obtained from self-concordant barriers relate to the max central path, as in the case of the log-barrier. From this, we deduce a proof of Theorem 1.7.

We follow the presentation of self-concordant barriers made by Renegar [52]. We consider a  $\mathcal{C}^2$  function  $f: D \rightarrow \mathbb{R}$  with open convex domain  $D \subseteq E$  where  $E = x^0 + L$  is an affine subspace of  $\mathbb{R}^n$ . We recall that, for  $x \in D$ , the gradient  $\nabla f(x)$  of  $f$  at  $x$  is the element of  $L$  such that for all  $h \in L$ ,  $\langle \nabla f(x), h \rangle = \lim_{t \rightarrow 0^+} \frac{f(x+th) - f(x)}{t}$ . Moreover, the Hessian  $\nabla^2 f(x)$  of  $f$  at  $x$  is the self-adjoint linear operator from  $L$  to itself such that for all  $h, k \in L$ ,  $\langle \nabla^2 f(x)h, k \rangle = \lim_{t \rightarrow 0^+} \frac{\langle \nabla f(x+tk) - \nabla f(x), h \rangle}{t}$ . We assume that  $\nabla^2 f(x)$  is positive definite, i.e.,  $\langle \nabla^2 f(x)h, h \rangle > 0$  for all  $h \in L$ ,  $h \neq 0$ . In particular,  $f$  is a strictly convex function on  $D$ . Given  $x \in D$ , we introduce  $\|\cdot\|_x$ , the norm over  $L$  associated with the Hessian  $\nabla^2 f(x)$ , i.e.,  $\|z\|_x := (\langle \nabla^2 f(x)z, z \rangle)^{1/2}$  for all  $z \in L$ . We define  $\mathcal{B}_x(y; r) := \{z \in D : \|y - z\|_x < r\}$ , the open ball with center  $y \in D$  and radius  $r \geq 0$  with respect to the norm  $\|\cdot\|_x$ .

The function  $f$  is (strongly nondegenerate) self-concordant if, for all  $x \in D$ , we have  $\mathcal{B}_x(x; 1) \subseteq D$ , and for all  $y \in \mathcal{B}_x(x; 1)$  and nonzero vector  $v \in L$ ,

$$1 - \|y - x\|_x \leq \frac{\|v\|_x}{\|v\|_y} \leq \frac{1}{1 - \|y - x\|_x}.$$

We define the *complexity value* as

$$\vartheta_f := \sup_{x \in D} \sup_{\|h\|_x \leq 1} \langle \nabla f(x), h \rangle^2.$$

If  $\vartheta_f < \infty$ , then  $f$  is called a *self-concordant barrier*.

Let  $f$  be self-concordant barrier over the (relative) interior of the polyhedron  $\mathcal{P}$ . The *central path* associated with  $f$  is the function that maps  $\mu > 0$  to the unique minimizer of the function  $x \mapsto \langle c, x \rangle + \mu f(x)$ . When  $f$  is (the restriction to  $E$  of) the log-barrier  $x \mapsto -\sum_{i=1}^n \log(x_i)$ , the scalar  $\mu$  is used as the parameter of the central path. In the case of general barriers, it is more convenient to parametrize the central path by the optimality gap  $g$ . This is made possible by the fact that, if any two points of the central path have the same optimality gap  $g$ , they both minimize the (strictly convex) function  $f$  over  $\{x \in \mathcal{P} : \langle c, x \rangle = g\}$ , thus they are equal. Following this, we denote by  $x^{\text{cp}(f)}(g)$  the unique point of the central path induced by the barrier  $f$  with optimality gap  $g > 0$ .

In the next statement, we generalize Lemma 1.2 to the case of the central path associated with the self-concordant barrier  $f$ .

**Proposition 10.1.** *For all  $g > 0$ , we have  $\frac{1}{2(2\vartheta_f+1)}x^{\text{m}}(g) < x^{\text{cp}(f)}(g) \leq x^{\text{m}}(g)$ .*

*Proof.* The right hand side inequality follows from  $x^{\text{cp}(f)}(g) \in \mathcal{P}_g$ . Using [52, Theorem 2.3.4] and the fact that  $\mu \nabla f(x^{\text{cp}(f)}(g)) + c \in L^\perp$  for some  $\mu > 0$ , we know that  $\mathcal{P}_g \subseteq \mathcal{B}_{x^{\text{cp}(f)}(g)}(x^{\text{cp}(f)}(g); 4\vartheta_f + 1)$ . Since  $\mathcal{P}_g \subseteq \mathbb{R}_{\geq 0}^n$ , [3, Lemma 4] shows that  $y < 2(2\vartheta_f + 1)x^{\text{cp}(f)}(g)$  for all  $y \in \mathcal{B}_{x^{\text{cp}(f)}(g)}(x^{\text{cp}(f)}(g); 4\vartheta_f + 1)$ . We deduce that  $x^{\text{m}}(g) < 2(2\vartheta_f + 1)x^{\text{cp}(f)}(g)$ .  $\square$

We are now ready to prove Theorem 1.7.

*Proof (Theorem 1.7).* Let  $x^{(0)}, \dots, x^{(T)}$  be the successive iterates of the IPM. We denote by  $g^{(0)}, \dots, g^{(T)}$  their respective optimality gap, where  $g^{(0)} \geq g_0$  and  $g^{(T)} \leq g_1$ .

Let  $i \in [n]$ . We claim that,  $\forall k \in [T]$ , the segment  $[(g^{(k-1)}, x_i^{(k-1)}), (g^{(k)}, x_i^{(k)})]$  is included in the neighborhood  $\mathcal{N}_i^{\text{mp}}(\theta')$  for  $\theta' = 1 - \frac{1-\theta}{2(2\vartheta_f+1)}$ . Indeed, for all  $\lambda \in [0, 1]$ , the point  $x = (1-\lambda)x^{(k-1)} + \lambda x^{(k)}$  has optimality gap  $g = (1-\lambda)g^{(k-1)} + \lambda g^{(k)}$ . As  $x \in \mathcal{P}$ , we have  $x_i \leq x_i^{\text{m}}(g)$ . Moreover, since  $x \geq (1-\theta)x^{\text{cp}(f)}(g)$ , Proposition 10.1 ensures that  $x_i > \frac{1-\theta}{2(2\vartheta_f+1)}x_i^{\text{m}}(g)$ . Hence,  $(g, x) \in \mathcal{N}_i^{\text{mp}}(\theta')$ . We deduce by Lemma 4.4 that  $T \geq \text{SLC}_{\theta', i}^{\text{p}}(g_1, g_0)$ .

Consequently, we have  $\sum_{i=1}^n \text{SLC}_{\theta', i}^{\text{p}}(g_1, g_0) \leq nT$ . By Theorem 1.4, the number  $T'$  of iterations performed by the algorithm **SLLS-IPM** is in

$$O \left( \sqrt{n} \log \left( \frac{n}{1-\theta} \right) \min \left\{ \sum_{i=1}^n \text{SLC}_{\theta, i}^{\text{p}}(g_1, g_0), \sum_{i=1}^n \text{SLC}_{\theta, i}^{\text{d}}(g_1, g_0) \right\} \right). \quad (126)$$

This implies that

$$T' = O \left( n^{1.5} \log \left( \frac{n\vartheta_f}{1-\theta} \right) T \right).$$

$\square$

Theorem 1.7 relates the iteration complexity of general barrier IPMs traversing the wide neighborhood  $\mathcal{N}^{\text{cp}(f)}(\theta)$  as defined in (6) to the iteration complexity of our IPM. Our wide neighborhood definition is however not entirely standard, and in particular, it does not map directly to the more standard neighborhoods such as those based on the Newton decrement. Furthermore, it does not obviously capture IPMs using primal-dual neighborhoods.

In [3, Section 4.1] however, it has been shown that all the neighborhoods of self-concordant central paths used in the literature (including that of the log-barrier central path like the projections of  $\mathcal{N}^2(\beta)$ ,  $0 < \beta < 1/4$ , and  $\mathcal{N}^{-\infty}(\theta)$ ,  $0 < \theta < 1$ , to primal variables) are actually contained in *multiplicative neighborhoods*, i.e., sets of form

$$\mathcal{M}^{\text{cp}(f)}(\underline{m}, \overline{m}) = \{x \in \mathcal{P} : \exists g \ \underline{m}x^{\text{cp}(f)}(g) \leq x \leq \overline{m}x^{\text{cp}(f)}(g)\} \quad (0 < \underline{m} \leq 1 \leq \overline{m}).$$

The latter turn out to be essentially equivalent to our definition of wide neighborhoods in (6). This is justified by the following lemma:

**Lemma 10.2.** *We have  $\mathcal{N}^{\text{cp}(f)}(\theta) \subseteq \mathcal{M}^{\text{cp}(f)}(1 - \theta, 2(2\vartheta_f + 1))$  for all  $\theta \in (0, 1)$ , and  $\mathcal{M}^{\text{cp}(f)}(\underline{m}, \overline{m}) \subseteq \mathcal{N}^{\text{cp}(f)}(\frac{\underline{m}}{2(2\vartheta_f+1)\overline{m}})$  for all  $0 < \underline{m} \leq 1 \leq \overline{m}$ .*

*Proof.* For the first statement, let  $\theta \in (0, 1)$ , and  $x \in \mathcal{N}^{\text{cp}(f)}(\theta)$ , and let  $g = \langle c, x \rangle - v^*$  denote the optimality gap of  $x$ . The inequality  $x \geq (1 - \theta)x^{\text{cp}(f)}(g)$  follows by the definition of  $\mathcal{N}^{\text{cp}(f)}(\theta)$ . We have  $x \leq x^{\text{m}}(g)$  since  $x \in \mathcal{P}$ , and so  $x \leq 2(2\vartheta_f + 1)x^{\text{cp}(f)}(g)$  by Proposition 10.1.

For the second statement, let  $0 < \underline{m} \leq 1 \leq \overline{m}$ , and let  $(\tilde{g}, x)$  be such that  $\underline{m}x^{\text{cp}(f)}(\tilde{g}) \leq x \leq \overline{m}x^{\text{cp}(f)}(\tilde{g})$ . Let  $s^* \geq 0$  denote a dual optimal solution to (LP). The optimality gap  $g = \langle c, x \rangle - v^* = \langle s^*, x \rangle$  of  $x$  satisfies  $g \leq \overline{m} \langle s^*, x^{\text{cp}(f)}(\tilde{g}) \rangle = \overline{m}\tilde{g}$  (since  $s^* \geq 0$ ). Then, we have

$$\begin{aligned} x &\geq \frac{\underline{m}}{2(2\vartheta_f + 1)}x^{\text{m}}(\tilde{g}) && \text{(by Proposition 10.1)} \\ &\geq \frac{\underline{m}}{2(2\vartheta_f + 1)\overline{m}}x^{\text{m}}(\overline{m}\tilde{g}) && \text{(by Lemma 4.3)} \\ &\geq \frac{\underline{m}}{2(2\vartheta_f + 1)\overline{m}}x^{\text{m}}(g) && \text{(as } g \leq \overline{m}\tilde{g}\text{)} \\ &\geq \frac{\underline{m}}{2(2\vartheta_f + 1)\overline{m}}x^{\text{cp}(f)}(g). && \text{(by Proposition 10.1)} \end{aligned}$$

□

As a consequence, the neighborhoods  $\mathcal{N}^{\text{cp}(f)}(\theta)$  are flexible enough to capture all known neighborhoods of central paths. Therefore, Theorem 1.7 indeed shows that our IPM is approximately optimal when compared against essentially any straight-line following IPM.

## References

- [1] X. Allamigeon, P. Benchimol, S. Gaubert, and M. Joswig. Log-barrier interior point methods are not strongly polynomial. *SIAM Journal on Applied Algebra and Geometry*, 2(1):140–178, 2018.
- [2] X. Allamigeon, P. Benchimol, S. Gaubert, and M. Joswig. What tropical geometry tells us about the complexity of linear programming. *SIAM Review*, 63(1):123–164, Jan. 2021.
- [3] X. Allamigeon, S. Gaubert, and N. Vandame. No self-concordant barrier interior point method is strongly polynomial. In *Proceedings of the 54th Annual ACM Symposium on Theory of Computing (STOC)*, pages 515–528, 2022.
- [4] K. Axiotis, A. Madry, and A. Vladu. Faster sparse minimum cost flow by electrical flow localization. In *Proceedings of the 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 528–539, 2022.
- [5] S. Axler. *Linear algebra done right*. Springer Nature, 2023.
- [6] S. J. Berkowitz. On computing the determinant in small parallel time using a small number of processors. *Information processing letters*, 18(3):147–150, 1984.
- [7] R. Bhatia. *Matrix Analysis*. Springer New York, 1997.
- [8] K. H. Borgwardt. *The Simplex Method: a probabilistic analysis*, volume 1. Springer Science & Business Media, 2012.
- [9] S. Bubeck and R. Eldan. The entropic barrier: a simple and optimal universal self-concordant barrier. arXiv preprint arXiv:1412.1587, 2014.
- [10] P. Businger and G. H. Golub. Linear least squares solutions by householder transformations. *Numerische Mathematik*, 7(3):269–276, June 1965.
- [11] S. L. Campbell and C. D. Meyer. *Generalized inverses of linear transformations*. SIAM, 2009.
- [12] T. F. Chan. Rank revealing qr factorizations. *Linear Algebra and its Applications*, 88–89:67–82, Apr. 1987.
- [13] S. Chandrasekaran and I. C. F. Ipsen. On rank-revealing factorisations. *SIAM Journal on Matrix Analysis and Applications*, 15(2):592–622, Apr. 1994.

- [14] L. Chen, R. Kyng, Y. P. Liu, R. Peng, M. P. Gutenberg, and S. Sachdeva. Maximum flow and minimum-cost flow in almost-linear time. In *Proceedings of the 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 612–623. IEEE, 2022.
- [15] S. Chewi. The entropic barrier is  $n$ -self-concordant. In *Geometric Aspects of Functional Analysis: Israel Seminar (GAFA) 2020-2022*, pages 209–222. Springer, 2023.
- [16] M. B. Cohen, Y. T. Lee, and Z. Song. Solving linear programs in the current matrix multiplication time. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 938–942, 2019.
- [17] D. Dadush and S. Huiberts. Smoothed analysis of the simplex method. In T. Roughgarden, editor, *Beyond the Worst-Case Analysis of Algorithms*. Cambridge University Press, Cambridge, 2020.
- [18] D. Dadush, S. Huiberts, B. Natura, and L. A. Végh. A scaling-invariant algorithm for linear programming whose running time depends only on the constraint matrix. In *Proceedings of the 52nd Annual ACM Symposium on Theory of Computing (STOC)*, pages 761–774, 2020.
- [19] D. Dadush, S. Huiberts, B. Natura, and L. A. Végh. A scaling-invariant algorithm for linear programming whose running time depends only on the constraint matrix. *Mathematical Programming*, pages 1–72, 2023.
- [20] S. I. Daitch and D. A. Spielman. Faster approximate lossy generalized flow via interior point algorithms. In *Proceedings of the 40th annual ACM symposium on Theory of Computing (STOC)*, pages 451–460, 2008.
- [21] A. Deza, E. Nematollahi, and T. Terlaky. How good are interior point methods? Klee–Minty cubes tighten iteration-complexity bounds. *Math. Program.*, 113(1):1–14, Oct. 2008.
- [22] I. Diakonikolas, C. Tzamos, and D. M. Kane. A strongly polynomial algorithm for approximate forster transforms and its application to halfspace learning. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, pages 1741–1754, 2023.
- [23] J. Edmonds. Systems of distinct representatives and linear algebra. *J. Res. Nat. Bur. Standards Sect. B*, 71(4):241–245, 1967.
- [24] Y. Gao, Y. P. Liu, and R. Peng. Fully dynamic electrical flows: Sparse maxflow faster than Goldberg–Rao. In *Proceedings of the 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 516–527. IEEE, 2022.
- [25] S. Gass and T. Saaty. The computational algorithm for the parametric objective function. *Naval Res. Logist. Quart.*, 2:39–45, 1955.
- [26] G. Golub. Numerical methods for solving linear least squares problems. *Numerische Mathematik*, 7(3):206–216, June 1965.
- [27] G. Golub, V. Klema, and G. Stewart. Rank degeneracy and least squares problems. In *Rep. STAN-CS-76-559*. Dept. Comput. Sci., Stanford University, 1976.
- [28] C. C. Gonzaga. Path-following methods for linear programming. *SIAM review*, 34(2):167–224, 1992.
- [29] R. C. Grinold. The payment scheduling problem. *Naval Research Logistics Quarterly*, 19:123–136, 1972.
- [30] M. Grötschel, L. Lovász, and A. Schrijver. *Geometric algorithms and combinatorial optimization*, volume 2. Springer Science & Business Media, 1988.
- [31] M. Gu and S. C. Eisenstat. Efficient algorithms for computing a strong rank-revealing qr factorization. *SIAM Journal on Scientific Computing*, 17(4):848–869, July 1996.
- [32] S. Kakahara, A. O. Ohara, and T. Tsuchiya. Information geometry and interior-point algorithms in semidefinite programs and symmetric cone programs. *Journal of Optimization Theory and Applications*, 157:749–780, 2013.

- [33] S. Kakihara, A. O. Ohara, and T. Tsuchiya. Curvature integrals and iteration complexities in SDP and symmetric cone programs. *Computational Optimization and Applications*, 57:623–665, 2014.
- [34] N. Karmarkar. A new polynomial-time algorithm for linear programming. In *Proceedings of the 16th Annual ACM Symposium on Theory of Computing (STOC)*, pages 302–311, 1984.
- [35] J. A. Kelner and D. A. Spielman. A randomized polynomial-time simplex algorithm for linear programming. In *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pages 51–60, 2006.
- [36] L. G. Khachiyan. A polynomial algorithm in linear programming. In *Doklady Akademii Nauk SSSR*, volume 244, pages 1093–1096, 1979.
- [37] V. Klee and G. J. Minty. How good is the simplex algorithm? In *Inequalities, III (Proc. Third Sympos., Univ. California, Los Angeles, Calif., 1969; dedicated to the memory of Theodore S. Motzkin)*, pages 159–175. Academic Press, New York, 1972.
- [38] G. Lan, R. D. Monteiro, and T. Tsuchiya. A polynomial predictor-corrector trust-region algorithm for linear programming. *SIAM Journal on Optimization*, 19(4):1918–1946, 2009.
- [39] Y. T. Lee and A. Sidford. Path finding methods for linear programming: Solving linear programs in  $\tilde{O}(\sqrt{\text{rank}})$  iterations and faster algorithms for maximum flow. In *Proceedings of the 55th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 424–433, 2014.
- [40] Y. T. Lee and A. Sidford. Solving linear programs with  $\tilde{O}(\sqrt{\text{rank}})$  linear system solves. arXiv preprint arXiv:1910.08033, 2019.
- [41] Y. T. Lee and M.-C. Yue. Universal barrier is  $n$ -self-concordant. *Mathematics of Operations Research*, 46(3):1129–1148, 2021.
- [42] A. Madry. Navigating central path with electrical flows: From flows to matchings, and back. In *Proceedings of the 54th Annual Symposium on Foundations of Computer Science*, pages 253–262. IEEE, 2013.
- [43] N. Megiddo. Is binary encoding appropriate for the problem-language relationship? *Theoretical Computer Science*, 19(3):337–341, 1982.
- [44] N. Megiddo, S. Mizuno, and T. Tsuchiya. A modified layered-step interior-point algorithm for linear programming. *Mathematical Programming*, 82(3):339–355, 1998.
- [45] S. Mizuno, M. Todd, and Y. Ye. On adaptive-step primal-dual interior-point algorithms for linear programming. *Mathematics of Operations Research*, 18:964–981, 11 1993.
- [46] R. D. Monteiro and T. Tsuchiya. A strong bound on the integral of the central path curvature and its relationship with the iteration-complexity of primal-dual path-following LP algorithms. *Mathematical Programming*, 115(1):105–149, 2008.
- [47] R. D. C. Monteiro and T. Tsuchiya. A variant of the Vavasis-Ye layered-step interior-point algorithm for linear programming. *SIAM Journal on Optimization*, 13(4):1054–1079, 2003.
- [48] R. D. C. Monteiro and T. Tsuchiya. A new iteration-complexity bound for the MTY predictor-corrector algorithm. *SIAM Journal on Optimization*, 15(2):319–347, 2005.
- [49] Y. Nesterov and A. Nemirovskii. *Interior-point polynomial algorithms in convex programming*. SIAM, 1994.
- [50] C. Pan. On the existence and computation of rank-revealing lu factorizations. *Linear Algebra and its Applications*, 316(1–3):199–222, Sept. 2000.
- [51] J. Renegar. A polynomial-time algorithm, based on Newton’s method, for linear programming. *Mathematical Programming*, 40(1-3):59–93, 1988.
- [52] J. Renegar. *A Mathematical View of interior point Methods in Convex Optimization*. Society for Industrial and Applied Mathematics, 2001.

- [53] C. Roos, T. Terlaky, and J.-P. Vial. *Interior point methods for linear optimization*. Springer Science & Business Media, 2005.
- [54] T. S. Shores. *Applied Linear Algebra and Matrix Analysis*. Springer, 2018.
- [55] S. Smale. Mathematical problems for the next century. *The Mathematical Intelligencer*, 20:7–15, 1998.
- [56] G. Sonnevend, J. Stoer, and G. Zhao. On the complexity of following the central path of linear programs by linear extrapolation II. *Mathematical Programming*, 52(1-3):527–553, 1991.
- [57] D. A. Spielman and S.-H. Teng. Nearly-linear time algorithms for graph partitioning, graph sparsification, and solving linear systems. In *Proceedings of the 36th Annual ACM Symposium on Theory of Computing (STOC)*, 2004.
- [58] D. A. Spielman and S.-H. Teng. Smoothed analysis of algorithms: Why the simplex algorithm usually takes polynomial time. *Journal of the ACM (JACM)*, 51(3):385–463, 2004.
- [59] V. Strassen. Vermeidung von divisionen. *Journal für die reine und angewandte Mathematik*, 264:184–202, 1973.
- [60] P. M. Vaidya. Speeding-up linear programming using fast matrix multiplication. In *Proceedings of the 30th Annual Symposium on Foundations of Computer Science*, pages 332–337, 1989.
- [61] J. van den Brand. A deterministic linear program solver in current matrix multiplication time. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 259–278. SIAM, 2020.
- [62] J. van den Brand, Y. T. Lee, Y. P. Liu, T. Saranurak, A. Sidford, Z. Song, and D. Wang. Minimum cost flows, MDPs, and L1-regression in nearly linear time for dense instances. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing (STOC)*, pages 859–869, 2021.
- [63] J. van den Brand, Y.-T. Lee, D. Nanongkai, R. Peng, T. Saranurak, A. Sidford, Z. Song, and D. Wang. Bipartite matching in nearly-linear time on moderately dense graphs. In *61st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 919–930, 2020.
- [64] J. van den Brand, Y. Tat Lee, A. Sidford, and Z. Song. Solving tall dense linear programs in nearly linear time. In *Proceedings of the 52nd Annual ACM Symposium on Theory of Computing (STOC)*, pages 775–788, 2020.
- [65] S. A. Vavasis and Y. Ye. Condition numbers for polyhedra with real number data. *Operations Research Letters*, 17:209–214, 06 1995.
- [66] S. A. Vavasis and Y. Ye. A primal-dual interior point method whose running time depends only on the constraint matrix. *Mathematical Programming*, 74(1):79–120, 1996.
- [67] Y. Ye. *Interior-Point Algorithms: Theory and Analysis*. John Wiley and Sons, New York, 1997.
- [68] N. Young. *An introduction to Hilbert space*. Cambridge university press, 1988.
- [69] M. Zong, Y. T. Lee, and M.-C. Yue. Short-step methods are not strongly polynomial-time. *Mathematical Programming*, July 2023.

## A Missing Proofs in Section 2.3

*Proof of Proposition 2.7.* We show the first part. In particular, we show that  $\ker(\text{ad}(T)) = V \cap \text{im}(T)^\perp$ :

$$\begin{aligned} \ker(\text{ad}(T)) &= \{y \in V : \langle \text{ad}(T)y, \text{ad}(T)y \rangle = 0\} = \{y \in V : \langle \text{ad}(T)y, u \rangle = 0, \forall u \in U\} \\ &= \{y \in V : \langle y, Tu \rangle = 0, \forall u \in U\} = V \cap \text{im}(T)^\perp. \end{aligned}$$

The second part now follows from the first part, applying the first part with  $T$  replaced by  $\text{ad}(T)$  and using that  $\text{ad}(\text{ad}(T)) = T$ .

For the third part, we must show  $\text{im}(T) = T(\text{im}(\text{ad}(T)))$  and  $\text{im}(\text{ad}(T)) = \text{ad}(T)(\text{im}(T))$ . Since  $\text{ad}(\text{ad}(T)) = T$ , it suffices to prove the first equality. For this purpose, by the first part, we have that

$$\text{im}(T) = T(U) = T(\text{im}(\text{ad}(T)) + \ker(T)) = T(\text{im}(\text{ad}(T))),$$

as needed. In particular, note that the above equality directly implies that  $\text{rank}(T) = \dim(\text{im}(T)) = \dim(T(\text{im}(\text{ad}(T)))) \leq \dim(\text{ad}(T))$ . Using again that  $\text{ad}(\text{ad}(T)) = T$ , we also have  $\text{rank}(\text{ad}(T)) \leq \text{rank}(T)$ , which proves the desired equality  $\text{rank}(T) = \text{rank}(\text{ad}(T))$ .  $\square$

*Proof of Proposition 2.11.* Let  $\mathbf{M}_{ij} = (e^i)^\top T(\Pi_U e^j)$ ,  $\forall i \in [m], j \in [n]$ . With this definition, clearly  $\mathbf{M}x = T(\Pi_U x)$ ,  $\forall x \in \mathbb{R}^n$ , and thus  $\mathbf{M}u = T(\Pi_U u) = T(u)$ ,  $\forall u \in U$ . We now check that  $\text{im}(\mathbf{M}^\top) \subseteq U$ . Since  $U^\perp = \ker(\Pi_U) \subseteq \ker(\mathbf{M})$  by construction, we see that  $\text{im}(\mathbf{M}^\top) = \ker(\mathbf{M})^\perp \subseteq (U^\perp)^\perp = U$  as needed.

We now show uniqueness: if  $\bar{\mathbf{M}}u = T(u)$ ,  $\forall u \in U$  and  $\text{im}(\bar{\mathbf{M}}^\top) \subseteq U$ , then  $\mathbf{M} = \bar{\mathbf{M}}$ . Let  $\mathbf{D} = \mathbf{M} - \bar{\mathbf{M}}$  and examine  $\mathbf{D}\mathbf{D}^\top e^i$  for  $i \in [m]$ . Since  $\mathbf{D}^\top e^i = \mathbf{M}^\top e^i - \bar{\mathbf{M}}^\top e^i \in U$ , we have  $\mathbf{D}\mathbf{D}^\top e^i = \mathbf{M}\mathbf{D}^\top e^i - \bar{\mathbf{M}}\mathbf{D}^\top e^i = T(\mathbf{D}^\top e^i) - T(\mathbf{D}^\top e^i) = \mathbf{0}_m$ . In particular,  $0 = (e^i)^\top \mathbf{D}\mathbf{D}^\top e^i = \|\mathbf{D}^\top e^i\|^2$ , and thus  $\mathbf{D}^\top e^i = \mathbf{0}_n$ ,  $\forall i \in [m]$ . In particular,  $\mathbf{D} = \mathbf{0}_{m \times n}$  and uniqueness follows. Therefore,  $\mathbf{M} = \mathcal{M}(T)$ , as needed.

The equality  $\mathcal{T}(\mathcal{M}(T)) = T|_U^{\mathbb{R}^m} \circ \Pi_U|_{\mathbb{R}^n}^U$  follows directly from the above construction, recalling that  $\mathcal{M}(T)(x) = T(\Pi_U(x))$ ,  $\forall x \in \mathbb{R}^n$ .

We now prove the furthermore. Since  $\text{im}(\mathcal{M}(T)^\top) \subseteq U$  by definition, we have  $U^\perp \subseteq \ker(\mathcal{M}(T))$ , and hence  $\text{im}(\mathcal{M}(T)) = \mathcal{M}(T)(U + U^\perp) = \mathcal{M}(T)(U) = T(U) = \text{im}(T)$ , where the second to last equality follows by definition of  $\mathcal{M}(T)$ . By Proposition 2.7, we have  $\mathbb{R}^m = \ker(\mathcal{M}(T)^\top) + \text{im}(\mathcal{M}(T)) = \ker(\mathcal{M}(T)^\top) + \text{im}(T)$ . In particular,  $\text{im}(\mathcal{M}(T)^\top) = \mathcal{M}(T)^\top(\ker(\mathcal{M}(T)^\top) + \text{im}(T)) = \mathcal{M}(T)^\top(\text{im}(T)) = \text{ad}(T)(\text{im}(T))$ , where the last equality follows by Remark 2.6 and  $\text{im}(T) \subseteq V$ . Since  $\ker(\text{ad}(T)) + \text{im}(T) = V$  by Proposition 2.7, we similarly get that  $\text{ad}(T)(\text{im}(T)) = \text{im}(\text{ad}(T))$ , as needed.  $\square$

## A.1 Approximate Singular Subspaces

We now give the proof of Lemma 2.26, which relates the approximate singular subspaces of an operator to those of its associated matrix.

*Proof of Lemma 2.26.* By Proposition 2.11, we have that  $\bar{T}(x) = T(\Pi_X(x))$ ,  $\forall x \in \mathbb{R}^n$ . Therefore, for all  $x \in X^\perp$  we have  $\bar{T}(x) = T(\Pi_X(x)) = T(\mathbf{0}_n) = \mathbf{0}_m$ , and hence  $X^\perp \subseteq \ker(\bar{T})$ .

We begin by proving (1). Firstly,  $\ker(T) = \ker(\bar{T}) \cap X \subseteq S \cap X$  by our assumption that  $\ker(\bar{T}) \subseteq S$ , which proves the first inclusion. Using that  $X^\perp \subseteq \ker(\bar{T}) \subseteq S$  and orthogonal decomposition, we get the desired first equality

$$\Pi_X(S) = X \cap (S + X^\perp) = X \cap S.$$

We now claim that  $(S \cap X) + X^\perp = S$ . Using orthogonal decomposition again, this follows by

$$S \subseteq \Pi_X(S) + X^\perp = (S \cap X) + X^\perp \subseteq S + S = S. \quad (127)$$

Using the above and  $X^\perp \subseteq \ker(\bar{T})$ , we get that  $\sigma_1(T|_{S \cap X}) = \sigma_1(\bar{T}|_{(S \cap X) + X^\perp})$  via the same argument as in Remark 2.23. This yields  $\sigma_1(T|_{S \cap X}) = \sigma_1(\bar{T}|_S)$  by (127).

We now prove (2). Firstly, using that  $\mathbb{R}^n = X + X^\perp$  and  $S = (S \cap X) + X^\perp$ , we have  $n = \dim(X) + \dim(X^\perp)$  and  $\dim(S) = \dim(S \cap X) + \dim(X^\perp)$ . In particular,

$$n - \dim(S) + 1 = \dim(X) - \dim(S \cap X) + 1. \quad (128)$$

Therefore, using  $\sigma_1(T|_{S \cap X}) = \sigma_1(\bar{T}|_S)$  from part (1), we get that  $S$  is a  $\varrho$ -approximate singular subspace for  $\bar{T}$  of dimension  $d \geq 0$  if and only if  $S \cap X$  is a  $\varrho$ -approximate singular subspace for  $T$  of dimension  $d - \dim(X^\perp)$ .

We now prove the moreover. By Proposition 2.19, recall that  $\sigma^+(T) = \sigma^+(\mathcal{M}(T)) = \sigma^+(\bar{T})$ . Thus,  $\sigma_{[\dim(X)]}(T) = \sigma_{[\dim(X)]}(\bar{T})$  and  $\sigma_{[n] \setminus [\dim(X)]}(\bar{T}) = \mathbf{0}_{[n] \setminus [\dim(X)]}$ . Therefore, for  $\tau \geq 0$ ,

$$\mathcal{C}_\sigma(\bar{T}, \tau) = \mathcal{C}_\sigma(T, \tau) + (n - \dim(X)) = \mathcal{C}_\sigma(T, \tau) + \dim(X^\perp).$$

$\square$

We now move to the proof of Lemma 2.27, which shows that the orthogonal complement of an approximate singular subspace is an approximate maximizer of (16) whenever there is a large enough gap in the singular values. The proof will require the following two helper lemmas.

**Lemma A.1** (Subspace Projection Adjoints). *Let  $S, U \subseteq \mathbb{R}^n$  be subspaces. Then,  $\Pi_U|_S^U = \text{ad}(\Pi_S|_U^S)$ . In particular,  $\sigma^+(\Pi_U|_S) = \sigma^+(\Pi_S|_U)$  and if  $\dim(U) = \dim(S)$ , then  $\sigma(\Pi_U|_S) = \sigma(\Pi_S|_U)$ .*

*Proof.* Let  $s \in S$  and  $u \in U$ . Then,

$$\begin{aligned} \langle s, \Pi_U|_S(u) \rangle &= \langle s, \Pi_S(u) \rangle = \langle \Pi_S(s), u \rangle = \langle s, u \rangle \\ &= \langle s, \Pi_U(u) \rangle = \langle \Pi_U(s), u \rangle = \langle \Pi_U|_S(s), u \rangle. \end{aligned}$$

This proves that  $\Pi_U|_S^U$  and  $\Pi_S|_U^S$  are adjoints of one another. The other statements follow from this and Proposition 2.18.  $\square$

**Lemma A.2** (Singular Subspace Proximity). *Let  $T: X \rightarrow Y$ ,  $X \subseteq \mathbb{R}^n, Y \subseteq \mathbb{R}^m$  be a linear operator with singular value decomposition  $\mathcal{M}(T) = \sum_{i=1}^{\text{rank}(T)} \sigma_i(T) y_i x_i^\top$  as in Definition 2.16. Let  $S \subseteq X$  be a  $\varrho$ -approximate singular subspace for  $T$  with  $\dim(\ker(T)) \leq \dim(S) < \dim(X)$  and let  $U = \text{span}(x_1, \dots, x_{\dim(X)-\dim(S)})^\perp \cap X$ . Then, for all  $s \in S \setminus \{\mathbf{0}_n\}$  we have that*

$$1 \geq \frac{\|\Pi_U(s)\|_2^2}{\|s\|_2^2} \geq 1 - \varrho^2 \frac{\sigma_{\dim(X)-\dim(S)+1}(T)^2}{\sigma_{\dim(X)-\dim(S)}(T)^2}.$$

*Proof.* The first inequality follows from properties of projections. We now prove the second inequality. Recalling that  $\text{rank}(T) + \dim(\ker(T)) = \dim(X)$  and using that  $\dim(\ker(T)) \leq \dim(S) < \dim(X)$ , we have  $1 \leq \dim(X) - \dim(S) \leq \text{rank}(T)$ . Let  $\bar{U} = \text{span}(x_1, \dots, x_{\dim(X)-\dim(S)}) = U^\perp \cap X$ . Note that  $T(\bar{U}) = \text{span}(y_1, \dots, y_{\dim(X)-\dim(S)}) \perp \text{span}(y_{\dim(X)-\dim(S)+1}, \dots, y_{\text{rank}(T)}) = T(U)$ . Therefore, for any  $s \in S \setminus \{\mathbf{0}_n\}$  we have that

$$\begin{aligned} \varrho^2 \sigma_{\dim(X)-\dim(S)+1}(T)^2 &\geq \frac{\|T(s)\|_2^2}{\|s\|_2^2} = \frac{\|T(\Pi_U(s))\|_2^2 + \|T(\Pi_{\bar{U}}(s))\|_2^2}{\|s\|_2^2} \\ &\geq \sigma_{\min}(T|_{\bar{U}})^2 \frac{\|\Pi_{\bar{U}}(s)\|_2^2}{\|s\|_2^2} = \sigma_{\dim(X)-\dim(S)}(T)^2 \frac{\|\Pi_{\bar{U}}(s)\|_2^2}{\|s\|_2^2}. \end{aligned} \tag{129}$$

Noting that  $\|s\|_2^2 = \|\Pi_U(s)\|_2^2 + \|\Pi_{\bar{U}}(s)\|_2^2$  and reordering the terms gives the result, where the condition  $\dim(X) > \dim(S) \geq \dim(\ker(T))$  ensures that  $\sigma_{\dim(X)-\dim(S)}(T) \geq \sigma_{\text{rank}(T)}(T) > 0$ .  $\square$

We are now ready to prove Lemma 2.27.

*Proof of Lemma 2.27.* If  $\dim(S) < \dim(\ker(T))$ , then  $\sigma_{\dim(X)-\dim(S)}(T) \leq \sigma_{\text{rank}(T)+1}(T) = 0$  and there is nothing to prove. So assume  $\dim(S) \geq \dim(\ker(T))$ . Let  $U$  be as in Lemma A.2. Recall that  $\bar{U} = U^\perp \cap X$  satisfies  $T(U) \perp T(\bar{U})$  and  $\sigma_{\min}(T|_{\bar{U}}) = \sigma_{\dim(X)-\dim(S)}(T)$  by construction. Furthermore,  $\dim(S) = \dim(U)$  and  $\dim(\bar{S}) = \dim(\bar{U}) = \dim(X) - \dim(S)$ . From here, we get that

$$\begin{aligned} \sigma_{\min}(T|_{\bar{S}})^2 &= \min_{\bar{s} \in \bar{S} \setminus \{\mathbf{0}\}} \frac{\|T(\bar{s})\|_2^2}{\|\bar{s}\|_2^2} \\ &\geq \min_{\bar{s} \in \bar{S} \setminus \{\mathbf{0}\}} \frac{\|T(\Pi_{\bar{U}}(\bar{s}))\|_2^2}{\|\bar{s}\|_2^2} && \text{(by } T(U) \perp T(\bar{U})\text{)} \\ &\geq \sigma_{\dim(X)-\dim(S)}(T)^2 \min_{\bar{s} \in \bar{S} \setminus \{\mathbf{0}\}} \frac{\|\Pi_{\bar{U}}(\bar{s})\|_2^2}{\|\bar{s}\|_2^2} && \text{(as } \sigma_{\min}(T|_{\bar{U}}) = \sigma_{\dim(X)-\dim(S)}(T)\text{)} \\ &= \sigma_{\dim(X)-\dim(S)}(T)^2 \left( 1 - \max_{\bar{s} \in \bar{S} \setminus \{\mathbf{0}\}} \frac{\|\Pi_U(\bar{s})\|_2^2}{\|\bar{s}\|_2^2} \right) && \text{(as } S \subseteq U + \bar{U} \text{ and } U \perp \bar{U}\text{)} \\ &= \sigma_{\dim(X)-\dim(S)}(T)^2 \left( 1 - \max_{u \in U \setminus \{\mathbf{0}\}} \frac{\|\Pi_S(u)\|_2^2}{\|u\|_2^2} \right) && \text{(by Lemma A.1)} \\ &= \sigma_{\dim(X)-\dim(S)}(T)^2 \min_{u \in U \setminus \{\mathbf{0}\}} \frac{\|\Pi_S(u)\|_2^2}{\|u\|_2^2} && \text{(as } U \subseteq S + \bar{S} \text{ and } S \perp \bar{S}\text{)} \\ &= \sigma_{\dim(X)-\dim(S)}(T)^2 \min_{s \in S \setminus \{\mathbf{0}\}} \frac{\|\Pi_U(s)\|_2^2}{\|s\|_2^2} && \text{(by Lemma A.1)} \\ &\geq \sigma_{\dim(X)-\dim(S)}(T)^2 \left( 1 - \varrho^2 \frac{\sigma_{\dim(X)-\dim(S)+1}(T)^2}{\sigma_{\dim(X)-\dim(S)}(T)^2} \right). && \text{(by Lemma A.2)} \end{aligned}$$

This proves the lemma.  $\square$