Trust Region Interior Point Methods: Optimal ℓ_2- and Faster Wide-Neighborhood Path Following

Daniel Dadush* Haoyuan Ma† Bento Natura‡ László A. Végh†

Abstract

We present improved running time and iteration complexities of interior point methods for linear programs parametrized by the straight line complexity, i.e., the minimum number of segments of any piecewise linear curve traversing a particular neighborhood of the central path. While the standard measure of progress is the reduction in duality gap, the straight line complexity provides a stronger instance-wise bound, reflecting the combinatorial structure of the problem.

Our first main result is a wide neighborhood interior point method whose running time is the wide-neighborhood straight line complexity times current matrix multiplication time, improving in essence a factor n over the algorithm by Allamigeon, Dadush, Loho, Natura, and Végh (SIAM J. Comput. 2025). The algorithm can be seen as a boosted version of the robust interior point methods of Cohen, Lee and Song (JACM 2021) and van den Brand (SODA 2020) that can reduce the gap by a polynomial factor in current matrix multiplication time. Our algorithm is also able to traverse any near-linear segments of the central path in current matrix multiplication time, independently of the length of the segment.

Our second main result focuses on interior point methods that stay in the narrow ℓ_2-neighborhood. We give a much stronger analysis of the ℓ_2-trust region interior point method introduced by Lan, Monteiro and Tsuchiya (SIAM J. Optim. 2009), showing that it is approximately instance optimal in this neighborhood: the number of iterations is within a constant factor of the lower bound.

A main ingredient in both methods are trust region subroutines with ℓ_∞ and ℓ_2-constraints, respectively. We develop fast and strongly polynomial algorithms for solving both these problems to high accuracy. In the ℓ_2-setting, this answers an open question by Lan, Monteiro and Tsuchiya.

*Centrum Wiskunde & Informatica (CWI), Amsterdam, The Netherlands.
†University of Bonn, Bonn, Germany.
‡Columbia University, New York, USA.

Contents

1 Introduction

Linear programming (LP) is one of the most fundamental problems in optimization and computer science, and the pursuit of efficient LP algorithms has been a major driving force in these disciplines. We write the problem in the standard primal-dual formulation

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

where the input is $\mathbf{A} \in \mathbb{R}^{m\times n}$ with $\mathrm{rk}(\mathbf{A}) = m \leq n$, $b \in \mathbb{R}^m$ and $c \in \mathbb{R}^n$. We refer to these as primal and dual programs, respectively, and we let $\mathcal{P} \coloneqq \{x \in \mathbb{R}^n : \mathbf{A}x = b\,,\ x \geq \mathbf{0}\}$ and $\mathcal{D} \coloneqq \{s \in \mathbb{R}^n : \exists\, y \text{ s.t. } \mathbf{A}^\top y + s = c, s \geq \mathbf{0}\}$ denote the primal and dual feasible regions.

Interior point methods (IPMs), first introduced by Karmarkar in 1984 [38], provide a rich family of LP algorithms that are efficient in theory as well as in practice. A popular class is formed by *path-following IPMs*, pioneered by Renegar [59], Megiddo [45], Kojima, Mizuno and Yoshise [39], and Nesterov and Nemirovski [55]. Primal-dual path following methods maintain a pair of strictly feasible primal and dual solutions in every iteration; the main work in each iteration corresponds to solving a linear system of equations. The worst case complexity estimate shows that the optimality gap is reduced by a constant factor in every $O(\sqrt{n})$ iterations.

The iterates are guided by the *central path*: for each value $\mu > 0$, consider the system

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

This system arises from the optimality characterization of $\min_{x:\mathbf{A}x=b} \langle c, x\rangle - \mu\sum_{i=1}^n \log x_i$; the latter term is the *logarithmic barrier function*. This has a unique solution $x^{\mathrm{cp}}(\mu), y^{\mathrm{cp}}(\mu), s^{\mathrm{cp}}(\mu)$; we denote $z^{\mathrm{cp}}(\mu) \coloneqq (x^{\mathrm{cp}}(\mu), s^{\mathrm{cp}}(\mu))$ and call it the central path point at μ. Note that the duality gap of $z^{\mathrm{cp}}(\mu)$ is $\langle c, x^{\mathrm{cp}}(\mu)\rangle - \langle b, y^{\mathrm{cp}}(\mu)\rangle = \langle x^{\mathrm{cp}}(\mu), s^{\mathrm{cp}}(\mu)\rangle = n\mu$. The points $\{z^{\mathrm{cp}}(\mu) : \mu > 0\}$ form a smooth algebraic curve in $\mathbb{R}^{2n}$, and for $\mu \searrow 0$, it converges to a pair of primal and dual optimal solutions $(x^\star, s^\star)$ to (LP); for a proof see e.g. [45].

From a practical perspective, according to Gondzio [26], the impressive features of IPM include *"[...] their low-degree polynomial worst-case complexity and an unrivalled ability to deliver optimal solutions in an almost constant number of iterations which depends very little, if at all, on the problem dimension."* According to the documentation of the solver MOSEK [1], the *"interior-point optimizer [...] tends to use between 20 and 100 iterations, almost independently of problem size"*.

From a theoretical perspective, as a culmination of a long line of research, Cohen, Lee, and Song [13] gave a randomized LP algorithm in *current matrix multiplication* time, by amortizing the work of computing subsequent iterates. Let us use ω to denote the matrix multiplication exponent, and α the dual matrix multiplication exponent;[1] for brevity, we denote $\tilde{\omega} \coloneqq \max\left\{\omega, 2 + \frac{1}{6}, 2.5 - \frac{\alpha}{2}\right\}$. For current best values $\omega \leq 2.371552$ and $\alpha \geq 0.321334$ [76], we have $\tilde{\omega} = \omega$. Thus, given a pair of solutions near the central path, the algorithm in [13] can reduce the duality gap by a factor ϱ in $n^{\tilde{\omega}+o(1)} \log(\varrho)$. Subsequently, van den Brand [71] gave a deterministic algorithm with the same guarantee; [37] improved on the second term in the definition of $\tilde{\omega}$ in the randomized setting. We refer to these algorithms as *'robust IPM methods'*; for an accessible tutorial, see [43]. For a rational input with bit-encoding length L, robust IPMs yield $n^{\tilde{\omega}+o(1)}L$ time algorithms for finding exact optimal primal and dual solutions to (LP). These algorithms still use short steps, and the $O(\sqrt{n})$ iteration complexity to reduce the gap by a constant factor remains the best known in the general setting.

However, for a particular instance, much fewer iterations may suffice. As an example, consider the simple LP $\min c^\top x$ s.t. $\mathbf{1}_n^\top x = 1$, $x \geq 0$, where $c_i = \varepsilon^{i-1}(1 - \varepsilon^{n-i})$, $i \in [n]$ for a very small $\varepsilon > 0$. Here, the primal-dual central path comprises n largely linear segments: letting $x_i \coloneqq (\varepsilon^i, \ldots, \varepsilon, \mathbf{1}_{n-i})/(n - i + \frac{\varepsilon(1-\varepsilon^i)}{(1-\varepsilon)})$, $s_i \coloneqq c + \varepsilon^i \mathbf{1}_n$ for $i \in \{0, \ldots, n-1\}$, and $(x_n, s_n) \coloneqq (e_n, c)$, the i^{th} segment, $i \in [n]$, essentially interpolates from (x_{i-1}, s_{i-1}) to (x_i, s_i). Starting from the near-central primal-dual solution (x_0, s_0), the algorithms [13, 71] would require $\tilde{O}(\sqrt{n}\log(1/\varepsilon))$ iterations and a total runtime $n^{\tilde{\omega}+o(1)} \log(1/\varepsilon)$ to reach the gap value ε. Even at this point the solution will be approximately (x_1, s_1), which is very far from the primal-dual optimal solution (e_n, c). Further, while such IPMs converge to an optimal solution, they do not reach an exact one and require a final rounding step when high enough accuracy is reached (in this example, less than ε^n), requiring $\tilde{O}(n^{1.5}\log(1/\varepsilon))$ iterations.

[1] That is, any two $n \times n$ matrices can be multiplied in $n^{\omega+o(1)}$ time, and any $n \times n$ and $n \times n^\alpha$ matrices can be multiplied in $n^{2+o(1)}$ time.

Hence, gap reduction may not be the only relevant progress measure for IPMs. The focus of this paper is on IPMs that can achieve *(near) instance optimal guarantees* with much fewer steps by detecting the combinatorial structure of the central path. In the above example, such an IPM would find an exact optimal solution in just $O(n)$ iterations. In order to describe this class of IPMs, we first describe the concept of central path neighborhoods.

Central path neighborhoods. The central path has robust advantageous properties that also hold for nearby points. For $z = (x, s) \in \mathcal{P} \times \mathcal{D}$, we define $\overline{\mu}(z) \coloneqq \langle x, s \rangle / n$ and refer to it as the *normalized duality gap* of z. One can define neighborhoods of the central path based on the vector $\frac{xs}{\overline{\mu}(z)} - \mathbf{1}_n$, called the *centrality error*. Throughout, $xs \in \mathbb{R}^n$ denotes the Hadamard product of the two vectors and $\mathbf{1}_n \in \mathbb{R}^n$ is the all ones vector. Two commonly used neighborhoods are

$$\begin{aligned} \mathcal{N}^2(\beta) &\coloneqq \left\{ z = (x, s) \in \mathcal{P} \times \mathcal{D} \colon \left\| \frac{xs}{\overline{\mu}(z)} - \mathbf{1}_n \right\|_2 \le \beta \right\} , \qquad \beta \in (0, 1) , \\ \mathcal{N}^{-\infty}(\beta) &\coloneqq \left\{ z = (x, s) \in \mathcal{P} \times \mathcal{D} \colon xs \ge (1 - \beta)\overline{\mu}(z)\mathbf{1}_n \right\} , \qquad \beta \in (0, 1) . \end{aligned} \tag{1}$$

$\mathcal{N}^2(\beta)$ is called the ℓ_2*-neighborhood* and $\mathcal{N}^{-\infty}(\beta)$ is called the *wide neighborhood*. Note that $\mathcal{N}^2(\beta) \subseteq \mathcal{N}^{-\infty}(\beta)$, but the latter neighborhood can be significantly larger. We let $\overline{\mathcal{N}}^2(\beta) := \mathrm{cl}(\mathcal{N}^2(\beta))$ and $\overline{\mathcal{N}}^{-\infty}(\beta) := \mathrm{cl}(\mathcal{N}^{-\infty}(\beta))$ denote the closure of the neighborhoods that also includes points $z = (x, s) \in \mathcal{P} \times \mathcal{D}$ with $\overline{\mu}(z) = 0$, i.e., optimal solutions. In general, by a *neighborhood* $\mathcal{N}$ we mean a subset of $\mathcal{P} \times \mathcal{D}$ that contains the central path $\{z^{\mathrm{cp}}(\mu) : \mu > 0\}$, and we let $\overline{\mathcal{N}}$ denote the closure of $\mathcal{N}$.

The ℓ_2-neighborhood is very amenable to theoretical analysis. Many classical IPMs, including the Primal-Dual Predictor-Corrector Method by Mizuno, Todd and Ye [50], stay throughout in the ℓ_2-neighborhood: each step corresponds to a linear segment inside this neighborhood. The robust IPMs [13, 71] use an intermediate neighborhood that is contained in the wide neighborhood, see Section 1.3. The wide neighborhood captures the trajectories generated by most IPMs in the literature: it follows from [6] that the trajectories of all algorithms based on a self-concordant barrier function lie in $\mathcal{N}^{-\infty}(1 - 1/(2\nu))$, where ν is the parameter of the barrier ($\nu \le \mathrm{poly}(n)$ for all major barriers).

Straight-line complexity bounds. Allamigeon, Benchimol, Gaubert, and Joswig [4] proposed a systematic way to lower bound the iteration complexity of IPMs. Consider an IPM where each step corresponds to a line segment inside a central path neighborhood $\mathcal{N}$. Then, we can get a lower bound on the iteration complexity as follows.

Definition (Straight Line Complexity)**.** Given a central path neighborhood $\mathcal{N}$ and $\mu_0 > \mu_1 \ge 0$, let $\mathrm{SLC}(\mathcal{N}, \mu_1, \mu_0)$ denote the *minimum* number of segments of any piecewise linear curve $\Gamma : [\mu_1, \mu_0] \to \overline{\mathcal{N}}$ that satisfies $\overline{\mu}(\Gamma(\mu)) = \mu$ for all $\mu_1 \le \mu \le \mu_0$.

Using this framework, they constructed a family of LPs in n variables and $O(n)$ constraints parametrized by a real parameter $t > 0$, such that for any $\beta > 0$, there exists an LP in the family and $t > 0$ with $\mathrm{SLC}(\mathcal{N}^{-\infty}(\beta), 1, t) \ge 2^n - 1$. This result was strengthened by Allamigeon, Gaubert and Vandame [6], who provided a family of combinatorial cubes (a.k.a. Klee–Minty cubes) having $2n$ constraints and n variables, achieving the same result.

Allamigeon, Dadush, Loho, Natura, and Végh [5] complemented the straight line complexity lower bound by an algorithmic result. They developed a *subspace layered least square (SLLS) IPM* that stays in the ℓ_2-neighborhood, while guaranteeing a number of iterations that matches the straight line complexity of the wide neighborhood up to a polynomial factor.

Theorem 1.1 ([5])**.** *There exists a path-following interior point method SLLS IPM with the following properties. Let $\theta \in (0, 1)$, $\beta \in (0, 1/6]$, and consider an instance of* (LP)*; let $\mu_0 > \mu_1 \ge 0$. Assume we are given a starting feasible solution $z \in \mathcal{N}^2(\beta)$ with $\mu(z) \le \mu_1$. Then, the SLLS IPM reaches an iterate z' with $\overline{\mu}(z') \le \mu_1$ within $O(n^{1.5}\beta^{-1}\log(n/(\beta(1-\theta)))\,\mathrm{SLC}(\mathcal{N}^{-\infty}(\theta), \mu_1, \mu_0))$ iterations. Every iteration corresponds to a line segment in $\mathcal{N}^2(\beta)$, and can be implemented in strongly polynomial time.*

We emphasize that only the existence of the piecewise linear curve is assumed and the algorithm does not have access to it. The key insight is that a long straight segment in the central path neighborhood enforces a certain 'polarization' of the corresponding part of the central path (see Section 1.2); the algorithm exploits this property. The IPM in [5] is the key component of the strongly polynomial algorithm for LPs with two nonzeros per column in [17]; see Section 1.5 for more background on 'combinatorial' IPMs and their applications for strongly polynomial computability.

1.1 Our contributions

Theorem 1.1 can be interpreted as a form of *approximate instance optimality* in the context of path following. Instance optimality is a very strong concept in beyond worst-case complexity, asserting that an algorithm is (up to a constant factor) on any input better than any other algorithm from a given class; see [62, Chapter 3], and we mention some examples in Section 1.5. Given the straight line complexity lower bound, one can define (approximate) instance optimality of path following IPMs in a neighborhood as follows.

Definition 1.2 (Near-optimality of Path-Following IPM)**.** Let $\mathcal{N} \subseteq \mathcal{N}'$ be two neighborhoods of the central path, and $\mu_0 > \mu_1 \geq 0$ be given. A primal-dual IPM is *α-optimal in $\mathcal{N}'$ with respect to $\mathcal{N}$* if it takes at most $\alpha\,\mathrm{SLC}(\mathcal{N}, \mu_1, \mu_0)$ many steps to arrive from any $z \in \mathcal{N}$ with $\overline{\mu}(z) = \mu$ to some $z_1 \in \mathcal{N}$ with $\overline{\mu}(z_1) \leq \mu_1$, and all iterates and the line segments between consecutive iterates lie in $\mathcal{N}'$. If $\mathcal{N} = \mathcal{N}'$, then we simply say that the algorithm is *α-optimal in $\mathcal{N}$*.

In these terms, the SLLS IPM in [5] is $O\left(n^{1.5} \log\left(n/(1-\theta)\right)\right)$-optimal in $\mathcal{N} = \mathcal{N}^{-\infty}(\theta)$. In fact, the iterates stay in the narrower $\mathcal{N}(\beta)$-neighborhood. We improve on this result in two directions, by exhibiting two algorithms, the *ℓ_∞-Trust Region IPM* (TRW-IPM) that is described in Section 3.1 as Algorithm 1, and the *ℓ_2-Trust Region IPM* (TR2-IPM) that is described in Section 4 as Algorithm 2. The first result significantly improves the running time bound in Theorem 1.1: we give an IPM whose running time is bounded by the $\mathcal{N}^{-\infty}(\theta)$ straight line complexity times the current matrix multiplication time.

Theorem 1.3 (Path Following in Current Matrix Multiplication Time)**.** *For any $\theta \in [1/8, 1)$, given $\mu_0 > \mu_1 \geq 0$ and a starting point $z_0 \in \mathcal{N}^{-\infty}(\theta)$ with $\overline{\mu}(z_0) \in [\mu_1, \mu_0]$, the algorithm TRW-IPM is $O(1)$-instance optimal in $\mathcal{N}^{-\infty}(\theta')$ with respect to $\mathcal{N}^{-\infty}(\theta)$, for $\theta' = 1 - ((1-\theta)/n)^{O(1)}$. The algorithm finds a solution $z_1 \in \overline{\mathcal{N}}^{-\infty}(\theta)$ with $\overline{\mu}(z_1) \leq \mu_1$ in randomized runtime $n^{\tilde{\omega}+o(1)} \log\left(\frac{1}{1-\theta}\right) \mathrm{SLC}(\mathcal{N}^{-\infty}(\theta), \mu_1, \mu_0)$, or deterministic runtime $O\left(\left(n^3 + n^{\tilde{\omega}+o(1)} \log\left(\frac{1}{1-\theta}\right)\right) \mathrm{SLC}(\mathcal{N}^{-\infty}(\theta), \mu_1, \mu_0)\right)$.*

The algorithm uses the deterministic Robust IPM by van den Brand [71] as black-box. The main improvement is that it can traverse in $n^{\tilde{\omega}+o(1)}$ time an *arbitrarily long* segment between μ and μ' of the central path assuming there exists a straight line segment in the wide neighborhood, instead of $n^{\tilde{\omega}+o(1)} \log(\mu/\mu')$ as guaranteed by the robust IPM [71]. The running time of the SLLS IPM in [5] is independent from $\log(\mu/\mu')$, but requires $\tilde{O}(n^{1.5})$ iterations, totaling to runtime $n^{\tilde{\omega}+1.5+o(1)}$. While the short steps in this algorithm could be replaced by a Robust IPM subroutine, it would still require n subspace layered least square steps, each taking n^{ω}. Hence, Theorem 1.4 improves by a factor n over [5].

We also note that the running time bound in Theorem 1.3 directly improves on the bounds of the Robust IPM. In fact, by choosing $\theta = 1 - 1/\mathrm{poly}(n)$, the straight line complexity of any segment between μ and μ' with $\mu/\mu' = \mathrm{poly}(n)$ has straight line complexity $O(1)$. This follows by Lemma 2.9 that shows that one can traverse any segment of multiplicative length $\gamma \in [0, 1]$ by a straight line in $\mathcal{N}^{-\infty}(1 - \gamma(1-\theta))$.

While the above result gives the strong theoretical guarantee that any straight segment of the wide neighborhood can be traversed essentially in current matrix multiplication time, the underlying Robust IPM used in each large iteration still performs $O(\sqrt{n} \log(n/(1-\theta)))$ short steps. The computational cost of these steps is amortized by using implicit updates and efficient data structures. In contrast, IPMs in practice are known to perform a very small number of iterations with long steps, each computed by solving a linear system. Towards a better theoretical explanation of this phenomenon, it is desirable to understand the best achievable iteration complexity of path following methods.

Our second main result focuses on ℓ_2-path following, a setting of many classical IPMs, and provides a much stronger instance optimality guarantee here. We show that the ℓ_2-Trust Region IPM (TR2-IPM), a natural and relatively simple method first proposed by Lan, Monteiro, and Tsuchiya [40] in 2009, is $O(1)$-optimal. The method combines standard affine scaling steps with special predictor steps that can traverse long straight segments.

Theorem 1.4 (Instance Optimality of the ℓ_2-Trust Region IPM)**.** *For any $\bar{\beta} \in (0, 2^{-8}]$, the algorithm TR2-IPM with the parameter choice $\beta = \bar{\beta}/82$ is $O(1)$-optimal in $\mathcal{N}^2(\bar{\beta})$. Namely, given any $\mu_0 > \mu_1 \geq 0$ and a starting point $z_0 \in \mathcal{N}^2(\bar{\beta})$ with $\overline{\mu}(z_0) \in [\mu_1, \mu_0]$, the TR2-IPM algorithm finds a solution $z_1 \in \overline{\mathcal{N}}^2(\bar{\beta})$ with $\overline{\mu}(z_1) \leq \mu_1$ in $O\left(\mathrm{SLC}(\mathcal{N}^2(\bar{\beta}), \mu_1, \mu_0)\right)$ many iterations. Further, all iterates and the line segments between consecutive iterates stay in $\overline{\mathcal{N}}^2(\bar{\beta})$. Every iteration of the algorithm can be implemented in strongly polynomial $O(n^3)$ time deterministically, or by an $\tilde{O}(n^{\omega})$ randomized algorithm.*

Both algorithms TRW-IPM and TR2-IPM use strongly polynomial singular value decomposition (SVD). The difference in the deterministic and randomized running times arise from this: while [22] yields a randomized $\tilde{O}(n^\omega)$ subroutine for SVD in strongly polynomial time, it would require $O(n^3)$ deterministic running time. For completeness, we give one deterministic subroutine based on Gaussian elimination with complete pivoting in Appendix C.1. All other parts of TRW-IPM and TR2-IPM are deterministic.

Comparison to the Subspace Layered Least Squares IPM. Since each iteration of [5] stays inside the $\mathcal{N}^2(\beta)$ neighborhood for $\beta \in (0, 1/6]$, it follows that the iteration complexity of TR2-IPM is asymptotically at least as good as the SLLS IPM. In fact, [4] identifies the trust-region direction as the 'ideal' movement direction along a polarized segment. It develops the subspace LLS step as a 'good enough' approximation of trust region. The paper states the main reason for not using the trust region step as the lack of a strongly polynomial implementation, which we also resolve in this paper.

We note that [5] also includes a stronger, amortized version of Theorem 1.1. A more fine-grained view on straight line complexity (SLC) is to define a separate bound for each variable i. The SLC in Theorem 1.1 is between the largest coordinate-wise SLC and the sum of these quantities. Theorem 1.4 in their paper shows that the number of SLLS steps can be bounded by the sum of coordinate-wise SLCs. Our bound in 1.3 can be shown to be at least as good as this, and can be a factor n better if the SLC is closer to the maximum coordinate-wise SLC.

1.2 From Polarization to Trust Region

Both of our algorithms can be understood through the lens of *polarization*. This was formally introduced in [5], but also used implicitly in previous work such as [40, 73]. Assume there is a straight line that goes from duality gap μ_0 to $\mu_1 < \mu_0$ inside the wide neighborhood of the central path. Then there exists partition (B, N) of the variable set $[n]$ such that for $i \in N$, $x_i(\mu)$ scales down approximately linearly with $\mu \in [\mu_1, \mu_0]$, whereas for $i \in B$, $x_i(\mu)$ changes only in a bounded way; the duals $s_i(\mu)$ exhibit the analogous behavior with B and N swapped. This is captured by the following definition.

Definition 1.5 (Wide Neighborhood Polarization)**.** For $\gamma \in (0, 1]$ and $0 \le \mu_1 < \mu_0$, we say that the central path is γ*-polarized* on $[\mu_1, \mu_0]$ with partition (B, N) if for any $\mu \in [\mu_1, \mu_0]$,

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

The definition implies the analogous bounds for $s^{\mathrm{cp}}(\mu)$ with the role of B and N swapped. It was shown in [5] that if there is a straight line segment in $\mathcal{N}^{-\infty}(\theta)$ between two points with normalized gaps μ_0 and μ_1, then the central path is $\frac{(1-\theta)^2}{16n^3}$-polarized for some partition (B, N) between μ_1 and μ_0 (see also Lemma 3.3).

From here, one can show (see Lemma 3.6) that if in the above setting $z = (x, s) \in \mathcal{N}^{-\infty}(\theta)$ with $\overline{\mu}(z) \le \mu_0$ and $z_1 \in \mathcal{N}^{-\infty}(\theta)$ denotes the endpoint of the straight line segment with $\overline{\mu}(z_1) = \mu_1$, then the 'ideal direction' $\Delta z^{\mathrm{id}} = (\Delta x^{\mathrm{id}}, \Delta s^{\mathrm{id}}) := (x_1 - x, s_1 - s)$ is feasible to the following system with $\ell = (1-\theta)^3/(16n^3)$ and $u = n/(1-\theta)$.

$$\begin{aligned} &\min \left\| \mathbf{1}_N + \frac{\Delta x_N}{x_N} \right\|_\infty && \min \left\| \mathbf{1}_B + \frac{\Delta s_B}{s_B} \right\|_\infty \\ &\text{s.t. } \ell\mathbf{1} \le \mathbf{1}_B + \frac{\Delta x_B}{x_B} \le u\mathbf{1} && \text{s.t. } \ell\mathbf{1} \le \mathbf{1}_N + \frac{\Delta s_N}{s_N} \le u\mathbf{1} \\ &\mathbf{A}\Delta x = \mathbf{0} && \mathbf{A}^\top \Delta y + \Delta s = \mathbf{0} \end{aligned} \qquad (\mathrm{TR}_\infty(B, N, \ell, u))$$

In TRW-IPM, we use the subroutine PathFollow to decrease the duality gap by a factor $\mathrm{poly}(n/(1-\theta))$ in $n^{\tilde{\omega}+o(1)}$ time by implementing a Robust IPM (see Theorem 2.25). After each call to this subroutine, we guess the partition (B, N), by observing the changes in the variables x_i and s_i during PathFollow. If we are currently on a long polarized segment, there is a simple way to guess the polarized partition (B, N): B is formed by the variables where the multiplicative change in x_i is smaller than the change in s_i. We then use a second subroutine WTR, discussed in Section 1.3, to find a direction $(\Delta x, \Delta s)$ by approximately solving $\mathrm{TR}_\infty(B, N, \ell, u)$ in $n^{\tilde{\omega}+o(1)}$, and compute a corresponding step length. Assuming we are on a long polarized segment, this is guaranteed to take us close to the end of this segment.

We now turn to the $O(1)$-instance optimal ℓ_2-Trust Region IPM. If we assume that a straight line segment exists between two points with normalized gaps μ_0 and μ_1 inside the $\mathcal{N}^2(\beta)$ neighborhood, then we can get stronger

ℓ_2-bounds on the change of primal and dual variables between μ_0 and μ_1. We call this notion *ℓ_2-polarization* and describe it in detail in Section 1.4.1. The corresponding ℓ_2-Trust Region problem is defined as

$$\begin{aligned} \min \left\| \frac{x_N + \Delta x_N}{x_N} \right\| \qquad\qquad & \min \left\| \frac{s_B + \Delta s_B}{s_B} \right\| \\ \text{s.t.} \left\| \frac{\Delta x_B}{x_B} \right\| \le \gamma \qquad\qquad & \text{s.t.} \left\| \frac{\Delta s_N}{s_N} \right\| \le \gamma \\ \mathbf{A}\Delta x = \mathbf{0} \qquad\qquad & \mathbf{A}^\top \Delta y + \Delta s = \mathbf{0} \end{aligned} \tag{$\mathrm{TR}_2(B,N,\gamma)$}$$

Thus, the primal trust region step Δx is trying to decrease the local norm of the coordinates in N by the maximum possible amount while only allowing a limited ℓ_2-local change in the coordinates in B.

Trust region steps are a key ingredient of TR2-IPM. As mentioned above, the algorithm itself is not new: it was first introduced by Lan, Monteiro, and Tsuchiya [40]. Their motivation was to develop a version of the Vavasis–Ye Layered Least Squares IPM [73] that is invariant under column scaling; see Section 1.5 for the context. In this paper, we derive the much stronger $O(1)$-instance optimality property.

The algorithm is a predictor-corrector method that uses two types of predictor steps. Starting from an iterate $z = (x, s) \in \mathcal{N}^2(\beta)$, the algorithm first computes the affine scaling direction $\Delta z^{\mathrm{a}} = (\Delta x^{\mathrm{a}}, \Delta s^{\mathrm{a}})$, a standard step in the literature (see Section 2.4). The algorithm then checks whether the largest step length α^{a} such that $z + \alpha^{\mathrm{a}}\Delta z^{\mathrm{a}} \in \mathcal{N}^2(O(\beta))$ is at least a constant (the standard analysis yields $\alpha^{\mathrm{a}} = \Omega(1/\sqrt{n})$, see Proposition 2.10). If $\alpha^{\mathrm{a}} \ge 1/4$, then the algorithm computes an alternative *trust region direction* $\Delta z^{\mathrm{TR}} = (\Delta x^{\mathrm{TR}}, \Delta s^{\mathrm{TR}})$ along with the maximal step length α^{TR}, and implements a trust region step $z + \alpha^{\mathrm{TR}}\Delta z^{\mathrm{TR}}$. Otherwise, the algorithm takes a standard affine scaling step $z + \alpha^{\mathrm{a}}\Delta z^{\mathrm{a}}$. In both cases, $O(1)$ many corrector steps are taken afterwards to move back to $\mathcal{N}^2(\beta)$.

The trust region steps are computed by solving $\mathrm{TR}_2(B, N, \gamma)$. This requires a partition (B, N), for which we use the *associated partition* with $B = \{i \in [n] \colon |\Delta x_i^{\mathrm{a}}/x_i| \le |\Delta s_i^{\mathrm{a}}/s_i|\}$. If the affine scaling step length α^{a} is sufficiently long, then the associated partition (B, N) reveals the true polarized partition (see Lemma 4.6).

It is also worth noting that the primal affine scaling step coincides with the trust region step for $N = [n]$ and $B = \emptyset$. The analogue holds for the dual affine scaling step with B and N swapped.

The proof of $O(1)$-optimality in the ℓ_2-neighborhood in Theorem 1.4 has two main ingredients. The first theorem asserts that the TR2-IPM algorithm terminates in $4\,\mathrm{SLC}(\mathcal{N}^2(\beta), \mu_1, \mu_0)$ iterations, and the iterates and line segments generated are guaranteed to stay in the wider $\mathcal{N}^2(82\beta)$-neighborhood.

Theorem 1.6. *Let $\beta \in (0, 1/128]$ and $\mu_0 > \mu_1 \ge 0$. Starting from any point $z_0 \in \mathcal{N}^2(\beta)$ with $\overline{\mu}(z_0) \in [\mu_1, \mu_0]$, the TR2-IPM algorithm finds a solution $z_1 \in \overline{\mathcal{N}}^2(\beta)$ with $\overline{\mu}(z_1) \le \mu_1$ in at most $4\,\mathrm{SLC}(\mathcal{N}^2(\beta), \mu_1, \mu_0)$ iterations, and all iterates and the line segments between consecutive iterates stay in $\mathcal{N}^2(82\beta)$.*

The second ingredient is the dependence of the straight-line complexity $\mathrm{SLC}(\mathcal{N}^2(\beta), \mu_1, \mu_0)$ on the parameter β. By definition, it is monotonically decreasing in β. Our next theorem establishes a stable property of ℓ_2-neighborhood in the sense that decreasing the neighborhood parameter only increases the straight line complexity by $O(1)$ times a constant power of the rate of decrease. This shows that moving into a narrower ℓ_2-neighborhood does not blow up the straight line complexity, and hence the iteration complexity of TR2-IPM. Whether wide neighborhood also admits such a stability property remains an interesting question.

Theorem 1.7. *There exists $C \ge 1$ such that for any $0 \le \beta \le \bar{\beta} \le 2^{-8}$, and $0 \le \mu_1 < \mu_0$,*

$$\mathrm{SLC}(\mathcal{N}^2(\bar{\beta}), \mu_1, \mu_0) \le \mathrm{SLC}(\mathcal{N}^2(\beta), \mu_1, \mu_0) \le O(1) \cdot \left(\frac{\bar{\beta}}{\beta}\right)^C \mathrm{SLC}(\mathcal{N}^2(\bar{\beta}), \mu_1, \mu_0)\,. \tag{2}$$

A canonical and optimal IPM in the ℓ_2-neighborhood. Path following IPMs in the ℓ_2-neighborhood are prevalent and core to the IPM family, see e.g. [16, 50, 51, 73]. We now argue that TR2-IPM can be seen as a canonical algorithm for path following in the ℓ_2-neighborhood.

Consider two subsequent iterates in the ℓ_2-neighborhood, $z, z^+ \in \mathcal{N}^2(\beta)$ with gap values $\mu = \mu(z) > \mu^+ = \mu(z^+)$, and the normalized step direction $\Delta z \coloneqq (z^+ - z)/(1 - \mu_1/\mu)$. As discussed in Section 1.4.1, if $[z^+, z]$ is long, i.e., $\mu_1/\mu \le 1/4$, then Δz is a feasible solution to the trust region program $\mathrm{TR}_2(B, N, \gamma)$ with respect to the polarized partition (Lemma 4.4), and a trust region step $z + \alpha^{\mathrm{TR}}\Delta z^{\mathrm{TR}}$ can reach the end of the segment with $\overline{\mu}(z + \alpha^{\mathrm{TR}}\Delta z^{\mathrm{TR}}) \le \mu_1$ and $[z + \alpha^{\mathrm{TR}}\Delta z^{\mathrm{TR}}, z] \subseteq \mathcal{N}^2(O(\beta))$ (Lemma 1.11). On the other hand, if $[z^+, z]$ is short, i.e., $\mu_1/\mu > 1/4$, then Δz is locally ℓ_2-close to the affine scaling step at z (Corollary 5.3). Moreover, an affine scaling step $z + \alpha^{\rm a}\Delta z^{\rm a}$ can reach the end of the segment with $\overline{\mu}(z + \alpha^{\rm a}\Delta z^{\rm a}) \le \mu_1$ and $[z + \alpha^{\rm a}\Delta z^{\rm a}, z] \subseteq \mathcal{N}^2(O(\beta))$ (Lemma 1.12).

Hence, one could replace each step of an ℓ_2-path following method by an affine scaling or trust region step to make at least the same progress in optimality gap while staying in an $\mathcal{N}^2(O(\beta))$ neighborhood; one can then return to $\mathcal{N}^2(\beta)$ in $O(1)$-corrector steps. Consequently, TR2-IPM makes the same progress as any other ℓ_2-path following algorithm with at most a constant increase in the number of corrector steps.

As far as we are aware, the only comparable results in terms of IPMs with near-optimal iteration complexity are those of Zhao and Stoer [80] and Zhao [79] who analyzed an ℓ_2-path-following method based on affine scaling steps in [50]. They proved that the number of affine-scaling iterations needed to traverse along a central path segment $[\mu_1, \mu_0]$ in $\mathcal{N}^2(\beta)$ can be upper bounded by a *Sonnevend curvature integral* [64] on the interval $[\mu_1, \mu_0]$ up to a factor of $O(1/\sqrt{\beta})$. Furthermore, when the Sonnevend curvature is uniformly lower bounded by a constant, they showed that the number of iterations is lower bounded by the same integral up to a factor of $O(\sqrt{\beta})$. Intuitively, this last result can be interpreted as saying that affine-scaling is essentially optimal when only short steps are possible. Our argument in Section 5 gives a direct proof of this fact. In Section 4.3, we complement this by the analysis of trust region steps, showing that they are essentially the best choice for a long step.

1.3 Solving the Trust Region Problems

We now discuss the subroutines for solving the Trust Region steps $\mathrm{TR}_2(B, N, \gamma)$ and $\mathrm{TR}_\infty(B, N, \ell, u)$. We consider the following general forms; both problems can be easily transformed to such a form. Let $\mathbf{B} \in \mathbb{R}^{m\times n}$ with $\mathrm{rk}(\mathbf{B}) = m$, $b \in \mathbb{R}^m$ and $I \cup J = [n]$ be a partition of the index set.

$$\begin{aligned} \min\ & \|y_J\|_2 \\ \|y_I\|_2 &\le 1\,, \\ \mathbf{B}y &= b\,, \end{aligned} \tag{TR-2}$$

$$\begin{aligned} \min\ & \|y_J\|_\infty \\ \|y_I\|_\infty &\le 1\,, \\ \mathbf{B}y &= b\,, \end{aligned} \tag{TR-max}$$

While the above two problems look similar, there are fundamental differences between them. We start by discussing (TR-2). Besides the IPM step $\mathrm{TR}_2(B, N, \gamma)$, this problem also arises in a broad range of other contexts; for example, as the classical ridge regression model in statistics. In the context of optimization, trust region subproblems arise as direction finding subroutines in trust region methods. We overview these applications in Section 7.5. Compared to ℓ_2-trust region, ℓ_∞-trust region problems are less understood and considered more challenging (see [15, Section 7.8]).

Using Lagrangian duality, this problem corresponds to a parametric minimum norm point problem with parameter λ (see (9) in Section 1.4.3). One can find the right value of λ using e.g., binary search. However, the possible value range is a priori the entire $\mathbb{R}_{\ge \mathbf{0}}$; one can obtain upper and lower bounds for λ using the binary encoding length of a rational input $(\mathbf{B}, b)$.

Lan, Monteiro, and Tsuchiya [40] gave such a *weakly polynomial* algorithm for solving the trust region steps, and left it as an open question to find a strongly polynomial implementation, i.e., where the number of arithmetic operations is $\mathrm{poly}(n)$.[2] We resolve this question in the affirmative. Moreover, we also give a strongly polynomial algorithm for approximately solving the significantly more challenging (TR-max) region problem.

Strongly polynomial solvability of (TR-2) was an important issue as the goal in [40] was to develop an algorithm whose running time is polynomial in n and the logarithm of condition number of $\mathbf{A}$, but independent of b and c. The affine scaling step as well as other IPM step directions, including the layered least square steps in [73] can be obtained by solving linear systems.

One cannot expect an exact optimal solution to (TR-2): the optimal solution may not even be rational for rational input.[3] While the optimal solution to (TR-max) is rational, solving it exactly in strongly polynomial time can be shown to imply a strongly polynomial LP algorithm. We define the notions of approximate solutions as follows; this slightly differs in the two cases.

Definition 1.8. Let $\delta \in (0, 1)$. We say that $y \in \mathbb{R}^n$ is a *δ-feasible solution* to (TR-2) if $\|y_I\| \le 1 + \delta$ and $\mathbf{B}y = b$, and *δ-feasible solution* to (TR-max), if $\|y_I\|_\infty \le 1 + \delta$ and $\mathbf{B}y = b$.

Let $\mathrm{OPT}_2 = \mathrm{OPT}_2(\mathbf{B}, b, I, J)$ and $\mathrm{OPT}_M = \mathrm{OPT}_M(\mathbf{B}, b, I, J)$ denote the optimum values of (TR-2) and (TR-max), respectively. We say that y is a *δ-optimal solution* to (TR-2) if it is δ-feasible and satisfies $\|y_J\| \le \mathrm{OPT}_2$. We say that $y \in \mathbb{R}^n$ is a *δ-optimal solution* to (TR-max), if it is δ-feasible and satisfies $\|y_J\|_\infty \le (1 + \delta)\mathrm{OPT}_M$.

[2] See Section 2.2 for the precise definition.

[3] E.g., let $\mathbf{B} = (1, 1, 1)$, $b = (2)$, $I = \{1, 2\}$, $J = \{3\}$.

Note that the δ-optimality in the ℓ_2-setting is much stronger, namely, we can achieve OPT_2 at a slight violation of the trust region constraint; whereas in the ℓ_∞-setting, the objective value is also slightly suboptimal. Getting the stronger guarantee of achieving OPT_M by at most δ violation on I in strongly polynomial time would again imply solving LP in strongly polynomial time, already in the case $I = \emptyset$. For the TR2-IPM algorithm, setting $\delta = 1/64$ already suffices; see Section 4.3

Theorem 1.9. *For $\delta \in (0,1)$, there exists an algorithm* TR2-SOLVE$(\mathbf{B}, b, I, J, \delta)$ *that finds a δ-optimal solution to* (TR-2) *or certifies infeasibility in strongly polynomial time. Let $n' := \min\{|I|, |J|\}$. The number of arithmetic operations is dominated by $O(\log(n') + \log\log(|I|/\delta))$ many linear system solves of size n, plus the time to get a $O\left(2^{2n'}\right)$ multiplicative approximation of the eigenvalues of a positive semidefinite matrix of size n'. Further,* TR2-SOLVE$(\mathbf{B}, b, I, J, \delta)$ *can be implemented in randomized $O(n^\omega(\log(n') + \log\log(|I|/\delta))$ or in deterministic $O\left((n')^3 + n^\omega(\log(n') + \log\log(|I|/\delta))\right)$ time.*

The algorithm requires a strongly polynomial subroutine for a crude multiplicative eigenvalue approximation. This can be done by strongly polynomial algorithms in $\tilde{O}(n^\omega)$ randomized or $O(n^3)$ deterministic time (see Section 2.8). The key idea is to use these eigenvalues as 'critical points' to speed up the search on the parameter λ, and showing that between two consecutive critical points, the parametric minimum norm point can be well-approximated by a simple function of λ; see an overview of the argument in Section 1.4.3.

We now turn to (TR-max). While for (TR-2), feasibility can be decided exactly by computing a projection, in the ℓ_∞-setting already the feasibility question $\mathbf{B}y = b$, $\|y_I\|_\infty \le 1$ corresponds to a general LP problem. We can however use the Robust IPM for approximate feasibility: either conclude infeasibility, or find a $y \in \mathbb{R}^n$ such that $\mathbf{B}y = b$, $\|y_I\|_\infty \le 1 + \delta$ in $n^{\tilde{\omega}+o(1)}\log(1/\delta)$ time (Theorem 2.25). Our main result in this context is that the same asymptotic running time bound applies for solving (TR-max).

Theorem 1.10. *For $\delta \in (0,1)$, there exists an algorithm* TRW-SOLVE$(\mathbf{B}, b, I, J, \delta)$ *that finds a δ-optimal solution to* (TR-max) *or certifies infeasibility in strongly polynomial time. The running time of the algorithm is randomized $n^{\tilde{\omega}+o(1)}\log(1/\delta)$ or deterministic $n^{\tilde{\omega}+o(1)}\log(1/\delta) + O(n^3)$.*

Similarly, for any fixed value λ, we can either find a feasible solution to $\mathbf{B}y = b$, $\|y_I\|_\infty \le 1+\delta$, $\|y_J\|_\infty \le (1+\delta)\lambda$ in $n^{\tilde{\omega}+o(1)}\log(1/\delta)$, or show that $\mathbf{B}y = b$, $\|y_I\|_\infty \le 1$, $\|y_J\|_\infty \le \lambda$ is infeasible. Thus, finding a δ-optimal solution to (TR-max) can be reduced to a parametric search over the objective value λ. Similarly to the ℓ_2-case, the range of possible λ values is weakly polynomial. To remove the dependence on the bit-complexity of $(\mathbf{B}, b)$, we again rely on an approximate singular value decomposition to narrow down the range of the optimal λ and exploit a simpler behaviour between two approximate singular values; see Section 1.4.4 for an overview.

1.4 Our techniques

We now give a more detailed technical overview of the paper. Section 1.4.1 is dedicated to the proof of Theorem 1.6 showing that the ℓ_2-Trust Region IPM is $O(1)$-optimal in $\mathcal{N}^2(\beta)$. We then outline the challenges and arguments for showing the dependence of the straight line complexity on the neighborhood parameter β i.e. the proof of Theorem 1.7, in Section 1.4.2. In Section 1.4.3, we explain the idea of solving (TR-2) in strongly polynomial time, and in Section 1.4.4, we give an overview of the algorithm for (TR-max). Theorem 1.3 then can be derived combining this implementation with the polarization arguments as described in Section 1.2.

1.4.1 Optimality of the Trust Region IPM

Assume that there exists a piecewise linear curve of T segments that stays entirely in $\mathcal{N}^2(\beta)$. Throughout, we focus on a single straight line segment $[z_0, z_1] \subseteq \mathcal{N}^2(\beta)$ between $z_0 = (x_0, s_0) \in \mathcal{N}^2(\beta)$ with $\overline{\mu}(z_0) = \mu$ and $z_1 = (x_1, s_1) \in \mathcal{N}^2(\beta)$ with $\overline{\mu}(z_1) = \mu_1 < \mu$. For the purpose of this overview, we make the simplifying assumption that $z = z_0$, i.e., our current iterate starts from the beginning at this line segment, and show that starting from z_0, the algorithm takes a single trust region or affine scaling step to reach a point $z^+ \in \mathcal{N}^2(O(\beta))$ with $\overline{\mu}(z^+) \le \mu_1$, and $[z^+, z_0] \in \mathcal{N}^2(O(\beta))$. While $[z_0, z_1]$ is not available for the algorithm, z_0 and z_1 behave similarly due to ℓ_2-local proximity (see Section 2.6), and hence the arguments extend to starting from a different point $z \in \mathcal{N}^2(\beta)$ with $\overline{\mu}(z_0) \le \mu$; we obtain a slightly larger $\mathcal{N}^2(O(\beta))$ neighborhood guarantee. We thus derive Lemmas 1.11 and 1.12; Theorem 1.6 then follows from the use of $O(1)$ many corrector steps to get back to $\mathcal{N}^2(\beta)$ for the next iteration.

We consider separately the cases that $\mu_1 \le \mu/4$, for which we say the straight line segment is *long*, and $\mu_1 \ge \mu/4$, for which we say the straight line segment is *short*. One could replace the threshold $1/4$ by any other constant: as long as $\mu_1/\mu \le c < 1$ for some constant c, Lemma 1.11 below on the trust region step length holds true, which means the use of a single trust region step can generate a trajectory that reduces the gap from μ to μ_1 while staying in $\mathcal{N}^2(O(\beta))$. The use of affine scaling step is needed only when $1 - \mu_1/\mu = o(1)$, i.e., the segment is *very short*. The threshold at 1/4 was chosen to obtain a suitable constant in $\mathcal{N}^2(O(\beta))$.

Definition (Curvature of a straight line)**.** For $z_0 = (x_0, s_0), z_1 = (x_1, s_1) \in \mathcal{P} \times \mathcal{D}$ with $\overline{\mu}(z_0) = \mu$ and $\overline{\mu}(z_1) = \mu_1 \leq \mu$, we define the *curvature* of the straight line between z_1 and z_0 to be

$$\kappa(z_1, z_0) := \frac{\|(x_1 - x_0)(s_1 - s_0)\|}{\left(\sqrt{\mu} + \sqrt{\mu_1}\right)^2} . \tag{3}$$

Note that $\kappa(z_0, z_1) = \kappa(z_1, z_0)$. The analysis for both long and short segments relies on Lemma 2.15, showing in particular that if $[z_1, z_0] \subseteq \mathcal{N}^2(\beta)$, then the line has a bounded curvature, namely $\kappa(z_1, z_0) \leq 2\beta$. Additionally, this quantity plays a key role in the proof of Theorem 1.7; this will be discussed in more detail in Section 1.4.2.

Long straight line segment and ℓ_2-polarization. When the straight line segment from z_0 to z_1 is long, the following lemma asserts the near-optimality of a trust region step.

Lemma 1.11. *For $\beta \in (0, 1/128]$, suppose there exist some $z_0 \in \mathcal{N}^2(\beta)$ with $\overline{\mu}(z_0) = \mu$ and $z_1 \in \mathcal{N}^2(\beta)$ with $\overline{\mu}(z_1) = \mu_1 \leq \mu/4$ such that $[z_1, z_0] \subseteq \mathcal{N}^2(\beta)$. Let $z \in \mathcal{N}^2(\beta)$ with $\overline{\mu}(z) = \mu$ be given, and Δz^{TR} be the trust region direction at z. Then, there exists a step-length $\alpha^{\mathrm{TR}} \in (0, 1]$ so that $\overline{\mu}(z + \alpha^{\mathrm{TR}} \Delta z^{\mathrm{TR}}) \leq \mu_1$ and $z + \alpha \Delta z^{\mathrm{TR}} \in \mathcal{N}^2(82\beta)$ for all $\alpha \in [0, \alpha^{\mathrm{TR}}]$.*

Let us denote $(\Delta x, \Delta s) := (x_1 - x_0, s_1 - s_0)$ as the full step from z_0 to z_1 which satisfies $\kappa(z_1, z_0) \leq 2\beta$ because $[z_1, z_0] \subseteq \mathcal{N}^2(\beta)$. This, together with the *residual equation*[4]

$$\left(\mathbf{1}_n + \frac{\Delta x}{x_0}\right)\left(\mathbf{1}_n + \frac{\Delta s}{s_0}\right) = \frac{x_1 s_1}{x_0 s_0} \approx \frac{\mu_1}{\mu} \cdot \mathbf{1}_n \tag{4}$$

when μ_1/μ is bounded away from 1 ensures the existence of a *polarized partition* $B \cup N = [n]$ such that by going from (x_0, s_0) to (x_1, s_1) while $(x_0 + \lambda \Delta x, s_0 + \lambda \Delta s) \in \mathcal{N}^2(\beta)$ for all $\lambda \in [0, 1]$, the primal variables in N must scale down massively while those in B barely change. For the dual variables the roles of B and N are swapped. We call this notion *ℓ_2-polarization.* More precisely, one can show

$$\max\left\{\left\|\frac{x_{0N} + \Delta x_N}{x_{0N}}\right\|, \left\|\frac{s_{0B} + \Delta s_B}{s_{0B}}\right\|\right\} \ll O(\beta) \quad \text{while} \quad \max\left\{\left\|\frac{\Delta x_B}{x_{0B}}\right\|, \left\|\frac{\Delta s_N}{s_{0N}}\right\|\right\} = O(\beta) . \tag{5}$$

Hence, conformal to our previous definition of the optimality of path-following IPM, one can say that ℓ_2-polarization is enforced for any long straight line segment in $\mathcal{N}^2(\beta)$ that reduces the normalized duality gap from μ to μ_1.

Trust region step. The idea behind the proof of Lemma 1.11 is to replace $(\Delta x, \Delta s)$ with an optimal feasible direction at z_0 that exhibits ℓ_2-polarization. This is precisely captured in the trust region program at z_0 with the polarized partition (B, N) and $\gamma = O(\beta)$. For convenience, we restate it with respect to z_0:

$$\begin{aligned} \min \ & \left\|\frac{x_{0N} + \Delta x_N}{x_{0N}}\right\| \\ \text{s.t.} \ & \left\|\frac{\Delta x_B}{x_{0B}}\right\| \leq \gamma \\ & \mathbf{A}\Delta x = \mathbf{0} \end{aligned} \qquad\qquad \begin{aligned} \min \ & \left\|\frac{s_{0B} + \Delta s_B}{s_{0B}}\right\| \\ \text{s.t.} \ & \left\|\frac{\Delta s_N}{s_{0N}}\right\| \leq \gamma \\ & \mathbf{A}^\top \Delta y + \Delta s = \mathbf{0} \end{aligned}$$

Let $\Delta z^{\mathrm{TR}} = (\Delta x^{\mathrm{TR}}, \Delta s^{\mathrm{TR}})$ denote its optimal solution. By definition, Δx^{TR} achieves a maximal multiplicative decrease on the coordinates in N under the condition that the coordinates in B change barely, as measured in the local ℓ_2-norm at x_0. Δs^{TR} achieves the same on the dual variables with B and N swapped. The maximal multiplicative decrease attained by $(\Delta x^{\mathrm{TR}}, \Delta s^{\mathrm{TR}})$ is large enough to ensure the existence of a step length $\alpha^{\mathrm{TR}} \in (0, 1]$, such that $\overline{\mu}(z_0 + \alpha^{\mathrm{TR}} \Delta z^{\mathrm{TR}}) \leq \mu_1$ and $z_0 + \alpha^{\mathrm{TR}} \Delta z^{\mathrm{TR}} \in \mathcal{N}^2(O(\beta))$ for all $\alpha \in [0, \alpha^{\mathrm{TR}}]$. This implies that ℓ_2-polarization alone is sufficient to obtain an *optimal* long straight line segment in $\mathcal{N}^2(O(\beta))$ that reduces the normalized duality gap from μ to μ_1.

[4]The $\approx$ here follows from $z_0, z_1 \in \mathcal{N}^2(\beta)$; see Section 4.2.

Short straight line segment and affine scaling step. As discussed in Section 1.2, the previous works by Zhao and Stoer [80] and Zhao [79] indicate that an affine scaling step should be near-optimal when the decrease in duality gap is small. We derive an explicit bound in the following lemma which asserts that when a straight line segment from z_0 to z_1 is short, an affine scaling step is indeed optimal in $\mathcal{N}^2(O(\beta))$ to traverse over the segment.

Lemma 1.12. *For $\beta \in (0, 1/128]$, suppose there exist some $z_0 \in \mathcal{N}^2(\beta)$ with $\overline{\mu}(z_0) = \mu$ and $z_1 \in \mathcal{N}^2(\beta)$ with $\overline{\mu}(z_1) = \mu_1 \geq \mu/4$ such that $[z_1, z_0] \subseteq \mathcal{N}^2(\beta)$. Let $z \in \mathcal{N}^2(\beta)$ with $\overline{\mu}(z) \leq \mu$ be given, and Δz^{a} be the affine scaling direction at z. Then, there exists a step-length $\alpha^{\mathrm{a}} \in [0, 1]$ so that $\overline{\mu}(z + \alpha^{\mathrm{a}}\Delta z^{\mathrm{a}}) \leq \mu_1$ and $z + \alpha\Delta z^{\mathrm{a}} \in \mathcal{N}^2(82\beta)$ for all $\alpha \in [0, \alpha^{\mathrm{a}}]$.*

The affine scaling step $z_0 + \alpha\Delta z^{\mathrm{a}}$ reduces the duality gap by exactly a factor of $1 - \alpha$ (see Section 2.4). Let us set $\alpha^* := 1 - \mu_1/\mu$ as the *ideal step length* for affine scaling step, so that $\overline{\mu}(z_0 + \alpha^*\Delta z^{\mathrm{a}}) = \mu_1$. In Section 5, we show that $z_0 + \alpha^*\Delta z^{\mathrm{a}} \in \mathcal{N}^2(O(\beta))$. Lemma 1.12 then follows for a choice that $\alpha^{\mathrm{a}} \in [\alpha^*, 1]$,

ℓ_2-local proximity of affine scaling step. The proof of $z_0 + \alpha^*\Delta z^{\mathrm{a}} \in \mathcal{N}^2(O(\beta))$ relies on the following property of the affine scaling step: for the direction $\Delta z = z_1 - z_0$, the local ℓ_2-norm of its difference to an affine scaling step with the ideal step length is $O(\kappa(z_1, z_0))$; see Lemma 2.16. This means that if $\kappa(z_1, z_0)$ is bounded, then taking the affine scaling step $z_0 + \alpha^*\Delta z^{\mathrm{a}}$ resembles taking the step $z_0 + \Delta z$, and the assumption $[z_1, z_0] \subseteq \mathcal{N}^2(\beta)$ guarantees that $\kappa(z_1, z_0) = O(\beta)$ by Lemma 2.15. We also point out that this approach does not work for long straight line segment, because in that case there is no lower bound on μ_1 in μ. As a result, the α^* could be arbitrarily close to 1, which means $z + \alpha^*\Delta z^{\mathrm{a}}$ resembles taking a full affine scaling step, and hence $z + \alpha^*\Delta z^{\mathrm{a}}$ would not stay in $\mathcal{N}^2(C\beta)$ for a fixed constant $C > 0$.

1.4.2 Curvature of straight lines in the ℓ_2-neighborhood

As aforementioned, for any two points $z_0, z_1 \in \mathcal{N}^2(\beta)$ with $\overline{\mu}(z_1) < \overline{\mu}(z_0)$, the centrality error of a point on the straight line between z_1 and z_0 depends on the curvature $\kappa(z_1, z_0)$ of the whole line $[z_1, z_0]$. In particular, when z_1, z_0 are both on the central path, it turns out that the maximum ℓ_2-norm of the centrality error achieved by any point on $[z_1, z_0]$ is exactly $\kappa(z_1, z_0)$; see Lemma 2.15. Motivated by this result, we introduce the following notion.

Path complexity and stability We define $\overline{\mathrm{SLC}}(\kappa, \mu_1, \mu_0)$, referred as path complexity, to be the minimum number of segments of any piecewise linear curves traversing from $z(\mu_0)$ to $z(\mu_1)$ such that all the breakpoints are on the central path and the curvature of each segment is at most the parameter β; see Definition 6.1. The use of $\overline{\mathrm{SLC}}(\beta, \mu_1, \mu_0)$ simplifies the analysis for the proof of Theorem 1.7 because the deviation from the central path of a straight line connecting two central path points is exactly the curvature of the line segment. Since $\overline{\mathrm{SLC}}(\kappa, \mu_1, \mu_0)$ considers a subclass of piecewise linear curves $\Gamma\colon [\mu_1, \mu_0] \to \mathcal{N}^2(\beta)$, it follows from their definitions that $\mathrm{SLC}(\mathcal{N}^2(\beta), \mu_1, \mu_0) \leq \overline{\mathrm{SLC}}(\beta, \mu_1, \mu_0)$.

The reverse direction of the above inequality also holds up to a constant factor of the curvature parameter. This is a consequence of the path stability of any straight line in the ℓ_2-neighborhood, namely if $[z_1, z_0] \subseteq \mathcal{N}^2(\beta)$, then mapping the endpoints to the central path points, we have $[z(\overline{\mu}(z_1)), z(\overline{\mu}(z_0))] \subseteq \mathcal{N}^2(8\beta)$; see Lemma 6.13. As a result, the piecewise linear curve attaining $\mathrm{SLC}(\mathcal{N}^2(\beta), \mu_1, \mu_0)$ can be converted to a piecewise linear curve staying in $\mathcal{N}^2(8\beta)$ that traverses from $z(\mu_0)$ to $z(\mu_1)$ with all breakpoints on the central path, showing that $\overline{\mathrm{SLC}}(8\beta, \mu_1, \mu_0) \leq \mathrm{SLC}(\mathcal{N}^2(\beta), \mu_1, \mu_0)$. Knowing how to relate path complexity to straight line complexity, it suffices to prove the following result which gives (2) but with respect to $\overline{\mathrm{SLC}}(\beta, \mu_1, \mu_0)$.

Theorem 1.13. *There exists a universal constant $C \geq 0$ such that for any $0 < \beta_1 \leq \beta_2 \leq 2^{-8}$ and $0 \leq \mu_1 < \mu_0$,*

$$\overline{\mathrm{SLC}}\left(\beta_1, \mu_1, \mu_0\right) = O\left(\left(\frac{\beta_2}{\beta_1}\right)^C \cdot \overline{\mathrm{SLC}}\left(\beta_2, \mu_1, \mu_0\right)\right). \tag{6}$$

For the proof of Theorem 1.13, we focus on a single straight line segment $[z(\mu_1), z(\mu_0)]$ on any piecewise linear curve with breakpoints on the central path and consider differently when this segment is either long or short.

Short segments. For the case of short segments (i.e., $\mu_1/\mu \geq 1/4$) we argue by using the proximity of $z(\mu)$ for $\mu \in (\mu_1, \mu_0)$ to the line segment $[z(\mu_1), z(\mu_0)]$ established in Section 2.6. In the short segment case, the aforementioned relationship between certain curvature integral of the central path and the number of iterations of standard path-following methods as in [52, 79, 80] is related to the statement in Theorem 1.13 when only short steps to $\mu_1 = \Omega(\mu_0)$ are possible. However, these results do not study precisely the same notion of $\overline{\mathrm{SLC}}$ as we do. Furthermore, the result by Monteiro and Tsuchiya [52] only applies in the limit as $\beta \to 0$, where they showed that the number of iterations of standard affine scaling algorithms converges to $\beta^{-1/2}$ times the curvature integral, and thus suggests an exponent $C = 1/2$ for $\mu_1 = \Omega(\mu_0)$. Our self-contained proof does not rely on the curvature integral, and hold for all sufficiently small constant β, not just in the limit.

Long segments. The analysis for long segments, i.e., $\mu_1 \ll \mu_0$, requires very different techniques. Intuitively, the analysis for short segments no longer applies, as the denominator in the norm term in the definition of curvature becomes very small when analyzing $\kappa(\mu_1, \mu)$ for some $\mu \in (\mu_1, \mu_0)$ with $\mu \ll \mu_0$, so curvature could increase rapidly along a segment. As discussed in Section 1.4.1, long segments exhibits ℓ_2-polarization, namely the variables can be partitioned into two sets B and N such that $x(\mu)_B/x(\mu_0)_B \approx \mathbf{1}_B$ and $x(\mu)_N/x(\mu_0)_N \approx \mu/\mu_0 \cdot \mathbf{1}_N$ for $\mu \in (\mu_1, \mu_0)$, and the analogue holds for $s(\mu)/s(\mu_0)$ with B and N swapped. Therefore, for $\mu \in (\mu_1, \mu_0)$ with $\mu \ll \mu_0$, the curvature $\kappa(z(\mu_1), z(\mu_0))$ can be approximated by the term

$$\begin{aligned}\kappa(z(\mu), z(\mu_0)) &= \frac{\|(x(\mu) - x(\mu_0))(s(\mu) - s(\mu_0))\|}{\left(\sqrt{\mu_0} + \sqrt{\mu}\right)^2} \\ &\approx \frac{\|(x(\mu)_B - x(\mu_0)_B)(\mathbf{0}_B - s(\mu_0)_B), (\mathbf{0}_N - x(\mu_0)_N)(s(\mu)_N - s(\mu_0)_N)\|}{\mu_0} \\ &= \left\| \left(\frac{x(\mu)_B}{x(\mu_0)_B} - \mathbf{1}_B, \frac{s(\mu)_N}{s(\mu_0)_N} - \mathbf{1}_N \right) \right\| =: \|\phi(\mu, \mu_0)\| \,.\end{aligned} \tag{7}$$

As such, we must investigate how $\|\phi(\cdot, \mu_0)\|$ evolves over $[\mu_1, \mu_0]$. Geometrically, we can consider the trajectory $\phi(\cdot, \mu_0) \subseteq \mathbb{R}^n$. For μ_0, the corresponding vector is $\phi(\mu_0, \mu_0) = (\mathbf{0}_B, \mathbf{0}_N)$, i.e., the origin. By the approximation in (7), for a fixed neighborhood width β, we can traverse $\phi(\cdot, \mu_0)$ up to a parameter μ for which $\|\phi(\mu, \mu_0)\| \approx \beta$. The goal of Theorem 1.13 is to show that for parameter 2β and its corresponding value η such that $\|\phi(\eta, \mu_0)\| \approx 2\beta$, the distance between $\phi(\eta, \mu_0)$ and $\phi(\mu, \mu_0)$ is smaller than the distance between $\phi(\eta, \mu_0)$ and the origin. A priori, this is unclear, as for example $\phi(\cdot, \mu_0)$ may have sharp turns and follow a spiral trajectory. However, we will show that this cannot happen and that in fact we obtain for *well-separated* $\mu_0 \ll \mu \ll \mu_1$ the approximate equality

$$\|\phi(\mu_1, \mu_0)\|_2^2 \approx \|\phi(\mu_1, \mu)\|_2^2 + \|\phi(\mu, \mu_0)\|_2^2 \,. \tag{8}$$

This will be the main ingredient for the proof of Theorem 1.13 for long segments.

1.4.3 Solving the ℓ_2-Trust Region Problem

We now give an overview of the proof of Theorem 1.9 on solving (TR-2). By Lagrangian duality, this can be formulated as a parametric search problem. Let

$$y(\lambda) := \operatorname*{arg\,min}_{\mathbf{B}y=b} \left(\|y_J\|^2 + \lambda\|y_I\|^2\right) \quad \text{and} \quad \psi(\lambda) := \|y_I(\lambda)\|^2 \,. \tag{9}$$

One can resolve the two extreme cases $\lambda = 0$ and ∞ by solving the related layered least square problems, which can be done efficiently as shown in [46, 73]. Assuming feasibility, for the root $\psi(\lambda^*) = 1$, $y(\lambda^*)$ is the optimal solution to (TR-2). We also note that for λ with $1 \le \psi(\lambda) < 1 + \delta$, $y(\lambda)$ is a δ-optimal solution to (TR-2). Hence, finding a δ-optimal solution to (TR-2) is equivalent to a root finding problem of $\psi(\lambda) - 1$ on $\mathbb{R}_{>\mathbf{0}}$.

There is an extensive literature in optimization and numerical analysis to solve quadratic trust region problems. Approaches include but are not limited to Moré-Sorensen method [54], hybrid Newton's method [77], Lanczos method [28], parameterized eigenvalue problem [61], and so on. We refer to the book [15] and the references within. As far as the authors are aware, the aspect of strongly polynomial solvability has only been highlighted by Lan, Monteiro and Tsuchiya [40], who raised this as an open question that would have been necessary to turn their algorithm strongly polynomial, assuming $\log \bar{\chi}^\star_A$ is polynomially bounded in n. In Section 7.5, we explain how (TR-2) relates to more general forms of the trust region problem, and the implications of our result on them.

The algorithm in [40] only required finding a 2-optimal solution, and used a binary search method. However, this requires lower and upper bounds on λ^*. One can find such bounds using the binary description of $(\mathbf{B}, b)$, but this gives a running time dependence on the encoding length of $(\mathbf{B}, b)$.

Root finding between eigenvalues. Our approach to give the strongly polynomial algorithm in Theorem 1.9 first uses a standard strategy in the literature, see e.g. [15, 54], which is to use the eigenvalue decomposition of an associated matrix to express ψ as a univariate analytic function:

$$\psi(\lambda) = \sum_{i=1}^{n} \frac{a_i^2}{(\beta_i + \lambda)^2} + C\,,$$

for $C, \beta_i \ge 0$; see (74). Then our algorithm follows a similar idea as the hybrid Newton's method by Ye [77]. At a high level, Ye's algorithm proceeds as follows: given an initial long interval where the root $\psi(\lambda^*) = 1$ lies, it is first partitioned into short intervals, such that if the root λ^* is contained in any one of these short intervals which would be found by binary search, then its lower endpoint satisfies Smale's criterion [63] using the structure of ψ and the narrowness of the interval. Hence, subsequent Newton's method applied to this interval exhibits quadratic convergence. Since the total number of the short intervals depends on the width of the initial long interval, this gives a weakly polynomial algorithm that can solve a general trust region problem to a high accuracy.

To move toward a strongly polynomial algorithm for our (TR-2), we take the initial long interval to be $\mathbb{R}_{>0}$, recalling $\lambda = 0, \infty$ can be resolved by two individual layered least square problems in [73], which also addresses the feasibility of (TR-2). We now highlight our procedure to shorten this infinitely long interval by making use of the β_i's which we identify as the *'critical points'* of ψ. Assume that $\lambda \in [\lambda^-, \lambda^+]$ such that for each critical point, either $\beta_i \ll \lambda^-$ or $\beta_i \gg \lambda^+$. Then, $\psi(\lambda) \approx p + q/\lambda^2$ on $[\lambda^-, \lambda^+]$ for some $p, q \ge 0$; see Lemma 7.12. Our algorithm first crudely approximates two adjacent critical points $\lambda_1 < \lambda_2$ such that $\psi(\lambda_1) > 1 \ge \psi(\lambda_2)$. If λ_2/λ_1 is $O(n2^{2n})/\mathrm{poly}(\delta)$ bounded, where the $O(n2^{2n})$ factor takes into account the approximation accuracy of the critical points from eigenvalue computation. We then apply binary search followed by Newton's method to find a value λ with $1 \le \psi(\lambda) < 1 + \delta$. As in [77], the use of the binary search is to provide a good initial guess for the subsequent Newton's method to ensure quadratic convergence. Otherwise, we use the above structural property of ψ to narrow down the search interval to $\lambda_1 \le \hat{\lambda}_1 \le \hat{\lambda}_2 \le \lambda_2$ with $\psi(\hat{\lambda}_1) > 1 \ge \psi(\hat{\lambda}_2)$ and $\hat{\lambda}_2/\hat{\lambda}_1$ is now $\mathrm{poly}(n/\delta)$ bounded. We then proceed with searching this interval as in the previous case.

We adopt a few techniques to conduct the above procedure in strongly polynomial time. We show that the critical points of ψ coincide with the positive eigenvalues of an associated matrix, for which multiplicative $O(n2^{2n})$-approximations can be computed in strongly polynomial time, and this suffices our purpose (see Section 2.8). The standard bisection scheme (e.g. in [15]) for hybrid Newton's method involves taking square roots which cannot be done in strongly polynomial time. We provide a square root free binary search subroutine (Algorithm 5) for root finding on bounded intervals. The square root issue again prevents us from solving $\psi(\lambda) \approx p + q/\lambda^2 = 1$ to get a δ-optimal solution directly when the search interval is long. We only make use of this functional approximation as a backstage structure of ψ to zoom into a $\mathrm{poly}(n/\delta)$-bounded interval that contains the root $\psi(\lambda^*) = 1$.

The quadratic convergence of Newton's method mentioned above allows us to show that when applied to root finding of $\psi - 1$, our $\textsc{HybridNewton}(\lambda_1, \lambda_2, \delta)$ terminates in $O(\log\log(\lambda_2/\lambda_1) + \log\log(1/\delta))$ many iterations. In the algorithm of Theorem 1.9, it is only conducted in interval $[\lambda_2, \lambda_1]$ with $\lambda_2/\lambda_1 = O(n2^{2n})/\mathrm{poly}(\delta)$.

1.4.4 Solving the ℓ_∞-Trust Region Problem

We consider (TR-max) also as a parametrized optimization problem. For each fixed parameter value λ, we have an ℓ_∞-regression problem: the subroutine $\mathrm{LM}\textsc{Guess}(\lambda, \delta)$ either finds a $y \in \mathbb{R}^n$ such that $\mathbf{B}y = b$, $\|y_I\|_\infty \le 1 + \delta$, $\|y_J\|_\infty \le (1 + \delta)\lambda$, or a Farkas certificate of infeasibility in $n^{\tilde{\omega}+o(1)}\log(1/\delta)$ (see Lemma 8.2). This is based on the adaptation of the Robust IPM, see Theorem 2.25. Our goal is to find values $\lambda/(1 + \delta/4) \le \lambda' \le \lambda$ such that $\mathrm{LM}\textsc{Guess}(\lambda, \delta/4)$ returns a feasible solution, whereas $\mathrm{LM}\textsc{Guess}(\lambda', \delta/4)$ returns a Farkas certificate showing that λ' is below the optimum value $\lambda^\star$ of (TR-max).

Similarly to the ℓ_2-setting, we identify *critical values* to reduce the search space. These are defined as the inverses σ_i^{-1} of the approximate singular values σ_i of the *lifting operator* $\mathbf{L} \in \mathbb{R}^{I \times J}$, see (83). For a vector $w \in \mathbb{R}^J$, $z = \mathbf{L}w \in \mathbb{R}^I$ is the minimum norm vector such that $\mathbf{B}_J w = \mathbf{B}_I z$, whenever such a vector exists. Besides the critical values, we also require an approximate partial singular value decomposition (see Definition 2.26) $\mathbf{U} \in \mathbb{R}^{\mathrm{rk}(\mathbf{L}) \times n}$ such that $\|\mathbf{U}x\|$ approximates $\|\mathbf{L}x\|$ up to a multiplicative factor $\varrho = 2^{O(n)}$.

Using binary search, we narrow the search interval to $[\lambda^-, \lambda^+]$ such that each critical value is either much smaller than λ^- or much larger than λ^+, and $\lambda^- \le \lambda^\star$, whereas $\mathrm{LM}\textsc{Guess}(\lambda^+, \delta/4)$ returns an approximately feasible solution $\bar{y}$. If λ^+/λ^- is polynomially bounded, the algorithm finds the right λ value using binary search.

If $\lambda^- \ll \lambda^+$, then we modify $\bar{y}$ by projecting out all 'cheap' parts of $\bar{y}_J$. By the choice of $[\lambda^-, \lambda^+]$, every approximate singular value of the lifting operator $\mathbf{L}$ is much larger than $1/\lambda^-$ or much smaller than $1/\lambda^+$. Projecting out the part of $\bar{y}_J$ in $\ker(\mathbf{B}_J)$ as well as the projection to the cheap subspace will only increase $\|\bar{y}_I\|_\infty$ by a small error, while possibly reducing $\|\bar{y}_J\|_\infty$ significantly, to a value $\hat{\lambda}$. We then show that $\hat{\lambda}/n \le \lambda^\star$, and thus the algorithm concludes with a binary search on $[\hat{\lambda}/n, \hat{\lambda}]$. The lower bound follows because, assuming $\hat{\lambda}/n > \lambda^-$, a better solution could only be improved by lifting a large norm vector in the 'expensive' subspace, thereby increasing $\|y_I\|_\infty$ well above 1.

1.5 Further related work

Several IPMs use barrier functions different from the logarithmic barrier function. For example, path-following IPMs based on the volumetric barrier function, which was introduced by Vaidya [70], were also developed (see e.g. [7]) for their desirable iteration complexity, despite the higher operation cost per iteration. Lee and Sidford [41, 42] designed a path-following IPM based on Lewis weights that achieved both $\tilde{O}(\sqrt{\mathrm{rk}(A)}L)$ iteration complexity and $\tilde{O}(1)$ linear system solves per iteration. All these barrier functions belong to a large class called self-concordant barrier functions, for which the notion of central path exists, and so does a theoretical universal barrier function [55]. As mentioned earlier, it follows from [6] that the trajectories of all algorithms based on a reasonable self-concordant barrier functions also lie in $\mathcal{N}^{-\infty}(1-\theta)$ for $\theta = 1/\mathrm{poly}(n)$, and therefore the straight line complexity bounds are applicable. On the other hand, it is currently unknown whether any analogous trust region step for these barrier functions exist.

Despite the empirical evidence suggesting that IPMs in very large neighborhoods of the central path are much more efficient in practice (see [47, 25, 36, 48, 14, 66]), already establishing theoretical running time bounds for wide neighborhood IPMs that match with ℓ_2-neighborhood IPMs has been regarded as a genuine challenge; see [60]. The analysis of Mizuno–Todd–Ye predictor corrector algorithm in wide neighborhood gives an $O(nL)$ iteration bound [50]. Hung and Ye [35] reduced this iteration bound to $O(n^{\frac{n+1}{2n}}L)$ with high-order correction. To further improve upon this, Sturm and Zhang [65] gave a first $O(\sqrt{n}L)$-iteration algorithm, in which second order corrections are used and the iterates lie in a somewhat complicated neighborhood. Later, Ai [2] gave a new wide neighborhood IPM achieving $O(\sqrt{n}L)$ bound. Peng, Terlaky and Zhao [58] also proposed an $O(\sqrt{n}\log(n)L)$ algorithm based on self-regular functions. Inspired by [2], Ai and Zhang [3] devised a wide neighborhood IPM that allows large update with also an $O(\sqrt{n}L)$ iteration bound. Following this line of research, Theorem 1.3 in this paper can be seen as moving the theoretical running time bound of wide neighborhood IPM closer to its actual performance in practice.

The SLLS IPM algorithm in [5] builds on a long line of work on *combinatorial interior point methods*. In a seminal 1996 paper, Vavasis and Ye [73] addressed the problem that affine scaling is not always aggressive enough to quickly traverse "long and straight" segments of the central path. They introduced the *layered least squares (LLS)* predictor step, which uses a combinatorial layering procedure to compute the predictor direction. Unlike most IPMs, this algorithm terminates with an exact optimal solution without any additional rounding step. The total number of iterations can be bounded by $O(n^{3.5}\log\bar{\chi}_{\mathbf{A}})$, where $\bar{\chi}_{\mathbf{A}}$ is the Dikin–Stuart–Todd condition number of the matrix. This result has remarkable implications in the context of strongly polynomial solvability of Linear Programming. This fundamental open question asks for exactly solving (LP) in $\mathrm{poly}(m,n)$ basic arithmetic operations. Tardos [67] gave an algorithm with $\mathrm{poly}(n,\log\Delta_{\mathbf{A}})$ running time, assuming $\mathbf{A}$ has integer entries and $\Delta_{\mathbf{A}}$ is the largest sub-determinant. This yields strongly polynomial algorithms for 'combinatorial LPs', such as network flow or multi-commodity flow problems, or packing and covering problems with small integer coefficients. The LLS IPM strengthens Tardos's result, since $\bar{\chi}_{\mathbf{A}}$ can be bounded by $\Delta_{\mathbf{A}}$; at the same time, this is a purely geometric condition number that does not require integrality and can be much smaller.

Improved LLS methods were given in [46, 51]. Monteiro and Tsuchiya [52] majorized the iteration bound of Mizuno-Todd-Ye predictor-corrector method by that of LLS IPMs. They show that the number of affine scaling steps to reduce the duality gap by a constant factor is asymptotically a $\beta^{-1/2}$-factor of a certain Sonnevend integral as $\beta \to 0$, which is strongly bounded by $\mathrm{poly}(n,\log\bar{\chi}^\star_{\mathbf{A}})$, where $\bar{\chi}^\star_{\mathbf{A}} := \inf\{\bar{\chi}_{\mathbf{AD}} : \mathbf{D} \in \mathcal{D}_{>\mathbf{0}}\}$ i.e. the best achievable value of $\bar{\chi}^\star_{\mathbf{A}}$ under column scaling. However, in contrast to many standard IPMs, the aforementioned LLS algorithms were not invariant under rescaling the columns of $\mathbf{A}$. Hence, the above works left the open question of finding a *scaling invariant LLS* algorithm; the running time dependence would thus improve to $\mathrm{poly}(n,\log\bar{\chi}^\star_{\mathbf{A}})$. The Trust Region IPM in the focus of our current work was introduced by Lan, Monteiro, and Tsuchiya [40] to tackle this challenge. The Trust Region IPM is scaling invariant, and they showed an $O(n^{3.5}\log\bar{\chi}^\star_{\mathbf{A}})$ iteration complexity bound. As mentioned above, the steps themselves were not strongly polynomial. A $\mathrm{poly}(n,\log\bar{\chi}^\star_{\mathbf{A}})$ LP algorithm was given by Dadush, Huiberts, Natura and Végh [16]. This is a scaling invariant LLS IPM method; a key insight relates the condition number $\bar{\chi}_{\mathbf{A}}$ to the *circuit imbalance measure* of the matrix, namely the ratio between the largest and smallest absolute value entries of minimal linear dependencies in $\ker(\mathbf{A})$.

The SLLS IPM—and consequently TR2-IPM—subsumes the above strongly polynomial results as it is approximately optimal with respect to the straight-line complexity lower bound. It was the key ingredient in the result by Dadush, Koh, Natura, Olver, and Végh [17] that gave the first strongly polynomial algorithm for LPs with two nonzero entries per column. The crux of the result is a strongly polynomial bound on the straight line complexity of such LPs. The TR2-IPM is also applicable to this problem and may ultimately lead to simpler algorithms.

Before [4, 6], the behavior of IPM trajectories in the neighborhood of Klee–Minty cubes has been studied in different contexts. For ℓ_2-path following, a lower bound $\Omega(n^{1/3})$ has been established for a broad class of algorithms [68, 69]. Deza, Nematollahi, Peyghami and Terlaky [20], and Deza, Nematollahi and Terlaky [21] proved further lower bounds for the number of iterations needed to decrease the gap by a constant factor. In terms of the above framework, an example of a Klee-Minty cube was given in [21] for which $\mathrm{SLC}(\mathcal{N}^2(\beta), 1, 2) = \Omega(\sqrt{n/\log^5(n)})$.

Very recently, Vladu [74] gave an IPM with a matching iteration bound $O(n^{1/3})$ for a class of M-matrix-based quadratic programs.

The concept of *instance optimality* was first introduced by Fagin, Lotem, and Naor [24] in the context of database aggregation. Other examples include various sorting algorithms, see [62, Chapter 3] and references within. A recent breakthrough by Haeupler, Hladík, Rozhoň, Tarjan, and Tětek [32] showed that Dijkstra's algorithm with an appropriate heap data structure is universally optimal, i.e., optimal for graphs with a fixed topology. Subsequently, they showed that bidirectional Dijkstra is instance optimal [33]. We note that our notion of instance optimality is weaker in the sense that it is restricted to a particular class of algorithms; on the other hand, we consider the very general linear programming problem.

1.6 Discussion and future directions

In this paper, we give an IPM that runs in current matrix multiplication times the wide neighborhood straight line complexity, and show that TR2-IPM algorithm is canonical and instance $O(1)$-optimal for ℓ_2-path following. To complete the picture of optimal path following, it would be desirable to characterize the straight line complexity by a curvature integral formula analogous to the Sonnevend curvature used in [79, 80] that essentially characterizes the number of iterations to traverse short segments of the central path. Another fundamental question is to understand the tradeoff in iteration complexity between the ℓ_2 and wide neighborhoods. Long step algorithms in the wide neighborhood algorithms perform much better in practice. Can we give tight upper and lower bounds between the optimal trajectory lengths in the two neighborhoods? Note that [5] implies that $\mathrm{SLC}\left(\mathcal{N}^2(\beta), \mu_1, \mu_0\right) \leq O\left(n^{1.5}\beta^{-1}\log(n/(\beta(1-\theta)))\,\mathrm{SLC}(\mathcal{N}^{-\infty}(\theta), \mu_1, \mu_0)\right)$. It remains open to improve this bound or show that it is the best possible.

It would also be interesting to further strengthen Theorem 1.3 to obtain an $O(1)$-optimal IPM in the wide neighborhood. This would need to involve a wide-neighborhood analogue of Theorem 1.7 to relate straight line complexities in different wide neighborhoods. Moreover, it would be desirable to replace the Robust IPM, that uses $O(\sqrt{n})$ steps in each call, by simpler steps. One challenge is that while affine scaling is naturally optimal on short straight segments of the central path, there is no such natural candidate in the wide neighborhood.

Finally, regarding ℓ_2-trust region, in Section 7.5 we show how the general convex form may be reduced to our setting (with certain restrictions). For the general (non-convex) setting, [77] gave a weakly polynomial algorithm. One might also attempt to obtain strongly polynomial algorithms for some nonconvex settings.

1.7 Roadmap and organization of the paper

Sections on the wide and narrow neighborhood algorithms can be read largely independently. Readers interested in TRW-IPM for path following in wide neighborhood are referred to Section 3 and Section 8 with preliminary results from Section 2.3 and Section 2.7. For TR2-IPM on path following in ℓ_2-neighborhood, the reader is referred to Section 4, Section 5 and Section 7 with preliminary results from Section 2.3-2.6, and Section 6 for completing the proof of instance optimality of TR2-IPM.

In more detail, the paper is organized as follows: in Section 2 we first introduce our notation, computational models for our algorithms, as well as basic results for IPMs including properties of duality gap, ℓ_2-neighborhood and wide neighborhood. Then, in Section 2.4 we describe the standard affine scaling and corrector steps and some basic properties. Section 2.5 studies the curvature of straight lines in the ℓ_2-neighborhood. Section 2.6 focuses on the results of the local proximity of points in ℓ_2-neighborhood and also wide neighborhood.

In Section 3, we define the wide neighborhood trust region direction, and then present and explain the algorithm TRW-IPM in Section 3.1. In Section 3.2 we formally define and derive the polarization of line segments in wide neighborhood. The analysis is given in Section 3.3, completing the proof of Theorem 1.3.

Moving on to Section 4, after defining ℓ_2-trust region direction, the algorithm TR2-IPM is presented and explained in Section 4.1. Section 4.2 is dedicated to formally deriving the notion of ℓ_2-polarization and finding the polarized partition. In Section 4.3, we show the near-optimality of trust region step for long straight line segment, proving Lemma 1.11. In Section 5, we show the near-optimality of affine scaling step for short straight line segment, proving Lemma 1.12, and hence Theorem 1.6.

Section 6 is dedicated to proving Theorem 1.7 and then the main result on path following in ℓ_2 neighborhood, Theorem 1.4 of the paper. We prove Theorem 1.13 separately for short segments in Section 6.1 and long segments in Section 6.2. At the end of this section, we combine these two results to prove Theorem 1.7 and hence finally Theorem 1.4 in Section 6.3.

Section 7 focuses on solving the ℓ_2-trust region problem. The Lagrangian function and critical points for (TR-2) are described in Section 7.1 and Section 7.2 respectively. The algorithm TR2-SOLVE($\mathbf{B}, b, I, J, \delta$) is given and explained in Section 7.3. Its correctness and running time, together with Theorem 1.9, are proven in Section 7.3 and 7.4. Its relation and applicability to general trust region problems are discussed in Section 7.5.

Section 8 focuses on solving the wide neighborhood trust region problem. The critical points of the lifting operator for (TR-max) are described in Section 8.1. The algorithm TRW-SOLVE($\mathbf{B}, b, I, J, \delta$) is given and explained in Section 8.2. Then, Theorem 1.10 is proven in Section 8.3.

2 Preliminaries

2.1 Notation

We let $\mathbb{R}_{>\mathbf{0}}$ denote the set of positive reals, $\mathbb{R}_{\geq\mathbf{0}}$ the set of nonnegative reals, and $\mathbb{N} = \{1, 2, \dots\}$ denote the natural numbers. For $n \in \mathbb{N}$, we let $[n] := \{1, 2, \dots, n\}$, and for $j > i \geq 1$, we use the notation $[i : j]$ to denote the set $\{i, ..., j\}$. The sets $B, N \subseteq [n]$ form a partition of $[n]$ if $B \cup N = [n]$ and $B \cap N = \emptyset$. 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. We let $\mathbf{1} = \mathbf{1}_n \in \mathbb{R}^n$ denote the all ones vector.

The inner product between two vectors is denoted by $\langle x, y \rangle = x^\top y$ for $x, y \in \mathbb{R}^n$. We use $\|x\| := \sqrt{\sum_{i=1}^n x_i^2}$ to denote the ℓ_2 norm of a vector. We further let $\|x\|_\infty := \max_{i\in[n]} |x_i|$ denote the ℓ_∞ norm, and $\|x\|_1 := \sum_{i=1}^n |x_i|$ denote the ℓ_1 norm.

For two vectors $a, b \in \mathbb{R}^n$, we let $[a, b] := \{\lambda a + (1 - \lambda)b : 0 \leq \lambda \leq 1\}$ denote the straight line between a and b. For $x, y \in \mathbb{R}^n$, we let $\operatorname{diag}(x) \in \mathbb{R}^{n\times n}$ denote the diagonal matrix with x on the diagonal, and we use the notation $xy \in \mathbb{R}^n$ for the Hadamard product $xy = \operatorname{diag}(x)\, y = (x_i y_i)_{i\in[n]}$. We let $y/x \in \mathbb{R}^n$ denote the vector $(y_i/x_i)_{i\in[n]}$ and x^{-1} denote the vector $(1/x_i)_{i\in[n]}$. For $S \subseteq [n]$, let $x_S \in \mathbb{R}^S$ denote the sub-vector induced by the set S.

For a matrix $\mathbf{X} \in \mathbb{R}^{m\times n}$, we let $\mathbf{X}^\top \in \mathbb{R}^{n\times m}$ denote the matrix transpose, and use $\mathbf{X}^\dagger \in \mathbb{R}^{n\times m}$ denote the Moore–Penrose pseudo-inverse of $\mathbf{X}$. When both $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{n\times n}$ are symmetric matrices, we write $\mathbf{X} \succeq \mathbf{0}$ if $\mathbf{X}$ is positive semidefinite and $\mathbf{X} \succeq \mathbf{Y}$ if $\mathbf{X} - \mathbf{Y}$ is positive semidefinite. For subsets $S \subseteq [m]$, $T \subseteq [n]$, we let $\mathbf{X}_{S,T} \in \mathbb{R}^{S\times T}$ be the submatrix induced by the rows in S and columns in T of $\mathbf{X}$. Unless specified otherwise, we simply use $\mathbf{X}_T$ to denote $\mathbf{X}_{[m],T}$, and $\mathbf{D}_T$ to denote $\mathbf{D}_{T,T}$ for any diagonal matrices $\mathbf{D}$.

For a domain $D \subseteq \mathbb{R}$, we let $\mathcal{C}^2(D)$ denote the set of all twice continuously differentiable functions defined over D. For $f \in \mathcal{C}^2(D)$, we denote its first derivative by f' and second derivative by f''.

2.2 Preliminaries on computation

Strongly polynomial computational models. We consider both the *real RAM model* and *Turing model* of computation. In the real RAM model, the input is given by K real numbers, and one can perform a sequence of elementary arithmetic operations $(+, -, \times, /)$ and comparisons $(\geq)$ on real numbers. We say that an algorithm is *polynomial in the real RAM model* if the number of elementary arithmetic operations and comparisons is bounded polynomially in the size K of the input; in the case of LP, this is $K = n \times m + n + m$.

In the Turing model, we consider a problem where the input is given by K integers; for LP, the input $(\mathbf{A}, b, c)$ is described by $K = 2(n \times m + n + m)$ integers representing the rational entries. An algorithm is *strongly polynomial in the Turing model* (see [29]), if it only performs $\operatorname{poly}(K)$ (in the LP case, this means $\operatorname{poly}(m, n)$) elementary arithmetic operations and comparisons as in the real model. Additionally, the bit-complexity of the numbers during the computations must remain polynomially bounded in the encoding length of the input. Equivalently, the algorithm must be PSPACE.

While a strongly polynomial algorithm in the Turing model implies a polynomial algorithm in the real RAM model, the converse is not necessarily the case: enforcing PSPACE may be challenging. The Robust IPMs [13, 43, 71], as well as the LLS IPMs [16, 46, 51, 53, 73], are polynomial in the real RAM model[5] whenever e.g., $\log(\bar{\chi}_\mathbf{A}) = \operatorname{poly}(n)$. Turning these algorithms into strongly polynomial ones in the Turing model is a nontrivial

[5]Some of these IPMs make use of square-root computations, and hence rely on the extended real model $(+, -, \times, /, \sqrt{\cdot})$.

challenge, as one needs to keep the bit-complexity of the iterates bounded using only the allowed operations $(+, -, \times, /)$; operations such as truncating bit representation are not allowed. For this purpose, Dadush, Koh, Natura, Olver, and Végh [17] developed a general technique. They introduced a strongly polynomial rounding step using Caratheodory decomposition arguments to convert a point in the central path neighborhood to one of bounded bit-complexity. Further, this is applicable irrespective of the particular iteration used by the algorithm. Their rounding technique can be used for a broad class of path following IPMs, but it is computationally expensive.

The TR2-Solve($\mathbf{B}, b, I, J, \delta$) algorithm can be easily implemented in the Turing model: the main computations are a Cholesky decomposition, and solving linear systems. It is also easy to argue that the value of the parameter λ, obtained using binary search and Newton's method, remains polynomially bounded in the input.

In contrast, the TRW-Solve($\mathbf{B}, b, I, J, \delta$) algorithm relies on the Robust IPM [71] and hence the claimed running time holds in the Real RAM model. This is also the case for the IPM algorithms TRW-IPM and TR2-IPM. Even though every iteration is strongly polynomial, the linear systems to compute the step-lengths involve the current iterate and thus the bit complexity could double in every $O(1)$ iterations. These could also be converted to algorithms in the Turing model using the rounding steps from [17], but this would involve a running time overhead.

Approximating square roots. Square-root computations are not allowed in the real RAM (or Turing) model; this is particularly challenging since IPMs often require them, particularly, for computing norms. In our algorithms, we also require Euclidean norms of vectors (see e.g., Lemma 4.7 and Lemma 5.5). However, we do not need exact values (and not even very high accuracy approximations). We can approximate vector norms in strongly polynomial time using Newton's method, and the observation that $\|x\|_1$ gives a rough starting approximation of $\|x\|$.

Proposition 2.1 (Square root computation). *For any given $x > 0$ and $\varepsilon > 0$, suppose $a > 0$ is given such that $a \le \sqrt{x}$, it takes $O\left(\log\left(\sqrt{x}/a\right) + \log\log(1/\varepsilon)\right)$ many arithmetic operations in strongly polynomial time to compute λ that satisfies $\sqrt{x} \le \lambda \le (1+\varepsilon)\sqrt{x}$.*

For computing the norm $\|x\|$ for $x \in \mathbb{R}^n$, we can compute a starting estimate $\|x\|_1$ using basic arithmetic. Noting that $\|x\|_1/\sqrt{n} \le \|x\| \le \|x\|_1$, and starting with a multiplicative binary search (see Algorithm 5), then Newton's method yields the following corollary.

Proposition 2.2 (ℓ_2-computation). *For any given $x \in \mathbb{R}^n$ and $\varepsilon > 0$, we can compute some λ that satisfies $\|x\| \le \lambda \le (1+\varepsilon)\|x\|$ in strongly polynomial time with $O\left(\log\log(n) + \log\log\left(1/\varepsilon\right)\right)$ many operations.*

In Appendix C, we show how a pivoted Cholesky factorization can be implemented in the square-root free real RAM model, avoiding approximate square roots during a standard Cholesky factorization.

2.3 Preliminaries on interior point methods

In (LP), the program in x will be referred to as the primal problem and the program in (y, s) as the dual program. Recall the notation

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

for the primal and dual feasible regions, and $\mathcal{P}_{>\mathbf{0}} \coloneqq \{x \in \mathcal{P} : x > \mathbf{0}\}$, $\mathcal{D}_{>\mathbf{0}} \coloneqq \{s \in \mathcal{D} : s > \mathbf{0}\}$. Throughout, we assume that (LP) is bounded and strictly feasible, so that $\mathcal{P}_{>\mathbf{0}}, \mathcal{D}_{>\mathbf{0}} \neq \varnothing$, and also $n \ge 2$. For the following two simple propositions, see e.g. [5].

Proposition 2.3. *Given $z = (x, s), z' = (x', s') \in \mathcal{P} \times \mathcal{D}$, we have*

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

Proposition 2.4 (Linearity of duality gap). *For $x^{(1)}, \dots, x^{(k)} \in \mathcal{P}$, $s^{(1)}, \dots, s^{(k)} \in \mathcal{D}$ 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*

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

The next two propositions follow directly from the definitions of the ℓ_2-neighborhood and wide neighborhood.

Proposition 2.5. *For $\beta \in [0, 1)$, let $z = (x, s) \in \mathcal{N}^2(\beta)$. Then $(1-\beta)\overline{\mu}(z)\mathbf{1} \le xs \le (1+\beta)\overline{\mu}(z)\mathbf{1}$.*

Proof. Let $\overline{\mu}(z) = \mu$. For all $i \in [n]$, $\left|\frac{x_i s_i}{\mu} - 1\right| \le \left\|\frac{xs}{\mu} - \mathbf{1}\right\| \le \beta$, so $(1-\beta)\mu \le x_i s_i \le (1+\beta)\mu$. □

Proposition 2.6. *For $\theta \in [0,1)$, let $z = (x,s) \in \mathcal{N}^{-\infty}(\theta)$, then $(1-\theta)\overline{\mu}(z)\mathbf{1} \le xs \le (n\theta + 1 - \theta)\overline{\mu}(z)\mathbf{1}$.*

Proof. Let $\overline{\mu}(z) = \mu$. Since $\sum_{i=1}^{n} x_i s_i = n\mu$ and $xs \ge (1-\theta)\mu\mathbf{1}$, for each $i \in [n]$, we have

$$x_i s_i = n\mu - \sum_{j\neq i} x_j s_j \le n\mu - \sum_{j \neq i}(1-\theta)\mu = n\mu - (n-1)(1-\theta)\mu = (n\theta + 1 - \theta)\mu\,.$$

□

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

Lemma 2.7 ([51, Lemma 4.4])**.** *For $\beta \in (0,1)$, let $z = (x,s) \in \mathcal{P}_{>\mathbf{0}} \times \mathcal{D}_{>\mathbf{0}}$ and $\mu' > 0$ satisfy that $\|xs - \mu'\mathbf{1}\| \le \beta\mu'$. Then,*

$$\left(1 - \frac{\beta}{\sqrt{n}}\right)\mu' \le \overline{\mu}(z) \le \left(1 + \frac{\beta}{\sqrt{n}}\right)\mu' \quad \text{and} \quad z \in \mathcal{N}^2\left(\frac{\beta}{1 - \beta/\sqrt{n}}\right).$$

A key property of the central path is near monotonicity, which was formulated in [73, Lemma 16]. In the following lemma, we show that near monotonicity also holds for points in the wide neighborhood. This will help us identify the polarized partition in TRW-IPM.

Lemma 2.8 (Near-monotonicity)**.** *For $\theta \in [0,1)$, let $z = (x,s) \in \mathcal{N}^{-\infty}(\theta)$. For any $z' = (x', s') \in \mathcal{N}^{-\infty}(\theta)$ with $\overline{\mu}(z') \le \frac{(1-\theta)^3}{4n}\overline{\mu}(z)$. we have $\frac{x_i'}{x_i} + \frac{s_i'}{s_i} \le \frac{n}{1-\theta}$.*

Proof. By Proposition 2.3, we get $\sum_{i=1}^{n}(x_i s_i' + s_i x_i') = n(\mu + \mu')$. For each $i \in [n]$, since $x_i s_i \ge (1-\theta)\mu$, we have

$$\frac{x_i'}{x_i} + \frac{s_i'}{s_i} = \sum_{i=1}^{n} \frac{x_i' s_i + x_i s_i'}{x_i s_i} - \sum_{j \neq i}\left(\frac{x_j'}{x_j} + \frac{s_j'}{s_j}\right) \le \frac{n}{1-\theta}\left(1 + \frac{\mu'}{\mu}\right) - \sum_{j\neq i}\left(\frac{x_j'}{x_j} + \frac{s_j'}{s_j}\right).$$

Then, using AM-GM inequality and Proposition 2.6,

$$\frac{x_i'}{x_i} + \frac{s_i'}{s_i} \le \frac{n}{1-\theta}\left(1 + \frac{\mu'}{\mu}\right) - 2\sum_{j\neq i}\sqrt{\frac{x_j' s_j'}{x_j s_j}} \le \frac{n}{1-\theta}\left(1 + \frac{\mu'}{\mu}\right) - 2(n-1)\sqrt{\frac{(1-\theta)\mu'}{(n\theta + 1 - \theta)\mu}}\,.$$

The function $f : [0,1] \to \mathbb{R}$ defined by $f(t) = \frac{n}{1-\theta}t - 2(n-1)\sqrt{\frac{1-\theta}{n\theta+1-\theta}t}$ is decreasing on $\left[0\,,\, \left(1 - \frac{1}{n}\right)^2 \frac{(1-\theta)^3}{n\theta+1-\theta}\right]$ and increasing on $\left[\left(1-\frac{1}{n}\right)^2 \frac{(1-\theta)^3}{n\theta+1-\theta}\,,\, 1\right]$. Since $0 \le \frac{\mu'}{\mu} \le \frac{(1-\theta)^3}{4n}$, $f(\frac{\mu'}{\mu}) \le f(0) = 0$, giving $\frac{x_i'}{x_i} + \frac{s_i'}{s_i} \le \frac{n}{1-\theta} + f(\frac{\mu'}{\mu}) \le \frac{n}{1-\theta}$. □

Finally, we prove that wide neighborhood is indeed big enough to contain the line segment from $z \in \mathcal{N}^{-\infty}(\theta)$ to $z_1 \in \mathcal{N}^{-\infty}(\theta)$ with $\overline{\mu}(z_1) \le \overline{\mu}(z)$. Taking z and z_1 to be any subsequent iterates in short step IPMs in e.g. [13, 50, 71], the straight line between them counts as 1 straight line complexity in $\mathcal{N}^{-\infty}\left(1 - \frac{\overline{\mu}(z_1)}{\overline{\mu}(z)}(1-\theta)\right)$.

Lemma 2.9. *For $\theta \in [0,1)$, let $z = (x,s) \in \mathcal{N}^{-\infty}(\theta)$ with $\overline{\mu}(z) = \mu$ and $z_1 = (x_1, s_1) \in \mathcal{N}^{-\infty}(\theta)$ with $\overline{\mu}(z_1) = \gamma\mu$ where $\gamma \in [0,1]$. Then, $[z_1, z] \subseteq \mathcal{N}^{-\infty}(1 - \gamma(1-\theta))$.*

Proof. Let $\lambda \in [0,1]$. By Proposition 2.4, $\overline{\mu}((1-\lambda)z + \lambda z_1) = (1-\lambda)\mu + \lambda\gamma\mu$. Using the AM-GM inequality,

$$\begin{aligned}
((1-\lambda)x + \lambda x_1)((1-\lambda)s + \lambda s_1) &\ge (xs)^{1-\lambda}(x_1 s_1)^{\lambda} \\
&\ge (1-\theta)\mu^{1-\lambda}(\gamma\mu)^{\lambda} \\
&\ge (1-\theta)\gamma\mu \\
&\ge (1-\theta)\gamma\overline{\mu}((1-\lambda)z + \lambda z_1)\,.
\end{aligned}$$

Hence, $[z_1, z] \subseteq \mathcal{N}^{-\infty}(1 - \gamma(1-\theta))$. □

In particular, when z and z_1 are on the central path, it can be seen from the above proof that the straight line between them counts as 1 straight line complexity in $\mathcal{N}^{-\infty}\left(1 - \frac{\overline{\mu}(z_1)}{\overline{\mu}(z)}\right)$.

2.4 Predictor-corrector step

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

$$\mathbf{A}\Delta x = \mathbf{0} \tag{10}$$

$$\mathbf{A}^\top \Delta y + \Delta s = \mathbf{0} \tag{11}$$

$$s\Delta x + x\Delta s = \nu\mu\mathbf{1} - xs \tag{12}$$

Predictor-corrector methods, such as the Mizuno–Todd–Ye Predictor-Corrector Algorithm [50], alternate between two types of steps. In *corrector steps*, we use $\nu = 1$, which gives the *centrality direction*, denoted as $\Delta z^{\mathrm{c}} = (\Delta x^{\mathrm{c}}, \Delta s^{\mathrm{c}})$ throughout. In *predictor steps*, we use $\nu = 0$. This direction is also called the *affine scaling direction*, and will be denoted as $\Delta z^{\mathrm{a}} = (\Delta x^{\mathrm{a}}, \Delta s^{\mathrm{a}})$ throughout. The following proposition summarizes the well-known properties of predictor and corrector steps.

Proposition 2.10. *For $\beta \in (0, 1)$, let $z = (x, s) \in \mathcal{N}^2(\beta)$.*

(i) Let Δz^{c} be the centrality direction at z. Then $\overline{\mu}(z + \Delta z^{\mathrm{c}}) = \overline{\mu}(z)$ and $z + \Delta z^{\mathrm{c}} \in \mathcal{N}^2(\frac{\sqrt{2}\beta^2}{4(1-\beta)})$.

(ii) Let $\alpha \in [0, 1]$ and Δz^{a} be the affine scaling direction at z, then $\overline{\mu}(z + \alpha\Delta z^{\mathrm{a}}) = (1 - \alpha)\overline{\mu}(z)$. Moreover, if $z + \alpha\Delta z^{\mathrm{a}} \in \mathcal{N}^2(\beta)$, then $z + \alpha'\Delta z^{\mathrm{a}} \in \mathcal{N}^2(\beta)$ for any $\alpha' \in [0, \alpha]$.

(iii) For any $k > 0$ such that $(k+1)\beta < 1$, let $\overline{\alpha} := \sup\left\{\alpha \in [0, 1] : z + \alpha'\Delta z^{\mathrm{a}} \in \mathcal{N}^2((k + 1)\beta),\ \forall\, \alpha' \in [0, \alpha]\right\}$. We define α as follows: if $\|\Delta x^{\mathrm{a}}\Delta s^{\mathrm{a}}\| = 0$, $\alpha := 1$, otherwise let α be the root on $\mathbb{R}_{>\mathbf{0}}$ of

$$\frac{\alpha^2}{1 - \alpha}\left\|\frac{\Delta x^{\mathrm{a}}\Delta s^{\mathrm{a}}}{\mu}\right\| = k\beta\,. \tag{13}$$

Then, $k\beta/\sqrt{n} \le \alpha \le \overline{\alpha}$.

Proof. For the first parts of (i) and(ii), see e.g. [50]. The second part of (i) is a result of [49, Lemma 1], and the second part of (ii) is [4, Lemma 2]. (iii) can be derived analogously following Lemma 4.17 and Theorem 4.18 in [78]. □

The following proposition explicitly states that the predictor and corrector steps can be computed in strongly polynomial time, because computing the steps amounts to solving linear systems based on the data x, s, μ and ν.

Proposition 2.11 (Step Formulas)**.** *Let $\mathbf{A} \in \mathbb{R}^{m\times n}$ with $\operatorname{rank}(\mathbf{A}) = m$, $z = (x, s) \in \mathcal{P}_{>\mathbf{0}} \times \mathcal{D}_{>\mathbf{0}}$ with $\overline{\mu}(z) = \mu$ and $t \in \mathbb{R}^n$. Then, the solution to $s\Delta x + x\Delta s = t$, $\Delta x \in \ker(\mathbf{A})$ and $\Delta s \in \operatorname{Im}(\mathbf{A}^\top)$ can be expressed as*

$$\begin{aligned}\Delta s &= \mathbf{A}^\top \left(\mathbf{A}\operatorname{diag}(x/s)\,\mathbf{A}^\top\right)^{-1}\mathbf{A}\frac{t}{s}\\ \Delta x &= \frac{t}{s} - \operatorname{diag}(x/s)\,\mathbf{A}^\top\left(\mathbf{A}\operatorname{diag}(x/s)\,\mathbf{A}^\top\right)^{-1}\mathbf{A}\frac{t}{s}.\end{aligned} \tag{14}$$

Moreover, both the affine scaling step $(\Delta x^{\mathrm{a}}, \Delta s^{\mathrm{a}})$ and corrector step $(\Delta x^{\mathrm{c}}, \Delta s^{\mathrm{c}})$, corresponding to $t = \nu\mu\mathbf{1}_n - xs$ for $\nu \in \{0, 1\}$ respectively can be computed in strongly polynomial time.

We now introduce some useful notation for the description and analysis of our TR2-IPM algorithm.

Definition 2.12 (Associated partition)**.** For $z = (x, s) \in \mathcal{N}^2(\beta)$, let $(\Delta x^{\mathrm{a}}, \Delta s^{\mathrm{a}})$ be the affine scaling step. Let us define the associated partition $\widetilde{B}_z \cup \widetilde{N}_z = [n]$ as

$$\widetilde{B}_z := \left\{i \in [n] : \left|\frac{\Delta x^{\mathrm{a}}_i}{x_i}\right| \le \left|\frac{\Delta s^{\mathrm{a}}_i}{s_i}\right|\right\}\,, \quad \widetilde{N}_z := [n] \setminus \widetilde{B}_z\,.$$

As a result of Proposition 2.11, the associated partition at z can be computed in strongly polynomial time.

Definition 2.13 (Normalized Iterates, Normalized and Local Search Direction)**.** For $z = (x, s) \in \mathcal{P}_{>\mathbf{0}} \times \mathcal{D}_{>\mathbf{0}}$, we let

$$\begin{aligned} \hat{\xi}(z) &:= \sqrt{\frac{xs}{\overline{\mu}(z)}} \in \mathbb{R}^n \,, \\ \hat{x} &:= \frac{x}{\hat{\xi}(z)} = \sqrt{\frac{x\overline{\mu}(z)}{s}} \in \mathbb{R}^n \,, \\ \hat{s} &:= \frac{s}{\hat{\xi}(z)} = \sqrt{\frac{s\overline{\mu}(z)}{x}} \in \mathbb{R}^n \,. \end{aligned} \tag{15}$$

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. Given a search direction $\Delta z = (\Delta x, \Delta s) \in \ker(\mathbf{A}) \times \operatorname{Im}(\mathbf{A}^\top)$ at z, we let

$$\Delta\bar{x} := \frac{\Delta x}{x} \quad \text{and} \quad \Delta\bar{s} := \frac{\Delta s}{s}\,,$$
$$\Delta\hat{x} := \frac{\Delta x}{\hat{x}} = \sqrt{\frac{s}{x\overline{\mu}(z)}}\,\Delta x \quad \text{and} \quad \Delta\hat{s} := \frac{\Delta s}{\hat{s}} = \sqrt{\frac{x}{s\overline{\mu}(z)}}\,\Delta s\,.$$

We call $\Delta\bar{z} := (\Delta\bar{x}, \Delta\bar{s})$ the *local search direction* at z and $\Delta\hat{z} := (\Delta\hat{x}, \Delta\hat{s})$ the *normalized search direction* at z.

$z = (x, s)$ falls on the central path if and only if $\hat{\xi}(z) = \mathbf{1}_n$. In this case, $\hat{x} = x$ and $\hat{s} = s$. When z is not on the central path, the variables $\hat{x}$ and $\hat{s}$ represent natural adjustments for points off the central path. The following result is a corollary of Proposition 2.5 which shows the ℓ_∞ and ℓ_2-proximity of $\hat{\xi}(z)$ to $\mathbf{1}_n$ for any $z \in \mathcal{N}^2(\beta)$.

Proposition 2.14. *For $\beta \in (0, 1)$, let $z \in \mathcal{N}^2(\beta)$ with $\overline{\mu}(z) = \mu$, and $\hat{\xi} := \hat{\xi}(z)$. Then $\sqrt{1-\beta}\cdot\mathbf{1}_n \le \hat{\xi} \le \sqrt{1+\beta}\cdot\mathbf{1}_n$. Moreover, for any $z_1 \in \mathcal{N}^2(\beta)$ with $\overline{\mu}(z_1) = \mu_1$, letting $\hat{\xi}_1 := \hat{\xi}(z_1)$, we have $\left\|\frac{\hat{\xi}_1^2}{\hat{\xi}^2} - \mathbf{1}_n\right\| \le \frac{2\beta}{1-\beta}$.*

Throughout, we will frequently consider local and normalized search directions. We refer to $\|\Delta x/x\|$ and $\|\Delta s/s\|$ as the *primal and dual local norms* of the search direction $\Delta z = (\Delta x, \Delta s)$ at $z = (x, s) \in \mathcal{P}_{>\mathbf{0}} \times \mathcal{D}_{>\mathbf{0}}$.

For the affine scaling direction ($\nu = 0$), (12) can be written in terms of the local affine scaling direction:

$$\Delta\bar{x}^{\mathrm{a}} + \Delta\bar{s}^{\mathrm{a}} = -\mathbf{1}_n\,, \tag{16}$$

and also in terms of the normalized affine scaling direction:

$$\Delta\hat{x}^{\mathrm{a}} + \Delta\hat{s}^{\mathrm{a}} = -\hat{\xi}, \tag{17}$$

which serves the purpose that now $\Delta\hat{x}^{\mathrm{a}} = \Delta x^{\mathrm{a}}/\hat{x}$ and $\Delta\hat{s}^{\mathrm{a}} = \Delta s^{\mathrm{a}}/\hat{s}$ are orthogonal vectors. Thus, (17) gives an orthogonal decomposition of $-\hat{\xi}$. As a result, $\Delta z^{\mathrm{a}} = (\Delta x^{\mathrm{a}}, \Delta s^{\mathrm{a}})$ can be seen as the optimal solutions of the following minimum-norm problems, see e.g. [5]

$$\Delta x^{\mathrm{a}} = \operatorname*{arg\,min}_{\mathbf{A}\Delta x = \mathbf{0}} \left\|\hat{x}^{-1}(x + \Delta x)\right\|\,, \quad \text{and} \quad \Delta s^{\mathrm{a}} = \operatorname*{arg\,min}_{\mathbf{A}^\top\Delta y + \Delta s = \mathbf{0}} \left\|\hat{s}^{-1}(s + \Delta s)\right\|\,. \tag{18}$$

In general, for any $\Delta z = (\Delta x, \Delta s) \in \ker(\mathbf{A}) \times \operatorname{Im}(\mathbf{A}^\top)$ at $z = (x, s)$, the primal and dual normalized search direction $\Delta\hat{x} = \Delta x/\hat{x}$ and $\Delta\hat{s} = \Delta s/\hat{s}$ are orthogonal to each other.

2.5 Curvature of a straight line

We next prove the two lemmas mentioned in Section 1.4.1. Given two points $z = (x, s), z_1 = (x_1, s_1) \in \mathcal{N}^2(\beta)$ with $\overline{\mu}(z) = \mu$ and $\overline{\mu}(z_1) = \mu_1$, $\mu_1 \le \mu$, we recall from (3) the *curvature* of the straight line between z_1 and z, defined as

$$\kappa(z_1, z) := \frac{\|(x_1 - x)(s_1 - s)\|}{\left(\sqrt{\mu} + \sqrt{\mu_1}\right)^2}\,.$$

This quantity determines how close the straight line $[z_1, z]$ is to the central path in the ℓ_2-neighborhood, by the following lemma. The final part shows that the curvature between central path points gives a tight bound on the neighborhood containing the line segment between these points, which plays a key role in proving Theorem 1.7.

Lemma 2.15. *For $\beta \in (0, 1)$, let $z = (x, s), z_1 = (x_1, s_1) \in \mathcal{N}^2(\beta)$ with $\overline{\mu}(z) = \mu$ and $\overline{\mu}(z_1) = \mu_1 \le \mu$.*

(i) *If* $\bar{z} = \lambda z_1 + (1-\lambda)z \in \mathcal{N}^2(\beta)$ *for all* $\lambda \in [0,1]$*, then* $\kappa(z_1, z) \leq 2\beta$*.*

(ii) *Conversely, if* $\kappa(z_1, z) \leq \beta$*, then* $\bar{z} = \lambda z_1 + (1-\lambda)z \in \mathcal{N}^2(2\beta)$ *for all* $\lambda \in [0,1]$*.*

(iii) *For* $0 < \mu_1 \leq \mu$*, it holds that*

$$\kappa\left(z(\mu_1), z(\mu)\right) = \sup\left\{ \left\| \frac{\bar{x}\bar{s}}{\overline{\mu}(\bar{z})} - \mathbf{1}_n \right\| \,:\, \bar{z} = (\bar{x}, \bar{s}) = \lambda z(\mu_1) + (1-\lambda) z(\mu),\, 0 \leq \lambda \leq 1 \right\}.$$

Proof. Let $\Delta z = (\Delta x, \Delta s) := (x_1 - x, s_1 - s)$. Thus, $(x_1, s_1) = (x + \Delta x, s + \Delta s)$, and $\lambda z_1 + (1-\lambda)z = z + \lambda \Delta z$. By Proposition 2.4, $\overline{\mu}(z + \lambda\Delta z) = (1-\lambda)\mu + \lambda\mu_1$. For $\bar{z} = \lambda z_1 + (1-\lambda)z$, we can write

$$\left\| \frac{\bar{x}\bar{s}}{\overline{\mu}(\bar{z})} - \mathbf{1}_n \right\| = \left\| \frac{(x+\lambda\Delta x)(s + \lambda \Delta s)}{(1-\lambda)\mu + \lambda\mu_1} - \mathbf{1}_n \right\| = \left\| \frac{(1-\lambda)(xs - \mu\mathbf{1}_n) + \lambda(x_1 s_1 - \mu_1 \mathbf{1}_n) - \lambda(1-\lambda)\Delta x \Delta s}{(1-\lambda)\mu + \lambda\mu_1} \right\|. \tag{19}$$

Let us start by showing part (iii). For $z = z(\mu)$ and $z_1 = z(\mu_1)$, the first two terms in the numerator are 0, thus

$$\left\| \frac{\bar{x}\bar{s}}{\overline{\mu}(\bar{z})} - \mathbf{1}_n \right\| = \frac{\lambda(1-\lambda)}{(1-\lambda)\mu + \lambda\mu_1} \left\| \Delta x \Delta s \right\|. \tag{20}$$

This expression is maximized at $\lambda = \frac{\sqrt{\mu}}{\sqrt{\mu} + \sqrt{\mu_1}}$. Taking the maximum value

$$\left\| \frac{\bar{x}\bar{s}}{\overline{\mu}(\bar{z})} - \mathbf{1}_n \right\| \leq \frac{1}{\left(\sqrt{\mu} + \sqrt{\mu_1}\right)^2} \left\| \Delta x \Delta s \right\| = \kappa(z_1, z),$$

which completes the proof of (iii).

For part (i), from (19) we get by the triangle inequality,

$$\begin{aligned} \left\| \frac{\bar{x}\bar{s}}{\overline{\mu}(\bar{z})} - \mathbf{1}_n \right\| &\geq \frac{-(1-\lambda)\|xs - \mu\mathbf{1}\| - \lambda\|x_1 s_1 - \mu_1 \mathbf{1}\| + \lambda(1-\lambda)\|\Delta x \Delta s\|}{(1-\lambda)\mu + \lambda\mu_1} \\ &\geq \frac{-(1-\lambda)\beta\mu - \lambda\beta\mu_1 + \lambda(1-\lambda)\|\Delta x \Delta s\|}{(1-\lambda)\mu + \lambda\mu_1} \\ &= \frac{\lambda(1-\lambda)}{(1-\lambda)\mu + \lambda\mu_1} \|\Delta x \Delta s\| - \beta. \end{aligned} \tag{21}$$

For the choice $\lambda = \frac{\sqrt{\mu}}{\sqrt{\mu} + \sqrt{\mu_1}}$, we get that

$$\left\| \frac{\bar{x}\bar{s}}{\overline{\mu}(\bar{z})} - \mathbf{1}_n \right\| \geq \kappa(z_1, z) - \beta,$$

implying $\kappa(z_1, z) \leq 2\beta$ since $\bar{z} \in \mathcal{N}^2(\beta)$ for any choice of λ.

Finally, for part (ii), we use the triangle inequality as in (21) to show

$$\begin{aligned} \left\| \frac{\bar{x}\bar{s}}{\overline{\mu}(\bar{z})} - \mathbf{1}_n \right\| &\leq \frac{(1-\lambda)\|xs - \mu\mathbf{1}_n\| + \lambda\|x_1 s_1 - \mu_1 \mathbf{1}_n\| + \lambda(1-\lambda)\|\Delta x \Delta s\|}{(1-\lambda)\mu + \lambda\mu_1} \\ &\leq \frac{\lambda(1-\lambda)}{(1-\lambda)\mu + \lambda\mu_1} \|\Delta x \Delta s\| + \beta \\ &\leq \kappa(z_1, z) + \beta \leq 2\beta. \end{aligned} \tag{22}$$

□

We now consider any feasible search direction $\Delta z = (\Delta x, \Delta s)$ at $z \in \mathcal{N}^2(\beta)$, such that by taking a full step, $\overline{\mu}(z + \Delta z) = \mu_1 \leq \mu$. The following lemma gives an upper bound in terms of $\kappa(z + \Delta z, z)$ for the local ℓ_2-distance between Δz and the affine scaling step $\alpha^* \Delta z^{\mathrm{a}}$ with the ideal step length $\alpha^* = 1 - \mu_1/\mu$.

Lemma 2.16. *For* $\beta \in (0,1)$*, let* $z = (x, s), z_1 = (x_1, s_1) \in \mathcal{N}^2(\beta)$ *with* $\overline{\mu}(z) = \mu$ *and* $\overline{\mu}(z_1) = \mu_1 \leq \mu$*. Let* $\nu := \mu_1/\mu \in [0,1]$*,* $\Delta z^{\mathrm{a}} = (\Delta x^{\mathrm{a}}, \Delta s^{\mathrm{a}})$ *be the affine scaling direction at* z*, and* $\Delta z = (\Delta x, \Delta s) := (x_1 - x, s_1 - s)$*. Then,*

$$\|(\Delta\hat{x} - (1-\nu)\Delta\hat{x}^{\mathrm{a}}\,,\, \Delta\hat{s} - (1-\nu)\Delta\hat{s}^{\mathrm{a}})\| \leq \frac{\kappa(z_1, z)\left(1 + \sqrt{\nu}\right)^2 + 2\beta\nu}{\sqrt{1-\beta}}.$$

Proof. We recall $\hat{\xi} = \sqrt{\frac{xs}{\mu}}$, and let $\hat{\xi}_1 := \sqrt{\frac{x_1 s_1}{\mu_1}}$. Noting that $\Delta\hat{x} - (1-\nu)\Delta\hat{x}^{\mathrm{a}}$ and $\Delta\hat{s} - (1-\nu)\Delta\hat{s}^{\mathrm{a}}$ are orthogonal vectors and that $\Delta\hat{x}^{\mathrm{a}} + \Delta\hat{s}^{\mathrm{a}} = -\hat{\xi}$,

$$\begin{aligned}\|(\Delta\hat{x} - (1-\nu)\Delta\hat{x}^{\mathrm{a}}\,,\, \Delta\hat{s} - (1-\nu)\Delta\hat{s}^{\mathrm{a}})\|^2 &= \|\Delta\hat{x} - (1-\nu)\Delta\hat{x}^{\mathrm{a}}\|^2 + \|\Delta\hat{s} - (1-\nu)\Delta\hat{s}^{\mathrm{a}}\|^2 \\ &= \left\|(1-\nu)\hat{\xi} + \Delta\hat{x} + \Delta\hat{s}\right\|^2 \\ &= \left\|\hat{\xi}^{-1}\left(\hat{\xi}^2 + \hat{\xi}\Delta\hat{x} + \hat{\xi}\Delta\hat{s} - \nu\hat{\xi}_1^2\right) + \nu\hat{\xi}^{-1}\left(\hat{\xi}_1^2 - \hat{\xi}^2\right)\right\|^2 \\ &\le \left(\left\|\hat{\xi}^{-1}\left(\hat{\xi}^2 + \hat{\xi}\Delta\hat{x} + \hat{\xi}\Delta\hat{s} - \nu\hat{\xi}_1^2\right)\right\| + \nu\left\|\hat{\xi}^{-1}\left(\hat{\xi}_1^2 - \hat{\xi}^2\right)\right\|\right)^2 .\end{aligned}$$

From $(x + \Delta x)(s + \Delta s) = x_1 s_1$ we get $\left(\hat{\xi} + \Delta\hat{x}\right)\left(\hat{\xi} + \Delta\hat{s}\right) = \nu\hat{\xi}_1^2$, so $\Delta\hat{x}\Delta\hat{s} = \nu\hat{\xi}_1^2 - \hat{\xi}^2 - \hat{\xi}\Delta\hat{x} - \hat{\xi}\Delta\hat{s}$. Hence,

$$\begin{aligned}\|(\Delta\hat{x} - (1-\nu)\Delta\hat{x}^{\mathrm{a}}\,,\, \Delta\hat{s} - (1-\nu)\Delta\hat{s}^{\mathrm{a}})\| &\le \left\|\hat{\xi}^{-1}\Delta\hat{x}\Delta\hat{s}\right\| + \nu\left\|\hat{\xi}^{-1}\left(\hat{\xi}_1^2 - \hat{\xi}^2\right)\right\| \\ &\le \left\|\hat{\xi}^{-1}\right\|_\infty \left\|\frac{\Delta x\Delta s}{\mu}\right\| + \nu\left\|\hat{\xi}^{-1}\right\|_\infty \left(\left\|\hat{\xi}_1^2 - \mathbf{1}\right\| + \left\|\hat{\xi}^2 - \mathbf{1}\right\|\right) \\ &\le \frac{\kappa(z_1, z)\left(1+\sqrt{\nu}\right)^2 + 2\beta\nu}{\sqrt{1-\beta}} ,\end{aligned}$$

where the final inequality uses the definition of $\kappa(z_1, z)$, Proposition 2.14 to bound $\left\|\hat{\xi}^{-1}\right\|_\infty \le 1/\sqrt{1-\beta}$, and the assumption $z, z_1 \in \mathcal{N}^2(\beta)$ to bound $\left\|\hat{\xi}_1^2 - \mathbf{1}\right\|, \left\|\hat{\xi}^2 - \mathbf{1}\right\| \le \beta$. □

2.6 Local proximity of points in ℓ_2-neighborhood

Given two general points in $\mathcal{N}^2(\beta)$ with the same duality gap, it turns out that being in the same ℓ_2-neighborhood ensures that they are close to each other. In this section, we show that both the ℓ_2 and the ℓ_∞-local norm of their difference are bounded by roughly 3β. This is based on the following upper bound by Gonzaga [27, Lemma 5.4] which establishes the local ℓ_2-proximity for a point $z \in \mathcal{N}^2(\beta)$ with $\overline{\mu}(z) = \mu$ to the corresponding central path point $z(\mu)$. We include a simple self-contained proof.

Lemma 2.17 (Gonzaga [27])**.** *For $\beta \in (0,1)$, let $z = (x, s) \in \mathcal{N}^2(\beta)$ with $\overline{\mu}(z) = \mu$. Then,*

$$\left\|\left(\frac{x - x(\mu)}{x}\,,\, \frac{s - s(\mu)}{s(\mu)}\right)\right\| \le \frac{\beta}{1-\beta} \quad \text{and} \quad \left\|\left(\frac{x - x(\mu)}{x(\mu)}\,,\, \frac{s - s(\mu)}{s}\right)\right\| \le \frac{\beta}{1-\beta} . \tag{23}$$

Proof. Let $\delta := \left\|\left(\frac{x - x(\mu)}{x}\,,\, \frac{s - s(\mu)}{s(\mu)}\right)\right\|$ denote the first expression in the statement. Observe that

$$\min_{i\in[n]} \frac{x_i}{x(\mu)_i} = \frac{1}{1 + \max_{i\in[n]}\left(\frac{x(\mu)_i}{x_i} - 1\right)} \ge \frac{1}{1 + \left\|\frac{x(\mu)}{x} - \mathbf{1}_n\right\|} \ge \frac{1}{1+\delta} . \tag{24}$$

Next, we derive the following lower bound on the centrality error of (x, s):

$$\begin{aligned}\left\|\frac{xs}{\mu} - \mathbf{1}_n\right\|^2 = \left\|\frac{xs}{\mu} - \frac{x}{x(\mu)} + \frac{x}{x(\mu)} - \mathbf{1}_n\right\|^2 &= \left\|\frac{x(s - s(\mu))}{\mu} + \frac{x - x(\mu)}{x(\mu)}\right\|^2 \\ &= \left\|\sqrt{\frac{x}{x(\mu)}}\left[\frac{\sqrt{x(\mu)x}(s - s(\mu))}{\mu} + \frac{x - x(\mu)}{\sqrt{x(\mu)x}}\right]\right\|^2 \\ &\ge \min_i \frac{x_i}{x(\mu)_i}\left\|\frac{\sqrt{x(\mu)x}(s - s(\mu))}{\mu} + \frac{x - x(\mu)}{\sqrt{x(\mu)x}}\right\|^2 \\ &\ge \frac{1}{1+\delta}\left\|\frac{\sqrt{x(\mu)x}(s - s(\mu))}{\mu} + \frac{x - x(\mu)}{\sqrt{x(\mu)x}}\right\|^2 ,\end{aligned}$$

using (24) in the last step. We now make use of the orthogonality of $\sqrt{x(\mu)x}(s - s(\mu))$ and $\frac{x - x(\mu)}{\sqrt{x(\mu)x}}$ to get

$$\begin{aligned}
\left\|\frac{xs}{\mu} - \mathbf{1}_n\right\|^2 &\geq \frac{1}{1+\delta}\left(\left\|\frac{\sqrt{x(\mu)x}(s - s(\mu))}{\mu}\right\|^2 + \left\|\frac{x - x(\mu)}{\sqrt{x(\mu)x}}\right\|^2\right) \\
&= \frac{1}{1+\delta}\left(\left\|\sqrt{\frac{x}{x(\mu)}} \cdot \frac{s - s(\mu)}{s(\mu)}\right\|^2 + \left\|\sqrt{\frac{x}{x(\mu)}} \cdot \frac{x - x(\mu)}{x}\right\|^2\right) \\
&\geq \frac{1}{1+\delta} \cdot \left(\min_i \frac{x_i}{x(\mu)_i}\right)\left(\left\|\frac{s - s(\mu)}{s(\mu)}\right\|^2 + \left\|\frac{x - x(\mu)}{x}\right\|^2\right) \\
&\overset{(24)}{\geq} \frac{\delta^2}{(1+\delta)^2}\,.
\end{aligned}$$

As $\beta \geq \|xs/\mu - \mathbf{1}_n\|$, we obtain $\beta \geq \delta/(1+\delta)$. Reordering the terms results in $\delta \leq \beta/(1-\beta)$ as desired. The other inequality follows analogously. ☐

Gonzaga's bound immediately yields the following bounds on the ℓ_∞-local distance between $z \in \mathcal{N}^2(\beta)$ and the central path point $z(\mu)$.

Proposition 2.18 ([51, Proposition 2.1])**.** *For $\beta \in (0,1)$, let $z = (x,s) \in \mathcal{N}^2(\beta)$ with $\overline{\mu}(z) = \mu$. Then for $i \in [n]$,*

$$(1-\beta)\, x_i \leq x_i(\mu) \leq \frac{x_i}{1-\beta} \quad \text{and} \quad (1-\beta)\, s_i \leq s_i(\mu) \leq \frac{s_i}{1-\beta}\,.$$

Using the previous two results, we are able to show that given any two points $z, z' \in \mathcal{N}^2(\beta)$ that have the same duality gap, their local ℓ_∞ and ℓ_2-distance are also close.

Lemma 2.19 ($\ell_{2,\infty}$-proximity)**.** *For $\beta \in (0,1)$, let $z = (x,s), z' = (x',s') \in \mathcal{N}^2(\beta)$ with $\overline{\mu}(z) = \overline{\mu}(z') = \mu$. Then*

$$(1-\beta)^2\mathbf{1}_n \leq \frac{x'}{x} \leq \frac{\mathbf{1}_n}{(1-\beta)^2} \quad \text{and} \quad (1-\beta)^2\mathbf{1}_n \leq \frac{s'}{s} \leq \frac{\mathbf{1}_n}{(1-\beta)^2}\,.$$

Moreover,

$$\left\|\left(\frac{x'-x}{x}, \frac{s'-s}{s}\right)\right\| \leq \frac{2\beta}{(1-\beta)^2} \quad \text{and} \quad \left\|\frac{x'-x}{x}\right\| + \left\|\frac{s'-s}{s}\right\| \leq \frac{2\sqrt{2}\beta}{(1-\beta)^2}\,.$$

Proof. By the previous proposition, for each $i \in [n]$, $x_i'(1-\beta) \leq x_i(\mu) \leq \frac{x_i}{1-\beta}$ so $(1-\beta)^2 \leq \frac{x_i'}{x_i} \leq \frac{1}{(1-\beta)^2}$ and the same bounds hold for $\frac{s_i'}{s_i}$. Using Proposition 2.18 and the AM-GM inequality,

$$\begin{aligned}
\left\|\left(\frac{x'-x}{x}, \frac{s'-s}{s}\right)\right\|^2 &= \left\|\frac{x'-x}{x}\right\|^2 + \left\|\frac{s'-s}{s}\right\|^2 \\
&\leq \left(\left\|\frac{x(\mu)}{x} \cdot \frac{x'}{x(\mu)} - \frac{x'}{x(\mu)}\right\| + \left\|\frac{x'}{x(\mu)} - \mathbf{1}_n\right\|\right)^2 + \left(\left\|\frac{s'}{s(\mu)} \cdot \frac{s(\mu)}{s} - \frac{s(\mu)}{s}\right\| + \left\|\frac{s(\mu)}{s} - \mathbf{1}_n\right\|\right)^2 \\
&\leq \left\|\frac{x'}{x(\mu)}\right\|_\infty^2 \left(\left\|\frac{x(\mu)}{x} - \mathbf{1}_n\right\| + \left\|\frac{x(\mu)}{x'} - \mathbf{1}_n\right\|\right)^2 + \left\|\frac{s(\mu)}{s}\right\|_\infty^2 \left(\left\|\frac{s'}{s(\mu)} - \mathbf{1}_n\right\| + \left\|\frac{s}{s(\mu)} - \mathbf{1}_n\right\|\right)^2 \\
&\leq \frac{2}{(1-\beta)^2}\left(\left\|\frac{x(\mu)}{x} - \mathbf{1}_n\right\|^2 + \left\|\frac{s}{s(\mu)} - \mathbf{1}_n\right\|^2 + \left\|\frac{x(\mu)}{x'} - \mathbf{1}_n\right\|^2 + \left\|\frac{s'}{s(\mu)} - \mathbf{1}_n\right\|^2\right) \\
&\leq \frac{4\beta^2}{(1-\beta)^4}
\end{aligned}$$

where in the last inequality we apply Lemma 2.17. Hence, $\left\|\left(\frac{x'-x}{x}, \frac{s'-s}{s}\right)\right\| \leq \frac{2\beta}{(1-\beta)^2}$. Finally,

$$\left(\left\|\frac{x'-x}{x}\right\| + \left\|\frac{s'-s}{s}\right\|\right)^2 \leq 2\left\|\left(\frac{x'-x}{x}, \frac{s'-s}{s}\right)\right\|^2 \leq \frac{8\beta^2}{(1-\beta)^4}$$

which gives $\left\|\frac{x'-x}{x}\right\| + \left\|\frac{s'-s}{s}\right\| \leq \frac{2\sqrt{2}\beta}{(1-\beta)^2}$. ☐

While we can also show proximity of two points in the same wide neighborhood, the bound on their ℓ_∞-local distance is $O(n/(1-\theta))$ which is weaker than in the ℓ_2-neighborhood.

Lemma 2.20. *For $\theta \in [0,1)$, let $z=(x,s), z'=(x',s') \in \mathcal{N}^{-\infty}(\theta)$ with $\overline{\mu}(z)=\overline{\mu}(z')$. Then*

$$\frac{1-\theta}{2n}x' \le x \le \frac{2n}{1-\theta}x' \quad \text{and} \quad \frac{1-\theta}{2n}s' \le s \le \frac{2n}{1-\theta}s'\,.$$

Proof. We only prove the inequalities on x, x', as the proof of the inequalities on s, s' is symmetric. Let $i \in [n]$, using Proposition 2.3, we have

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

Analogously, we get $\frac{x'_i}{x_i} \ge \frac{1-\theta}{2n}$ which gives $\frac{x_i}{x'_i} \le \frac{2n}{1-\theta}$. □

2.7 Fast IPM solvers

We now state the properties of the Robust IPM by van den Brand [71] and show how it can be used in our algorithm. The following potential is used to define the central path neighborhood in these methods. We use the version as described in the tutorial [43]. Let $\Phi : \mathbb{R}^n \to \mathbb{R}$ be defined as

$$\Phi(r) \coloneqq \sum_{i=1}^n \cosh(\lambda r_i) = \frac{1}{2}\sum_{i=1}^n (\mathrm{e}^{\lambda r_i} + \mathrm{e}^{-\lambda r_i})\,, \tag{25}$$

where $\lambda \coloneqq 16\lceil \log(40n)\rceil$. We get the following simple bound relating this to the wide neighborhood, see e.g., [43, Lemma 15].

Lemma 2.21. *Let $x, s \in \mathbb{R}^n_{>\mathbf{0}}$, $t \ge 0$. If $\Phi(xs/t - \mathbf{1}_n) \le 16n$, then $\|xs/t - \mathbf{1}_n\|_\infty \le 1/16$, and consequently, $(x,s) \in \mathcal{N}^{-\infty}(1/8)$.*

Proof. For every $i \in [n]$, $\frac{1}{2}\mathrm{e}^{\lambda|r_i|} \le \Phi(r)$, and therefore $\|r\|_\infty \le \log(2\Phi(r))/\lambda$. The first claim follows by the choice of λ. This in particular implies $|\overline{\mu}(z)/t - 1| \le 1/16$ for $z=(x,s)$, implying $z \in \mathcal{N}^{-\infty}(1/8)$. □

For the following theorem from [71], see [43, Lemma 24].

Theorem 2.22 (Robust IPM)**.** *There exists a deterministic algorithm that for an instance of (LP), $\bar z = (\bar x, \bar s) \in \mathcal{P} \times \mathcal{D}$ and $\bar t > 0$ such that $\Phi(\bar x \bar s/\bar t - \mathbf{1}_n) \le 16n$, and $\varrho \in (0,1)$, outputs $(x^{\mathrm{out}}, s^{\mathrm{out}}) \in \mathcal{P} \times \mathcal{D}$ and $t^{\mathrm{out}} \le \varrho t$ such that $\Phi(x^{\mathrm{out}}s^{\mathrm{out}}/t^{\mathrm{out}} - \mathbf{1}_n) \le 16n$ in time*

$$O\left(n^{\tilde\omega + o(1)} \log\left(\frac{1}{\varrho}\right)\right)\,.$$

We show that the same guarantee holds if the starting point is only assumed to be in the wide neighborhood $\mathcal{N}^{-\infty}(\theta)$. The reduction is given in Appendix A, using an extended system, following the same lines as [43, Theorem 11].

Theorem 2.23. *Let $\theta \in (1/8, 1)$. Consider an instance of (LP). There exists a subroutine* PathFollow(x,s,ϱ) *that, given $(\bar x, \bar s) \in \mathcal{N}^{-\infty}(\theta)$ with $\overline{\mu}(\bar x, \bar s) = \bar\mu$ and $\varrho \in (0,1)$, outputs $(x^{\mathrm{out}}, s^{\mathrm{out}}) \in \mathcal{N}^{-\infty}(3/4)$ with $\overline{\mu}(x^{\mathrm{out}}, s^{\mathrm{out}}) \le \varrho\bar\mu$ in time $O\left(n^{\tilde\omega+o(1)}\log\left(\frac{1}{\varrho(1-\theta)}\right)\right)$.*

Next, we show how to use the Robust IPM method to approximately solve LP with ℓ_∞-norm constraint.

Definition 2.24 (Approximate ℓ_∞-regression solver)**.** An approximate ℓ_∞ regression solver is an algorithm that, given a matrix $\mathbf{B} \in \mathbb{R}^{m\times n}$, a vector $b \in \mathrm{Im}(\mathbf{B})$, and a parameter $\delta > 0$, either

(i) returns a vector $y \in \mathbb{R}^n$ with $\mathbf{B}y = b$ and $\|y\|_\infty \le 1+\delta$, or

(ii) returns a vector $z \in \mathbb{R}^m$ with $\|\mathbf{B}^\top z\|_1 < \langle b, z\rangle$, certifying that no vector y with $\mathbf{B}y = b$ and $\|y\|_\infty \le 1$ exists.

Note that the second outcome is a Farkas certificate. In particular, for any y with $\mathbf{B}y = b$ we have that $\langle b, z\rangle = y^\top \mathbf{B}^\top z \le \|y\|_\infty \|\mathbf{B}^\top z\|_1 < \|y\|_\infty \langle b, z\rangle$, showing $\|y\|_\infty > 1$.

Theorem 2.25. *There exists an approximate ℓ_∞-regression solver* LMFeas$(\mathbf{B}, b, \delta)$ *satisfying Definition 2.24 that runs in time $O\left(n^{\tilde\omega+o(1)}\log\left(\frac{1}{\delta}\right)\right)$.*

The proof is given in Appendix A; we give a brief sketch. We first compute the minimum-norm point $\bar y = \mathbf{B}^\dagger b$. If $\|\bar y\|_2 > \sqrt{n}$, we may conclude infeasibility, and find a Farkas certificate as in *(ii)*. Otherwise, we can use $\bar y$ to initialize the IPM for $\min \alpha$ s.t. $\mathbf{B}y = b$, $\|y\|_\infty \le 1+\alpha$ with $\alpha = O(n)$.

2.8 Singular value decompositions

The eigenvalue problem has always been of great interest and has been extensively studied in different communities. Various algorithms with different approaches, computational models and notions of approximation were proposed to tackle this problem, see e.g. [8, 9, 10, 57]. In particular, the randomized algorithm in [9] achieves diagonalization of general matrices with high probability in nearly matrix multiplication time. For general non-Hermitian matrices, it is difficult to design a globally convergent and numerically stable algorithm to approximate the eigenvalues and eigenvectors; see [19]. In contrast, fast and stable eigenvalue algorithms that globally converge are well-known for Hermitian matrices (e.g. [75]) with further speed-up possible, for example, $\tilde{O}(n^\omega)$ running time if only the largest eigenvalue is needed [44].

For our purpose to compute the critical points (see Section 1.4.3), we require an algorithm to approximate only the eigenvalues of a symmetric positive semi-definite matrix. However, as opposed to the above algorithms, such approximation needs to be multiplicative, and the algorithm needs to be implementable in strongly polynomial time in order to fit in our overall computation model. The requirement is captured by the following definition.

Definition 2.26 (Partial SVD)**.** Let $\mathbf{A} \in \mathbb{R}^{m\times n}$ with $m \leq n$. A matrix $\mathbf{U} \in \mathbb{R}^{\mathrm{rk}(\mathbf{A})\times n}$ is called a partial ϱ-SVD of $\mathbf{A}$ for some parameter $\varrho \geq 1$ if $\mathbf{U}$ has orthogonal rows and for all $x \in \mathbb{R}^n$ we have that

$$\varrho^{-1}\|\mathbf{U}x\|_2 \leq \|\mathbf{A}x\|_2 \leq \varrho\|\mathbf{U}x\|_2\,. \tag{26}$$

Our definition of a partial SVD resembles a spectral approximation with the additional requirement of an orthogonal basis whose basis vectors correspond approximately to singular vectors of $\mathbf{A}$ and whose basis vector norms correspond approximately to the singular values of $\mathbf{A}$.

Since the development of Francis's QR algorithm, there is a long line of research on the use of rank-revealing QR decomposition via column pivoting to multiplicatively approximate the eigenvalues. One of the earliest such algorithms was proposed by Chan [11], then algorithms with tighter approximation and better implementation were developed in [12, 30, 56]. However, these algorithms generally require the computation of Householder reflectors or a certain singular vector, which cannot be done in strongly polynomial time. More recently, a randomized strongly polynomial algorithm was provided by Diakonikolas, Tzamos and Kane [22] to compute a $(1+\varepsilon)$-approximate singular value decomposition, by running power iterations on a randomized initial basis. Allamigeon, Dadush, Loho, Natura, and Végh in [5] also gave a deterministic algorithm for $(1+\varepsilon)$-singular value decomposition which was needed in the subroutine of their IPM. Their algorithm is based on Gram-Schmidt orthogonalization with a special pivoting rule which replaces the role of Householder reflectors, and has a strongly polynomial $\tilde{O}(n^5)$ running time. While strongly polynomial, the running time of their algorithm is significantly inferior to that of the weakly polynomial algorithms commonly referenced in the literature. Their approach can be characterized as a pivoted QR orthogonalization, wherein the primary computational bottleneck arises from the pivoting step, the selection of which requires complexity of order $\tilde{O}(n^4)$. The substantial computational overhead associated with pivoting emerges from the necessity, at each iteration, to compute projections of every column yet to be pivoted onto the orthogonal complement of the set of remaining columns.

It is noteworthy that their algorithm could alternatively be reformulated through repeated matrix inversion, incurring a computational cost of $\tilde{O}(n^\omega)$ per iteration. Nevertheless, achieving a running time better than $\tilde{O}(n^{\omega+1})$ via this matrix inversion strategy appears unlikely.

In Appendix C.1, we prove Lemma C.1 by presenting a $O(n^3)$ variant of the algorithm in [5], using pivoted Cholesky factorization. It gives a multiplicative $\varrho = O(n2^{2n})$-approximation of the eigenvalues of a symmetric positive definite matrix in strongly polynomial time. The running time is comparable to standard QR, Cholesky and LDLT factorizations, if no fast matrix multiplication is used. In particular, at each iteration, the algorithm applies the Schur complement computation in the row and column with largest diagonal entry. Beyond operating on the symmetric Gram matrix, the primary conceptual distinction from [5] lies in the pivot selection criterion: our method selects pivots according to descending singular values, whereas [5] selects pivots based on ascending singular values.

Although one could boost this result and get an $(1+\varepsilon)$-approximation by applying this algorithm to a power of the matrix, we note that already the weak approximation suffices for our purposes, as the running time dependence on ϱ in TR2-SOLVE$(\mathbf{B}, b, I, J, \delta)$ is $\log\log\varrho$ (see Lemma 7.5).

For the ℓ_∞-regression problem (Section 1.4.4) however, we also do require approximations to the singular vectors. To this end, we show in Appendix C the following result, which is indeed achieved by boosting the method of Lemma C.1 via matrix powers.

Lemma 2.27 (Deterministic strongly polynomial partial SVD)**.** *There exists an $\tilde{O}(n^3 \max\{1, \log(1/\varepsilon)\})$ deterministic algorithm that given $m \times n$ matrix $\mathbf{A}$ with $m \leq n$ and $\varepsilon > 0$ computes a partial $(1+\varepsilon)$-SVD $\mathbf{U}$ of $\mathbf{A}$.*

Using the randomization technique of Diakonikolas, Tzamos, and Kane [22] a sped up version of the algorithm can be achieved.

Lemma 2.28 (Randomized strongly polynomial partial SVD)**.** *There exists an $\tilde{O}(n^\omega \max\{1, \log(1/\varepsilon)\})$ randomized algorithm that given $m \times n$ matrix $\mathbf{A}$ with $m \le n$ and $\varepsilon > 0$ computes a partial $(1+\varepsilon)$-SVD $\mathbf{U}$ of $\mathbf{A}$.*

3 The Wide Neighborhood Trust Region Interior Point Method

We present and explain the algorithm TRW-IPM in Section 3.1. Then, we derive wide neighborhood polarization in Section 3.2. Section 3.3 focuses on the analysis of TRW-IPM and proving Theorem 1.3.

3.1 Description of the Algorithm

We first formally restate the wide neighborhood trust region subproblem $\mathrm{TR}_\infty(B, N, \ell, u)$ from Section 1.2.

Definition 3.1. Let $z = (x, s) \in \mathcal{N}^{-\infty}(\theta)$, $B \cup N = [n]$ be a partition and $u, \ell > 0$. We define the (B, N, ℓ, u)-trust region direction at z as the optimal solution $(\Delta x^{\mathrm{TR}}, \Delta s^{\mathrm{TR}}, \Delta y^{\mathrm{TR}})$ of the primal and dual trust region problems

$$\begin{aligned} \min\ & \left\| \mathbf{1}_N + \frac{\Delta x_N}{x_N} \right\|_\infty \\ \text{s.t. } & \ell \mathbf{1} \le \mathbf{1}_B + \frac{\Delta x_B}{x_B} \le u\mathbf{1} \\ & \mathbf{A}\Delta x = \mathbf{0} \end{aligned} \qquad \begin{aligned} \min\ & \left\| \mathbf{1}_B + \frac{\Delta s_B}{s_B} \right\|_\infty \\ \text{s.t. } & \ell \mathbf{1} \le \mathbf{1}_N + \frac{\Delta s_N}{s_N} \le u\mathbf{1} \\ & \mathbf{A}^\top \Delta y + \Delta s = \mathbf{0} \end{aligned} \tag{$\mathrm{TR}_\infty(B, N, \ell, u)$}$$

This can be reduced to an instance of (TR-max), hence $(\Delta x^{\mathrm{TR}}, \Delta s^{\mathrm{TR}}, \Delta y^{\mathrm{TR}})$ can be computed by Theorem 1.10, we deferred the statement and its proof to the end of this section.

We now present our Wide Neighborhood Trust Region IPM in Algorithm 1. It uses two steps: PathFollow implementing the Robust IPM as in Theorem 2.23, and WTR for the Trust Region step as in Proposition 3.9. We repeatedly call PathFollow to decrease the normalized gap μ by a factor $\varrho = (1-\theta)^9/(256n^9)$. After each call, we guess the partition $\overline{B} \cup \overline{N} = [n]$ by setting $\overline{B}$ as the set of indices where the value of x_i decreased by at most a factor $\frac{(1-\theta)^4}{4n^4}$ during this step. We then compute a Trust Region step with $(\overline{B}, \overline{N})$, along with a corresponding step length. If this step decreases the normalized gap by at least ϱ, we take it. Otherwise, we discard this trust region step, and make another call to PathFollow instead.

Algorithm 1: TRW-IPM

Input : An instance of (LP) with constraint matrix $\mathbf{A} \in \mathbb{R}^{m \times n}, \mathrm{rk}(\mathbf{A}) = m$, $\theta \in (0,1)$, an initial iterate $(x^0, s^0) \in \mathcal{N}^{-\infty}(\theta)$, and $\mu_1 \ge 0$.

Output: $(x, s) \in \overline{\mathcal{N}}^{-\infty}(\theta)$ satisfying $\overline{\mu}(z_1) \le \mu_1$

$(x, s) \leftarrow (x^0, s^0)$;
$\varrho \leftarrow \frac{(1-\theta)^9}{256n^9}$;
$(x, s) \leftarrow (x^0, s^0)$;
while $\overline{\mu}(x, s) > \mu_1$ **do**
  $(\bar{x}, \bar{s}) \leftarrow (x, s)$;
  $(x', s') \leftarrow$ PathFollow(x, s, ϱ) ;
  $\overline{B} \leftarrow \left\{ i \in [n] : x'_i \ge \frac{(1-\theta)^4}{4n^4} \bar{x}_i \right\}$, $\overline{N} \leftarrow [n] \setminus \overline{B}$;
  $(\Delta x^{\mathrm{TR}}, \Delta s^{\mathrm{TR}}) \leftarrow \mathrm{WTR}(\overline{B}, \overline{N})$;
  $\varepsilon \leftarrow \max\left\{ \left\| \mathbf{1}_N + \frac{\Delta x_N^{\mathrm{TR}}}{x_N} \right\|_\infty, \left\| \mathbf{1}_B + \frac{\Delta s_B^{\mathrm{TR}}}{s_B} \right\|_\infty \right\}$; $\alpha^{\mathrm{TR}} \leftarrow 1/(1+2\varepsilon)$;
  if $\overline{\mu}(x + \alpha^{\mathrm{TR}} \Delta x^{\mathrm{TR}}, s + \alpha^{\mathrm{TR}} \Delta s^{\mathrm{TR}}) \le \varrho \overline{\mu}(x, s)$ **then**
    $(x, s) \leftarrow (x + \alpha^{\mathrm{TR}} \Delta x^{\mathrm{TR}}, s + \alpha^{\mathrm{TR}} \Delta s^{\mathrm{TR}})$;
  else
    $(x, s) \leftarrow$ PathFollow(x, s, ϱ);
return (x, s);

3.2 Wide neighborhood polarization

Recall Definition 1.5 on polarization of the central path in the wide neighborhood. We next define γ-polarization of a line segment.

Definition 3.2 (Polarization of line segment)**.** For $\gamma \in (0,1]$ and $z_0 = (x_0, s_0), z_1 = (x_1, s_1) \in \mathcal{N}^{-\infty}(\theta)$ with $\overline{\mu}(z_1) \le \overline{\mu}(z_0)$, we say the straight line $[z_1, z_0]$ is γ-polarized if there exists a partition $B \cup N = [n]$ such that for all $z \in [z_1, z_0]$,

$$x_B \ge \gamma \cdot x_{0B} \quad \text{and} \quad s_N \ge \gamma \cdot s_{0N}\,.$$

The next result is from [5, Lemma 3.6]. For completeness, we include the proof in Appendix B.

Lemma 3.3. *For $\theta \in [0,1)$, let $z_0 = (x_0, s_0), z_1 = (x_1, s_1) \in \mathcal{N}^{-\infty}(\theta)$ with $\overline{\mu}(z_1) < \overline{\mu}(z_0)$. If the straight line $[z_1, z_0] \subseteq \mathcal{N}^{-\infty}(\theta)$, then there exists a partition $B \cup N = [n]$ such that $[z_0, z_1]$ is $\frac{1-\theta}{4n}$-polarized with partition (B, N). Moreover, for any $z_0' \in \mathcal{N}^{-\infty}(\theta)$ with $\overline{\mu}(z_0') = \overline{\mu}(z_0)$ and any $z_1' \in \mathcal{N}^{-\infty}(\theta)$ with $\overline{\mu}(z_1') = \overline{\mu}(z_1)$, the straight lines $[z_1', z_0']$ are $\frac{(1-\theta)^3}{16n^3}$-polarized with (B, N).*

To prove that the trust region direction can take us near the end of the straight line segment, we make use of the local ideal direction in the next definition.

Definition 3.4. For any $z = (x, s) \in \mathcal{N}^{-\infty}(\theta)$ with $\overline{\mu}(z) = \mu$ and $\mu_1 \in (0, \mu)$, the *local ideal direction from z to μ_1* is defined as $\Delta z^{\mathrm{id}} = (\Delta x^{\mathrm{id}}, \Delta s^{\mathrm{id}}) := (x_1 - x, s_1 - s)$, where $z_1 = (x_1, s_1) \in \mathcal{P}_{>\mathbf{0}} \times \mathcal{D}_{>\mathbf{0}}$ is the unique vector such that $\frac{x_1 s_1}{xs} = \frac{\mu_1}{\mu}\mathbf{1}$.

Proposition 3.5. *For any $z = (x, s) \in \mathcal{N}^{-\infty}(\theta)$ with $\overline{\mu}(z) = \mu$ and $\mu_1 \in (0, \mu)$, there exists a unique point $z_1 = (x_1, s_1) \in \mathcal{N}^{-\infty}(\theta)$ such that $\frac{x_1 s_1}{xs} = \frac{\mu_1}{\mu}\mathbf{1}$.*

Proof. Let $w = \frac{xs}{\mu} \in \mathbb{R}^n$. Then we take $(x_1, s_1) = (x(\mu_1), s(\mu_1))$ as the unique w-central path point that satisfies the following system:

$$\begin{aligned} \mathbf{A}x(\mu_1) = b\,,\ \ x(\mu_1) > \mathbf{0}\,, \\ \mathbf{A}^\top y(\mu_1) + s(\mu_1) = c\,,\ \ s(\mu_1) > \mathbf{0}\,, \\ x(\mu_1)s(\mu_1) = \mu_1 w\,. \end{aligned}$$

Such $(x(\mu_1), s(\mu_1))$ exists and is unique; see [31, 45]. Then, $x_1 s_1 = \mu_1 xs/\mu \ge (1-\theta)\mu_1\mathbf{1}$, showing $z_1 \in \mathcal{N}^{-\infty}(\theta)$. ☐

Our next key lemma connects polarization to the trust region problem. We show that if there exists a long polarized segment, then the ideal direction gives a sufficiently good solution to $\mathrm{TR}_\infty\left(B, N, \frac{(1-\theta)^3}{16n^3}, \frac{n}{1-\theta}\right)$.

Lemma 3.6. *For $\theta \in (0,1]$, suppose there exist some $z_0 \in \mathcal{N}^{-\infty}(\theta)$ with $\overline{\mu}(z_0) = \mu$ and $z_1 \in \mathcal{N}^{-\infty}(\theta)$ with $\overline{\mu}(z_1) = \mu_1 \le \frac{(1-\theta)^3}{4n}\mu$ such that $[z_1, z_0] \subseteq \mathcal{N}^{-\infty}(\theta)$. Let $z \in \mathcal{N}^{-\infty}(\theta)$ with $\overline{\mu}(z) = \mu$, and let $B \cup N = [n]$ denote the polarized partition as in Lemma 3.3.*

(i) The local ideal direction from z to μ_1 is feasible to $\mathrm{TR}_\infty\left(B, N, \frac{(1-\theta)^3}{16n^3}, \frac{n}{1-\theta}\right)$, namely

$$\frac{(1-\theta)^3}{16n^3}\mathbf{1}_B \le \mathbf{1}_B + \frac{\Delta x_B^{\mathrm{id}}}{x_B} \le \frac{n}{1-\theta}\mathbf{1}_B \quad \text{and} \quad \frac{(1-\theta)^3}{16n^3}\mathbf{1}_N \le \mathbf{1}_N + \frac{\Delta s_N^{\mathrm{id}}}{s_N} \le \frac{n}{1-\theta}\mathbf{1}_N\,. \tag{27}$$

(ii) The objective value attained by $(\Delta x^{\mathrm{id}}, \Delta s^{\mathrm{id}})$ in $\mathrm{TR}_\infty\left(B, N, \frac{(1-\theta)^3}{16n^3}, \frac{n}{1-\theta}\right)$ satisfies

$$\max\left\{\left\|\mathbf{1}_N + \frac{\Delta x_N^{\mathrm{id}}}{x_N}\right\|_\infty, \left\|\mathbf{1}_B + \frac{\Delta s_B^{\mathrm{id}}}{s_B}\right\|_\infty\right\} \le \frac{16n^3\mu_1}{(1-\theta)^3\mu}\,. \tag{28}$$

Proof. **Part (i).** Let $z_1 = (x_1, s_1)$ satisfy $\frac{x_1 s_1}{xs} = \frac{\mu_1}{\mu}\mathbf{1}$. By Lemma 3.3, the straight line $[z_1, z]$ is $\frac{(1-\theta)^3}{16n^3}$-polarized, which gives $x_{1B} \ge \frac{(1-\theta)^3}{16n^3}x_B$ and $s_{1N} \ge \frac{(1-\theta)^3}{16n^3}s_N$. Hence, $\mathbf{1}_B + \frac{\Delta x_B^{\mathrm{id}}}{x_B} \ge \frac{(1-\theta)^3}{16n^3}\mathbf{1}_B$ and $\mathbf{1}_N + \frac{\Delta s_N^{\mathrm{id}}}{s_N} \ge \frac{(1-\theta)^3}{16n^3}\mathbf{1}_N$. The upper bounds follow from Lemma 2.8, where we take $(x', s') = (x + \Delta x^{\mathrm{id}}, s + \Delta s^{\mathrm{id}})$.

Part (ii). Using $\frac{x_1 s_1}{xs} = \frac{\mu_1}{\mu}\mathbf{1}$ and $\frac{n}{1-\theta}s_N \ge s_{1N} \ge \frac{(1-\theta)^3}{16n^3}s_N$, we have

$$\left\|\mathbf{1}_N + \frac{\Delta x_N^{\mathrm{id}}}{x_N}\right\|_\infty = \left\|\frac{x_{1N}}{x_N}\right\|_\infty = \frac{\mu_1}{\mu}\left\|\frac{s_N}{s_{1N}}\right\|_\infty \le \frac{16n^3\mu_1}{(1-\theta)^3\mu}\,. \tag{29}$$

The same bound holds for $\left\|\mathbf{1}_B + \frac{\Delta s_B^{\mathrm{id}}}{s_B}\right\|_\infty$ by an analogous derivation. ☐

3.3 Analysis

We are now ready to prove the key lemma to Theorem 1.3, which is that a trust region step reaches near the end of the current polarized segment while staying in $\mathcal{N}^{-\infty}(\theta')$ for a larger θ' value.

Lemma 3.7. *For $\theta \in (0,1]$, suppose there exist some $z_0 \in \mathcal{N}^{-\infty}(\theta)$ with $\overline{\mu}(z_0) = \mu$ and $z_1 \in \mathcal{N}^{-\infty}(\theta)$ with $\overline{\mu}(z_1) = \mu_1 \leq \frac{(1-\theta)^3}{64n^3}\mu$ such that $[z_1, z_0] \subseteq \mathcal{N}^{-\infty}(\theta)$. Let $z \in \mathcal{N}^{-\infty}(\theta)$ with $\overline{\mu}(z) = \mu$, and $\Delta z^{\mathrm{TR}} = (\Delta x^{\mathrm{TR}}, \Delta s^{\mathrm{TR}})$ be the trust region direction at z. Then for $\alpha^{\mathrm{TR}} = \frac{1}{1+2\varepsilon} \in [0,1]$ where*

$$\varepsilon := \max\left\{ \left\|\mathbf{1}_N + \frac{\Delta x_N^{\mathrm{TR}}}{x_N}\right\|_\infty, \left\|\mathbf{1}_B + \frac{\Delta s_B^{\mathrm{TR}}}{s_B}\right\|_\infty \right\},$$

$\overline{\mu}(z + \alpha^{\mathrm{TR}}\Delta z^{\mathrm{TR}}) \leq \frac{32n^4}{(1-\theta)^4}\mu_1$ and $z + \alpha^{\mathrm{TR}}\Delta z^{\mathrm{TR}} \in \mathcal{N}^{-\infty}\left(1 - \frac{(1-\theta)^5}{72n^4}\right)$.

Proof. Since $(\Delta x^{\mathrm{TR}}, \Delta s^{\mathrm{TR}})$ is the optimal solution to $\mathrm{TR}_\infty\left(B, N, \frac{(1-\theta)^3}{16n^3}, \frac{n}{1-\theta}\right)$, we have

$$\begin{aligned}
\overline{\mu}(z + \alpha^{\mathrm{TR}}\Delta z^{\mathrm{TR}}) &= (x + \alpha^{\mathrm{TR}}\Delta x^{\mathrm{TR}})^\top(s + \alpha^{\mathrm{TR}}\Delta s^{\mathrm{TR}})/n \\
&= (1-\alpha^{\mathrm{TR}})\mu + \alpha^{\mathrm{TR}}(x + \Delta x^{\mathrm{TR}})^\top(s + \Delta s^{\mathrm{TR}})/n \\
&= (1-\alpha^{\mathrm{TR}})\mu + \alpha^{\mathrm{TR}}\sum_{i=1}^n \frac{x_i s_i}{n}(1 + \Delta\bar{x}_i^{\mathrm{TR}})(1 + \Delta\bar{s}_i^{\mathrm{TR}}) \\
&\leq \left(1 - \frac{1}{1+2\varepsilon}\right)\mu + \frac{n}{1-\theta}\cdot\frac{\varepsilon}{1+2\varepsilon}\mu \leq \frac{2n\varepsilon}{1-\theta}\mu\,.
\end{aligned} \tag{30}$$

Since $\varepsilon \leq \frac{16n^3\mu_1}{(1-\theta)^3\mu}$, we get $\overline{\mu}(z + \alpha^{\mathrm{TR}}\Delta z^{\mathrm{TR}}) \leq \frac{32n^4}{(1-\theta)^4}\mu_1$. Let $i \in B$, then from $1 + \Delta\bar{x}_i^{\mathrm{TR}} \geq \frac{(1-\theta)^3}{16n^3}$ we get

$$1 + \alpha^{\mathrm{TR}}\Delta\bar{x}_i^{\mathrm{TR}} \geq 1 - \alpha^{\mathrm{TR}}\left(1 - \frac{(1-\theta)^3}{16n^3}\right) \geq \frac{(1-\theta)^3}{16n^3(1+2\varepsilon)}$$

and from $1 + \Delta\bar{s}_i^{\mathrm{TR}} \geq -\varepsilon$ we get

$$1 + \alpha^{\mathrm{TR}}\Delta\bar{s}_i^{\mathrm{TR}} \geq 1 - \alpha^{\mathrm{TR}}(1+\varepsilon) = 1 - \frac{1+\varepsilon}{1+2\varepsilon} = \frac{\varepsilon}{1+2\varepsilon}\,.$$

Then, we use $z \in \mathcal{N}^{-\infty}(\theta)$ and (30) to obtain

$$\begin{aligned}
\frac{(x_i + \alpha^{\mathrm{TR}}\Delta x_i^{\mathrm{TR}})(s_i + \alpha^{\mathrm{TR}}\Delta s_i^{\mathrm{TR}})}{\overline{\mu}(z + \alpha^{\mathrm{TR}}\Delta z^{\mathrm{TR}})} &= \frac{x_i s_i(1 + \alpha^{\mathrm{TR}}\Delta\bar{x}_i^{\mathrm{TR}})(1 + \alpha^{\mathrm{TR}}\Delta\bar{s}_i^{\mathrm{TR}})}{\overline{\mu}(z + \alpha^{\mathrm{TR}}\Delta z^{\mathrm{TR}})} \\
&\geq \frac{(1-\theta)\mu\cdot\frac{(1-\theta)^3\varepsilon}{16n^3(1+2\varepsilon)^2}}{\frac{2n\varepsilon}{1-\theta}\mu} \\
&\geq \frac{(1-\theta)^5}{32n^4}\cdot\frac{1}{(1+2\varepsilon)^2} \geq \frac{(1-\theta)^5}{72n^4}\,,
\end{aligned} \tag{31}$$

where the last inequality uses $\varepsilon \leq \frac{16n^3\mu_1}{(1-\theta)^3\mu} \leq \frac{1}{4}$. By an analogous derivation, we can get (31) for $i \in N$. Therefore,

$$(x + \alpha^{\mathrm{TR}}\Delta x^{\mathrm{TR}})(s + \alpha^{\mathrm{TR}}\Delta s^{\mathrm{TR}}) \geq \frac{(1-\theta)^5}{72n^4}\overline{\mu}(z + \alpha^{\mathrm{TR}}\Delta z^{\mathrm{TR}})\mathbf{1}$$

which shows that $z + \alpha^{\mathrm{TR}}\Delta z^{\mathrm{TR}} \in \mathcal{N}^{-\infty}\left(1 - \frac{(1-\theta)^5}{72n^4}\right)$. □

We now show that if the current point is far enough along the polarized segment, then the choice of $(\overline{B}, \overline{N})$ in the algorithm correctly reveals the polarized partition (B, N).

Lemma 3.8. *For $\theta \in [0,1)$, suppose $[z_0, z_1] \subseteq \mathcal{N}^{-\infty}(\theta)$ is γ-polarized with partition (B, N). Let $\bar{z} \in \mathcal{N}^{-\infty}(\theta)$ with $\overline{\mu}(\bar{z}) = \bar{\mu} \leq \frac{(1-\theta)^3}{4n}\overline{\mu}(z_0)$ and $z \in \mathcal{N}^{-\infty}(\theta)$ with $\overline{\mu}(z) = \mu \geq \overline{\mu}(z_1)$ such that $\mu \leq \frac{(1-\theta)^7\gamma^2}{16n^7}\bar{\mu}$. Then for*

$$\overline{B} := \left\{ i \in [n] : x_i \geq \frac{(1-\theta)^3\gamma}{4n^3}\bar{x}_i \right\}, \quad \overline{N} := \left\{ i \in [n] : s_i \geq \frac{(1-\theta)^3\gamma}{4n^3}\bar{s}_i \right\}.$$

we have $(\overline{B}, \overline{N}) = (B, N)$.

Proof. Since (B, N) is a partition, it suffices to prove that $B \subseteq \overline{B}$, $N \subseteq \overline{N}$, and $\overline{B} \cap \overline{N} = \varnothing$. Let z', $\bar{z}' \in [z_0, z_1]$ such that $\overline{\mu}(z') = \overline{\mu}(z)$ and $\overline{\mu}(\bar{z}') = \overline{\mu}(\bar{z})$. For $i \in B$, since $[z_0, z_1]$ is γ-polarized with (B, N), $x_i' \geq \gamma x_{0i}$. By Lemma 2.8, $\bar{x}_i' \leq \frac{n}{1-\theta} x_{0i}$. Hence, we get $x_i' \geq \frac{(1-\theta)\gamma}{n} \bar{x}_i'$. Then, using Lemma 2.20,

$$x_i \geq \frac{1-\theta}{2n} x_i' \geq \frac{(1-\theta)^2 \gamma}{2n^2} \bar{x}_i' \geq \frac{(1-\theta)^3 \gamma}{4n^3} \bar{x}_i$$

which shows that $i \in \overline{B}$, and $B \subseteq \overline{B}$. By an analogous derivation, we can show that $N \subseteq \overline{N}$. Suppose there exists $i \in \overline{B} \cap \overline{N}$, then $x_i s_i \geq \frac{(1-\theta)^6 \gamma^2}{16n^6} \bar{x}_i \bar{s}_i$. Since $z, \bar{z} \in \mathcal{N}^{-\infty}(\theta)$, $\frac{\bar{x}_i \bar{s}_i}{x_i s_i} \geq \frac{(1-\theta)\bar{\mu}}{(n\theta + 1 - \theta)\mu}$ by Proposition 2.6, which gives

$$\mu > \frac{(1-\theta)\bar{\mu}}{n} \cdot \frac{x_i s_i}{\bar{x}_i \bar{s}_i} \geq \frac{(1-\theta)^7 \gamma^2}{16n^7} \bar{\mu}$$

leading to a contradiction. □

We now show that the trust region step can be computed to a high accuracy that satisfy our purpose.

Proposition 3.9. *There exists an algorithm* $\mathrm{WTR}(B, N)$ *that in time* $O(n^{\tilde{\omega}+o(1)} \log(1/(1-\theta)))$ *computes solutions* $(\Delta x^{\mathrm{TR}}, \Delta s^{\mathrm{TR}})$ *to* $\mathrm{TR}_\infty\left(B, N, \frac{(1-\theta)^3}{16n^3}, \frac{n}{1-\theta}\right)$ *that satisfy*

$$\frac{(1-\theta)^3}{32n^3} \mathbf{1}_B \leq \mathbf{1}_B + \frac{\Delta x_B^{\mathrm{TR}}}{x_B} \leq \frac{3n}{2(1-\theta)} \mathbf{1}_B\,, \quad \text{and} \quad \frac{(1-\theta)^3}{32n^3} \mathbf{1}_N \leq \mathbf{1}_N + \frac{\Delta s_N^{\mathrm{TR}}}{s_N} \leq \frac{3n}{2(1-\theta)} \mathbf{1}_N\,.$$

and $\max\left\{\left\|\mathbf{1}_N + \frac{\Delta x_N^{\mathrm{TR}}}{x_N}\right\|_\infty, \left\|\mathbf{1}_B + \frac{\Delta s_B^{\mathrm{TR}}}{s_B}\right\|_\infty\right\} \leq \frac{17n^3 \mu_1}{(1-\theta)^3 \mu}$.

In the same setting as Lemma 3.7, let $z \in \mathcal{N}^{-\infty}(\theta)$ *with* $\overline{\mu}(z) = \mu$, *and* $\Delta z^{\mathrm{TR}} = (\Delta x^{\mathrm{TR}}, \Delta s^{\mathrm{TR}})$ *be the trust region direction at* z *computed by* $\mathrm{WTR}(B, N)$. *Then for* $\alpha^{\mathrm{TR}} = \frac{1}{1+2\varepsilon} \in [0, 1]$ *where*

$$\varepsilon := \max\left\{\left\|\mathbf{1}_N + \frac{\Delta x_N^{\mathrm{TR}}}{x_N}\right\|_\infty, \left\|\mathbf{1}_B + \frac{\Delta s_B^{\mathrm{TR}}}{s_B}\right\|_\infty\right\},$$

$\overline{\mu}(z + \alpha^{\mathrm{TR}} \Delta z^{\mathrm{TR}}) \leq \frac{34n^4}{(1-\theta)^4} \mu_1$ *and* $z + \alpha^{\mathrm{TR}} \Delta z^{\mathrm{TR}} \in \mathcal{N}^{-\infty}\left(1 - \frac{(1-\theta)^5}{144n^4}\right)$.

Proof. For the primal program, we use the variable substitution

$$y = \left(\frac{x_N + \Delta x_N}{x_N}, \frac{2}{u - \ell}\left(\frac{u + \ell}{2} \mathbf{1}_B - \frac{x_B + \Delta x_B}{x_B}\right)\right),$$

and accordingly define the system $\mathbf{B} y = \bar{b}$ with

$$\mathbf{B} = \left(\mathbf{A}_N \mathrm{diag}(x_N), \frac{\ell - u}{2} \mathbf{A}_B \mathrm{diag}(x_B)\right), \bar{b} = \mathbf{A}_N x_N - \frac{u + \ell - 2}{2} \mathbf{A}_B x_B\,,$$

and let $I = B$, $J = N$. Then, $\min \|y_J\|_\infty$ s.t. $\mathbf{B} y = \bar{b}$, $\|y_I\|_\infty \leq 1$ is equivalent to the primal system, and the solution can be easily transformed to a solution Δx. Setting $\delta = \frac{(1-\theta)^4}{16n^4}$, a δ-optimal solution to this system can be computed using Theorem 1.10, and it satisfies the required error bounds by following Lemma 3.6. For the second part of the statement, we use the bounds obtained from the first part throughout the proof of Lemma 3.7. □

Theorem 1.3 (Path Following in Current Matrix Multiplication Time)**.** *For any* $\theta \in [1/8, 1)$, *given* $\mu_0 > \mu_1 \geq 0$ *and a starting point* $z_0 \in \mathcal{N}^{-\infty}(\theta)$ *with* $\overline{\mu}(z_0) \in [\mu_1, \mu_0]$, *the algorithm* TRW-IPM *is* $O(1)$*-instance optimal in* $\mathcal{N}^{-\infty}(\theta')$ *with respect to* $\mathcal{N}^{-\infty}(\theta)$, *for* $\theta' = 1 - ((1-\theta)/n)^{O(1)}$. *The algorithm finds a solution* $z_1 \in \overline{\mathcal{N}}^{-\infty}(\theta)$ *with* $\overline{\mu}(z_1) \leq \mu_1$ *in randomized runtime* $n^{\tilde{\omega}+o(1)} \log\left(\frac{1}{1-\theta}\right) \mathrm{SLC}(\mathcal{N}^{-\infty}(\theta), \mu_1, \mu_0)$, *or deterministic runtime* $O\left(\left(n^3 + n^{\tilde{\omega}+o(1)} \log\left(\frac{1}{1-\theta}\right)\right) \mathrm{SLC}(\mathcal{N}^{-\infty}(\theta), \mu_1, \mu_0)\right)$.

Proof. By Proposition 3.9, all iterates of the algorithm stay in $\mathcal{N}^{-\infty}(\theta')$ for $\theta' = 1 - (1-\theta)^4/(144n^3)$; note that $\log(1/(1-\theta')) = O(\log(n/(1-\theta)))$. After each run of PathFollow, the current iterate is in $\mathcal{N}^{-\infty}(1/8) \subseteq \mathcal{N}^{-\infty}(\theta)$ by the assumption $\theta \geq 1/8$.

It suffices to show that, given any straight line segment $[z_1', z_0'] \subseteq \mathcal{N}^{-\infty}(\theta)$ with $\overline{\mu}(z_0') = \mu_0'$ and $\overline{\mu}(z_1') = \mu_1'$ where $\mu_1 \le \mu_1' \le \mu_0' \le \mu_0$, after the first iteration of Algorithm 1 which gives an iterate $\bar{z}$ with $\overline{\mu}(\bar{z}) = \bar{\mu} \le \mu_0'$, an iterate $z^+ = (x^+, s^+)$ with $\overline{\mu}(z^+) \le \mu_1'$ is reached within at most three iterations.

Every iteration reduces the gap by at least a factor $\varrho = \frac{(1-\theta)^9}{256n^9}$. Hence, the claim is immediate if $\mu_1'/\mu_0' \ge \varrho^4$. Otherwise, consider the first iteration of PathFollow subsequent to the iterate $\bar{z}$, which gives the current iterate z. Since the straight line $[z_1', z_0']$ is $\frac{1-\theta}{4n}$-polarized by Lemma 3.3, the conditions of Lemma 3.8 hold: $\bar{\mu} \le \varrho\mu_0' < \frac{(1-\theta)^3}{4n}\mu_0'$ and $\overline{\mu}(z) \le \varrho\bar{\mu} = \frac{(1-\theta)^9}{256n^9}\bar{\mu}$. Hence, $(\overline{B}, \overline{N})$ identified in the algorithm coincides with the polarized partition (B, N) of the line segment $[z_0', z_1']$.

By Proposition 3.9, $\overline{\mu}(z + \alpha^{\mathrm{TR}}\Delta z^{\mathrm{TR}}) \le \frac{34n^4}{(1-\theta)^4}\mu_1$. By line 10 of Algorithm 1, we take this step if $\overline{\mu}(z + \Delta z^{\mathrm{TR}}) \le \varrho\overline{\mu}(z)$. In this case, taking z^+ to be the subsequent PathFollow iterate, we get $\overline{\mu}(z^+) \le \varrho \cdot \frac{34n^4\mu_1'}{(1-\theta)^4} < \mu_1'$. If we do not take this step, then instead we use a PathFollow iterate by line 13 of the algorithm, which has normalized duality gap at most $\frac{34n^4}{(1-\theta)^4}\mu_1'$. Again taking z^+ to be its subsequent PathFollow iterate, we also get $\overline{\mu}(z^+) \le \varrho \cdot \frac{34n^4\mu_1'}{(1-\theta)^4} < \mu_1'$. Thus we obtain an iterate z^+ with $\overline{\mu}(z^+) \le \mu_1'$ in at most four iterations for $[z_1', z_0']$.

By Lemma 2.9, the straight line segments stay in $\mathcal{N}^{-\infty}(\theta')$, where $\theta' = 1 - ((1-\theta)/n)^{O(1)}$, and the running time bound is obtained using the bounds in Theorem 2.23 and Proposition 3.9. □

4 The ℓ_2-Trust Region Interior Point Method

In Section 4.1, we give a formal description of the ℓ_2-Trust Region IPM algorithm. Sections 4.2 and 4.3 prove Lemma 1.11 on near-optimality of the trust region step for long segments.

4.1 Description of the algorithm

To start with, we restate the definition of trust region direction $\Delta z^{\mathrm{TR}} = (\Delta x^{\mathrm{TR}}, \Delta s^{\mathrm{TR}})$ for convenience.

Definition 4.1. Let $z = (x, s) \in \mathcal{N}^2(\beta)$, $B \cup N = [n]$ be a partition and $\gamma > 0$. We define the (B, N, γ)-ℓ_2-trust region direction at z as the optimal solution $(\Delta x^{\mathrm{TR}}, \Delta s^{\mathrm{TR}}, \Delta y^{\mathrm{TR}})$ of the following primal and dual trust region problem:

$$\begin{aligned}
&\min \left\| \mathbf{1}_N + \frac{\Delta x_N}{x_N} \right\| && \min \left\| \mathbf{1}_B + \frac{\Delta s_B}{s_B} \right\| \\
&\text{s.t.} \left\| \frac{\Delta x_B}{x_B} \right\| \le \gamma && \text{s.t.} \left\| \frac{\Delta s_N}{s_N} \right\| \le \gamma \\
&\mathbf{A}\Delta x = \mathbf{0} && \mathbf{A}^\top \Delta y + \Delta s = \mathbf{0}
\end{aligned} \tag{$\mathrm{TR}_2(B,N,\gamma)$}$$

The trust region constraints are given in the local norms of the search direction at z. By definition, the primal trust region direction achieves a maximal multiplicative decrease on the coordinates in N subject to the condition that the coordinates in B can only move within the imposed trust-region radius, as measured in the local ℓ_2-norm. The dual trust-region direction achieves the same on the dual side with the role of N and B swapped.

The optimal solution to $\mathrm{TR}_2(B, N, \gamma)$ exists and can be uniquely defined. By strong convexity, the vectors Δx_N (resp. Δs_B) are always unique. If Δx_B (resp. Δx_N) is not unique, then it follows that the primal (resp. dual) program has an optimal solution where the trust region constraint is not binding. In this case, we pick the unique solution among the optimal ones where $\|\Delta x_B/x_B\|$ (resp. $\|\Delta s_N/s_N\|$) is minimal; see [40] and our Definition 7.1. We note this corresponds to solving an instance of layered least squares problem in [73].

Proposition 4.2. *$\mathrm{TR}_2(B, N, \gamma)$ can be reformulated into two instances of the trust region problem (TR-2).*

Proof. We only consider the primal trust region problem. Let $\mathbf{P} \in \mathbb{R}^{n\times n}$ be the column permutation matrix such that $\mathbf{AP} = \begin{bmatrix} \mathbf{A}_B & \mathbf{A}_N \end{bmatrix}$. Then taking $I = B$, $J = N$, $y_I = \frac{\Delta x_B}{\gamma x_B}$, $y_J = \mathbf{1}_N + \frac{\Delta x_N}{x_N}$, $b = \mathbf{A}_N x_N$ and $\mathbf{B} = \mathbf{AP}\begin{bmatrix} \operatorname{diag}(\gamma x_B) & \\ & \operatorname{diag}(x_N) \end{bmatrix}$, we reach the form of (TR-2). □

Algorithm 2: TR2-IPM

Input : An instance of (LP) with constraint matrix $\mathbf{A} \in \mathbb{R}^{m\times n}, \mathrm{rk}(\mathbf{A}) = m$, and initial iterate $(x^0, s^0) \in \mathcal{P} \times \mathcal{D}$, $\beta \in (0, 1/12)$, $\mu_1 \geq 0$.

Output: $(x, s) \in \overline{\mathcal{N}}^2(\beta)$ satisfying $\overline{\mu}(x, s) \leq \mu_1$

1 $(x, s) \leftarrow (x^0, s^0)$;
2 **while** $\overline{\mu}(x, s) > \mu_1$ **do**
3   **while** $\left\|\frac{xs}{\overline{\mu}(x,s)} - \mathbf{1}\right\|^2 > \beta^2$ **do**
4   Compute $(\Delta x^{\mathrm{c}}, \Delta s^{\mathrm{c}})$ at (x, s) according to Proposition 2.11;
5         $(x, s) \leftarrow (x + \Delta x^{\mathrm{c}}, s + \Delta s^{\mathrm{c}})$;
6   Compute $(\Delta x^{\mathrm{a}}, \Delta s^{\mathrm{a}})$ at (x, s) according to Proposition 2.11;
7   **if** $\frac{81}{16}\left\|\frac{\Delta x^{\mathrm{a}}\Delta s^{\mathrm{a}}}{\overline{\mu}(x,s)}\right\|^2 \leq 80^2\beta^2$ `// check if affine scaling step length is at least 3/4`
8   **then**
9         $\widetilde{B}_z \leftarrow \left\{i \in [n] : \left|\frac{\Delta x_i^{\mathrm{a}}}{x_i}\right| \leq \left|\frac{\Delta s_i^{\mathrm{a}}}{s_i}\right|\right\}$, $\widetilde{N}_z \leftarrow [n] \setminus \widetilde{B}_z$;
10   Compute $(\Delta x^{\mathrm{TR}}, \Delta s^{\mathrm{TR}})$ at (x, s) by solving $\mathrm{TR}_2(\widetilde{B}_z, \widetilde{N}_z, 12\beta)$ according to Theorem 1.9 with $\delta = \frac{1}{64}$;
11   Compute $\alpha^{\mathrm{TR}} \in (0, 1]$ for $(\Delta x^{\mathrm{TR}}, \Delta s^{\mathrm{TR}})$ according to Lemma 4.7 ;
12        $(x, s) \leftarrow (x + \alpha^{\mathrm{TR}}\Delta x^{\mathrm{TR}}, s + \alpha^{\mathrm{TR}}\Delta s^{\mathrm{TR}})$;
13   **else**
14   Compute $\alpha^{\mathrm{a}} \in (0, 1]$ for $(\Delta x^{\mathrm{a}}, \Delta s^{\mathrm{a}})$ according to Lemma 5.5;
15        $(x, s) \leftarrow (x + \alpha^{\mathrm{a}}\Delta x^{\mathrm{a}}, s + \alpha^{\mathrm{a}}\Delta s^{\mathrm{a}})$;
16 **return** (x, s);

Description of ℓ_2-TR-IPM algorithm. We now formally describe our ℓ_2-Trust Region Interior Point algorithm. We take a starting iterate (x^0, s^0) of primal and dual feasible solutions. While $\overline{\mu}(x, s) > \mu_1$, we repeat the following. First, we perform corrector steps until reaching $\mathcal{N}^2(\beta)$. The number of these corrector steps depends on (x^0, s^0) in the very first iteration; in all subsequent iterations, this will be bounded as $O(1)$ as we need to move from $\mathcal{N}^2(82\beta)$ back to $\mathcal{N}^2(\beta)$.

Once we reached $(x, s) \in \mathcal{N}^2(\beta)$, the algorithm starts with computing the affine scaling direction Δz^{a} at z. If it is possible to move by a step length of at least $3/4$ in this affine scaling direction (checked by line 7), then we identify the associated partition $\widetilde{B}_z \cup \widetilde{N}_z = [n]$ and hence compute the corresponding trust region direction Δz^{TR} by solving $\mathrm{TR}(\widetilde{B}_z, \widetilde{N}_z, 12\beta)$. To follow, a sufficiently long trust region step length $\alpha^{\mathrm{TR}} \in (0, 1]$ is computed, and then a trust region step $z + \alpha^{\mathrm{TR}}\Delta z^{\mathrm{TR}}$ is taken. For such a step, α^{TR} and Δz^{TR} guarantee $z + \alpha^{\mathrm{TR}}\Delta z^{\mathrm{TR}} \in \mathcal{N}^2(82\beta)$. If we cannot move by step length of $3/4$ in the affine scaling direction, the algorithm instead computes an affine scaling step length $\alpha^{\mathrm{a}} \in (0, 3/4]$; again, the step $z + \alpha^{\mathrm{a}}\Delta z^{\mathrm{a}}$ remains in $\mathcal{N}^2(82\beta)$. We then move on to the next iteration.

The step formulae for affine scaling and corrector step are given in Proposition 2.11. The trust region direction is computed from solving $\mathrm{TR}(\widetilde{B}_z, \widetilde{N}_z, 12\beta)$ using a strongly polynomial subroutine which we discuss in Section 7. We give a procedure on how to compute the step lengths α^{a} and α^{TR} that satisfy the mentioned properties in strongly polynomial time in Lemma 4.7 and Lemma 5.5 respectively.

4.2 ℓ_2-polarization and trust region direction

Towards proving Lemma 1.11, we now formally derive the phenomenon of ℓ_2-polarization exhibited by the long straight line in $\mathcal{N}^2(\beta)$ between z_0 with $\overline{\mu}(z) = \overline{\mu}(z_0) = \mu$ and $\overline{\mu}(z_1) = \mu_1 \leq \mu/4$, as assumed in the Lemma. Our ultimate goal is to analyze the trust region step with respect to the associated partition $(\widetilde{B}_z, \widetilde{N}_z)$ at the current iterate z. First, we show the existence of *'polarized'* partition (B, N) that one can derive from the *'ideal direction'* pointing from z to the end of the assumed straight line segment z_1. We subsequently show that these two partitions coincide (Lemma 4.6). We formally define these concepts as follows.

Definition 4.3 (Polarization and ideal direction)**.** Suppose there exists some $z_0 \in \mathcal{N}^2(\beta)$ with $\overline{\mu}(z_0) = \mu$ and $z_1 \in \mathcal{N}^2(\beta)$ with $\overline{\mu}(z_1) = \mu_1 \leq \mu/4$ such that $[z_1, z_0] \subseteq \mathcal{N}^2(\beta)$. We let $\Delta z_0 = (\Delta x_0, \Delta s_0) := (x_1 - x_0, s_1 - s_0)$ and define the *polarized partition* to be

$$N := \left\{ i \in [n] : \left| \frac{\Delta x_{0i}}{x_{0i}} \right| \geq \left| \frac{\Delta s_{0i}}{s_{0i}} \right| \right\} , \, B := [n] \setminus N \, .$$

For any $z \in \mathcal{N}^2(\beta)$ with $\overline{\mu}(z) = \mu$, we define the *ideal direction* to be $\Delta z^{\mathrm{id}} = (\Delta x^{\mathrm{id}}, \Delta s^{\mathrm{id}}) := (x_1 - x, s_1 - s)$. Further, we define the *rescaled ideal direction* to be

$$\Delta z^{\mathrm{Rid}} = (\Delta x^{\mathrm{Rid}}, \Delta s^{\mathrm{Rid}}) := \frac{\mu}{\mu - \mu_1}(x_1 - x, s_1 - s) \, . \tag{32}$$

We note that the rescaling normalizes Δz^{id} in a way that for $\alpha \in [0,1]$, $\overline{\mu}(z + \alpha \Delta z^{\mathrm{Rid}}) = \mu(1-\alpha)$. We now show that Δz^{Rid} is feasible to $\mathrm{TR}_2(B, N, 12\beta)$ at z, with the objective value being $O(\beta\mu_1/\mu)$.

Lemma 4.4. *For $\beta \in (0, 1/128]$, suppose there exist some $z_0 \in \mathcal{N}^2(\beta)$ with $\overline{\mu}(z_0) = \mu$ and $z_1 \in \mathcal{N}^2(\beta)$ with $\overline{\mu}(z_1) = \mu_1 \leq \mu/4$ such that $[z_1, z_0] \subseteq \mathcal{N}^2(\beta)$. Let (B, N) be the polarized partition, and $z \in \mathcal{N}^2(\beta)$ with $\overline{\mu}(z) = \mu$.*

(i) $(\Delta x^{\mathrm{Rid}}, \Delta s^{\mathrm{Rid}})$ is feasible to $\mathrm{TR}(B, N, 12\beta)$ at z with

$$\max \left\{ \left\| \frac{\Delta x_B^{\mathrm{Rid}}}{x_B} \right\| , \left\| \frac{\Delta s_N^{\mathrm{Rid}}}{s_N} \right\| \right\} \leq 12\beta \, . \tag{33}$$

(ii) The objective value attained by $(\Delta x^{\mathrm{Rid}}, \Delta s^{\mathrm{Rid}})$ in $\mathrm{TR}(B, N, 12\beta)$ at z satisfies

$$\max \left\{ \left\| \mathbf{1}_N + \frac{\Delta x_N^{\mathrm{Rid}}}{x_N} \right\| , \left\| \mathbf{1}_B + \frac{\Delta s_B^{\mathrm{Rid}}}{s_B} \right\| \right\} \leq 16\beta\mu_1/\mu \, . \tag{34}$$

For the proof, recall the notation $\Delta \bar{x}_i = \Delta x_i / x_i$ and $\Delta \bar{x}_i = \Delta x_i / x_i$, which we accordingly use also for $\Delta \bar{x}_{0i}, \Delta \bar{s}_{0i}, \Delta \bar{x}_{1i}, \Delta \bar{s}_{1i}$. Further, recall $\hat{\xi} = \hat{\xi}(z) = \sqrt{\frac{xs}{\mu}}$. Throughout, we denote $\hat{\xi}_1 := \hat{\xi}(z_1) = \sqrt{\frac{x_1 s_1}{\mu_1}}$. A key tool in the proof is to analyze the expressions $(1 + \Delta \bar{x}_{0i})(1 + \Delta \bar{s}_{0i}) = x_{1i}s_{1i}/(x_{0i}s_{0i}) \approx \mu_1/\mu_0 < 1/4$ for $i \in [n]$. We will further show that $|\Delta \bar{x}_{0i} \Delta \bar{s}_{0i}|$ is small. From these bounds, we derive strong upper bounds on $\min\{|\Delta \bar{x}_{0i}|, |\Delta \bar{s}_{0i}|\}$.

Proposition 4.5. *Suppose $a, b \in \mathbb{R}$ satisfy $|a| \geq |b|$ and $a + b \leq 0$. Let $\gamma := ab$ and $\nu := (1+a)(1+b)$. If $\nu \geq 0$ and $\gamma < 1$, then $(\nabla |a|)(\nu, \gamma) \leq \mathbf{0}$. Moreover, if $\nu \in \left[0, \frac{5}{18}\right]$ and $|\gamma| \leq \frac{1}{12}$, then $|b| \leq \frac{3}{2}|\gamma|$.*

Proof. We have $(1+a)(1+\gamma/a) = \nu$, giving a quadratic equation $a^2 + (1 - \nu + \gamma)a + \gamma = 0$. Solving it yields

$$a = \frac{-(1-\nu+\gamma) \pm \sqrt{(1-\nu+\gamma)^2 - 4\gamma}}{2} \quad \text{and} \quad b = \frac{-(1-\nu+\gamma) \mp \sqrt{(1-\nu+\gamma)^2 - 4\gamma}}{2} \, . \tag{35}$$

Since $1 - \nu + \gamma = -a - b \geq 0$ and $|a| \geq |b|$, we have $|a| = \frac{1-\nu+\gamma+\sqrt{(1-\nu+\gamma)^2-4\gamma}}{2}$. Then,

$$\frac{\partial |a|}{\partial \nu} = -\frac{1}{2} - \frac{1 - \nu + \gamma}{2\sqrt{(1-\nu+\gamma)^2 - 4\gamma}} < 0 \, .$$

Also, as $\nu \geq 0$, $(1 + \nu - \gamma)^2 \geq (1 - \nu + \gamma)^2 - 4\gamma$. This together with $1 + \nu - \gamma > 0$ give

$$\frac{\partial |a|}{\partial \gamma} = \frac{1}{2} - \frac{1 + \nu - \gamma}{2\sqrt{(1-\nu+\gamma)^2 - 4\gamma}} \leq 0 \, .$$

As a result, if $\nu \leq \frac{5}{18}$ and $\gamma \leq \frac{1}{12}$, then $|a| \geq \frac{1}{2}\left(\frac{13}{18} + \frac{1}{12} + \sqrt{\left(\frac{13}{18} + \frac{1}{12}\right)^2 - \frac{1}{3}}\right) > \frac{2}{3}$. Hence, $|b| = \frac{|\gamma|}{|a|} \leq \frac{3}{2}|\gamma|$. ☐

Proof of Lemma 4.4. **Part (i).** By Lemma 2.15, $\kappa(z_1, z_0) = \frac{\|\Delta x_0 \Delta s_0\|}{(\sqrt{\mu}+\sqrt{\mu_1})^2} \leq 2\beta$. Using Proposition 2.5 and the assumption $\mu_1/\mu \leq 1/4$, we get

$$\|\Delta \bar{x}_0 \Delta \bar{s}_0\| \leq \frac{2\beta}{1-\beta}\left(1 + \sqrt{\frac{\mu_1}{\mu}}\right)^2 \leq \frac{9\beta}{2(1-\beta)} < \frac{1}{12} \, .$$

Looking at each coordinate $i \in [n]$ of $(x_0 + \Delta x_0)(s_0 + \Delta s_0) = x_1 s_1$, where $\frac{x_1 s_1}{x_0 s_0} \le \frac{1+\beta}{1-\beta} \cdot \frac{\mu_1}{\mu} \cdot \mathbf{1}_n$, we get

$$0 \le (1 + \Delta \bar{x}_{0i})(1 + \Delta \bar{s}_{0i}) = \frac{x_{1i} s_{1i}}{x_{0i} s_{0i}} \le \frac{1+\beta}{1-\beta} \cdot \frac{1}{4} \le \frac{5}{18} .$$

Let $i \in N$ and $\gamma_i := \Delta \bar{x}_{0i} \Delta \bar{s}_{0i}$. The conditions in Proposition 4.5 hold and thus we obtain $|\Delta \bar{s}_{0i}| \le \frac{3}{2}|\gamma_i|$. Hence, $\|\Delta \bar{s}_{0N}\| \le \frac{3}{2}\|\Delta \bar{x}_{0N} \Delta \bar{s}_{0N}\| \le \frac{27\beta}{4(1-\beta)}$. Using Lemma 2.19,

$$\begin{aligned}
\left\| \frac{\Delta s_N^{\mathrm{id}}}{s_N} \right\| &= \left\| \frac{s_{1N} - s_{0N} + s_{0N} - s_N}{s_N} \right\| \\
&\le \left\| \frac{s_{1N} - s_{0N}}{s_{0N}} \cdot \frac{s_{0N}}{s_N} \right\| + \left\| \frac{s_{0N} - s_N}{s_N} \right\| \\
&\le \left\| \frac{s_{0N}}{s_N} \right\|_\infty \cdot \|\Delta \bar{s}_{0N}\| + \left\| \frac{s_{0N} - s_N}{s_N} \right\| \\
&\le \frac{27\beta}{4(1-\beta)^3} + \frac{2\beta}{(1-\beta)^2} \le 9\beta .
\end{aligned} \tag{36}$$

Therefore, $\left\| \frac{\Delta s_N^{\mathrm{Rid}}}{s_N} \right\| = \frac{\mu}{\mu - \mu_1} \left\| \frac{\Delta s_N^{\mathrm{id}}}{s_N} \right\| \le 12\beta$ and the same bound holds for $\left\| \frac{\Delta x_B^{\mathrm{Rid}}}{x_B} \right\|$ by an analogous derivation.

Part (ii). Using $\frac{x_1 s_1}{xs} = \frac{\mu_1 \hat{\xi}_1^2}{\mu \hat{\xi}^2}$, we first have

$$\begin{aligned}
\left\| \mathbf{1}_N + \frac{\Delta x_N^{\mathrm{Rid}}}{x_N} \right\| &= \left\| \mathbf{1}_N + \frac{\mu}{\mu - \mu_1} \frac{x_{1N} - x_N}{x_N} \right\| \\
&= \left\| \frac{\mu}{\mu - \mu_1} \cdot \frac{x_{1N}}{x_N} - \frac{\mu_1}{\mu - \mu_1} \mathbf{1}_N \right\| \\
&= \frac{\mu_1}{\mu - \mu_1} \left\| \frac{\hat{\xi}_{1N}^2}{\hat{\xi}_N^2} \cdot \frac{s_N}{s_{1N}} - \mathbf{1}_N \right\| \\
&\le \frac{4\mu_1}{3\mu} \left(\left\| \frac{\hat{\xi}_{1N}^2}{\hat{\xi}_N^2} \left(\frac{s_N}{s_{1N}} - \mathbf{1}_N \right) \right\| + \left\| \frac{\hat{\xi}_{1N}^2}{\hat{\xi}_N^2} - \mathbf{1}_N \right\| \right) .
\end{aligned} \tag{37}$$

The bound $\|\Delta s_N^{\mathrm{id}}/s_N\| \le 9\beta$ shown in (36) gives $(1 - 9\beta)\mathbf{1}_N \le s_{1N}/s_N \le (1 + 9\beta)\mathbf{1}_N$. Then, using also Proposition 2.14,

$$\left\| \frac{\hat{\xi}_{1N}^2}{\hat{\xi}_N^2} \left(\frac{s_N}{s_{1N}} - \mathbf{1}_N \right) \right\| = \left\| \frac{\hat{\xi}_{1N}^2}{\hat{\xi}_N^2} \left(\frac{s_{1N} - s_N}{s_N} \cdot \frac{s_N}{s_{1N}} \right) \right\| \le \left\| \frac{\hat{\xi}_1^2}{\hat{\xi}^2} \right\|_\infty \left\| \frac{s_N}{s_{1N}} \right\|_\infty \left\| \frac{\Delta s_N^{\mathrm{id}}}{s_N} \right\| \le \frac{9\beta(1+\beta)}{(1-9\beta)(1-\beta)} .$$

Again by Proposition 2.14, $\left\| \frac{\hat{\xi}_{1N}^2}{\hat{\xi}_N^2} - \mathbf{1}_N \right\| \le \frac{2\beta}{1-\beta}$. Therefore, following (37), we obtain

$$\left\| \mathbf{1}_N + \frac{\Delta x_N^{\mathrm{Rid}}}{x_N} \right\| \le \frac{4\mu_1}{3\mu} \left(\frac{9\beta(1+\beta)}{(1-9\beta)(1-\beta)} + \frac{2\beta}{1-\beta} \right) \le 16\beta \cdot \frac{\mu_1}{\mu} .$$

The same bound holds for $\left\| \mathbf{1}_B + \frac{\Delta s_B^{\mathrm{Rid}}}{s_B} \right\|$ by an analogous derivation. □

The above result shows that the maximum possible degree of cancellation for primal variables in N and dual variables in B achieved by the trust region direction of $\mathrm{TR}_2(B, N, 12\beta)$ satisfies (34). As we will see in Lemma 4.7, this is important in deriving the suitably small size of the residual and curvature of the trust region direction.

We recall that only the existence of the long straight line segment in $\mathcal{N}^2(\beta)$ from z_0 to z_1 is assumed, hence we do not have the access to the polarized partition at z_0. However, the next lemma asserts that under the assumption of the long straight line segment, the polarized partition can be revealed by the associated partition of the affine scaling direction at any $z \in \mathcal{N}^2(\beta)$ with $\overline{\mu}(z) = \overline{\mu}(z_0)$. As a result, $\mathrm{TR}_2(B, N, 12\beta)$ coincides with $\mathrm{TR}_2(\widetilde{B}_z, \widetilde{N}_z, 12\beta)$, which is the key to the computability of trust region direction. We first recall that the normalized rescaled ideal direction is defined as

$$(\Delta \hat{x}^{\mathrm{Rid}}, \Delta \hat{s}^{\mathrm{Rid}}) := \left(\sqrt{\frac{s}{x\mu}} \Delta x^{\mathrm{Rid}}, \sqrt{\frac{x}{s\mu}} \Delta s^{\mathrm{Rid}} \right) .$$

Lemma 4.6. *Under the same setting of Lemma 4.4, the polarized partition (B, N) coincides with the associated partition $(\widetilde{B}_z, \widetilde{N}_z)$ at z.*

Proof. Using Proposition 2.14 and the previous two lemmas, we first have

$$\begin{aligned}
\left\|\Delta\hat{x}^{\mathrm{a}} - \Delta\hat{x}^{\mathrm{Rid}} + \Delta\hat{s}^{\mathrm{a}} - \Delta\hat{s}^{\mathrm{Rid}}\right\|^2 &= \left\|\hat{\xi} + \Delta\hat{x}^{\mathrm{Rid}} + \Delta\hat{s}^{\mathrm{Rid}}\right\|^2 \\
&= \left\|\hat{\xi}_N + \Delta\hat{x}_N^{\mathrm{Rid}} + \Delta\hat{s}_N^{\mathrm{Rid}}\right\|^2 + \left\|\hat{\xi}_B + \Delta\hat{x}_B^{\mathrm{Rid}} + \Delta\hat{s}_B^{\mathrm{Rid}}\right\|^2 \\
&\le \left(\left\|\hat{\xi}_N + \Delta\hat{x}_N^{\mathrm{Rid}}\right\| + \left\|\Delta\hat{s}_N^{\mathrm{Rid}}\right\|\right)^2 + \left(\left\|\hat{\xi}_B + \Delta\hat{s}_B^{\mathrm{Rid}}\right\| + \left\|\Delta\hat{x}_B^{\mathrm{Rid}}\right\|\right)^2 \\
&\le (1+\beta)\left(\left\|\mathbf{1}_N + \Delta\bar{x}_N^{\mathrm{Rid}}\right\| + \left\|\Delta\bar{s}_N^{\mathrm{Rid}}\right\|\right)^2 + (1+\beta)\left(\left\|\mathbf{1}_B + \Delta\bar{s}_B^{\mathrm{Rid}}\right\| + \left\|\Delta\bar{x}_B^{\mathrm{Rid}}\right\|\right)^2 \\
&\le 2(1+\beta)\left(16\beta\mu_1/\mu + 12\beta\right)^2 .
\end{aligned}$$

Using the orthogonality of $\Delta\hat{x}^{\mathrm{a}} - \Delta\hat{x}^{\mathrm{Rid}}$ and $\Delta\hat{s}^{\mathrm{a}} - \Delta\hat{s}^{\mathrm{Rid}}$ and $\left\|\hat{\xi}^{-1}\right\|_\infty \le \frac{1}{\sqrt{1-\beta}}$, we have

$$\begin{aligned}
\left\|\Delta\bar{x}^{\mathrm{a}} - \Delta\bar{x}^{\mathrm{Rid}}\right\|^2 + \left\|\Delta\bar{s}^{\mathrm{a}} - \Delta\bar{s}^{\mathrm{Rid}}\right\|^2 &\le \left\|\hat{\xi}^{-1}\right\|_\infty^2 \left(\left\|\Delta\hat{x}^{\mathrm{a}} - \Delta\hat{x}^{\mathrm{Rid}}\right\|^2 + \left\|\Delta\hat{s}^{\mathrm{a}} - \Delta\hat{s}^{\mathrm{Rid}}\right\|^2\right) \\
&= \left\|\hat{\xi}^{-1}\right\|_\infty^2 \cdot \left\|\Delta\hat{x}^{\mathrm{a}} - \Delta\hat{x}^{\mathrm{Rid}} + \Delta\hat{s}^{\mathrm{a}} - \Delta\hat{s}^{\mathrm{Rid}}\right\|^2 \\
&\le \frac{2(1+\beta)}{1-\beta}\left(16\beta\mu_1/\mu + 12\beta\right)^2 .
\end{aligned}$$

which gives

$$\max\left\{\left\|\Delta\bar{x}^{\mathrm{a}} - \Delta\bar{x}^{\mathrm{Rid}}\right\|, \left\|\Delta\bar{s}^{\mathrm{a}} - \Delta\bar{s}^{\mathrm{Rid}}\right\|\right\} \le \sqrt{\frac{2(1+\beta)}{1-\beta}}\left(16\beta\mu_1/\mu + 12\beta\right) \le 23\beta .$$

As a result, by Lemma 4.4,

$$\left\|\Delta\bar{x}_B^{\mathrm{a}}\right\| \le \left\|\Delta\bar{x}_B^{\mathrm{a}} - \Delta\bar{x}_B^{\mathrm{Rid}}\right\| + \left\|\Delta\bar{x}_B^{\mathrm{Rid}}\right\| \le 23\beta + 12\beta = 35\beta$$

and we can derive the same bound analogously for $\|\Delta\bar{s}_N^{\mathrm{a}}\|$. Finally, we recall that $\Delta\bar{x}^{\mathrm{a}} + \Delta\bar{s}^{\mathrm{a}} = -\mathbf{1}_n$, hence for any $i \in B$, using also the assumption $\beta \le 1/128$, we have

$$|\Delta\bar{x}_i^{\mathrm{a}}| \le 35\beta < 1 - 35\beta \le 1 - |\Delta\bar{x}_i^{\mathrm{a}}| \le |1 + \Delta\bar{x}_i^{\mathrm{a}}| = |\Delta\bar{s}_i^{\mathrm{a}}| .$$

By a symmetric argument, we also have $|\Delta\bar{s}_i^{\mathrm{a}}| \le |\Delta\bar{x}_i^{\mathrm{a}}|$ for any $i \in N$. Therefore, $(B, N) = (\widetilde{B}_z, \widetilde{N}_z)$. ☐

4.3 Near-optimality of the ℓ_2-Trust Region Step

We are now ready to establish the key component to prove Lemma 1.12, which is that the trust region direction of $\mathrm{TR}_2(\widetilde{B}_z, \widetilde{N}_z, 12\beta)$ has $O(\beta\mu_1/\mu)$ residual and $O(\beta)$ curvature local to z (defined below). Based on this, we can find a step length $\bar{\alpha}^{\mathrm{TR}} \in (0, 1]$ that satisfies both $\overline{\mu}(z + \bar{\alpha}^{\mathrm{TR}}\Delta z^{\mathrm{TR}}) \le \mu_1$ because of the small residual, and $z + \bar{\alpha}^{\mathrm{TR}}\Delta z^{\mathrm{TR}} \in \mathcal{N}^2(O(\beta))$ because of the small curvature.

On the other hand, the exact $\bar{\alpha}^{\mathrm{TR}}$ requires ℓ_2-norm computation. We will make use of $\|\cdot\|_\infty \le \|\cdot\| \le \|\cdot\|_1 \le n\|\cdot\|_\infty$ and our strongly polynomial binary search procedure (Algorithm 5) to compute a step length $\alpha^{\mathrm{TR}} \in (0, 1]$ that is close enough to $\bar{\alpha}^{\mathrm{TR}}$. This computed α^{TR} is used in our TR2-IPM algorithm instead. It is slightly greater than $\bar{\alpha}^{\mathrm{TR}}$ to ensure $\overline{\mu}(z + \alpha^{\mathrm{TR}}\Delta z^{\mathrm{TR}}) \le \mu_1$ while $z + \alpha^{\mathrm{TR}}\Delta z^{\mathrm{TR}}$ ends up in a slightly larger ℓ_2-neighborhood.

Lemma 4.7. *For $\beta \in (0, 1/128]$, suppose there exist some $z_0 \in \mathcal{N}^2(\beta)$ with $\overline{\mu}(z_0) = \mu$ and $z_1 \in \mathcal{N}^2(\beta)$ with $\overline{\mu}(z_1) = \mu_1 \le \mu/4$ such that $[z_1, z_0] \subseteq \mathcal{N}^2(\beta)$. For any $z = (x, s) \in \mathcal{N}^2(\beta)$ with $\overline{\mu}(z) \le \mu$, let $\Delta z^{\mathrm{TR}} = \left(\Delta x^{\mathrm{TR}}, \Delta s^{\mathrm{TR}}\right)$ be the trust-region direction at z of $\mathrm{TR}_2(\widetilde{B}_z, \widetilde{N}_z, 12\beta)$.*

(i) We have the following bounds on the residual and curvature of a full trust region step local to z:

$$r := \left\|\frac{(x + \Delta x^{\mathrm{TR}})(s + \Delta s^{\mathrm{TR}})}{xs}\right\| \le 25\beta\mu_1/\mu \quad \text{and} \quad v := \left\|\frac{\Delta x^{\mathrm{TR}}\Delta s^{\mathrm{TR}}}{xs}\right\| \le 18\beta . \tag{38}$$

(ii) *It takes $O(\log\log n)$ many arithmetic operations in strongly polynomial time to compute a trust region step length $\alpha^{\mathrm{TR}} \in (0,1]$ that satisfies $z + \alpha^{\mathrm{TR}}\Delta z^{\mathrm{TR}} \in \mathcal{N}^2(82\beta)$ and $\overline{\mu}(z+\alpha^{\mathrm{TR}}\Delta z^{\mathrm{TR}}) \le \mu_1$.*

Proof. **Part (i).** By the definition of $(\Delta x^{\mathrm{TR}}, \Delta s^{\mathrm{TR}})$, we use (33) and (34) to get

$$\begin{aligned}
\left\|\frac{(x+\Delta x^{\mathrm{TR}})(s+\Delta s^{\mathrm{TR}})}{xs}\right\|^2 &= \left\|\left(\mathbf{1}_n + \Delta\bar{x}^{\mathrm{TR}}\right)\left(\mathbf{1}_n + \Delta\bar{s}^{\mathrm{TR}}\right)\right\|^2 \\
&\le \left\|\left(\mathbf{1}_N + \Delta\bar{x}_N^{\mathrm{TR}}\right)\left(\mathbf{1}_N + \Delta\bar{s}_N^{\mathrm{TR}}\right)\right\|^2 + \left\|\left(\mathbf{1}_B + \Delta\bar{x}_B^{\mathrm{TR}}\right)\left(\mathbf{1}_B + \Delta\bar{s}_B^{\mathrm{TR}}\right)\right\|^2 \\
&\le \left(1+\left\|\Delta\bar{s}_N^{\mathrm{TR}}\right\|\right)^2\left\|\mathbf{1}_N + \Delta\bar{x}_N^{\mathrm{Rid}}\right\|^2 + \left(1+\left\|\Delta\bar{x}_B^{\mathrm{TR}}\right\|\right)^2\left\|\mathbf{1}_B + \Delta\bar{s}_B^{\mathrm{Rid}}\right\|^2 \\
&\le 2\left(1+12\beta\right)^2\left(16\beta\mu_1/\mu\right)^2 .
\end{aligned}$$

Hence, $r \le \sqrt{2}(1+12\beta)\cdot 16\beta\mu_1/\mu \le 25\beta\mu_1/\mu$. Similarly,

$$\begin{aligned}
\left\|\frac{\Delta x^{\mathrm{TR}}\Delta s^{\mathrm{TR}}}{xs}\right\|^2 &= \left\|\Delta\bar{x}^{\mathrm{TR}}\Delta\bar{s}^{\mathrm{TR}}\right\|^2 \\
&\le \left\|\Delta\bar{x}_N^{\mathrm{TR}}\Delta\bar{s}_N^{\mathrm{TR}}\right\|^2 + \left\|\Delta\bar{x}_B^{\mathrm{TR}}\Delta\bar{s}_B^{\mathrm{TR}}\right\|^2 \\
&\le \left(1+\left\|\mathbf{1}_N+\Delta\bar{x}_N^{\mathrm{Rid}}\right\|\right)^2\left\|\Delta\bar{s}_N^{\mathrm{TR}}\right\|^2 + \left(1+\left\|\mathbf{1}_B+\Delta\bar{s}_B^{\mathrm{Rid}}\right\|\right)^2\left\|\Delta\bar{x}_B^{\mathrm{TR}}\right\|^2 \\
&\le 2\left(1+16\beta\mu_1/\mu\right)^2\left(12\beta\right)^2 .
\end{aligned}$$

Since $\mu_1 \le \mu/4$, we have $v \le \sqrt{2}(1+4\beta)\cdot 12\beta \le 18\beta$.

Part (ii). We use Proposition 2.2 to first compute $\hat{r}$ such that $\hat{r} \le r \le 1.01\hat{r}$ in $O(\log\log n)$ time. We then set

$$\alpha^{\mathrm{TR}} := \frac{1}{1+\frac{\hat{r}}{36\beta}} \in (0,1]$$

with $\alpha^{\mathrm{TR}} = 1$ if and only if $r = 0$. Using Part (i) and $\frac{\alpha^{\mathrm{TR}}}{1-\alpha^{\mathrm{TR}}} = \frac{36\beta}{\hat{r}} < \frac{1.01\cdot 36\beta}{r}$, we have

$$\begin{aligned}
&\left\|\frac{\left(x+\alpha^{\mathrm{TR}}\Delta x^{\mathrm{TR}}\right)\left(s+\alpha^{\mathrm{TR}}\Delta s^{\mathrm{TR}}\right)}{\left(1-\alpha^{\mathrm{TR}}\right)\mu} - \mathbf{1}_n\right\| \\
&= \left\|\frac{(1-\alpha^{\mathrm{TR}})xs + \alpha^{\mathrm{TR}}(x+\Delta x^{\mathrm{TR}})(s+\Delta s^{\mathrm{TR}}) - \alpha^{\mathrm{TR}}(1-\alpha^{\mathrm{TR}})\Delta x^{\mathrm{TR}}\Delta s^{\mathrm{TR}}}{\left(1-\alpha^{\mathrm{TR}}\right)\mu} - \mathbf{1}_n\right\| \\
&\le \left\|\frac{xs}{\mu} - \mathbf{1}_n\right\| + \frac{\alpha^{\mathrm{TR}}}{1-\alpha^{\mathrm{TR}}}\left\|\frac{(x+\Delta x^{\mathrm{TR}})(s+\Delta s^{\mathrm{TR}})}{xs}\right\|\cdot\left\|\frac{xs}{\mu}\right\|_\infty + \alpha^{\mathrm{TR}}\left\|\frac{\Delta x^{\mathrm{TR}}\Delta s^{\mathrm{TR}}}{xs}\right\|\cdot\left\|\frac{xs}{\mu}\right\|_\infty \\
&< \beta + 36.5\beta(1+\beta) + 18\beta(1+\beta) \le 56\beta\,.
\end{aligned} \tag{39}$$

We now apply Lemma 2.7 with 56β and $n \ge 2$, obtaining that $z + \alpha^{\mathrm{TR}}\Delta z^{\mathrm{TR}} \in \mathcal{N}^2\left(\frac{56\beta}{1-56\beta/\sqrt{n}}\right) \subseteq \mathcal{N}^2(82\beta)$, and

$$\overline{\mu}(z+\alpha^{\mathrm{TR}}\Delta z^{\mathrm{TR}}) \le \left(1+\frac{56\beta}{\sqrt{n}}\right)(1-\alpha^{\mathrm{TR}})\mu \le \left(1+\frac{56\beta}{\sqrt{2}}\right)\frac{r\mu}{36\beta} \overset{(38)}{\le} \left(1+\frac{56\beta}{\sqrt{2}}\right)\frac{25\mu_1}{36} \le \mu_1\,.$$

□

Proof of Lemma 1.11. Under the assumption of Lemma 1.11, $\mathrm{TR}_2(B,N,12\beta)$ coincides with $\mathrm{TR}_2(\widetilde{B}_z,\widetilde{N}_z,12\beta)$ by Lemma 4.6. As the optimal solution to $\mathrm{TR}_2(\widetilde{B}_z,\widetilde{N}_z,12\beta)$, $(\Delta x^{\mathrm{TR}},\Delta s^{\mathrm{TR}})$ satisfies (33) and (34) by Lemma 4.4, and hence (38) by Lemma 4.7. If we compute α^{TR} as in Part (ii) of Lemma 4.7, then we have $z+\alpha^{\mathrm{TR}}\Delta z^{\mathrm{TR}} \in \mathcal{N}^2(82\beta)$ and $\overline{\mu}(z+\alpha^{\mathrm{TR}}\Delta z^{\mathrm{TR}}) \le \mu_1$. One can use (39) to verify that $z+\alpha\Delta z^{\mathrm{TR}} \in \mathcal{N}^2(82\beta)$ for any $\alpha \in [0,\alpha^{\mathrm{TR}}]$. □

Remark 4.8. In our TR2-IPM algorithm, we compute a δ-optimal solution to $\mathrm{TR}_2(\widetilde{B}_z,\widetilde{N}_z,12\beta)$ (see Section 7) and use it for our $(\Delta x^{\mathrm{TR}},\Delta s^{\mathrm{TR}})$ in the algorithm, instead of the actual optimal solution. Hence, our $(\Delta x^{\mathrm{TR}},\Delta s^{\mathrm{TR}})$ admits at most a $(1+\delta)$-factor of violation of the trust region constraint in (33), while attaining an objective value that is at least as good as in (34). One can then verify that (38) still holds, and hence Lemma 4.7 and Lemma 1.11 remain true if we replace the actual trust region direction with our computed $(\Delta x^{\mathrm{TR}},\Delta s^{\mathrm{TR}})$ when $\delta \le 1/64$. This is because (38) is in fact not tight for the actual trust region direction and it allows an extra multiplicative factor of $65/64$ on $\|\Delta\bar{x}_B^{\mathrm{TR}}\|$ and $\|\Delta\bar{s}_N^{\mathrm{TR}}\|$.

5 Near-optimality of Affine Scaling Step

In this section we prove Lemma 1.12 and hence Theorem 1.6. We let $\alpha^* = 1 - \mu_1/\mu$ denote the ideal step length, and assume the existence of a short straight line in $\mathcal{N}^2(\beta)$ from z_0 with $\overline{\mu}(z_0) = \mu$ to z_1 with $\overline{\mu}(z_1) = \mu_1 \geq \mu/4$. As described in Section 1.4.1, our strategy to prove the near-optimality of affine scaling step is to show that $\kappa(z + \alpha^* \Delta z^{\mathrm{a}}, z) = O(\beta)$. We first recall the residual equation (4) in this case satisfies

$$\frac{1-\beta}{4(1+\beta)} \cdot \mathbf{1}_n \leq \frac{\mu_1(1-\beta)}{\mu(1+\beta)} \cdot \mathbf{1}_n \leq \left(\mathbf{1}_n + \frac{\Delta x_0}{x_0}\right)\left(\mathbf{1}_n + \frac{\Delta s_0}{s_0}\right) = \frac{x_1 s_1}{x_0 s_0} \leq \frac{\mu_1(1+\beta)}{\mu(1-\beta)} \cdot \mathbf{1}_n \leq \frac{1+\beta}{1-\beta} \cdot \mathbf{1}_n$$

by the use of Proposition 2.5 and $\mu/4 \leq \mu_1 < \mu$. Again, we look at each index $i \in [n]$ for a bivariate quadratic equation $(1 + \Delta\bar{x}_{0i})(1 + \Delta\bar{s}_{0i}) = \frac{x_{1i}s_{1i}}{x_{0i}s_{0i}}$ where $\frac{1-\beta}{4(1+\beta)} \leq \frac{x_{1i}s_{1i}}{x_{0i}s_{0i}} \leq \frac{1+\beta}{1-\beta}$. For $(\Delta x_0, \Delta s_0)$ which is a short step from z_0 to z_1, the next result gives a universal bound on the $|1 + \Delta\bar{x}_{0i}|$ and $|1 + \Delta\bar{s}_{0i}|$ i.e. the size of the primal and dual residual relative to z_0, for each coordinate $i \in [n]$. As the second part, it also provides a universal bound on $|\Delta\bar{x}_{0i}|$ and $|\Delta\bar{s}_{0i}|$ i.e. the size of the primal and dual local step at z_0, for each coordinate $i \in [n]$. Both bounds are in terms of the upper and lower bounds of $\frac{x_{1i}s_{1i}}{x_{0i}s_{0i}}$ and $\Delta\bar{x}_{0i}\Delta\bar{s}_{0i}$.

Proposition 5.1. *Suppose $a, b \in \mathbb{R}$ satisfy $|a| \geq |b|$, $\nu := (1+a)(1+b) \in [0, \bar{\nu}]$ and $\gamma := ab \in [-1, 1]$.*

(i) Then,

$$\max\{|1+a|\,,\,|1+b|\} \leq \frac{1 + \bar{\nu} + |\gamma| + \sqrt{(|\gamma| + \bar{\nu} - 1)^2 + 4|\gamma|}}{2}\,.$$

(ii) If ν_0 satisfies $\nu_0 \leq \nu$ and $|\gamma| < \frac{1-\nu_0-(\bar{\nu}-1)}{2}$, then $|a| \leq \frac{1-\nu_0-|\gamma|+\sqrt{(1-\nu_0-|\gamma|)^2+4|\gamma|}}{2}$ which is increasing in $|\gamma|$.

Proof. **Part (i).** Solving $(1+a)(1+\gamma/a) = \nu$ for a we obtain (35). Then,

$$1 + a = \frac{1 + \nu - \gamma \pm \sqrt{(1-\nu+\gamma)^2 - 4\gamma}}{2} \quad \text{and} \quad 1 + b = \frac{1 + \nu - \gamma \mp \sqrt{(1-\nu+\gamma)^2 - 4\gamma}}{2}\,.$$

Since $1 + \nu - \gamma \geq 0$, $\max\{|1+a|, |1+b|\} \leq \frac{1+\nu-\gamma+\sqrt{(1-\nu+\gamma)^2-4\gamma}}{2}$ and this upper bound is decreasing in γ, as $\gamma \leq 1$ and $\nu \geq 0$. Hence,

$$\max\{|1+a|,\,|1+b|\} \leq \frac{1 + \nu + |\gamma| + \sqrt{(1-\nu-|\gamma|)^2 + 4|\gamma|}}{2} := h(\nu, \gamma)$$

where $\frac{\partial h(\nu,\gamma)}{\partial \nu} = \frac{1}{2}\left(1 - \frac{1-\nu-|\gamma|}{\sqrt{(1-\nu-|\gamma|)^2+4|\gamma|}}\right) \geq 0$. Therefore, using $\nu \leq \bar{\nu}$ we arrive at the result.

Part (ii). We first assume $1 - \nu + \gamma \geq 0$. By (35), $|a| = \frac{1-\nu+\gamma+\sqrt{(1-\nu+\gamma)^2-4\gamma}}{2}$ and $(\nabla|a|)(\nu, \gamma) \leq \mathbf{0}$ in this case. Then, using $\nu \geq \nu_0$, $|a| \leq \frac{1-\nu_0-|\gamma|+\sqrt{(1-\nu_0-|\gamma|)^2+4|\gamma|}}{2}$.

We now assume $1 - \nu + \gamma < 0$. In this case $a = \frac{-(1-\nu+\gamma)+\sqrt{(1-\nu+\gamma)^2-4\gamma}}{2} > 0$ by (35). $1 - \nu + \gamma < 0$ gives

$$\frac{\partial a}{\partial \nu} = \frac{1}{2} - \frac{1 - \nu + \gamma}{2\sqrt{(1-\nu+\gamma)^2 - 4\gamma}} > 0\,.$$

Since $\nu \geq 0$ and $\gamma < 1$, $1 + \nu - \gamma > 0$. Then,

$$\frac{\partial a}{\partial \gamma} = -\frac{1}{2} - \frac{1 + \nu - \gamma}{2\sqrt{(1-\nu+\gamma)^2 - 4\gamma}} < 0\,.$$

Using $\nu \leq \bar{\nu}$, we get $a \leq \frac{\bar{\nu}-1+|\gamma|+\sqrt{(\bar{\nu}-1+|\gamma|)^2+4|\gamma|}}{2}$. Since $\bar{\nu} - 1 + |\gamma| \leq 1 - \nu_0 - |\gamma|$ by the assumption, we have $|a| \leq \frac{1-\nu_0-|\gamma|+\sqrt{(1-\nu_0-|\gamma|)^2+4|\gamma|}}{2}$. The monotonicity of the upper bound is because its first derivative with respect to γ is positive when $\gamma \geq 0$ and negative for $\gamma < 0$. $\square$

By the local ℓ_2-proximity between $z_0, z \in \mathcal{N}^2(\beta)$ with $\overline{\mu}(z) = \overline{\mu}(z_0)$ where z is regarded as the current iterate of TR2-IPM, we can derive $\kappa(z + \Delta z^{\mathrm{id}}, z) = \kappa(z_0 + \Delta z_0, z_0) + O(\beta)$ by the previous result. Then, Lemma 2.16 suggests that $\kappa(z + \alpha^* \Delta z^{\mathrm{a}}, z) = \kappa(z_0 + \Delta z_0, z_0) + O(\beta)$. We recall $\hat{\xi} = \hat{\xi}(z) = \sqrt{\frac{xs}{\mu}}$ and denote $\hat{\xi}_0 := \hat{\xi}(z_0)$, $\hat{\xi}_1 := \hat{\xi}(z_1)$.

Proposition 5.2. *For $\beta \in (0, 1/128]$, suppose there exists some $z_0 \in \mathcal{N}^2(\beta)$ with $\overline{\mu}(z_0) = \mu$ and $z_1 \in \mathcal{N}^2(\beta)$ with $\overline{\mu}(z_1) = \mu_1 \geq \mu/4$ such that $[z_1, z_0] \subseteq \mathcal{N}^2(\beta)$. Let $\nu := \frac{\mu_1}{\mu} \in \left[\frac{1}{4}, 1\right]$, $z \in \mathcal{N}^2(\beta)$ with $\overline{\mu}(z) = \mu$ and $\Delta z^{\mathrm{id}} = (\Delta x^{\mathrm{id}}, \Delta s^{\mathrm{id}})$ be the ideal direction. Then, we have $\kappa(z + \Delta z^{\mathrm{id}}, z) \leq \frac{21\beta}{2(1+\sqrt{\nu})^2}$.*

Proof. By Lemma 2.15, $\kappa(z_0 + \Delta z_0, z_0) = \frac{\|\Delta x_0 \Delta s_0\|}{(\sqrt{\mu_0}+\sqrt{\mu_1})^2} \leq 2\beta$ which gives $\left\|\frac{\Delta x_0 \Delta s_0}{\mu}\right\| \leq 2\beta(1+\sqrt{\nu})^2 \leq 8\beta$. We first rewrite $\frac{\Delta x^{\mathrm{id}} \Delta s^{\mathrm{id}}}{\mu}$ as

$$\begin{aligned}\frac{\Delta x^{\mathrm{id}} \Delta s^{\mathrm{id}}}{\mu} &= \frac{(x_1 - x_0 + x_0 - x)(s_1 - s_0 + s_0 - s)}{\mu} \\ &= \frac{\Delta x_0 \Delta s_0}{\mu} + \Delta \bar{x}_0 \cdot \frac{x_0 s_0}{\mu} \cdot \frac{s_0 - s}{s_0} + \Delta \bar{s}_0 \cdot \frac{x_0 s_0}{\mu} \cdot \frac{x_0 - x}{x_0} + \frac{(x_0 - x)(s_0 - s)}{\mu}.\end{aligned} \tag{40}$$

Without loss of generality, we assume that $\max\{\|\Delta \bar{x}_0\|_\infty, \|\Delta \bar{s}_0\|_\infty\} = \|\Delta \bar{x}_0\|_\infty$ and it is attained at some $j \in [n]$. By Proposition 2.5, $(1 + \Delta \bar{x}_{0j})(1 + \Delta \bar{s}_{0j}) = \frac{x_{1j} s_{1j}}{x_{0j} s_{0j}} = \nu \frac{\hat{\xi}_{1j}^2}{\hat{\xi}_{0j}^2} \in \left[\frac{1-\beta}{4(1+\beta)}, \frac{1+\beta}{1-\beta}\right]$ and $|\Delta \bar{x}_{0j} \Delta \bar{s}_{0j}| \leq \|\Delta \bar{x}_0 \Delta \bar{s}_0\| \leq \frac{8\beta}{1-\beta}$. For $\beta \in (0, 1/128]$, $\frac{8\beta}{1-\beta} \leq \frac{1}{2}\left(1 - \frac{1-\beta}{4(1+\beta)} - \left(\frac{1+\beta}{1-\beta} - 1\right)\right)$. Hence, Proposition 5.1(ii) can be applied to get

$$|\Delta \bar{x}_{0j}| \leq \frac{1}{2}\left(1 - \frac{1-\beta}{4(1+\beta)} - \frac{8\beta}{1-\beta} + \sqrt{\left(1 - \frac{1-\beta}{4(1+\beta)} - \frac{8\beta}{1-\beta}\right)^2 + \frac{32\beta}{1-\beta}}\right) < \frac{4}{5}.$$

This together with Lemma 2.19 gives

$$\begin{aligned}\left\|\Delta \bar{x}_0 \frac{x_0 s_0}{\mu} \frac{s_0 - s}{s_0} + \Delta \bar{s}_0 \frac{x_0 s_0}{\mu} \frac{x_0 - x}{x_0}\right\| &\leq \left\|\hat{\xi}_0^2 \Delta \bar{x}_0 \frac{s_0 - s}{s_0}\right\| + \left\|\hat{\xi}_0^2 \Delta \bar{s}_0 \frac{x_0 - x}{x_0}\right\| \\ &\leq \max\left\{\left\|\hat{\xi}_0^2 \Delta \bar{x}_0\right\|_\infty, \left\|\hat{\xi}_0^2 \Delta \bar{s}_0\right\|_\infty\right\}\left(\left\|\frac{x - x_0}{x_0}\right\| + \left\|\frac{s - s_0}{s_0}\right\|\right) \\ &\leq \frac{4(1+\beta)}{5} \cdot \frac{2\sqrt{2}\beta}{(1-\beta)^2} = \frac{8\sqrt{2}(1+\beta)\beta}{5(1-\beta)^2}.\end{aligned}$$

For the last term in (40), we have $\left\|\frac{(x_0 - x)(s_0 - s)}{\mu}\right\| \leq \frac{1}{2}\left\|\hat{\xi}_0^2\right\|_\infty \left\|\left(\frac{x - x_0}{x_0}, \frac{s - s_0}{s_0}\right)\right\|^2 \leq \frac{2(1+\beta)\beta^2}{(1-\beta)^4}$ where we use AM-GM inequality and Lemma 2.19. Finally, following (40)

$$\kappa(z + \Delta z^{\mathrm{id}}, z) = \frac{1}{(1+\sqrt{\nu})^2}\left\|\frac{\Delta x^{\mathrm{id}} \Delta s^{\mathrm{id}}}{\mu}\right\| \leq \frac{1}{(1+\sqrt{\nu})^2}\left(8\beta + \frac{8\sqrt{2}(1+\beta)\beta}{5(1-\beta)^2} + \frac{2(1+\beta)\beta^2}{(1-\beta)^4}\right) \leq \frac{21\beta}{2(1+\sqrt{\nu})^2}.$$

□

Proposition 5.2 enables us to show that the ideal direction from z to z_1 is locally ℓ_2-close to the affine scaling step at z with ideal step length.

Corollary 5.3. *Under the same conditions as Proposition 5.2,*

$$\left\|\left(\Delta \bar{x}^{\mathrm{id}} - \alpha^* \Delta \bar{x}^{\mathrm{a}}, \Delta \bar{s}^{\mathrm{id}} - \alpha^* \Delta \bar{s}^{\mathrm{a}}\right)\right\| \leq 13\beta.$$

Proof. Using Lemma 2.16 and then Proposition 5.2, we get

$$\begin{aligned}\left\|\left(\Delta \bar{x}^{\mathrm{Rid}} - \alpha^* \Delta \bar{x}^{\mathrm{a}}, \Delta \bar{s}^{\mathrm{id}} - \alpha^* \Delta \bar{s}^{\mathrm{a}}\right)\right\| &\leq \left\|\hat{\xi}^{-1}\right\|_\infty \left\|\left(\Delta \hat{x}^{\mathrm{id}} - (1-\nu)\Delta \hat{x}^{\mathrm{a}}, \Delta \hat{s}^{\mathrm{id}} - (1-\nu)\Delta \hat{s}^{\mathrm{a}}\right)\right\| \\ &\leq \frac{\kappa(z_1, z)\left(1+\sqrt{\nu}\right)^2 + 2\beta\nu}{1-\beta} \\ &\leq \frac{10.5\beta + 2\beta}{1-\beta} \leq 13\beta.\end{aligned} \tag{41}$$

□

As mentioned in Section 1.4.1, in order to prove Lemma 1.12, we want to show $\alpha^{*2}\|\Delta x^{\mathrm{a}}\Delta s^{\mathrm{a}}\|/\mu = O(\beta)$. We recall the notation for the normalized direction at z is

$$\Delta\hat{x} = \sqrt{\frac{s}{x\mu}}\,\Delta x = \hat{x}^{-1}\Delta x \quad \text{and} \quad \Delta\hat{s} = \sqrt{\frac{x}{s\mu}}\,\Delta s = \hat{s}^{-1}\Delta s\,.$$

The main argument is to bound $\alpha^{*2}\|\Delta x^{\mathrm{a}}\Delta s^{\mathrm{a}}\|/\mu$ by

$$(1+o(1))\|(\alpha^*\Delta\hat{x}^{\mathrm{a}} - \Delta\hat{x}^{\mathrm{id}}\,,\, \alpha^*\Delta\hat{s}^{\mathrm{a}} - \Delta\hat{s}^{\mathrm{id}})\| + \text{residuals}$$

and then apply Lemma 2.16 and Proposition 5.2. The residuals are shown to be $O(\beta)$ using local ℓ_2-proximity.

Lemma 5.4. *For $\beta \in (0, 1/128]$, suppose there exists some $z_0 \in \mathcal{N}^2(\beta)$ with $\overline{\mu}(z_0) = \mu$ and $z_1 \in \mathcal{N}^2(\beta)$ with $\overline{\mu}(z_1) = \mu_1 \geq \mu/4$ such that $[z_1, z_0] \subseteq \mathcal{N}^2(\beta)$. Let $z \in \mathcal{N}^2(\beta)$ with $\overline{\mu}(z) = \mu$ and $\Delta z^{\mathrm{a}} = (\Delta x^{\mathrm{a}}, \Delta s^{\mathrm{a}})$ be the affine scaling direction at z, then*

$$\left(1 - \frac{\mu_1}{\mu}\right)^2 \left\|\frac{\Delta x^{\mathrm{a}}\Delta s^{\mathrm{a}}}{\mu}\right\| \leq 20\beta\,.$$

Proof. For notational convenience, we replace α^* with α in this proof. As $(x + \Delta x^{\mathrm{id}})(s + \Delta s^{\mathrm{id}}) = x_1 s_1 = \mu_1\hat{\xi}_1^2$, $\left(\hat{\xi} + \Delta\hat{x}^{\mathrm{id}}\right)\left(\hat{\xi} + \Delta\hat{s}^{\mathrm{id}}\right) = (1-\alpha)\hat{\xi}_1^2$. Let $\eta := (1-\alpha)\left(\hat{\xi}_1^2 - \hat{\xi}^2\right)$. Since $(x+\alpha\Delta x^{\mathrm{a}})(s+\alpha\Delta s^{\mathrm{a}}) = (1-\alpha)xs + \alpha^2\Delta x^{\mathrm{a}}\Delta s^{\mathrm{a}}$,

$$\begin{aligned}
\alpha^2\Delta\hat{x}^{\mathrm{a}}\Delta\hat{s}^{\mathrm{a}} &= \left(\hat{\xi} + \alpha\Delta\hat{x}^{\mathrm{a}}\right)\left(\hat{\xi} + \alpha\Delta\hat{s}^{\mathrm{a}}\right) - (1-\alpha)\hat{\xi}^2 \\
&= \left(\hat{\xi} + \Delta\hat{x}^{\mathrm{id}} + \alpha\Delta\hat{x}^{\mathrm{a}} - \Delta\hat{x}^{\mathrm{id}}\right)\left(\hat{\xi} + \Delta\hat{s}^{\mathrm{id}} + \alpha\Delta\hat{s}^{\mathrm{a}} - \Delta\hat{s}^{\mathrm{id}}\right) - (1-\alpha)\hat{\xi}^2 \\
&= \eta + \left(\hat{\xi} + \Delta\hat{s}^{\mathrm{id}}\right)\left(\alpha\Delta\hat{x}^{\mathrm{a}} - \Delta\hat{x}^{\mathrm{id}}\right) + \left(\hat{\xi} + \Delta\hat{x}^{\mathrm{id}}\right)\left(\alpha\Delta\hat{s}^{\mathrm{a}} - \Delta\hat{s}^{\mathrm{id}}\right) + \left(\alpha\Delta\hat{x}^{\mathrm{a}} - \Delta\hat{x}^{\mathrm{id}}\right)\left(\alpha\Delta\hat{s}^{\mathrm{a}} - \Delta\hat{s}^{\mathrm{id}}\right)\,.
\end{aligned} \tag{42}$$

We first note that $\|\eta\| \leq \left\|\hat{\xi}_1^2 - \mathbf{1}\right\| + \left\|\hat{\xi}^2 - \mathbf{1}\right\| \leq 2\beta$. Recall $(\hat{x}, \hat{s}) = \left(\sqrt{\frac{x\mu}{s}}, \sqrt{\frac{s\mu}{x}}\right)$. Bounding the second term requires the orthogonality of $\alpha\hat{\Delta}x^{\mathrm{a}} - \Delta\hat{x}^{\mathrm{id}}$ and $\alpha\Delta\hat{s}^{\mathrm{a}} - \Delta\hat{s}^{\mathrm{id}}$. We first have

$$\left(\hat{\xi} + \Delta\hat{s}^{\mathrm{id}}\right)\left(\alpha\Delta\hat{x}^{\mathrm{a}} - \Delta\hat{x}^{\mathrm{id}}\right) + \left(\hat{\xi} + \Delta\hat{x}^{\mathrm{id}}\right)\left(\alpha\Delta\hat{s}^{\mathrm{a}} - \Delta\hat{s}^{\mathrm{id}}\right) = \hat{s}^{-1}s(\mu_1)\left(\alpha\Delta\hat{x}^{\mathrm{a}} - \Delta\hat{x}^{\mathrm{id}}\right) + \hat{x}^{-1}x(\mu_1)\left(\alpha\Delta\hat{s}^{\mathrm{a}} - \Delta\hat{s}^{\mathrm{id}}\right) + R$$

where $R := \left(\hat{\xi} + \Delta\hat{s}^{\mathrm{id}} - \hat{s}^{-1}s(\mu_1)\right)\left(\alpha\Delta\hat{x}^{\mathrm{a}} - \Delta\hat{x}^{\mathrm{id}}\right) + \left(\hat{\xi} + \Delta\hat{x}^{\mathrm{id}} - \hat{x}^{-1}x(\mu_1)\right)\left(\alpha\Delta\hat{s}^{\mathrm{a}} - \Delta\hat{s}^{\mathrm{id}}\right)$ denotes the residual. For the first term, we recall $(x_1, s_1) = (x + \Delta x^{\mathrm{id}}, s + \Delta s^{\mathrm{id}})$. Then, by Lemma 2.16 and Proposition 2.18:

$$\begin{aligned}
&\left\|\hat{s}^{-1}s(\mu_1)\left(\alpha\Delta\hat{x}^{\mathrm{a}} - \Delta\hat{x}^{\mathrm{id}}\right) + \hat{x}^{-1}x(\mu_1)\left(\alpha\Delta\hat{s}^{\mathrm{a}} - \Delta\hat{s}^{\mathrm{id}}\right)\right\|^2 \\
&= \left\|\hat{s}^{-1}s(\mu_1)\left(\alpha\Delta\hat{x}^{\mathrm{a}} - \Delta\hat{x}^{\mathrm{id}}\right)\right\|^2 + \left\|\hat{x}^{-1}x(\mu_1)\left(\alpha\Delta\hat{s}^{\mathrm{a}} - \Delta\hat{s}^{\mathrm{id}}\right)\right\|^2 \\
&\leq \max\left\{\left\|\hat{x}^{-1}x(\mu_1)\right\|_\infty^2, \left\|\hat{s}^{-1}s(\mu_1)\right\|_\infty^2\right\}\left(\left\|\alpha\Delta\hat{x}^{\mathrm{a}} - \Delta\hat{x}^{\mathrm{id}}\right\|^2 + \left\|\alpha\Delta\hat{s}^{\mathrm{a}} - \Delta\hat{s}^{\mathrm{id}}\right\|^2\right) \\
&= \max\left\{\left\|\hat{\xi}\,\frac{x(\mu_1)}{x_1}\frac{x_1}{x}\right\|_\infty, \left\|\hat{\xi}\,\frac{s(\mu_1)}{s_1}\frac{s_1}{s}\right\|_\infty\right\}^2 \left\|\left(\alpha\Delta\hat{x}^{\mathrm{a}} - \Delta\hat{x}^{\mathrm{id}}\,,\, \alpha\Delta\hat{s}^{\mathrm{a}} - \Delta\hat{s}^{\mathrm{id}}\right)\right\|^2 \\
&\leq \frac{1+\beta}{(1-\beta)^2}\max\left\{\left\|\mathbf{1} + \Delta\bar{x}^{\mathrm{id}}\right\|_\infty, \left\|\mathbf{1} + \Delta\bar{s}^{\mathrm{id}}\right\|_\infty\right\}^2\left(\frac{\kappa(z + \Delta z^{\mathrm{id}}, z)(1+\sqrt{\nu})^2 + 2\beta(1-\alpha)}{\sqrt{1-\beta}}\right)^2
\end{aligned} \tag{43}$$

where $\kappa(z + \Delta z^{\mathrm{id}}, z) \leq \frac{21\beta}{2(1+\sqrt{\nu})^2}$ by Proposition 5.2. Let $\max\left\{\left\|\mathbf{1} + \Delta\bar{x}^{\mathrm{id}}\right\|_\infty, \left\|\mathbf{1} + \Delta\bar{s}^{\mathrm{id}}\right\|_\infty\right\}$ be attained at some $j \in [n]$ and without loss of generality, we suppose that $\left|\Delta\bar{x}_j^{\mathrm{id}}\right| \geq \left|\Delta\bar{s}_j^{\mathrm{id}}\right|$. Since $\left(1 + \Delta\bar{x}_j^{\mathrm{id}}\right)\left(1 + \Delta\bar{s}_j^{\mathrm{id}}\right) = \frac{x_{1j}s_{1j}}{x_j s_j} \leq \frac{1+\beta}{1-\beta}$ by Proposition 2.5, if we apply Proposition 5.1(i) for $\beta \in (0, 1/128]$,

$$\max\left\{\left\|\mathbf{1} + \Delta\bar{x}^{\mathrm{id}}\right\|_\infty, \left\|\mathbf{1} + \Delta\bar{s}^{\mathrm{id}}\right\|_\infty\right\} \leq \frac{1}{2}\left(1 + \frac{1+\beta}{1-\beta} + \frac{21\beta}{2} + \sqrt{\left(\frac{21\beta}{2} + \frac{1+\beta}{1-\beta} - 1\right)^2 + 42\beta}\right) \leq 1 + 4\sqrt{\beta}\,. \tag{44}$$

Then, following (43),

$$\left\|\hat{s}^{-1}s(\mu_1)\left(\alpha\Delta\hat{x}^{\mathrm{a}} - \Delta\hat{x}^{\mathrm{id}}\right) + \hat{x}^{-1}x(\mu_1)\left(\alpha\Delta\hat{s}^{\mathrm{a}} - \Delta\hat{s}^{\mathrm{id}}\right)\right\| \leq \frac{\sqrt{1+\beta}}{(1-\beta)^{\frac{3}{2}}}\left(1 + 4\sqrt{\beta}\right)\cdot 12.5\beta\,.$$

What remains is to bound the residual R. We note

$$\hat{\xi} + \Delta\hat{x}^{\mathrm{id}} - \hat{x}^{-1}x(\mu_1) = \hat{x}^{-1}\left(x + \Delta x^{\mathrm{id}} - x(\mu_1)\right) = \hat{\xi}\,\frac{x_1}{x}\frac{x_1 - x(\mu_1)}{x_1} = \hat{\xi}(\mathbf{1} + \Delta\bar{x}^{\mathrm{id}})\frac{x_1 - x(\mu_1)}{x_1}\,.$$

Similarly, $\hat{\xi} + \Delta\hat{s}^{\mathrm{id}} - \hat{s}^{-1}s(\mu_1) = \hat{\xi}\left(\mathbf{1} + \Delta\bar{s}^{\mathrm{id}}\right)\frac{s_1 - s(\mu_1)}{s_1}$. Then, using (44), Lemma 2.16 and 2.17,

$$\begin{aligned}
\|R\| &\le \left\|\left(\hat{\xi} + \Delta\hat{s}^{\mathrm{id}} - \hat{s}^{-1}s(\mu_1)\right)\left(\alpha\Delta\hat{x}^{\mathrm{a}} - \Delta\hat{x}^{\mathrm{id}}\right)\right\| + \left\|\left(\hat{\xi} + \Delta\hat{x}^{\mathrm{id}} - \hat{x}^{-1}x(\mu_1)\right)\left(\alpha\Delta\hat{s}^{\mathrm{a}} - \Delta\hat{s}^{\mathrm{id}}\right)\right\| \\
&\le \max\left\{\left\|\hat{\xi}\left(\mathbf{1} + \Delta\bar{x}^{\mathrm{id}}\right)\frac{s_1 - s(\mu_1)}{s_1}\right\|_\infty, \left\|\hat{\xi}\left(\mathbf{1} + \Delta\bar{s}^{\mathrm{id}}\right)\frac{x_1 - x(\mu_1)}{x_1}\right\|_\infty\right\} \cdot \sqrt{2}\left\|\left(\alpha\Delta\hat{x}^{\mathrm{a}} - \Delta\hat{x}^{\mathrm{id}}, \alpha\Delta\hat{s}^{\mathrm{a}} - \Delta\hat{s}^{\mathrm{id}}\right)\right\| \\
&\le \sqrt{2}\frac{\sqrt{1+\beta}(1+4\sqrt{\beta})}{(1-\beta)^{\frac{3}{2}}} \cdot 12.5\beta^2\,.
\end{aligned}$$

We bound the final term in (42) using the orthogonality between $\alpha\Delta\hat{x}^{\mathrm{a}} - \Delta\hat{x}^{\mathrm{id}}$ and $\alpha\Delta\hat{s}^{\mathrm{a}} - \Delta\hat{s}^{\mathrm{id}}$ and Lemma 2.16:

$$\left\|\left(\alpha\Delta\hat{x}^{\mathrm{a}} - \Delta\hat{x}^{\mathrm{id}}\right)\left(\alpha\Delta\hat{s}^{\mathrm{a}} - \Delta\hat{s}^{\mathrm{id}}\right)\right\| \le \frac{\left\|\left(\alpha\Delta\hat{x}^{\mathrm{a}} - \Delta\hat{x}^{\mathrm{id}}, \alpha\Delta\hat{s}^{\mathrm{a}} - \Delta\hat{s}^{\mathrm{id}}\right)\right\|^2}{2} \le \frac{625\beta^2}{8(1-\beta)}\,.$$

Finally, applying all the bounds derived to (42) for $\beta \in (0, 1/128]$,

$$\begin{aligned}
&\left(1 - \frac{\mu_1}{\mu}\right)^2\left\|\frac{\Delta x^{\mathrm{a}}\Delta s^{\mathrm{a}}}{\mu}\right\| = \alpha^2\left\|\Delta\hat{x}^{\mathrm{a}}\Delta\hat{s}^{\mathrm{a}}\right\| \\
&\le \|\eta\| + \left\|\hat{s}^{-1}s(\mu_1)\left(\alpha\Delta\hat{x}^{\mathrm{a}} - \Delta\hat{x}^{\mathrm{id}}\right) + \hat{x}^{-1}x(\mu_1)\left(\alpha\Delta\hat{s}^{\mathrm{a}} - \Delta\hat{s}^{\mathrm{id}}\right)\right\| + \|R\| + \left\|\left(\alpha\Delta\hat{x}^{\mathrm{a}} - \Delta\hat{x}^{\mathrm{id}}\right)\left(\alpha\Delta\hat{s}^{\mathrm{a}} - \Delta\hat{s}^{\mathrm{id}}\right)\right\| \\
&\le 2\beta + 12.5\beta\frac{\sqrt{1+\beta}}{(1-\beta)^{\frac{3}{2}}}(1+\sqrt{2}\beta)\left(1+4\sqrt{\beta}\right) + \frac{625\beta^2}{8(1-\beta)} \\
&\le 20\beta\,.
\end{aligned}$$

□

Based on Lemma 5.4 and Proposition 2.10(iii), we can find a step length $\bar{\alpha}^{\mathrm{a}} \ge \alpha^*$ while $z + \bar{\alpha}^{\mathrm{a}}\Delta z^{\mathrm{a}} \in \mathcal{N}^2(O(\beta))$. On the other hand, computing the exact $\bar{\alpha}^{\mathrm{a}}$ is a root finding of (13). Since TR2-IPM only takes affine scaling step when $\bar{\alpha}^{\mathrm{a}} \le 3/4$ (checked in line 7, can be done without square root) and $\bar{\alpha}^{\mathrm{a}}$ is bounded below, we can employ our strongly polynomial binary search procedure (Algorithm 5, described in Section 7.4) to compute a step length $\alpha^{\mathrm{a}} \in (0, 3/4]$ that is close enough to $\bar{\alpha}^{\mathrm{a}}$, which would be used in TR2-IPM instead. It is greater than $\bar{\alpha}^{\mathrm{a}}$ to ensure $\overline{\mu}(z + \alpha^{\mathrm{a}}\Delta z^{\mathrm{a}}) \le \mu_1$ while $z + \alpha^{\mathrm{a}}\Delta z^{\mathrm{a}}$ stays in a slightly larger ℓ_2-neighborhood.

Lemma 5.5. *For $\beta \in (0, 1/128]$, suppose there exists some $z_0 \in \mathcal{N}^2(\beta)$ with $\overline{\mu}(z_0) = \mu$ and $z_1 \in \mathcal{N}^2(\beta)$ with $\overline{\mu}(z_1) = \mu_1 \ge \mu/4$ such that $[z_1, z_0] \subseteq \mathcal{N}^2(\beta)$. Let $z \in \mathcal{N}^2(\beta)$ with $\overline{\mu}(z) = \mu$ and $\Delta z^{\mathrm{a}} = (\Delta x^{\mathrm{a}}, \Delta s^{\mathrm{a}})$ be the affine scaling direction at z.*

(i) Let $\bar{\alpha}^{\mathrm{a}}$ be defined as follows: if $\|\Delta x^{\mathrm{a}}\Delta s^{\mathrm{a}}\| = 0$ then $\bar{\alpha}^{\mathrm{a}} := 1$, and otherwise let $\bar{\alpha}^{\mathrm{a}}$ be the root on $\mathbb{R}_{>\mathbf{0}}$ of $\frac{\alpha^2}{1-\alpha}\left\|\frac{\Delta x^{\mathrm{a}}\Delta s^{\mathrm{a}}}{\mu}\right\| = 80\beta$. Then $z + \bar{\alpha}^{\mathrm{a}}\Delta z^{\mathrm{a}} \in \mathcal{N}^2(81\beta)$ and $\overline{\mu}(z + \bar{\alpha}^{\mathrm{a}}\Delta z^{\mathrm{a}}) \le \mu_1$.

(ii) If $\bar{\alpha}^{\mathrm{a}} \le \frac{3}{4}$, it takes $O(\log\log(n/\beta))$ many arithmetic operations in strongly polynomial time to compute an affine scaling step length $\alpha^{\mathrm{a}} \in \left(0, \frac{3}{4}\right]$ that satisfies $z + \alpha^{\mathrm{a}}\Delta z^{\mathrm{a}} \in \mathcal{N}^2(82\beta)$ and $\overline{\mu}(z + \alpha^{\mathrm{a}}\Delta z^{\mathrm{a}}) \le \mu_1$.

Proof. **Part (i).** Applying Proposition 2.10(iii) with $k = 80$, we have $\bar{\alpha}^{\mathrm{a}} \in (0, 1]$ and $z + \bar{\alpha}^{\mathrm{a}}\Delta z^{\mathrm{a}} \in \mathcal{N}^2(81\beta)$. The exact formula for $\bar{\alpha}^{\mathrm{a}}$ by solving the quadratic equation is

$$\bar{\alpha}^{\mathrm{a}} = \sqrt{\frac{80^2\beta^2\mu^2}{4\|\Delta x^{\mathrm{a}}\Delta s^{\mathrm{a}}\|^2} + \frac{80\beta\mu}{\|\Delta x^{\mathrm{a}}\Delta s^{\mathrm{a}}\|}} - \frac{40\beta\mu}{\|\Delta x^{\mathrm{a}}\Delta s^{\mathrm{a}}\|}\,. \tag{45}$$

By Lemma 5.4, $\alpha^{*2}\left\|\frac{\Delta x^{\mathrm{a}}\Delta s^{\mathrm{a}}}{\mu}\right\| \le 20\beta$. Since $g(x) := \sqrt{\frac{x^2}{4} + x} - \frac{x}{2}$ is increasing on $\mathbb{R}_{>\mathbf{0}}$ and $\alpha^* \in \left[0, \frac{3}{4}\right]$, it follows from (45) that $\bar{\alpha}^{\mathrm{a}} \ge \sqrt{4\alpha^{*4} + 4\alpha^{*2}} - 2\alpha^{*2} \ge \alpha^*$. Then by Proposition 2.10(ii),

$$\overline{\mu}(z + \bar{\alpha}^{\mathrm{a}}\Delta z^{\mathrm{a}}) = (1 - \bar{\alpha}^{\mathrm{a}})\mu \le (1 - \alpha^*)\mu = \mu_1\,.$$

Part (ii). We first note that $\bar{\alpha}^{\mathrm{a}}$ is also the unique root on $\mathbb{R}_{>\mathbf{0}}$ of

$$\frac{\alpha^4}{(1-\alpha)^2}\left\|\frac{\Delta x^{\mathrm{a}}\Delta s^{\mathrm{a}}}{\mu}\right\|^2 = 80^2\beta^2\,.$$

Let $f(\alpha) := 80^2\beta^2 - \frac{\alpha^4}{(1-\alpha)^2}\left\|\frac{\Delta x^{\mathrm{a}}\Delta s^{\mathrm{a}}}{\mu}\right\|^2$ which is strictly decreasing on $(0,1)$. Since $\bar{\alpha}^{\mathrm{a}} > 20\beta/n$ by Proposition 2.10(iii), we have $f(20\beta/n) > 0$. We know $f(3/4) \le 0$ by the assumption, and the evaluation of any $f(\alpha)$ can be done in strongly polynomial time. Then by Lemma 7.16, BinarySearch $(f, 20\beta/n, 3/4, 0.002)$ (Algorithm 5) returns $\hat{\alpha}$ with $f(\hat{\alpha}) > 0 \ge f(1.002\hat{\alpha})$ using $O(\log\log(n/\beta))$ many operations in strongly polynomial time. Hence, $\hat{\alpha} < \bar{\alpha}^{\mathrm{a}} \le 1.002\hat{\alpha}$. We take $\alpha^{\mathrm{a}} := \min\{1.002\hat{\alpha}, 3/4\}$. By part (i), $\overline{\mu}(z+\alpha^{\mathrm{a}}\Delta z^{\mathrm{a}}) = (1-\alpha^{\mathrm{a}})\mu \le (1-\bar{\alpha}^{\mathrm{a}})\mu \le \mu_1$, and

$$\begin{aligned}\left\|\frac{(x+\alpha^{\mathrm{a}}\Delta x^{\mathrm{a}})(s+\alpha^{\mathrm{a}}\Delta s^{\mathrm{a}})}{(1-\alpha^{\mathrm{a}})\mu} - \mathbf{1}_n\right\| &\le \left\|\frac{xs}{\mu} - \mathbf{1}_n\right\| + \frac{{\alpha^{\mathrm{a}}}^2}{1-\alpha^{\mathrm{a}}}\left\|\frac{\Delta x^{\mathrm{a}}\Delta s^{\mathrm{a}}}{\mu}\right\| \\ &\le \beta + \frac{{\bar{\alpha}^{\mathrm{a}}}^2}{1-\bar{\alpha}^{\mathrm{a}}}\left\|\frac{\Delta x^{\mathrm{a}}\Delta s^{\mathrm{a}}}{\mu}\right\| \cdot \frac{1.002^2}{1-\frac{0.002\bar{\alpha}^{\mathrm{a}}}{1-\bar{\alpha}^{\mathrm{a}}}} \\ &< \beta + 80\beta\cdot\frac{1.002^2}{1-0.006} < 82\beta\,.\end{aligned}$$

□

Proof of Lemma 1.12. Under the assumption of Lemma 1.12, we apply Lemma 5.4 to get ${\alpha^*}^2\left\|\Delta x^{\mathrm{a}}\Delta s^{\mathrm{a}}\right\|/\mu \le 20\beta$. Taking $\alpha^{\mathrm{a}} \in (0, 3/4]$ as in Lemma 5.5(ii), we have $z+\alpha^{\mathrm{a}}\Delta z^{\mathrm{a}} \in \mathcal{N}^2(82\beta)$ and $\overline{\mu}(z+\alpha^{\mathrm{a}}\Delta z^{\mathrm{a}}) \le \mu_1$. We also have $z+\alpha\Delta z^{\mathrm{a}} \in \mathcal{N}^2(82\beta)$ for any $\alpha \in [0,\alpha^{\mathrm{a}}]$, by Proposition 2.10(iii). □

We end this section and also the analysis of our TR2-IPM by proving Theorem 1.6.

Proof of Theorem 1.6. Let $\Gamma : [\mu_1,\mu_0] \to \overline{\mathcal{N}}^2(\beta)$ be the piecewise linear curve that satisfies $\overline{\mu}(\Gamma(\mu)) = \mu$ for all $\mu \in [\mu_1,\mu_0]$ with the minimum number of linear segments $T = \mathrm{SLC}(\mathcal{N}^2(\beta),\mu_0,\mu_1)$. Each segment is either a long or a short straight line that stays entirely in $\mathcal{N}^2(\beta)$. Starting from any $z \in \mathcal{N}^2(\beta)$ with $\overline{\mu}(z) \in [\mu_1,\mu_0]$, there exists $z_0 \in \mathcal{N}^2(\beta)$ with $\overline{\mu}(z_0) = \overline{\mu}(z)$ on one of the straight line segments. Let us denote the endpoint on this segment that has the smaller duality gap to be z_1 with $\overline{\mu}(z_1) < \overline{\mu}(z)$. Then our TR2-IPM either takes a trust region step $z+\alpha^{\mathrm{TR}}\Delta z^{\mathrm{TR}}$ or an affine scaling step $z+\alpha^{\mathrm{a}}\Delta z^{\mathrm{a}}$ from z. By Lemma 1.11 and 1.12, in both cases, we obtain a point $z+\alpha\Delta z \in \mathcal{N}^2(82\beta)$ with $\overline{\mu}(z+\alpha\Delta z) \le \overline{\mu}(z_1)$. Then, starting from $z+\alpha\Delta z$, the algorithm keeps taking corrector steps until a new iterate in $\mathcal{N}^2(\beta)$ is obtained (line 3). One can verify using Proposition 2.10(i) that it takes at most 3 corrector steps to reach $z^+ \in \mathcal{N}^2(\beta)$ with $\overline{\mu}(z^+) = \overline{\mu}(z+\alpha\Delta z) \le \overline{\mu}(z_1)$. Such z^+ is used as the new iterate. Therefore, for each of the T straight line segments in $[\mu_1,\mu_0]$ the TR2-IPM algorithm takes 1 predictor step, which is either a trust region or an affine scaling step, followed by at most 3 corrector steps to traverse over the segment. Hence, the algorithm uses at most $4T$ steps to terminate and generates at most $4T$ straight line segments. By Lemma 1.11 and 1.12 again, all the straight line between any two consecutive iterates stay in $\mathcal{N}^2(82\beta)$. □

6 Curvature inequalities

Recall the definition of the curvature $\kappa(z_1,z)$ for $z_1, z \in \mathcal{P}\times\mathcal{D}$ from (3). In this section, we focus on the curvature between central path points. Hence, for $0 \le \eta \le \mu$, we use the shorthand $\kappa(\eta,\mu) := \kappa(z(\eta), z(\mu))$, that is,

$$\kappa(\eta,\mu) = \frac{\left\|\left(x(\eta)-x(\mu)\right)\left(s(\eta)-s(\mu)\right)\right\|_2}{\left(\sqrt{\mu}+\sqrt{\eta}\right)^2}\,. \tag{46}$$

It will be useful to use the equivalent formulation

$$\kappa(\eta,\mu) = \frac{1}{\left(1+\sqrt{\frac{\eta}{\mu}}\right)^2}\left\|\frac{x(\eta)-x(\mu)}{x(\mu)}\,\frac{s(\eta)-s(\mu)}{s(\mu)}\right\|_2 \tag{47}$$

Recall Lemma 2.15(iii), where we showed

$$\kappa\left(\eta,\mu\right) = \sup\left\{\left\|\frac{\bar{x}\bar{s}}{\overline{\mu}(\bar{z})} - \mathbf{1}_n\right\|_2 \,:\, \bar{z} = (\bar{x},\bar{s}) = \lambda z(\eta) + (1-\lambda)z(\mu),\, 0\le\lambda\le 1\right\}. \tag{48}$$

A natural definition for this section is therefore the following:

Definition 6.1 (Path complexity)**.** For $\mu, \kappa > 0$ and $K \in \mathbb{Z}_+$, we define $\overline{\mathrm{SLC}}(\eta, \mu, \kappa)$ to be the smallest number K such that there exists as sequence $\eta = \nu_0 \leq \cdots \leq \nu_k = \mu$ such that consecutive points have curvature at most κ. That is,

$$\overline{\mathrm{SLC}}(\kappa, \eta, \mu) \coloneqq \min \left\{ K : \exists (\nu_0, \ldots, \nu_K) : \nu_0 = \eta, \nu_K = \mu, \max_{i \in [K-1]} \kappa(\nu_{i+1}, \nu_i) \leq \kappa \right\} . \tag{49}$$

$\overline{\mathrm{SLC}}(\kappa, \eta, \mu)$ essentially captures with three parameters how many segments we require to go down the path from parameter μ to η while only moving from one central path point to another and never leaving the κ-neighborhood of the central path.

The central theorem of this section is the following:

Theorem 6.2 (Universal Curvature Theorem)**.** *There exists a universal constant C such that for any $0 \leq \eta \leq \mu$ with $\kappa(\eta, \mu) \leq 2^{-8}$ we have that for any $\bar{\kappa} \leq \kappa(\eta, \mu)$*

$$\overline{\mathrm{SLC}}\left(\bar{\kappa}, \eta, \mu\right) \leq O\left(\left(\frac{\kappa(\eta, \mu)}{\bar{\kappa}} \right)^C \right) . \tag{50}$$

Corollary 6.3. *There exists a universal constant C such that for any $\bar{\kappa} \leq \kappa \leq 2^{-8}$ we have that*

$$\overline{\mathrm{SLC}}\left(\bar{\kappa}, \eta, \mu\right) \leq O\left(\left(\frac{\kappa}{\bar{\kappa}} \right)^C \overline{\mathrm{SLC}}\left(\kappa, \eta, \mu\right) \right) . \tag{51}$$

We will prove Theorem 6.2 in the following two sections (Sections 6.1 and 6.2). The proof strategy in both is remarkably different, however the case distinction is comparable to the optimality proof of the trust region algorithm in previous sections. In Section 6.1 we will show that Theorem 6.2 holds for a single step whenever $\eta = \Omega(1)$. Intuitively, what we will exploit there is that the denominator in the norm term in the definition of curvature is large enough. The main idea is that if we start at $(x(1), s(1))$ and consider the convex combination $(\bar{x}, \bar{s})$ of $(x(1), s(1))$ and $x((\eta), s(\eta))$ with parameter $\mu \in (\eta, 1)$, then Gonzaga's lemma (Lemma 2.17) allows us to bound the distance of $(\bar{x}, \bar{s})$ to the central path point $(x(\mu), s(\mu))$ in terms of the distance of $(x(\eta), s(\eta))$ to the central path point $(x(1), s(1))$. This proximity will allow us to prove the main theorem Theorem 6.4 of the next section.

In Section 6.2 we use a very different strategy. There, we exploit that for the very long segments, i.e., $\eta = o(1)$ we have that the curvature $\kappa(\eta, 1)$ is well-approximated by terms $\|(x(\eta)_B - x(1)_B)/x(1)_B\|_2$ and $\|(s(\eta)_N - s(1)_N)/s(1)_N\|_2$ for some partition $B \cup N = [n]$. It turns out that these two terms fulfill very useful identities, most importantly Lemma 6.11. This again will lead us to the desired inequality Theorem 6.12 for long steps.

We combine these two results in Theorem 6.2 and provide a proof at the end of this section.

6.1 Curvature bounds on short segments

Let us now fix a gap of 1 and normalize such that $x(1) = \mathbf{1}_n$ and $s(1) = \mathbf{1}_n$. This is without loss of generality because one can always rescale $\mathcal{P}$ and $\mathcal{D}$ accordingly. For some gaps $0 \leq \eta \leq \mu \leq 1$ we let $\alpha = \frac{1-\mu}{1-\eta}$. We further define $\bar{x} = \alpha x(\eta) + (1 - \alpha)x(1)$ and $\bar{s} = \alpha s(\eta) + (1 - \alpha)s(1)$. Then $(\bar{x}, \bar{s})$ have by linearity of gap a normalized optimality gap of μ (Proposition 2.4).

Let us first establish some useful facts and inequalities.

It will be helpful to bound

$$\frac{1 - \alpha}{\mu} = \frac{1 - (1 - \mu)/(1 - \eta)}{\mu} = \frac{1 - \eta - (1 - \mu)}{\mu(1 - \eta)} = \frac{\mu - \eta}{\mu(1 - \eta)} = \frac{1 - \frac{\eta}{\mu}}{1 - \eta} \leq 1. \tag{52}$$

Another useful inequality is the following: For fixed $\eta < 1$ and variable $\mu \in [\eta, 1]$ we have that

$$\alpha + \frac{1 - \alpha}{\mu} = \frac{1 - \mu + 1 - \frac{\eta}{\mu}}{1 - \eta} = \frac{2 - \mu - \frac{\eta}{\mu}}{1 - \eta} \leq \frac{2(1 - \sqrt{\eta})}{1 - \eta} = \frac{2}{1 + \sqrt{\eta}} ,$$

where in the inequality we used that $2\sqrt{\eta} \leq \mu + \eta/\mu$ by the AM-GM inequality.

Furthermore, we have with (52) that

$$\alpha^2 + \frac{(1 - \alpha)^2}{\mu} \leq \alpha^2 + (1 - \alpha) = 1 - \alpha(1 - \alpha) \leq 1. \tag{53}$$

We will also repeatedly use the following identity for $a \neq 0$:

$$\frac{1}{a} - 1 = 1 - a + \frac{(a-1)^2}{a}\,. \tag{54}$$

Having established these inequalities and identities, we can now prove a curvature bound via Gonzaga's lemma.

Theorem 6.4 (Curvature bounds via Gonzaga's lemma)**.** *For any $\eta \leq \mu \leq \nu$ such that $\kappa(\eta,\nu) \leq 2^{-4}$ we have that*

$$\kappa(\eta,\mu) + \kappa(\mu,\nu) \leq \frac{5}{4}\left(1 + \frac{2^3}{(1+\sqrt{\frac{\eta}{\nu}})^3}\right)\kappa(\eta,\nu) \tag{55}$$

In particular, we have that $\kappa(\mu,\nu), \kappa(\eta,\mu) \leq 2^4\kappa(\eta,\nu)$.

Proof. Let us fix $\nu = 1$ and assume w.l.o.g. by rescaling that $x(1) = s(1) = \mathbf{1}_n$. As above, let $\bar{x} = (1-\alpha)x(1) + \alpha x(\eta) = (1-\alpha)\mathbf{1}_n + \alpha x(\eta)$ and $\bar{s} = (1-\alpha)s(1) + \alpha s(\eta) = (1-\alpha)\mathbf{1}_n + \alpha s(\eta)$ and let us define $\beta := \left\|\frac{\bar{x}\bar{s}}{\mu} - \mathbf{1}_n\right\|$. Note that from the above can conclude that $\beta \leq \kappa(\eta,1) \leq 2^{-4}$.

With Lemma 2.17 we have

$$\left\|\frac{\bar{x}}{x(\mu)}\right\|_\infty \leq 1 + \left\|\frac{\bar{x}}{x(\mu)} - \mathbf{1}_n\right\|_\infty \leq 1 + \frac{\beta}{1-\beta} = \frac{1}{1-\beta}$$

Another application of Gonzaga's bound Lemma 2.17 shows that

$$\left\|\left(\frac{\bar{x}-x(\mu)}{x(\mu)}, \frac{\bar{s}-s(\mu)}{s(\mu)}\right)\right\| \leq \max\left\{1, \left\|\frac{\bar{x}}{x(\mu)}\right\|_\infty\right\}\left\|\left(\frac{\bar{x}-x(\mu)}{\bar{x}}, \frac{\bar{s}-s(\mu)}{s(\mu)}\right)\right\| \leq \frac{1}{1-\beta}\frac{\beta}{1-\beta} = \frac{\beta}{(1-\beta)^2}\,. \tag{56}$$

The assumption $\kappa(\eta,1) \leq 2^{-4}$ furthermore implies that $\max\{\|s(\mu)_i - \mathbf{1}_n\|_\infty, \|x(\mu)_i - \mathbf{1}_n\|_\infty\} \leq 1$. This can be seen by an application of Proposition 5.1 part (i), but we can also see this directly with the following argument: Assume that $|s(\mu)_i - 1| > 1$. Then as $s(\mu)_i \geq 0$ this implies that $s(\mu)_i > 2$.

We can now compute that

$$\kappa(\mu,1) \geq \frac{1}{(1+\sqrt{\mu})^2}\left|(x(\mu)_i - 1)(s(\mu)_i - 1)\right| \geq \frac{1}{(1+\sqrt{\mu})^2}\left|\frac{\mu}{s(\mu)_i} - 1\right| > \frac{1}{(1+\sqrt{\mu})^2}\left|\frac{\mu}{2} - 1\right| = \frac{1-\frac{\mu}{2}}{(1+\sqrt{\mu})^2} \geq \frac{1}{8}\,,$$

where the last inequality follows as the function of μ is minimized at $\mu = 1$. This is a contradiction to our assumption on $\kappa(\mu,1)$, hence $|s(1)_i - s(\mu)_i| \leq 1$.

This also implies that $\frac{x(\eta)_i}{\eta} = \frac{1}{s(\eta)_i} \geq \frac{1}{2}$ and analogously $\frac{s(\eta)_i}{\eta} \geq \frac{1}{2}$ for all $i \in [n]$.

From here, note that

$$\bar{x}_i = (1-\alpha)x(1)_i + \alpha x(\eta)_i \geq (1-\alpha)1 + \alpha\frac{\eta}{2} \geq \frac{1}{2}\left((1-\alpha)\cdot 1 + \alpha\cdot\eta\right) = \frac{\mu}{2}$$

and analogously $\bar{s}_i \geq \frac{\mu}{2}$ for all $i \in [n]$. These are very loose bounds, but are sufficient for our purposes.

Now, by proximity

$$x(\mu)_i \geq \bar{x} - |\bar{x}_i - x(\mu)_i| \geq \bar{x}_i - \bar{x}_i\left\|\frac{\bar{x}-x(\mu)}{\bar{x}}\right\|_\infty \geq \bar{x}_i\left(1 - \frac{\beta}{1-\beta}\right) \geq \frac{1-2\beta}{2(1-\beta)}\mu\,, \tag{57}$$

where the penultimate inequality follows again from Lemma 2.17. Analogously, we get $s(\mu)_i \geq \frac{1-2\beta}{2(1-\beta)}\mu$. We now have that

$$\begin{aligned}
&(1+\sqrt{\mu})^2\kappa(\mu,1) + \left(1+\sqrt{\tfrac{\eta}{\mu}}\right)^2\kappa(\eta,\mu) \\
&= \left\|\frac{x(\mu)-x(1)}{x(1)} \circ \frac{s(\mu)-s(1)}{s(1)}\right\|_2 + \left\|\frac{x(\eta)-x(\mu)}{x(\mu)} \circ \frac{s(\eta)-s(\mu)}{s(\mu)}\right\|_2 \\
&\leq \left\|\frac{\bar{x}-x(1)}{x(1)} \circ \frac{\bar{s}-s(1)}{s(1)}\right\|_2 + \left\|\frac{x(\mu)-\bar{x}}{x(1)} \circ \frac{s(\mu)-\bar{s}}{s(1)}\right\|_2 + \left\|\frac{x(\mu)-\bar{x}}{x(1)} \circ \frac{\bar{s}-s(1)}{s(1)}\right\|_2 + \left\|\frac{\bar{x}-x(1)}{x(1)} \circ \frac{s(\mu)-\bar{s}}{s(1)}\right\|_2 \\
&\quad + \left\|\frac{\bar{x}-x(\eta)}{x(\mu)} \circ \frac{\bar{s}-s(\eta)}{s(\mu)}\right\|_2 + \left\|\frac{x(\mu)-\bar{x}}{x(\mu)} \circ \frac{s(\mu)-\bar{s}}{s(\mu)}\right\|_2 + \left\|\frac{x(\mu)-\bar{x}}{x(\mu)} \circ \frac{\bar{s}-s(\eta)}{s(\mu)}\right\|_2 + \left\|\frac{\bar{x}-x(\mu)}{x(\mu)} \circ \frac{s(\mu)-\bar{s}}{s(\mu)}\right\|_2\,,
\end{aligned} \tag{58}$$

which follows from the triangle inequality. We can now separately bound the terms in the above inequality as follows:

$$\left\|\frac{\bar{x}-x(1)}{x(1)} \circ \frac{\bar{s}-s(1)}{s(1)}\right\|_2 = \left\|\frac{\alpha(x(1)-x(\eta))}{x(1)} \circ \frac{\alpha(s(1)-s(\eta))}{s(1)}\right\|_2 = \alpha^2\left(1+\sqrt{\eta}\right)^2 \kappa(\eta,1)$$

$$\left\|\frac{\bar{x}-x(\eta)}{x(\mu)} \circ \frac{\bar{s}-s(\eta)}{s(\mu)}\right\|_2 = \mu^{-1}\left\|\frac{(1-\alpha)(x(1)-x(\eta))}{x(1)} \circ \frac{(1-\alpha)(s(1)-s(\eta))}{s(1)}\right\|_2 = \mu^{-1}(1-\alpha)^2\left(1+\sqrt{\eta}\right)^2 \kappa(\eta,1)$$

Furthermore, using proximity of $(\bar{x},\bar{s})$ to $(x(\mu),s(\mu))$ we get using (56) that

$$\begin{aligned}
\left\|\frac{x(\mu)-\bar{x}}{x(1)} \circ \frac{s(\mu)-\bar{s}}{s(1)}\right\|_2 + \left\|\frac{x(\mu)-\bar{x}}{x(\mu)} \circ \frac{s(\mu)-\bar{s}}{s(\mu)}\right\|_2 &= \mu\left\|\frac{x(\mu)-\bar{x}}{x(\mu)} \circ \frac{s(\mu)-\bar{s}}{s(\mu)}\right\|_2 + \left\|\frac{x(\mu)-\bar{x}}{x(\mu)} \circ \frac{s(\mu)-\bar{s}}{s(\mu)}\right\|_2 \\
&\le (1+\mu)\left\|\frac{x(\mu)-\bar{x}}{x(\mu)}\right\|_2 \cdot \left\|\frac{s(\mu)-\bar{s}}{s(\mu)}\right\|_2 \\
&\le \frac{(1+\mu)\beta^2}{(1-\beta)^4}
\end{aligned}$$

We can bound further terms in (58) (using (57)) as follows:

$$\begin{aligned}
\left\|\frac{x(\mu)-\bar{x}}{x(1)} \circ \frac{\bar{s}-s(1)}{s(1)}\right\|_2 &\le \mu\left\|\frac{x(\mu)-\bar{x}}{x(\mu)}\right\|_2 \left\|\frac{\alpha(s(\eta)-s(1))}{s(\mu)}\right\|_\infty \\
&\le \mu\left\|\frac{x(\mu)-\bar{x}}{x(\mu)}\right\|_2 \left\|\frac{\alpha(s(\eta)-s(1))}{s(1)}\right\|_\infty \left\|s(\mu)^{-1}\right\|_\infty \\
&\le \mu \cdot \frac{\beta}{(1-\beta)^2} \cdot \alpha \cdot \frac{2(1-\beta)}{(1-2\beta)\mu} \\
&= \frac{2\alpha\beta}{(1-\beta)(1-2\beta)}
\end{aligned}$$

and analogously

$$\left\|\frac{\bar{x}-x(1)}{x(1)} \circ \frac{s(\mu)-\bar{s}}{s(1)}\right\|_2 \le \frac{2\alpha\beta}{(1-\beta)(1-2\beta)}.$$

Similarly, we can bound

$$\begin{aligned}
\left\|\frac{x(\mu)-\bar{x}}{x(\mu)} \circ \frac{\bar{s}-s(\eta)}{s(\mu)}\right\|_2 &\le \left\|\frac{x(\mu)-\bar{x}}{x(\mu)}\right\|_2 \left\|\frac{(1-\alpha)(s(\eta)-s(1))}{s(\mu)}\right\|_\infty \\
&\le \frac{\beta}{(1-\beta)^2} \cdot (1-\alpha) \cdot \frac{2(1-\beta)}{(1-2\beta)\mu} = \frac{2(1-\alpha)\beta}{(1-\beta)(1-2\beta)\mu},
\end{aligned}$$

and symmetrically

$$\left\|\frac{\bar{x}-x(\mu)}{x(\mu)} \circ \frac{s(\mu)-\bar{s}}{s(\mu)}\right\|_2 \le \frac{2(1-\alpha)\beta}{(1-\beta)(1-2\beta)\mu}.$$

Putting everything together into (58), and using the preliminary facts and inequalities established at the beginning of this section we have that

$$\begin{aligned}
&\left(1+\sqrt{\mu}\right)^2 \kappa(\mu,1) + \left(1+\sqrt{\tfrac{\eta}{\mu}}\right)^2 \kappa(\eta,\mu) \\
&\le \left(\alpha^2 + \frac{(1-\alpha)^2}{\mu}\right)\left(1+\sqrt{\eta}\right)^2 \kappa(\eta,1) + (1+\mu)\frac{\beta^2}{(1-\beta)^4} + \left(\alpha + \frac{1-\alpha}{\mu}\right)\frac{4\beta}{(1-\beta)(1-2\beta)} \\
&\le \left(1+\sqrt{\eta}\right)^2 \kappa(\eta,1) + (1+\mu)\frac{\beta^2}{(1-\beta)^4} + \frac{2}{1+\sqrt{\eta}}\frac{4\beta}{(1-\beta)(1-2\beta)} \\
&\le \left(1+\sqrt{\eta}\right)^2 \kappa(\eta,1) + (1+\mu)\frac{\kappa(\eta,1)^2}{(1-\kappa(\eta,1))^4} + \frac{1}{(1-2\kappa(\eta,1))(1-\kappa(\eta,1))}\frac{8}{1+\sqrt{\eta}}\kappa(\eta,1) \\
&= \left(1+\sqrt{\eta}\right)^2 \kappa(\eta,1)\left(1 + \frac{1+\mu}{(1+\sqrt{\eta})^2}\frac{\kappa(\eta,1)}{(1-\kappa(\eta,1))^4} + \frac{1}{(1-2\kappa(\eta,1))(1-\kappa(\eta,1))}\frac{8}{(1+\sqrt{\eta})^3}\right)
\end{aligned}$$

Dividing by $(1+\sqrt{\eta})^2$ (which is smaller than both $(1+\sqrt{\mu})^2$ and $(1+\sqrt{\frac{\eta}{\mu}})^2$) and using that $\kappa(\eta,1) \le 2^{-4}$ we obtain that

$$\kappa(\mu,1) + \kappa(\eta,\mu) \le \frac{5}{4}\left(1 + \frac{2^3}{(1+\sqrt{\eta})^3}\right)\kappa(\eta,1)\,, \tag{59}$$

which proves the main inequality of the lemma. For the last part note that the terms on the left side of (59) are nonnegative and that $2^3/(1+\sqrt{\eta})^3 \le 2^3$ and so $\max\{\kappa(\mu,1),\kappa(\eta,\mu)\} \le \frac{5}{4}\left(1+2^3\right)\kappa(\eta,1) \le 2^4\kappa(\eta,1)$.

□

Proposition 6.5. *If $\kappa(\eta,1) \le 2^{-8}$, then $\kappa(\mu,\nu) \le 2^8\kappa(\eta,1)$ for any $\eta \le \mu \le \nu \le 1$.*

Proof. First, we use Theorem 6.4 to show that $\kappa(\mu,1) \le 2^4\kappa(\eta,1) \le 2^{-4}$ and then again apply Theorem 6.4 to show that $\kappa(\mu,\nu) \le 2^4\kappa(\mu,1)$, which gives the desired bound. □

Proposition 6.6. *Let $\eta := \overline{\mathrm{SLC}}(1,\kappa,1)$ for some $\kappa < 2^{-4}$ and assume that $\eta \ge \frac{7}{8}$. Then, for any $\bar\kappa < \kappa$, we have that*

$$\overline{\mathrm{SLC}}(\eta,\mu,\bar\kappa) \le 2\left(\frac{\kappa}{\bar\kappa}\right)^C \cdot \overline{\mathrm{SLC}}(\eta,\mu,\kappa) \tag{60}$$

for some universal constant $C > 0$.

Proof. Theorem 6.4 implies that for any $\mu \in [\eta,1]$ we have that

$$\kappa(\eta,\mu) + \kappa(\mu,1) \le \frac{5}{4}\left(1 + \frac{2^3}{\left(1+\sqrt{\frac{7}{8}}\right)^3}\right)\kappa(\eta,1) \le (2-2^{-5})\kappa(\eta,1)\,.$$

By definition, it is easy to see that $\kappa(\cdot,1)$ and $\kappa(\eta,\cdot)$ are continuous functions. Furthermore, $\kappa(1,1) = \kappa(\eta,\eta) = 0$. Hence, there must exists $\mu \in [\eta,1]$ for which $\kappa(\mu,1) = \kappa(\eta,\mu)$. For this μ, we have that

$$\max\{\kappa(\mu,1),\kappa(\eta,\mu)\} = \frac{\kappa(\mu,1)+\kappa(\eta,\mu)}{2} \le (1-2^{-4})\kappa(\eta,1)\,.$$

Iterating this argument, we get that there exists a constant $C > 0$ such that for any $\bar\kappa \in (0,\kappa]$ we have that

$$\begin{aligned}\overline{\mathrm{SLC}}(\bar\kappa,\eta,\mu) &\le 2^{\lceil \log_{1-2^{-4}}(\bar\kappa/\kappa)\rceil}\overline{\mathrm{SLC}}(\kappa,\eta,\mu)\\ &= 2^{\lceil \log_2(\kappa/\bar\kappa)/\log_2(1/(1-2^{-4}))\rceil}\overline{\mathrm{SLC}}(\kappa,\eta,\mu)\\ &\le 2\left(\frac{\kappa}{\bar\kappa}\right)^C\overline{\mathrm{SLC}}(\kappa,\eta,\mu)\end{aligned}$$

for $C = 1/\log_2(1/(1-2^{-4}))$ as desired. □

Proposition 6.7. *Let $\kappa \le 2^{-8}$ and $\bar\kappa \le \kappa$. Then,*

$$\overline{\mathrm{SLC}}(\bar\kappa,\eta,\mu) \le 2\left(1+\log\left(\frac{\mu}{\eta}\right)\right)\left(\frac{\kappa}{\bar\kappa}\right)^C \cdot \overline{\mathrm{SLC}}(\kappa,\eta,\mu), \tag{61}$$

for some universal constant $C > 0$.

Proof. Let $0 \le \eta < \mu$ be minimal such that $\overline{\mathrm{SLC}}(\eta,\mu,\kappa) = 1$, i.e. $\kappa(\eta,\mu) \le \kappa$. Consider the points $\mu_i = (7/8)^i$ for $i = 1,\dots,k$ for $k = \lceil \log_{7/8}(\eta/\mu)\rceil$. Then, by Proposition 6.5 we conclude that $\kappa(\mu_{i+1},\mu_i) \le 2^4\kappa \le 2^{-4}$ for all $i = 1,\dots,k-1$. Hence, we can apply Proposition 6.6 to conclude that for the constant C from Proposition 6.6 we have that

$$\overline{\mathrm{SLC}}(\bar\kappa,\mu_{i+1},\mu_i) \le 2\left(\frac{\kappa}{\bar\kappa}\right)^C \cdot \overline{\mathrm{SLC}}\left(\kappa(\mu_{i+1},\mu_i),\mu_{i+1},\mu_i\right) = 2\left(\frac{\kappa}{\bar\kappa}\right)^C\,.$$

The result follows as

$$\overline{\mathrm{SLC}}(\bar\kappa,\eta,\mu) \le \sum_{i=1}^{k-1}\overline{\mathrm{SLC}}(\bar\kappa,\mu_{i+1},\mu_i) \le \sum_{i=1}^{k-1} 2\left(\frac{\kappa}{\bar\kappa}\right)^C = 2\lceil \log_{7/8}(\eta/\mu)\rceil\left(\frac{\kappa}{\bar\kappa}\right)^C\,.$$

□

The above theory shows that we can bound the number of steps as a function of $\kappa/\bar\kappa$, whenever η is not too small. In the following section, we extend the analysis to arbitrarily small η.

6.2 Curvature bounds on long segments

As in the previous section fix $\nu = 1$ and let $\eta \leq \mu \leq 1$. Fix a partition $B \cup N = [n]$ and define the vector

$$\phi(\eta,\mu) = \left(\frac{x(\eta)_B}{x(\mu)_B} - \mathbf{1}_B, \frac{s(\eta)_N}{s(\mu)_N} - \mathbf{1}_N\right) \tag{62}$$

We do not identify the chosen partition in the definition of ϕ as we will throughout use the partition $B \cup N = [n]$ for which $\phi(\eta,\mu) \ll 1$, which for $\eta \ll \mu$ will be unique if it exists. In particular, it will always be clear from context.

The motivation to introduce the vector $\phi(\eta,\mu)$ is multifold. Firstly, it is exactly the term whose ℓ_2-norm one aims to bound in the trust region program. While the current version of the trust region program solves a program on B and N separately, it clearly approximates the ℓ_2-norm within a factor of 2. Furthermore, we will show that if $\eta \ll \mu$ and the partition $B \cup N$ is chosen correctly, then up to scaling the norm of $\phi(\eta,\mu)$ is exactly the curvature $\kappa(\eta,\mu)$. Secondly, we are able to show strong triangle-type inequalities for the norm of the vectors $\phi(\eta,\mu)$, $\phi(\mu,1)$ and $\phi(\eta,1)$, which will allow us to derive the analogue to Theorem 6.4 when the gap $\eta \ll 1$ is small.

We will now show the first property, that $\phi(\eta,\mu)$ is a good approximation of the curvature $\kappa(\eta,\mu)$ up to scaling.

Lemma 6.8. *Let $x(1) = s(1) = \mathbf{1}_n$, let $\mu \leq 2^{-8}$ and the partition chosen such that $B = \{i \in [n] : |1 - x(\mu)_i| \leq |1 - s(\mu)_i|\}$ and $N = [n] \setminus B$. Further, assume that $\kappa(\mu,1) \leq 2^{-6}$. Then, we have that*

$$(1-8\mu)(1+\sqrt{\mu})^4\kappa(\mu,1)^2 \leq (1-\mu)\left\|\phi(\mu,1)\right\|_2^2 \leq (1+8\mu)(1+\sqrt{\mu})^4\kappa(\mu,1)^2 \tag{63}$$

Proof. For any $i \in B$ we have that $x(\mu)_i s(\mu)_i = \mu$ and so

$$|1 - x(\mu)_i| \leq \sqrt{|1-x(\mu)_i| \cdot |1 - s(\mu)_i|} \leq (1+\sqrt{\mu})\sqrt{\kappa(\mu,1)}$$

and therefore,

$$s(\mu)_i = \frac{\mu}{x(\mu)_i} \leq \frac{\mu}{1-|1-x(\mu)_i|} \leq \frac{\mu}{1-(1+\sqrt{\mu})\sqrt{\kappa(\mu,1)}} \leq \frac{\mu}{1-\left(1+\sqrt{2^{-8}}\right)\sqrt{\kappa(\mu,1)}} \leq 2\mu \tag{64}$$

From here, we have that

$$\begin{aligned}
|(x(\mu)_i - 1)(s(\mu)_i - 1) - (\mu-1)(x(\mu)_i - 1)| &= |(x(\mu)_i - 1)(s(\mu)_i - 1 - (\mu - 1))| \\
&= |(x(\mu)_i - 1)\left((x(\mu)_i - 1)s(\mu)_i\right)| \\
&= \left|(x(\mu)_i - 1)^2(s(\mu)_i - 1)^2\frac{s(\mu)_i}{(s(\mu)_i - 1)^2}\right| \\
&\leq (x(\mu)_i - 1)^2(s(\mu)_i - 1)^2\frac{2\mu}{(1-2\mu)^2} \\
&\leq 4\mu(x(\mu)_i - 1)^2(s(\mu)_i - 1)^2\,,
\end{aligned}$$

where the last inequality used that $\mu \leq 2^{-8}$. Therefore,

$$\begin{aligned}
&\left|(1+\sqrt{\mu})^4\,\kappa(\mu,1)^2 - (1-\mu)\left\|\phi(\mu,1)\right\|_2^2\right| \\
&= \left|\sum_{i\in B}\left((x(\mu)_i - 1)^2(s(\mu)_i - 1)^2 - (1-\mu)^2(x(\mu)_i - 1)^2\right) + \sum_{i\in N}\left((x(\mu)_i - 1)^2(s(\mu)_i - 1)^2 - (1-\mu)^2(s(\mu)_i - 1)^2\right)\right| \\
&\leq \left|\sum_{i\in B}\left((x(\mu)_i - 1)(s(\mu)_i - 1) + (1-\mu)(x(\mu)_i - 1)\right)\left((x(\mu)_i - 1)(s(\mu)_i - 1) - (1-\mu)(x(\mu)_i - 1)\right)\right| \\
&\quad + \left|\sum_{i\in N}\left((x(\mu)_i - 1)(s(\mu)_i - 1) - (1-\mu)(s(\mu)_i - 1)\right)\left((x(\mu)_i - 1)(s(\mu)_i - 1) + (1-\mu)(s(\mu)_i - 1)\right)\right| \\
&\leq \left(\max_{i\in[n]}|(x_i(\mu) - 1)(s_i(\mu) - 1)| + (1-\mu)\max_{i\in[n]}\{|1-x_i(\mu)|, |1-s_i(\mu)|\}\right)\sum_{i\in[n]}4\mu(x(\mu)_i - 1)^2(s(\mu)_i - 1)^2 \\
&\leq 2\cdot\sum_{i\in[n]}4\mu(x(\mu)_i - 1)^2(s(\mu)_i - 1)^2 \\
&= 8\mu(1+\sqrt{\mu})^4\kappa(\mu,1)^2\,.
\end{aligned}$$

The lemma follows.

□

The following power series identity will turn out to be useful.

Lemma 6.9. *For any $0 \le \eta \le \mu$ assume there is a partition $B \cup N = [n]$ such that $\|\phi(\eta,\mu)\|_\infty < 1$. Then,*

$$\langle \phi(\eta,\mu), \mathbf{1}_n \rangle = -\frac{\frac{\eta}{\mu}}{1-\frac{\eta}{\mu}} \sum_{k=2}^{\infty} (-1)^k \left\langle \phi(\eta,\mu)^k, \mathbf{1}_n \right\rangle \tag{65}$$

Proof. By orthogonality we have that $0 = \langle x(\mu) - x(\eta), s(\mu) - s(\eta) \rangle$. Using that $x(\eta)s(\eta) = \eta \mathbf{1}_n$ and $x(\mu)s(\mu) = \mu \mathbf{1}_n$ we can therefore write that

$$\sum_{i \in [n]} \frac{x(\eta)_i}{x(\mu)_i} + \sum_{i \in [n]} \frac{s(\eta)_i}{s(\mu)_i} - \frac{\eta+\mu}{\mu} n = 0$$

Using the partition $B \cup N = [n]$ we can rewrite the above as

$$\begin{aligned} 0 &= \sum_{i \in B} \left(\frac{x(\eta)_i}{x(\mu)_i} - 1 \right) + \sum_{i \in N} \left(\frac{s(\eta)_i}{s(\mu)_i} - 1 \right) + \frac{\eta}{\mu} \sum_{i \in N} \left(\frac{\mu x(\eta)_i}{\eta x(\mu)_i} - 1 \right) + \frac{\eta}{\mu} \sum_{i \in B} \left(\frac{\mu s(\eta)_i}{\eta s(\mu)_i} - 1 \right) \\ &= \sum_{i \in B} \left(\frac{x(\eta)_i}{x(\mu)_i} - 1 \right) + \sum_{i \in N} \left(\frac{s(\eta)_i}{s(\mu)_i} - 1 \right) + \frac{\eta}{\mu} \sum_{i \in N} \left(\frac{s(\mu)_i}{s(\eta)_i} - 1 \right) + \frac{\eta}{\mu} \sum_{i \in B} \left(\frac{x(\mu)_i}{x(\eta)_i} - 1 \right) \end{aligned} \tag{66}$$

Using for $a \in (0,2)$ the identity $a^{-1} = \sum_{k=0}^{\infty} (1-a)^k = \sum_{k=0}^{\infty} (-1)^k (a-1)^k$, we obtain that

$$\sum_{i \in N} \frac{s(\mu)_i}{s(\eta)_i} + \sum_{i \in B} \frac{x(\mu)_i}{x(\eta)_i} = \sum_{k=0}^{\infty} (-1)^k \left(\sum_{i \in N} \left(\frac{s(\eta)_i}{s(\mu)_i} - 1 \right)^k + \sum_{i \in B} \left(\frac{x(\eta)_i}{x(\mu)_i} - 1 \right)^k \right)$$

and therefore by subtracting n on both sides

$$\sum_{i \in N} \left(\frac{s(\mu)_i}{s(\eta)_i} - 1 \right) + \sum_{i \in B} \left(\frac{x(\mu)_i}{x(\eta)_i} - 1 \right) = \sum_{k=1}^{\infty} (-1)^k \left(\sum_{i \in N} \left(\frac{s(\eta)_i}{s(\mu)_i} - 1 \right)^k + \sum_{i \in B} \left(\frac{x(\eta)_i}{x(\mu)_i} - 1 \right)^k \right)$$

Plugging this into (66), taking out the term for $k = 1$ and dividing by $1 - \eta/\mu$ gives the result. □

We will now prove a useful bound for $\phi(\eta, 1)$.

Lemma 6.10. *For any $0 \le \eta \le \mu$ assume there is a partition $B \cup N = [n]$ such that $\|\phi(\eta,\mu)\|_\infty < 1$ we have for the corresponding vector $\phi(\eta,\mu)$ that*

$$\left| \langle \phi(\eta,\mu), \mathbf{1}_n \rangle + \frac{\frac{\eta}{\mu}}{1-\frac{\eta}{\mu}} \|\phi(\eta,\mu)\|_2^2 \right| \le \frac{\frac{\eta}{\mu}}{1-\frac{\eta}{\mu}} \left(\|\phi(\eta,\mu)\|_3^3 + \frac{\|\phi(\eta,\mu)\|_4^4}{1 - \|\phi(\eta,\mu)\|_\infty} \right) \tag{67}$$

Proof. Direct consequence of Lemma 6.9. □

We are ready for the main lemma for $\phi(\cdot,\cdot)$ of this section.

Lemma 6.11. *Let $\phi(\cdot,\cdot)$ as before. Then,*

$$\left| \frac{1+\eta}{1-\eta} \|\phi(\eta,1)\|_2^2 - \frac{1+\mu}{1-\mu} \|\phi(\mu,1)\|_2^2 - \frac{1+\frac{\eta}{\mu}}{1-\frac{\eta}{\mu}} \|\phi(\eta,\mu)\|_2^2 \right| = O\left(\frac{\bar{\phi}^3}{(1-\bar{\phi})^2} \right), \tag{68}$$

where $\bar{\phi} := \max\{\|\phi(\eta,1)\|_2, \|\phi(\mu,1)\|_2, \|\phi(\eta,\mu)\|_2\}$.

Proof. We have that

$$\begin{aligned} \left\| \frac{x(\eta)_B - \mathbf{1}_B}{x(\mu)_B} \right\|_2^2 &= \left\| \frac{x(\eta)_B - x(\mu)_B - (\mathbf{1}_B - x(\mu)_B)}{x(\mu)_B} \right\|_2^2 \\ &= \left\| \frac{x(\eta)_B - x(\mu)_B}{x(\mu)_B} \right\|_2^2 + \left\| \frac{x(\mu)_B - \mathbf{1}_B}{x(\mu)_B} \right\|_2^2 - 2 \left\langle \frac{x(\mu)_B - x(\eta)_B}{x(\mu)_B}, \frac{x(\mu)_B - \mathbf{1}_B}{x(\mu)_B} \right\rangle \\ &= \|\phi(\eta,\mu)_B\|_2^2 + \|\phi(1,\mu)_B\|_2^2 - 2 \left\langle \frac{x(\mu)_B - x(\eta)_B}{x(\mu)_B}, \frac{x(\mu)_B - \mathbf{1}_B}{x(\mu)_B} \right\rangle \end{aligned}$$

We use again (54) to rewrite the last term as

$$\left\langle \frac{x(\mu)_B - x(\eta)_B}{x(\mu)_B}, \frac{x(\mu)_B - \mathbf{1}_B}{x(\mu)_B} \right\rangle = \left\langle \frac{x(\mu)_B - x(\eta)_B}{x(\mu)_B}, x(\mu)_B - \mathbf{1}_B - \frac{(x(\mu)_B - \mathbf{1}_B)^2}{x(\mu)_B} \right\rangle$$

where we can further write

$$\begin{aligned}
\left\langle \frac{x(\mu)_B - x(\eta)_B}{x(\mu)_B}, x(\mu)_B - \mathbf{1}_B \right\rangle &= \left\langle \frac{x(\eta)_B}{x(\mu)_B} - \mathbf{1}_B, \mathbf{1}_B \right\rangle + \langle x(\mu)_B - x(\eta)_B, \mathbf{1}_B \rangle \\
&= \left\langle \frac{x(\eta)_B}{x(\mu)_B} - \mathbf{1}_B, \mathbf{1}_B \right\rangle + \langle x(\mu)_B - x(1)_B, \mathbf{1}_B \rangle - \langle x(\eta)_B - x(1)_B, \mathbf{1}_B \rangle \\
&= \langle \phi(\eta,\mu)_B, \mathbf{1}_B \rangle + \langle \phi(\mu,1)_B, \mathbf{1}_B \rangle - \langle \phi(\eta,1)_B, \mathbf{1}_B \rangle
\end{aligned}$$

and therefore

$$\left\| \frac{x(\eta)_B - \mathbf{1}_B}{x(\mu)_B} \right\|_2^2 = \|\phi(\eta,\mu)_B\|_2^2 + \|\phi(1,\mu)_B\|_2^2 - 2\left(\langle \phi(\eta,\mu)_B, \mathbf{1}_B \rangle + \langle \phi(\mu,1)_B, \mathbf{1}_B \rangle - \langle \phi(\eta,1)_B, \mathbf{1}_B \rangle\right) + O(\bar{\phi}^3)$$

Using the same inequalities for s_N we obtain that

$$\begin{aligned}
&\left\| \frac{x(\eta)_B - \mathbf{1}_B}{x(\mu)_B} \right\|_2^2 + \left\| \frac{s(\eta)_N - \mathbf{1}_N}{s(\mu)_N} \right\|_2^2 \\
&= \|\phi(\eta,\mu)\|_2^2 + \|\phi(1,\mu)\|_2^2 - 2\left(\langle \phi(\eta,\mu), \mathbf{1}_n \rangle + \langle \phi(\mu,1), \mathbf{1}_n \rangle - \langle \phi(\eta,1), \mathbf{1}_n \rangle + O(\bar{\phi}^3)\right) \\
&= \|\phi(\eta,\mu)\|_2^2 + \|\phi(1,\mu)\|_2^2 - 2\left(-\frac{\frac{\eta}{\mu}}{1-\frac{\eta}{\mu}} \|\phi(\eta,\mu)\|_2^2 - \frac{\frac{\mu}{1}}{1-\frac{\mu}{1}} \|\phi(\mu,1)\|_2^2 + \frac{\frac{\eta}{1}}{1-\frac{\eta}{1}} \|\phi(\eta,1)\|_2^2 + O(\bar{\phi}^3) \right)
\end{aligned}$$

where the last step used Lemma 6.10 to rewrite the inner products in terms of the norms of $\phi(\cdot,\cdot)$. We will also use the estimate that

$$\begin{aligned}
&\left\| \frac{x(\eta)_B - \mathbf{1}_B}{x(\mu)_B} \right\|_2^2 + \left\| \frac{s(\eta)_N - \mathbf{1}_N}{s(\mu)_N} \right\|_2^2 - \|\phi(\eta,1)\|_2^2 \\
&= \sum_{i\in B} \left(\frac{x(\eta)_i - 1}{x(\mu)_i} \right)^2 + \sum_{i \in N} \left(\frac{s(\eta)_i - 1}{s(\mu)_i} \right)^2 - \sum_{i\in B} \left(\frac{x(\eta)_i - 1}{1} \right)^2 - \sum_{i\in N} \left(\frac{s(\eta)_i - 1}{1} \right)^2 \\
&= \sum_{i\in B} \left(\frac{1}{x(\mu)_i^2} - 1 \right) \left(\frac{x(\eta)_i - 1}{1} \right)^2 + \sum_{i\in N} \left(\frac{1}{s(\mu)_i^2} - 1 \right) \left(\frac{s(\eta)_i - 1}{1} \right)^2 \\
&= O\left(\|\phi(\eta,1)\|_2^2 \frac{\|\phi(\mu,1)\|_2}{(1 - \|\phi(\mu,1)\|_2)^2} \right)
\end{aligned}$$

Combining all the above we obtain the estimate that

$$\left| \frac{1+\eta}{1-\eta} \|\phi(\eta,1)\|_2^2 - \frac{1+\mu}{1-\mu} \|\phi(\mu,1)\|_2^2 - \frac{1+\frac{\eta}{\mu}}{1-\frac{\eta}{\mu}} \|\phi(\eta,\mu)\|_2^2 \right| = O\left(\frac{\bar{\phi}^3}{(1-\bar{\phi})^2} \right),$$

where $\bar{\phi} \coloneqq \max\{\|\phi(\eta,1)\|_2, \|\phi(\mu,1)\|_2, \|\phi(\eta,\mu)\|_2\}$. This proves the lemma. □

Theorem 6.12. *Assume that $\mu \le 2^{-8}$ and $\eta \le 2^{-8}\mu$ and $\kappa(\eta,1) \le 2^{-8}$. Then, we have that*

$$\kappa(\eta,1)^2 \ge \frac{7}{8}\left(\kappa(\eta,\mu)^2 + \kappa(\mu,1)^2\right) \tag{69}$$

Proof. Using Lemmas 6.8 and 6.11 we can rewrite the curvature triangle inequality as follows:

$$\begin{aligned}
\kappa(\eta,1)^2 &\geq \frac{1}{(1+8\eta)(1+\sqrt{\eta})^4}(1-\eta)\,\|\phi(\eta,1)\|_2^2 \\
&\geq \frac{1}{(1+8\eta)(1+\sqrt{\eta})^4}(1-\eta)\frac{1-\eta}{1+\eta}\left(\frac{1+\mu}{1-\mu}\|\phi(\mu,1)\|_2^2+\frac{1+\frac{\eta}{\mu}}{1+\frac{\eta}{\mu}}\|\phi(\eta,\mu)\|_2^2-O\left(\frac{\bar{\phi}^3}{(1-\bar{\phi})^2}\right)\right) \\
&\geq \frac{1}{(1+8\eta)(1+\sqrt{\eta})^4}(1-\eta)\frac{1-\eta}{1+\eta}\left(\frac{1+\mu}{1-\mu}\frac{(1-8\mu)(1+\sqrt{\mu})^4}{1-\mu}\kappa(\mu,1)^2+\frac{1+\frac{\eta}{\mu}}{1-\frac{\eta}{\mu}}\frac{(1-8\frac{\eta}{\mu})(1+\sqrt{\frac{\eta}{\mu}})^4}{1-\frac{\eta}{\mu}}\kappa(\eta,\mu)^2\right) \\
&\quad -O\left(\frac{\bar{\phi}^3}{(1-\bar{\phi})^2}\right)
\end{aligned}$$

We still need to bound $\bar{\phi}$. Note that by Lemma 6.8 and the assumption on μ we have that

$$\bar{\phi} \leq 2\max\{\kappa(\eta,1),\kappa(\mu,1),\kappa(\eta,\mu)\}.$$

Furthermore, by Theorem 6.4 we have that $\max\{\kappa(\eta,\mu),\kappa(\mu,1)\} \leq 2^4\kappa(\eta,1)$. This gives overall that $\bar{\phi} \leq 2^5\kappa(\eta,1)$ and so by choice of $\kappa(\eta,1) \leq 2^{-8}$ the term involving $\bar{\phi}^3$ essentially vanishes and so

$$\kappa(\eta,1)^2 \geq \frac{31}{32}\frac{1-8\max(\eta/\mu,\mu)}{1+8\eta}\left(\kappa(\mu,1)^2+\kappa(\eta,\mu)^2\right)\,,$$

proving the theorem. ☐

We are ready for the proof of the main theorem of this section.

Proof of Theorem 6.2. Let $\eta < 1$ be minimal such that for fixed κ we have $\overline{\mathrm{SLC}}(\kappa,\eta,1) = 1$. If $\eta \geq 2^{-16}$ then we can apply Proposition 6.7 to get the desired bound. Otherwise, let $\mu^- = 2^8\eta$ and let $\mu^+ = 2^{-8}$. We then have that $\eta \leq \mu^- \leq \mu^+ \leq 1$. If $\kappa(\eta,\mu^+) \leq \kappa(\mu^+,1)$, then we can apply Theorem 6.12 to observe that

$$\kappa(\eta,\mu^+)^2 \leq \frac{\kappa(\eta,\mu^+)^2+\kappa(\mu^+,1)^2}{2} \leq \frac{8}{2\cdot 7}\kappa(\eta,1)^2 = \frac{4}{7}\kappa(\eta,1)^2\,.$$

Combined with the use of Proposition 6.7 applied to $\kappa(\mu^+,1)$ we obtain the bound

$$\begin{aligned}
\overline{\mathrm{SLC}}(\sqrt{4/7}\kappa(\eta,1),\eta,1) &\leq \overline{\mathrm{SLC}}(\sqrt{4/7}\kappa(\eta,1),\eta,\mu^+)+\overline{\mathrm{SLC}}(\sqrt{4/7}\kappa(\eta,1),\mu^+,1) \\
&\leq 1+2\left(1+\log\left(\frac{1}{\mu^+}\right)\right)\left(\frac{\kappa(\eta,1)}{\sqrt{4/7}\kappa(\eta,1)}\right)^{O(1)}\overline{\mathrm{SLC}}(\kappa(\eta,1),\mu^+,1) \\
&= O\left(\overline{\mathrm{SLC}}(\kappa(\eta,1),\mu^+,1)\right) \\
&= O\left(\overline{\mathrm{SLC}}(\kappa(\eta,1),\eta,1)\right) \\
&= O(1)\,.
\end{aligned}$$

We can argue analogously in the case that $\kappa(\eta,\mu^-) \geq \kappa(\mu^-,1)$ to then observe that

$$\overline{\mathrm{SLC}}(\sqrt{4/7}\kappa(\eta,1),\eta,1) = O\left(\overline{\mathrm{SLC}}(\kappa(\eta,1),\eta,\mu^-)\right) = O\left(\overline{\mathrm{SLC}}(\kappa(\eta,1),\eta,1)\right) = O(1)\,.$$

$\kappa(\mu^-,1) \leq \sqrt{4/7}\kappa(\eta,1)$ and again apply Proposition 6.7 to obtain the same bound. The remaining case is when $\kappa(\eta,\mu^+) > \kappa(\mu^+,1)$ and $\kappa(\eta,\mu^-) < \kappa(\mu^-,1)$. In this case, by continuity of curvature there must by the intermediate value theorem exist $\mu \in [\mu^-,\mu^+]$ such that $\kappa(\eta,\mu) = \kappa(\mu,1)$. For this value we can apply Theorem 6.12 to obtain that

$$\max\{\kappa(\eta,\mu),\kappa(\mu,1)\} = \frac{\kappa(\eta,\mu)+\kappa(\mu,1)}{2} \leq \sqrt{\frac{8}{2\cdot 7}}\kappa(\eta,1) = \sqrt{\frac{4}{7}}\kappa(\eta,1)\,,$$

which would allow us to conclude that

$$\overline{\mathrm{SLC}}(\sqrt{4/7}\kappa(\eta,1),\eta,1) \leq \overline{\mathrm{SLC}}(\sqrt{4/7}\kappa(\eta,1),\eta,\mu)+\overline{\mathrm{SLC}}(\sqrt{4/7}\kappa(\eta,1),\mu,1) = 2 = O(1)$$

as before. Overall, we concluded in all cases that

$$\overline{\mathrm{SLC}}(\sqrt{4/7}\kappa(\eta,1),\eta,1) = O(1)\,.$$

Iteratively applying this bound proves the theorem. ☐

6.3 Proof of the main theorem

In this section, we prove Theorem 1.7 based on Theorem 6.2, and hence derive our main theorem (Theorem 1.4). In the first step, we need to relate the quantity $\overline{\mathrm{SLC}}(\beta, \mu_1, \mu_0)$ to the straight line complexity $\mathrm{SLC}(\mathcal{N}^2(\beta), \mu_1, \mu_0)$. According to Definition 6.1 and Lemma 2.15(iii), $\overline{\mathrm{SLC}}(\beta, \mu_1, \mu_0)$ is the minimum number of pieces of a piecewise linear curve in $\mathcal{N}^2(\beta)$ between μ_1 and μ_0 that has all breakpoints on the central path. Thus, by definition $\mathrm{SLC}(\mathcal{N}^2(\beta), \mu_1, \mu_0) \le \overline{\mathrm{SLC}}(\beta, \mu_1, \mu_0)$. We now show the approximate reverse direction as follows.

Lemma 6.13 (Path stability)**.** *For $\beta \in (0, 1/128]$, let $z = (x, s) \in \mathcal{N}^2(\beta)$ with $\overline{\mu}(z) = \mu$, $z_1 = (x_1, s_1) \in \mathcal{N}^2(\beta)$ with $\overline{\mu}(z_1) = \mu_1 \le \mu$ and $[z_1, z] \subseteq \mathcal{N}^2(\beta)$. Then we have $[z(\mu_1), z(\mu)] \subseteq \mathcal{N}^2(8\beta)$. Consequently,*

$$\overline{\mathrm{SLC}}(8\beta, \mu_1, \mu_0) \le \mathrm{SLC}(\mathcal{N}^2(\beta), \mu_1, \mu_0)\,.$$

Proof. By Lemma 2.15, $\kappa(z_1, z) \le 2\beta$, which gives $\|(x_1 - x)(s_1 - s)\| \le 2\beta(\sqrt{\mu} + \sqrt{\mu_1})^2$. Hence,

$$\begin{aligned} \|x_1 s + x s_1 - (\mu + \mu_1)\mathbf{1}_n\| &\le \|x_1 s + x s_1 - x_1 s_1 - x s\| + \|x_1 s_1 + x s - (\mu + \mu_1)\mathbf{1}_n\| \\ &\le \|(x_1 - x)(s_1 - s)\| + \|x_1 s_1 - \mu_1 \mathbf{1}_n\| + \|x s - \mu \mathbf{1}_n\| \\ &\le 2\beta(\sqrt{\mu} + \sqrt{\mu_1})^2 + \beta(\mu + \mu_1)\,. \end{aligned} \tag{70}$$

If we now consider $(x(\mu), s(\mu))$ and $(x(\mu_1), s(\mu_1))$, we have

$$\begin{aligned} &\|(x(\mu_1) - x(\mu))(s(\mu_1) - s(\mu))\| \\ &= \|x(\mu_1)s(\mu) + x(\mu)s(\mu_1) - (\mu + \mu_1)\mathbf{1}_n\| \\ &\le \left\| \left(\frac{x(\mu_1)s(\mu)}{x_1 s} - \mathbf{1}_n \right) x_1 s + \left(\frac{x(\mu)s(\mu_1)}{x s_1} - \mathbf{1}_n \right) x s_1 \right\| + \|x_1 s + x s_1 - (\mu + \mu_1)\mathbf{1}_n\| \\ &\le \max\{\|x_1 s\|_\infty, \|x s_1\|_\infty\} \left(\left\| \frac{x(\mu_1)s(\mu)}{x_1 s} - \mathbf{1}_n \right\| + \left\| \frac{x(\mu)s(\mu_1)}{x s_1} - \mathbf{1}_n \right\| \right) + 2\beta(\sqrt{\mu} + \sqrt{\mu_1})^2 + \beta(\mu + \mu_1) \end{aligned} \tag{71}$$

Since $(x, s), (x_1, s_1) \in \mathcal{P}_{>\mathbf{0}} \times \mathcal{D}_{>\mathbf{0}}$, $\max\{\|x_1 s\|_\infty, \|x s_1\|_\infty\} \le \|x_1 s + x s_1\|_\infty \le 2\beta(\sqrt{\mu} + \sqrt{\mu_1})^2 + \beta(\mu + \mu_1) + \mu + \mu_1$ from (70). Using Lemma 2.17,

$$\begin{aligned} \left\| \frac{x(\mu)s(\mu_1)}{x s_1} - \mathbf{1}_n \right\| &\le \left\| \left(\frac{x(\mu)}{x} - \mathbf{1}_n \right) \left(\frac{s(\mu_1)}{s_1} - \mathbf{1}_n \right) \right\| + \left\| \frac{x(\mu)}{x} - \mathbf{1}_n \right\| + \left\| \frac{s(\mu_1)}{s_1} - \mathbf{1}_n \right\| \\ &\le \frac{\beta^2}{(1-\beta)^2} + \frac{2\beta}{1-\beta} \le \frac{9\beta}{4}\,, \end{aligned}$$

and the same bound holds analogously for $\left\| \frac{x(\mu_1)s(\mu)}{x_1 s} - \mathbf{1}_n \right\|$. Therefore, using also $\mu + \mu_1 \le (\sqrt{\mu} + \sqrt{\mu_1})^2$, we get

$$\begin{aligned} \|(x(\mu_1) - x(\mu))(s(\mu_1) - s(\mu))\| &\le 4.5\beta \cdot (2\beta(\sqrt{\mu} + \sqrt{\mu_1})^2 + \beta(\mu + \mu_1) + \mu + \mu_1) + 2\beta(\sqrt{\mu} + \sqrt{\mu_1})^2 + \beta(\mu + \mu_1) \\ &\le (7.5\beta + 13.5\beta^2)(\sqrt{\mu} + \sqrt{\mu_1})^2 \le 8\beta(\sqrt{\mu} + \sqrt{\mu_1})^2\,. \end{aligned}$$

Hence, $\kappa(z(\mu_1), z(\mu)) \le 8\beta$, and the result follows from Lemma 2.15(iii). The final conclusion follows by taking the optimal piecewise linear curve in the definition of $\mathrm{SLC}(\mathcal{N}^2(\beta), \mu_1, \mu_0)$, and moving all breakpoints to the central path. By the above, this gives a piecewise linear curve with the same number of breakpoints with $\kappa(z, z') \le 8\beta$ for each segment. Hence, the bound follows. □

Proof of Theorem 1.7. From Theorem 6.2, we have $\overline{\mathrm{SLC}}(\beta_1, \mu_1, \mu) \le O\left(\left(\frac{8\beta_2}{\beta_1}\right)^C\right) \cdot \overline{\mathrm{SLC}}(8\beta_2, \mu_1, \mu)$ for some absolute constant $C \ge 0$. Combining with Lemma 6.13, we get

$$\mathrm{SLC}(\mathcal{N}^2(\beta_1), \mu_1, \mu_0) \le \overline{\mathrm{SLC}}(\beta_1, \mu_1, \mu_0) \le \bar{C} 8^C \left(\frac{\beta_2}{\beta_1}\right)^C \cdot \overline{\mathrm{SLC}}(8\beta_2, \mu_1, \mu_0) \le \bar{C} 8^C \left(\frac{\beta_2}{\beta_1}\right)^C \cdot \mathrm{SLC}(\mathcal{N}^2(\beta_2), \mu_1, \mu_0)\,,$$

where $\bar{C} > 1$ is a universal constant. □

We end this section by proving our main theorem, showing that the TR2-IPM is indeed optimal in $\mathcal{N}^2(\beta)$.

Proof of Theorem 1.4. Recall that in order to obtain a path-following method in $\mathcal{N}^2(\bar{\beta})$, we use TR2-IPM for $\beta = \bar{\beta}/82$. Given a starting $z_0 \in \mathcal{N}^2(\bar{\beta}) = \mathcal{N}^2(82\beta)$ with $\overline{\mu}(z_0) \in [\mu_1, \mu_0]$, it takes $O(1)$ many corrector steps to reach $z_0' \in \mathcal{N}^2(\beta)$ with $\overline{\mu}(z_0') = \overline{\mu}(z_0)$. Using Theorem 1.6, TR2-IPM takes at most $\mathrm{SLC}(\mathcal{N}^2(\beta), \mu_1, \mu_0)$ many iterations to reach $z_1 \in \mathcal{N}^2(\beta)$ with $\overline{\mu}(z_1) \le \mu_1$ and all iterates and the line segments between consecutive iterates stay in $\mathcal{N}^2(\bar{\beta})$. Then, the theorem follows from $\mathrm{SLC}(\mathcal{N}^2(\beta), \mu_1, \mu_0) = O(\mathrm{SLC}(\mathcal{N}^2(\bar{\beta}), \mu_1, \mu_0))$ by Theorem 1.7. Finally, the running time bound of each iteration follows by Theorem 1.9 and Proposition 2.11. □

7 Solving the ℓ_2-Trust Region Problem

We restate the ℓ_2-trust region problem (TR-2) here for convenience.

$$\min \|y_J\|^2 \quad \text{s.t.} \quad \mathbf{B}y = b,\ \|y_I\|_2 \le 1, \tag{TR-2}$$

We will explain its relation to the general trust region problem and its applications in Section 7.5. Our main goal of this section is to prove Theorem 1.9. We recall that $y \in \mathbb{R}^n$ is a δ-optimal solution of (TR-2) if it satisfies $\mathbf{B}y = b$, $\|y_I\|^2 < 1+\delta$ and $\|y_J\|^2 \le \mathrm{OPT}_2$.

Theorem 1.9. *For $\delta \in (0,1)$, there exists an algorithm* TR2-SOLVE$(\mathbf{B}, b, I, J, \delta)$ *that finds a δ-optimal solution to (TR-2) or certifies infeasibility in strongly polynomial time. Let $n' := \min\{|I|, |J|\}$. The number of arithmetic operations is dominated by $O(\log(n') + \log\log(|I|/\delta))$ many linear system solves of size n, plus the time to get a $O\left(2^{2n'}\right)$ multiplicative approximation of the eigenvalues of a positive semidefinite matrix of size n'. Further,* TR2-SOLVE$(\mathbf{B}, b, I, J, \delta)$ *can be implemented in randomized $O(n^\omega(\log(n') + \log\log(|I|/\delta))$ or in deterministic $O\left((n')^3 + n^\omega(\log(n') + \log\log(|I|/\delta))\right)$ time.*

7.1 The Lagrangian

For the description and analysis of TR2-SOLVE$(\mathbf{B}, b, I, J, \delta)$, we introduce the following potential function.

Definition 7.1 (The Lagrangian)**.** Let $\mathbf{B} \in \mathbb{R}^{m\times n}$ with $\mathrm{rk}(\mathbf{B}) = m$, $b \in \mathbb{R}^m$ and $I \cup J = [n]$ be a partition. For $\lambda > 0$, we define

$$y(\lambda) := \arg\min_{\mathbf{B}y=b} \|y_J\|^2 + \lambda\|y_I\|^2 .$$

We extend $y(.)$ to the limit cases $\lambda = 0$ and $\lambda = \infty$ as follows. Let $\mathcal{L}_J := \arg\min_{\mathbf{B}y=b} \|y_J\|^2$ and $\mathcal{L}_I := \arg\min_{\mathbf{B}y=b} \|y_I\|^2$, and we define

$$y(0) := \arg\min_{y\in\mathcal{L}_J} \|y_I\|^2 , \quad y(\infty) := \arg\min_{y\in\mathcal{L}_I} \|y_J\|^2 . \tag{72}$$

For $\lambda \in \mathbb{R}_{\ge \mathbf{0}} \cup \{\infty\}$, we denote $\psi(\lambda) := \|y_I(\lambda)\|^2$.

Note that (TR-2) is feasible if and only if $\psi(\infty) \le 1$ because $\psi(\infty) = \min_{y:\mathbf{B}y=b} \|y_I\|^2$, hence the feasibility of (TR-2) can be checked by computing $y(\infty)$. From now on we assume that $\psi(\infty) \le 1$. Further, if $\psi(0) \le 1$ then $y(0)$ is clearly an optimal solution to (TR-2). The next lemma asserts the correctness of the definition in (72) and the strongly polynomial computability of $y(0)$ and $y(\infty)$.

Lemma 7.2. *The definitions of $y(0)$ and $\psi(0)$, $y(\infty)$ and $\psi(\infty)$ give continuous extensions of $y(\lambda)$ and $\psi(\lambda)$ at $\lambda = 0$ and $\lambda = \infty$ respectively. The vectors $y(0)$ and $y(\infty)$ can be computed via linear system solves.*

Proof. The first half of the result follows from [40, Lemma 2.10]. Since both $y(0)$ and $y(\infty)$ in (72) correspond to the solutions of two layered least square problems as in [73], they can be formulated as solving linear systems of equations. See Lemma 5.2 of [40] and Section 4.1 of [46]. □

We will prove the following monotonicity properties; the proof follows from the form of $\psi(\lambda)$ derived in (74).

Lemma 7.3. *$\psi(\lambda)$ is strictly convex and strictly decreasing on $\mathbb{R}_{\ge \mathbf{0}}$. For any $0 < \lambda_1 < \lambda_2$, $\lambda_1^2\psi(\lambda_1) < \lambda_2^2\psi(\lambda_2)$.*

Throughout, we let λ^* denote the unique value such that $\psi(\lambda^*) = 1$. Thus, $y(\lambda^*)$ is optimal to (TR-2) by Lagrange duality. Our TR2-SOLVE$(\mathbf{B}, b, I, J, \delta)$ algorithm essentially searches for λ such that $1 \le \psi(\lambda) < 1+\delta$. Then it follows that $y(\lambda)$ is a δ-optimal solution, as we will show at the end of Section 7.3.

Let $\mathbf{D}(\lambda) \in \mathbb{R}^{n\times n}$ be the diagonal matrix with entries $\mathbf{D}(\lambda)_{ii} = 1$ for any $i \in I$ and $\mathbf{D}(\lambda)_{ii} = \lambda$ for any $i \in J$.

Lemma 7.4. *For $\lambda > 0$, we can write*

$$y(\lambda) = \mathbf{D}(\lambda)\mathbf{B}^\top \left(\mathbf{B}\mathbf{D}(\lambda)\mathbf{B}^\top\right)^{-1} b \quad \text{and} \quad \psi(\lambda) = \left\|\mathbf{B}_I^\top \left(\mathbf{B}\mathbf{D}(\lambda)\mathbf{B}^\top\right)^{-1} b\right\|^2 .$$

Proof. Note that we can equivalently write $y(\lambda) = \arg\min_{\mathbf{B}y=b} \|y_I\|^2 + \frac{1}{\lambda}\|y_J\|^2$ as $\lambda > 0$. Let $z = \mathbf{D}(\lambda)^{-\frac{1}{2}}y$. Since $\mathbf{B}\mathbf{D}(\lambda)^{\frac{1}{2}}$ has full row rank, the least norm solution to $\mathbf{B}\mathbf{D}(\lambda)^{\frac{1}{2}}z = b$ is $z^* = \mathbf{D}(\lambda)^{\frac{1}{2}}\mathbf{B}^\top \left(\mathbf{B}\mathbf{D}(\lambda)\mathbf{B}^\top\right)^{-1} b$. Hence, $y(\lambda) = \mathbf{D}(\lambda)^{\frac{1}{2}}z^* = \mathbf{D}(\lambda)\mathbf{B}^\top \left(\mathbf{B}\mathbf{D}(\lambda)\mathbf{B}^\top\right)^{-1} b$, and $y_I(\lambda) = \mathbf{B}_I^\top \left(\mathbf{B}\mathbf{D}(\lambda)\mathbf{B}^\top\right)^{-1} b$. □

The first ingredient of our algorithm is formulated in the following lemma.

Lemma 7.5. *There exists an algorithm, given in* HybridNewton$(\lambda_1, \lambda_2, \delta)$*, such that given* $\delta > 0$ *and* $0 < \lambda_1 < \lambda_2$ *with* $\psi(\lambda_1) \geq 1 \geq \psi(\lambda_2)$*, it returns a value* $\lambda \in [\lambda_1, \lambda_2]$ *such that* $\psi(\lambda) \in [1, 1+\delta)$*. The algorithm runs in strongly polynomial time using* $O\left(\log\log\left(\frac{\lambda_2}{\lambda_1}\right) + \log\log\left(\frac{1}{\delta}\right)\right)$ *iterations, and in each iteration the number of operations is dominated by* $O(1)$ *linear system solves of size* $\min\{m, |I|, |J|\}$*.*

We describe this subroutine in Section 7.4, which follows a similar approach of standard hybrid Newton's method in the literature, see e.g. [15, 77]. It first performs a geometric binary search to narrow down the search interval by finding $\hat{\lambda} \in [\lambda_1, \lambda_2)$ with $\psi(\hat{\lambda}) \geq 1 \geq \psi(1.1\hat{\lambda})$. Starting from $\hat{\lambda}$, we then perform Newton's method for root finding of $\psi(\lambda) = 1$, which exhibits quadratic convergence as now we start sufficiently close to λ^*.

7.2 The critical points

We obtain a strongly polynomial algorithm by first finding suitable values λ_1 and λ_2 with $\psi(\lambda_1) \geq 1 \geq \psi(\lambda_2)$ such that $\lambda_2/\lambda_1 = 2^{\mathrm{poly(n)}}/\mathrm{poly}(\delta)$ and then applying HybridNewton$(\lambda_1, \lambda_2, \delta)$ as in Lemma 7.5.

In order to obtain such values, we define *critical points* that are helpful in describing the function $\psi(\lambda)$. For this, we follow the standard approach in [15] to express ψ as an analytic function in λ.

Throughout, let $\mathbf{H} \coloneqq \mathbf{B}\mathbf{B}^\top \in \mathbb{R}^{m\times m}$. Since $\mathrm{rk}(\mathbf{B}) = m$, we have $\mathbf{H} \succ \mathbf{0}$, and hence its principal matrix square root $\mathbf{H}^{-\frac{1}{2}}$ exists.[6] Consider an orthogonal eigenvalue decomposition of the following matrix:

$$\mathbf{H}^{-\frac{1}{2}}\mathbf{B}_J\mathbf{B}_J^\top\mathbf{H}^{-\frac{1}{2}} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^\top .$$

Here, $\mathbf{V} \in \mathbb{R}^{m\times m}$ is orthogonal, $\mathbf{\Lambda}$ is diagonal with entries $(\mu_k)_{k=1}^m$; these are the eigenvalues of $\mathbf{H}^{-\frac{1}{2}}\mathbf{B}_J\mathbf{B}_J^\top\mathbf{H}^{-\frac{1}{2}}$ in non-increasing order. We also immediately get $\mathbf{H}^{-\frac{1}{2}}\mathbf{B}_I\mathbf{B}_I^\top\mathbf{H}^{-\frac{1}{2}} = \mathbf{I}_m - \mathbf{H}^{-\frac{1}{2}}\mathbf{B}_J\mathbf{B}_J^\top\mathbf{H}^{-\frac{1}{2}} = \mathbf{V}(\mathbf{I}_m - \mathbf{\Lambda})\mathbf{V}^\top$ as its orthogonal eigenvalue decomposition with the eigenvalues being $(1-\mu_k)_{k=1}^m$.

Proposition 7.6. *For all* $k \in [m]$*,* $0 \leq \mu_k \leq 1$*.*

Proof. This follows from applying the monotonicity of eigenvalue to the following Loewner order:

$$\mathbf{0} \preceq \mathbf{H}^{-\frac{1}{2}}\mathbf{B}_J\mathbf{B}_J\mathbf{H}^{-\frac{1}{2}} \preceq \mathbf{H}^{-\frac{1}{2}}\mathbf{B}_J\mathbf{B}_J\mathbf{H}^{-\frac{1}{2}} + \mathbf{H}^{-\frac{1}{2}}\mathbf{B}_I\mathbf{B}_I\mathbf{H}^{-\frac{1}{2}} = \mathbf{I}_m .$$

□

To express $\psi(\lambda)$ as a univariate function in λ, we first derive the following identity:

$$\left(\mathbf{B}\mathbf{D}(\lambda)\mathbf{B}^\top\right)^{-1} = \mathbf{H}^{-\frac{1}{2}}\left(\mathbf{H}^{-\frac{1}{2}}\mathbf{B}_I\mathbf{B}_I^\top\mathbf{H}^{-\frac{1}{2}} + \lambda\mathbf{H}^{-\frac{1}{2}}\mathbf{B}_J\mathbf{B}_J^\top\mathbf{H}^{-\frac{1}{2}}\right)^{-1}\mathbf{H}^{-\frac{1}{2}} = \mathbf{H}^{-\frac{1}{2}}\mathbf{V}\left(\mathbf{I} - \mathbf{\Lambda} + \lambda\mathbf{\Lambda}\right)^{-1}\mathbf{V}^\top\mathbf{H}^{-\frac{1}{2}} . \quad (73)$$

Then, following Lemma 7.4 and defining $w \coloneqq \mathbf{V}^\top\mathbf{H}^{-\frac{1}{2}}b \in \mathbb{R}^m$, we get

$$\begin{aligned}\psi(\lambda) &= \left\|\mathbf{B}_I^\top\left(\mathbf{B}\mathbf{D}(\lambda)\mathbf{B}^\top\right)^{-1}b\right\|^2 \\ &= \left\|\mathbf{B}_I^\top\mathbf{H}^{-\frac{1}{2}}\mathbf{V}(\mathbf{I} - \mathbf{\Lambda} + \lambda\mathbf{\Lambda})^{-1}\mathbf{V}^\top\mathbf{H}^{-\frac{1}{2}}b\right\|^2 \\ &= b^\top\mathbf{H}^{-\frac{1}{2}}\mathbf{V}\mathrm{diag}\left(\frac{1-\mu_k}{(1-\mu_k+\lambda\mu_k)^2}\right)\mathbf{V}^\top\mathbf{H}^{-\frac{1}{2}}b \\ &= \sum_{k:0<\mu_k<1}\frac{(1-\mu_k)w_k^2}{(1-\mu_k+\lambda\mu_k)^2} + \sum_{k:\mu_k=0} w_k^2 .\end{aligned} \quad (74)$$

The above form of $\psi(\lambda)$ yields the proof of Lemma 7.3. As indicated in (74), the values $\frac{1-\mu_k}{\mu_k}$ for $\mu_k \in (0,1)$ are critical in determining the behavior of $\psi(\lambda)$. This is captured in the following definition.

Definition 7.7 (Critical Points)**.** Let $\mathcal{P} \coloneqq \{k \in [m] : \mu_k \in (0,1)\}$ and $\beta_k \coloneqq \frac{1-\mu_k}{\mu_k}$ for each $k \in \mathcal{P}$. We call the β_k values *critical points* of ψ on $\mathbb{R}_{>\mathbf{0}}$

[6] We note that $\mathbf{H}^{-\frac{1}{2}}$ does not need to be computed in the algorithm.

Approximating the critical points. We next show that the critical points coincide with the non-zero eigenvalues of two associated matrices, and explain how the critical points can be approximately computed.

Lemma 7.8. *$(\beta_k)_{k\in\mathcal{P}}$ coincides with the positive spectrum of*

$$\mathbf{M}_1 := \left(\mathbf{B}^\dagger \mathbf{B}_J\right)^\dagger \left(\mathbf{I} - \mathbf{B}^\dagger \mathbf{B}_J \left(\mathbf{B}^\dagger \mathbf{B}_J\right)^\top\right) \left(\mathbf{B}^\dagger \mathbf{B}_J\right)^{\dagger^\top} \in \mathbb{R}^{|J|\times|J|}$$

and also the reciprocal of the positive spectrum of

$$\mathbf{M}_2 := \left(\mathbf{B}^\dagger \mathbf{B}_I\right)^\dagger \left(\mathbf{I} - \mathbf{B}^\dagger \mathbf{B}_I \left(\mathbf{B}^\dagger \mathbf{B}_I\right)^\top\right) \left(\mathbf{B}^\dagger \mathbf{B}_I\right)^{\dagger^\top} \in \mathbb{R}^{|I|\times|I|}.$$

Both $\mathbf{M}_1$ and $\mathbf{M}_2$ are positive semidefinite.

Lemma 7.8 is a result of the following proposition.

Proposition 7.9. *Let $\mathbf{X} \in \mathbb{R}^{n\times p}$, $n \geq p$ and the singular values of $\mathbf{X}$ be $(\sigma_i)_{i=1}^p$. Then the eigenvalues of the matrix $\mathbf{X}^\dagger \left(\mathbf{I} - \mathbf{X}\mathbf{X}^\top\right) \mathbf{X}^{\dagger^\top} \in \mathbb{R}^{p\times p}$ consist of exactly $\frac{1-\sigma_i^2}{\sigma_i^2}$ for any $\sigma_i > 0$, and 0 for any $\sigma_i = 0$.*

Proof. Let $r = \mathrm{rk}(\mathbf{X})$ and $\mathbf{X} = \mathbf{U}_r \Sigma_r \mathbf{V}_r^\top$ be its singular value decomposition where $\mathbf{U}_r \in \mathbb{R}^{n\times r}$ and $\mathbf{V}_r \in \mathbb{R}^{p\times r}$ are orthonormal, and $\mathbf{\Sigma}_r \succ \mathbf{0}$ is diagonal with the positive singular values of $\mathbf{X}$. Then $\mathbf{X}^\dagger = \mathbf{V}_r \mathbf{\Sigma}_r^{-1} \mathbf{U}_r^\top$. Let $\mathbf{V}_\perp \in \mathbb{R}^{p\times(p-r)}$ be the orthogonal complement of $\mathbf{V}_r$, then

$$\mathbf{X}^\dagger \left(\mathbf{I} - \mathbf{X}\mathbf{X}^\top\right) \mathbf{X}^{\dagger^\top} = \mathbf{V}_r \mathbf{\Sigma}_r^{-1} \mathbf{U}_r^\top \left(\mathbf{I} - \mathbf{U}_r \mathbf{\Sigma}_r^2 \mathbf{U}_r^\top\right) \mathbf{U}_r \mathbf{\Sigma}_r^{-1} \mathbf{V}_r^\top = \begin{bmatrix} \mathbf{V}_r & \mathbf{V}_\perp \end{bmatrix} \begin{bmatrix} \mathbf{\Sigma}_r^{-2}\left(\mathbf{I}_r - \mathbf{\Sigma}_r^2\right) & \\ & \mathbf{0}_{p-r} \end{bmatrix} \begin{bmatrix} \mathbf{V}_r^\top \\ \mathbf{V}_\perp^\top \end{bmatrix}$$

which is an orthogonal eigenvalue decomposition of the matrix, and hence the result follows. □

Proof of Lemma 7.8. $\mathbf{M}_1, \mathbf{M}_2 \succeq \mathbf{0}$ follows from Proposition 7.6 and 7.9. Using $\mathrm{eig}(\mathbf{Y}\mathbf{Z}) = \mathrm{eig}(\mathbf{Z}\mathbf{Y})$, the positive spectrum of $\mathbf{H}^{-\frac{1}{2}} \mathbf{B}_J \mathbf{B}_J^\top \mathbf{H}^{-\frac{1}{2}}$ coincides with the positive spectrum of $\mathbf{B}_J^\top (\mathbf{B}\mathbf{B}^\top)^{-1} \mathbf{B}_J = \left(\mathbf{B}^\dagger \mathbf{B}_J\right)^\top \mathbf{B}^\dagger \mathbf{B}_J$. Hence, applying Proposition 7.9 with $\mathbf{X} = \mathbf{B}^\dagger \mathbf{B}_J$, the spectrum of $\mathbf{M}_1$ consists exactly $\frac{1-\sigma_i^2}{\sigma_i^2} = \frac{1-\mu_i}{\mu_i}$ for any $\mu_i > 0$, and 0 for any $\mu_i = 0$. Similarly, since the positive spectrum of $\mathbf{H}^{-\frac{1}{2}} \mathbf{B}_I \mathbf{B}_I^\top \mathbf{H}^{-\frac{1}{2}}$ coincide with that of $\left(\mathbf{B}^\dagger \mathbf{B}_I\right)^\top \mathbf{B}^\dagger \mathbf{B}_I$, if we apply Proposition 7.9 now with $\mathbf{X} = \mathbf{B}^\dagger \mathbf{B}_I$, the spectrum of $\mathbf{M}_2$ consists exactly $\frac{1-\sigma_i^2}{\sigma_i^2} = \frac{1-(1-\mu_i)}{1-\mu_i} = \frac{\mu_i}{1-\mu_i}$ for any $\mu_i < 1$ and 0 for any $\mu_i = 1$. □

We are now ready to describe how we compute in strongly polynomial time a sequence of approximate critical points that are multiplicatively close to the actual critical points.

Lemma 7.10. *For $\varrho > 1$, it takes one pivoted Cholesky factorization of size $\min\{|I|, |J|\}$ to compute ϱ-approximate eigenvalues $(\hat{\beta}_k)_{k\in\mathcal{P}}$ such that $\varrho^{-1}\hat{\beta}_k \leq \beta_k \leq \varrho\hat{\beta}_k$ for critical points $(\beta_k)_{k\in\mathcal{P}}$.*

Proof. If $|I| \geq |J|$, we compute the pivoted Cholesky factorization of $\mathbf{M}_1$ as in Lemma C.1, which gives approximate eigenvalues $(\hat{\beta}_k)_{k\in\mathcal{P}}$ such that $\varrho^{-1}\hat{\beta}_k \leq \beta_k \leq \varrho\hat{\beta}_k$ by Corollary C.5. If $|J| \geq |I|$ we compute the pivoted Cholesky factorization of $\mathbf{M}_2$ instead which gives approximate eigenvalues $(1/\hat{\beta}_k)_{\in\mathcal{P}}$ such that $\varrho^{-1}\hat{\beta}_k \leq \beta_k \leq \varrho\hat{\beta}_k$. Hence, the overall running time is dominated by one pivoted Cholesky factorization for a matrix of size $\min\{|I|, |J|\}$. □

Remark 7.11. We use ϱ to denote the approximation factor from computing the eigenvalues to allow the use of algorithms other than the pivoted Cholesky factorization that give better multiplicative approximation.

7.3 The TR-Solve Algorithm

In this subsection, we fully describe our TR2-SOLVE$(\mathbf{B}, b, I, J, \delta)$ algorithm; the pseudocode is given in Algorithm 3. We explain how the algorithm finds a δ-optimal solution to (TR-2) and also the number of arithmetic operations it needs in strongly polynomial time, and hence prove Theorem 1.9.

The TR2-SOLVE$(\mathbf{B}, b, I, J, \delta)$ algorithm first checks the values $\psi(\infty)$ and $\psi(0)$. When $\psi(\infty) > 1$, (TR-2) is infeasible. When $\psi(0) < 1 + \delta$, we return $y(0)$ as a δ-optimal solution to (TR-2). For the rest of the algorithm, $\psi(0) \geq 1 + \delta$ and $\psi(\infty) \leq 1$. We compute the approximate critical points $\hat{\beta}_k$ as in Lemma 7.10, and let $\mathcal{P}^*$ include 0, ∞, and $\nu\hat{\beta}_k/\varrho$, $\varrho\hat{\beta}_k/\nu$ for all the $\hat{\beta}_k$ computed; we set $\nu = \delta/4$. We then find consecutive $\lambda_1 < \lambda_2$ in $\mathcal{P}^*$ such that $\psi(\lambda_1) \geq 1 \geq \psi(\lambda_2)$. If $\psi(\lambda_1) < 1 + \delta$, we can terminate with $y = y(\lambda_1)$. If $\lambda_2 \leq \varrho^2\lambda_1/\nu^2$ i.e. λ_1 and λ_2 are

Algorithm 3: TR2-Solve($\mathbf{B}, b, I, J, \delta$)

Input : An instance of (TR-2) with constraint matrix $\mathbf{B} \in \mathbb{R}^{m\times n}$, $\text{rk}(\mathbf{B}) = m$, $b \in \mathbb{R}^m$, a partition $I \cup J = [n]$, $\delta \in (0,1)$ and a subroutine to compute ϱ-approximate eigenvalues.

Output: a δ-optimal solution $y \in \mathbb{R}^n$ to (TR-2) or infeasibility conclusion.

1 Compute $y(\infty)$ and $y(0)$;
2 **if** $\psi(\infty) > 1$ **then return** Infeasible and Terminate;
3 **if** $\psi(0) < 1+\delta$ **then return** $y = y(0)$ and Terminate;
4 $\nu \leftarrow \delta/4$;
5 Compute the ϱ-approximate critical points $(\hat\beta_k)_{k\in\mathcal{P}}$ according to Lemma 7.10;
6 $\mathcal{P}^* \leftarrow \left\{\frac{\nu\hat\beta_k}{\varrho}, \frac{\varrho\hat\beta_k}{\nu} : k \in \mathcal{P}\right\} \cup \{0,\infty\}$;
7 Use binary search in $\mathcal{P}^*$ to find two consecutive values $\lambda_1 < \lambda_2$ that $\psi(\lambda_1) > 1 \geq \psi(\lambda_2)$;
8 **if** $\psi(\lambda_1) < 1+\delta$ **then return** $y = y(\lambda_1)$ and Terminate;
9 **if** $\lambda_2 \leq \varrho^2\lambda_1/\nu^2$ **then** $\lambda \leftarrow$ HybridNewton($\lambda_1, \lambda_2, \delta$) ;
10 **else**
11 $\quad\hat\lambda_1 \leftarrow \max\left\{\lambda_1, \frac{\lambda_1\|y_I(\lambda_1)\|_\infty}{2}\right\}$, $\hat\lambda_2 \leftarrow \min\left\{\lambda_2, \frac{2\lambda_1\|y_I(\lambda_1)\|_1}{\delta}\right\}$;
12 $\quad$**if** $\psi(\hat\lambda_2) > 1$ **then** $\lambda \leftarrow \hat\lambda_2$;
13 $\quad$**else** $\lambda \leftarrow$ HybridNewton($\hat\lambda_1, \hat\lambda_2, \delta$);
14 **return** $y = y(\lambda)$

'close', HybridNewton($\lambda_1, \lambda_2, \delta$) is called, and Lemma 7.5 asserts that it returns a value λ with $1 \leq \psi(\lambda) < 1+\delta$ in strongly polynomial time.

The remaining case is when $\lambda_2 > \varrho^2\lambda_1/\nu^2$ i.e. λ_1 and λ_2 are far away. We then define new values $\hat\lambda_1$ and $\hat\lambda_2$ in line 11. Clearly, $\hat\lambda_2/\hat\lambda_1 \leq 4|I|/\delta$. The next two lemmas guarantee that $\psi(\hat\lambda_1) > 1$ and $\psi(\hat\lambda_2) < 1+\delta$. Therefore, we can again call HybridNewton($\lambda_1, \lambda_2, \delta$) on $[\hat\lambda_1, \hat\lambda_2]$ to find a value λ with $1 \leq \psi(\lambda) < 1+\delta$.

The key part of the analysis is based on that no critical points can lie between $\lambda_1\nu$ and λ_2/ν by the construction of $\mathcal{P}^*$. This enables us to approximate $\psi(\lambda)$ with a simpler function $p + q/\lambda^2$ on $[\lambda_1, \lambda_2]$.

Lemma 7.12. *Let $\nu \in (0,1)$ and $0 \leq \lambda_1 < \lambda_2$. Assume that for all critical points $(\beta_k)_{k\in\mathcal{P}}$, either $\beta_k \geq \frac{\lambda_2}{\nu}$ or $\beta_k \leq \lambda_1\nu$. Then there exists $p, q \geq 0$ such that*

$$\frac{1}{(1+\nu)^2}\left(p + \frac{q}{\lambda^2}\right) \leq \psi(\lambda) \leq p + \frac{q}{\lambda^2}, \quad \forall\lambda \in [\lambda_1, \lambda_2].$$

Further, if $\lambda_1 = 0$, then $q = 0$ must hold.

Proof. For any $\beta_k \geq \lambda_2$, we have $\beta_k\nu \geq \lambda_2 \geq \lambda$, so $\frac{1}{(\beta_k+\lambda)^2} \in \left[\frac{1}{(1+\nu)^2\beta_k^2}, \frac{1}{\beta_k^2}\right]$. For any $\beta_k \leq \lambda_1$, we have $\beta_k \leq \lambda_1\nu \leq \lambda\nu$, so $\frac{1}{(\beta_k+\lambda)^2} \in \left[\frac{1}{\lambda^2(1+\nu)^2}, \frac{1}{\lambda^2}\right]$. Since there is no β_k contained in $\left(\lambda_1\nu, \frac{\lambda_2}{\nu}\right)$, following (74), we have

$$\begin{aligned}\psi(\lambda) &= \sum_{k:0<\mu_k<1} \frac{(1-\mu_k)w_k^2}{\mu_k^2(\beta_k+\lambda)^2} + \sum_{k:\mu_k=0} w_k^2 \\ &= \sum_{k:\lambda_2\leq\beta_k<\infty} \frac{\beta_k w_k^2}{\mu_k(\beta_k+\lambda)^2} + \sum_{k:0<\beta_k\leq\lambda_1} \frac{\beta_k w_k^2}{\mu_k(\beta_k+\lambda)^2} + \sum_{k:\mu_k=0} w_k^2 \\ &\in \sum_{k:\lambda_2\leq\beta_k<\infty} \frac{w_k^2}{\mu_k\beta_k}\left[\frac{1}{(1+\nu)^2}, 1\right] + \sum_{k:0<\beta_k\leq\lambda_1} \frac{\beta_k w_k^2}{\mu_k\lambda^2}\left[\frac{1}{(1+\nu)^2}, 1\right] + \sum_{k:\mu_k=0} w_k^2 .\end{aligned} \tag{75}$$

Taking $p := \sum_{k:\lambda_2\leq\beta_k<\infty} \frac{w_k^2}{1-\mu_k} + \sum_{k:\mu_k=0} w_k^2$ and $q := \sum_{k:0<\beta_k\leq\lambda_1} \frac{\beta_k w_k^2}{\mu_k}$, we have $\frac{1}{(1+\nu)^2}\left(p + \frac{q}{\lambda^2}\right) \leq \psi(\lambda) \leq p + \frac{q}{\lambda^2}$. It also follows that if $\lambda_1 = 0$, then $q = 0$. ☐

This lemma justifies the choice of $\hat\lambda_1$ and $\hat\lambda_2$ in line 11 of the algorithm.

Lemma 7.13. *Let $\delta \in (0,1)$, $\nu \in (0, \delta/4]$ and $0 \le \lambda_1 < \lambda_2$. Assume that for all critical points $(\beta_k)_{k\in\mathcal{P}}$, either $\beta_k \ge \frac{\lambda_2}{\nu}$ or $\beta_k \le \lambda_1\nu$. Assume further that $\psi(\lambda_1) \ge 1+\delta$ and $\psi(\lambda_2) \le 1$. Then we must have $\lambda_1 > 0$. Further, for*

$$\hat{\lambda}_1 := \max\left\{\lambda_1, \frac{\lambda_1 \|y_I(\lambda_1)\|_\infty}{2}\right\} \quad \text{and} \quad \hat{\lambda}_2 := \min\left\{\lambda_2, \frac{2\lambda_1\|y_I(\lambda_1)\|_1}{\delta}\right\},$$

we have $\psi(\hat{\lambda}_1) > 1$, $\psi(\hat{\lambda}_2) < 1+\delta$, and $1 \le \frac{\hat{\lambda}_2}{\hat{\lambda}_1} \le \frac{4|I|}{\delta}$.

Proof. Let $\gamma_1 := \psi(\lambda_1)$ and $\gamma_2 := \psi(\lambda_2)$. We first assume that $\lambda_1 = 0$. By the previous lemma, $q = 0$, and thus $p/(1+\nu)^2 \le \psi(\lambda) \le p$ for all $\lambda \in [\lambda_1, \lambda_2]$. Since $\gamma_2 \le 1$, it follows that $p \le (1+\nu)^2\gamma_2 < 1+\delta$. This is a contradiction to $p \ge \gamma_1 \ge 1+\delta$. Hence, $\lambda_1 > 0$.

To show $\psi(\hat{\lambda}_1) > 1$, we define $\lambda_1' := \lambda_1\sqrt{\frac{\gamma_1}{1+\delta}}$. The bound clearly holds when $\hat{\lambda}_1 = \lambda_1$ so we assume that $\hat{\lambda}_1 = \lambda_1\|y_I(\lambda_1)\|_\infty/2$. Since

$$\lambda_1' = \frac{\lambda_1\|y_I(\lambda_1)\|}{\sqrt{1+\delta}} \ge \frac{\lambda_1\|y_I(\lambda_1)\|_\infty}{2} = \hat{\lambda}_1\,,$$

it suffices to show $\psi(\lambda_1') > 1$, then $\psi(\hat{\lambda}_1) > 1$ follows from the monotonicity of $\psi(.)$. We first show that $\lambda_1' \in [\lambda_1, \lambda_2]$. Since $\gamma_1 \ge 1+\delta$, we have $\lambda_1' \ge \lambda_1$. Also, by Lemma 7.3 and $\gamma_2 \le 1$,

$$\frac{\lambda_2^2}{\lambda_1^2} > \frac{\gamma_1}{\gamma_2} > \frac{\gamma_1}{1+\delta} = \frac{\lambda_1'^2}{\lambda_1^2}$$

giving $\lambda_1' < \lambda_2$. Hence, we can now apply Lemma 7.12 to get

$$\begin{aligned}
\psi(\lambda_1') &\ge \frac{1}{(1+\nu)^2}\left(p + \frac{q}{\lambda_1'^2}\right)\\
&= \frac{p(1+\delta)}{(1+\nu)^2\gamma_1} + \frac{1}{(1+\nu)^2}\cdot\frac{q}{\lambda_1'^2} + \frac{p}{(1+\nu)^2}\left(1 - \frac{1+\delta}{\gamma_1}\right)\\
&\ge \frac{1+\delta}{(1+\nu)^2\gamma_1}\left(p + \frac{q}{\lambda_1^2}\right)\\
&\ge \frac{1+\delta}{(1+\nu)^2} > 1\,.
\end{aligned}$$

To show $\psi(\hat{\lambda}_2) < 1+\delta$, we define $\lambda_2' := 2\lambda_1\sqrt{\gamma_1/\delta}$. The bound clearly holds when $\hat{\lambda}_2 = \lambda_2$ so we assume that $\hat{\lambda}_2 = 2\lambda_1\|y_I(\lambda_1)\|_1/\delta$. Since

$$\lambda_2' = \frac{2\lambda_1\|y_I(\lambda_1)\|}{\sqrt{\delta}} \le \frac{2\lambda_1\|y_I(\lambda_1)\|_1}{\delta} = \hat{\lambda}_2\,,$$

it again suffices to show that $\psi(\lambda_2') < 1+\delta$ by the monotonicity of $\psi(.)$. Since $p, q \ge 0$, $p \le \gamma_2(1+\nu)^2 \le (1+\nu)^2$ by Lemma 7.12. We have $\lambda_2' < \hat{\lambda}_2 \le \lambda_2$ and also $\lambda_2' > \lambda_1$ as $\gamma_1 \ge 1+\delta$, so we can again apply Lemma 7.12 to get

$$\begin{aligned}
\psi(\lambda_2') \le p + \frac{q}{\lambda_2'^2} &= p + \frac{\delta q}{4\lambda_1^2\gamma_1}\\
&\le p + \frac{\delta}{4\gamma_1}\left(p + \frac{q}{\lambda_1^2}\right)\\
&\le (1+\nu)^2 + \frac{\delta}{4\gamma_1}(1+\nu)^2\gamma_1\\
&\le (1+\nu)^2(1+\delta/4) < 1+\delta\,.
\end{aligned}$$

□

Proof of Theorem 1.9. TR2-Solve$(\mathbf{B}, b, I, J, \delta)$ algorithm either returns Infeasible when $\psi(\infty) > 1$, or $y = y(\lambda)$ for some $\lambda \ge 0$. We first claim that $\lambda_1 > 0$ in the algorithm. Indeed, if $\lambda_1 = 0$, no $(\beta_k)_{k\in\mathcal{P}}$ can lie between $\lambda_1\nu$ and λ_2/ν. This gives a contradiction to Lemma 7.13. Hence, $y(0)$ being returned can only occur in line 3, and it is clearly δ-optimal to (TR-2) in this case.

If $\lambda > 0$ is returned, then $1 \le \psi(\lambda) < 1+\delta$ by Lemma 7.5. Hence, $y(\lambda)$ is δ-feasible to (TR-2) with $\|y_I(\lambda)\|^2 \ge 1$. Let y^* denote the optimal solution of (TR-2). Then, by the optimality of $y(\lambda)$ and feasibility of y^*,

$$\|y_J(\lambda)\|^2 + \lambda \le \|y_J(\lambda)\|^2 + \lambda\|y_I(\lambda)\|^2 \le \|y_J^*\|^2 + \lambda\|y_I^*\|^2 \le \mathrm{OPT}_2 + \lambda \tag{76}$$

which gives $\|y_J(\lambda)\|^2 \le \mathrm{OPT}_2$, and hence $y(\lambda)$ is a δ-optimal solution to (TR-2).

$\psi(\lambda)$ is always computed via linear system solve by Lemma 7.4. Hence, Lemma 7.5 and Lemma C.1 together shows that TR2-Solve$(\mathbf{B}, b, I, J, \delta)$ runs in strongly polynomial time. In particular, HybridNewton$(\lambda_1, \lambda_2, \delta)$ is conducted either in the interval $[\lambda_1, \lambda_2]$ with $\lambda_2 \le \varrho^2\lambda_1/\nu^2$ which requires $O(\log\log(\varrho/\delta))$ many iteration, or in the interval $[\hat{\lambda}_1, \hat{\lambda}_2]$ with $\hat{\lambda}_2/\hat{\lambda}_1 \le 4|I|/\delta$ which requires $O(\log\log(|I|/\delta))$ many iteration. We always have $\varrho \le \min\{|I|,|J|\}2^{2\min\{|I|,|J|\}}$ because of Corollary C.5 for the pivoted Cholesky factorization. Hence, the running time stated in Theorem 1.9 follows. ☐

7.4 The HybridNewton subroutine

In this subsection, we present the subroutine HybridNewton$(\lambda_1, \lambda_2, \delta)$ and prove Lemma 7.5. Recall that for a general function $f \in \mathcal{C}^2(\mathbb{R})$, if an interval $[\lambda_1, \lambda_2]$ such that $f(\lambda_1) \ge f(\lambda_2)$ and an initial guess $\lambda^0 \in [\lambda_1, \lambda_2]$ are given, Newton's method for root finding computes the iterates by

$$\lambda^{t+1} = \lambda^t - \frac{f(\lambda^t)}{f'(\lambda^t)}, \quad t \ge 0$$

until a certain convergence criterion is met. The main part of our subroutine is to apply Newton's method to $\psi(\lambda) - 1 = 0$ in an interval $[\lambda_1, \lambda_2]$ that satisfies $\psi(\lambda_1) \ge 1+\delta$ and $\psi(\lambda_2) \le 1$. We first prove that ψ' on $\mathbb{R}_{>\mathbf{0}}$ can be computed by solving a linear system, so that Newton's method can be implemented in strongly polynomial time.

Lemma 7.14. *For $\lambda > 0$,*

$$\psi'(\lambda) = -2b^\top \left(\mathbf{B}\mathbf{D}(\lambda)\mathbf{B}^\top\right)^{-1} \mathbf{B}_J\mathbf{B}_J^\top \left(\mathbf{B}\mathbf{D}(\lambda)\mathbf{B}^\top\right)^{-1} \mathbf{B}_I\mathbf{B}_I^\top \left(\mathbf{B}\mathbf{D}(\lambda)\mathbf{B}^\top\right)^{-1} b\,. \tag{77}$$

Both $\lambda^3|\psi'(\lambda)|$ and $\lambda^4\psi''(\lambda)$ are strictly increasing on $\mathbb{R}_{>\mathbf{0}}$.

Proof. We differentiate $\psi(\lambda)$ via (74) to get

$$\psi'(\lambda) = -2\sum_{k:0<\mu_k<1} \frac{\mu_k(1-\mu_k)w_k^2}{(1-\mu_k+\lambda\mu_k)^3} \quad \text{and} \quad \psi''(\lambda) = 6\sum_{k:0<\mu_k<1} \frac{\mu_k^2(1-\mu_k)w_k^2}{(1-\mu_k+\lambda\mu_k)^4}\,. \tag{78}$$

Hence, for any $0 < \lambda_1 < \lambda_2$, we have $\lambda_1^3|\psi'(\lambda_1)| < \lambda_2^3|\psi'(\lambda_2)|$ and $\lambda_1^4\psi''(\lambda_1) < \lambda_2^4\psi''(\lambda_2)$. Using (73), we obtain

$$\begin{aligned}
& -2b^\top \left(\mathbf{B}\mathbf{D}(\lambda)\mathbf{B}^\top\right)^{-1} \mathbf{B}_J\mathbf{B}_J^\top \left(\mathbf{B}\mathbf{D}(\lambda)\mathbf{B}^\top\right)^{-1} \mathbf{B}_I\mathbf{B}_I^\top \left(\mathbf{B}\mathbf{D}(\lambda)\mathbf{B}^\top\right)^{-1} b \\
=& -2b^\top\mathbf{H}^{-\frac{1}{2}}\mathbf{V}\left(\mathbf{I}-\mathbf{\Lambda}+\lambda\mathbf{\Lambda}\right)^{-1}\mathbf{V}^\top\mathbf{V}\mathbf{\Lambda}\mathbf{V}^\top\mathbf{V}\left(\mathbf{I}-\mathbf{\Lambda}+\lambda\mathbf{\Lambda}\right)^{-1}\mathbf{V}^\top\mathbf{V}(\mathbf{I}-\mathbf{\Lambda})\mathbf{V}^\top\mathbf{V}\left(\mathbf{I}-\mathbf{\Lambda}+\lambda\mathbf{\Lambda}\right)^{-1}\mathbf{V}^\top\mathbf{H}^{-\frac{1}{2}}b \\
=& -2b^\top\mathbf{H}^{-\frac{1}{2}}\mathbf{V}\mathrm{diag}\left(\frac{\mu_k(1-\mu_k)}{(1-\mu_k+\lambda\mu_k)^3}\right)\mathbf{V}^\top\mathbf{H}^{-\frac{1}{2}}b \\
=& -2\sum_{k:0<\mu_k<1}\frac{\mu_k(1-\mu_k)w_k^2}{(1-\mu_k+\lambda\mu_k)^3}\,.
\end{aligned}$$

☐

We note that both $\log(-\psi')$ and $\log(\psi'')$ are Lipschitz continuous on $\mathbb{R}_{>\mathbf{0}}$, showing the multiplicative stability of ψ' and ψ''. This suggests that Newton's method should have quadratic convergence when close to the root, by Smale's criterion [63]. We give a simple self-contained proof below. Throughout we let λ^* denote the root of $\psi(\lambda) - 1 = 0$ in $[\lambda_1, \lambda_2]$, and λ^t for $t \ge 0$ denote the iterates of Newton's method.

Lemma 7.15. *For $t \ge 0$, if $\lambda^0 \in [\lambda_1, \lambda^*)$ satisfies $\lambda^* \le 1.1\lambda^0$;*

(i) The iterates λ^t satisfy $\lambda^t \in [\lambda^0, \lambda^*)$, $\lambda^t \uparrow \lambda^*$. Moreover,

$$0 \le \frac{\lambda^* - \lambda^{t+1}}{\lambda^*} \le \frac{3}{2}\left(\frac{\lambda^*}{\lambda^0}\right)^4\left(\frac{\lambda^* - \lambda^t}{\lambda^*}\right)^2. \tag{79}$$

Algorithm 4: HybridNewton$(\lambda_1, \lambda_2, \delta)$

Input : $0 < \lambda_1 < \lambda_2$ such that $\psi(\lambda_1) > 1 \geq \psi(\lambda_2)$, $\delta \in (0,1)$
Output: $\lambda \in [\lambda_1, \lambda_2]$ that satisfies $1 \leq \psi(\lambda) < 1+\delta$.

$\hat{\lambda} \leftarrow$ BinarySearch$(\psi - 1, \lambda_1, \lambda_2, 0.1)$;
$t \leftarrow 0$, $\lambda^0 \leftarrow \hat{\lambda}$;
while $\psi(\lambda^t) \geq 1+\delta$ **do**
  Compute $\psi'(\lambda^t)$ by Lemma 7.14;
  $\lambda^{t+1} \leftarrow \lambda^t - \frac{\psi(\lambda^t)-1}{\psi'(\lambda^t)}$;
  $t \leftarrow t+1$;
return λ^t;

(ii) For $t \geq \log\log\left(\frac{4}{\delta}\right)$, $1 < \psi(\lambda^t) < 1+\delta$.

Proof. **Part (i).** $\lambda^t \in [\lambda^0, \lambda^*)$ and $\lambda^t \uparrow \lambda^*$ follows from the strict convexity of ψ. By Taylor's Theorem, for some $\xi \in (\lambda^t, \lambda^*)$,

$$\psi(\lambda^*) - 1 = \psi(\lambda^t) + \psi'(\lambda^t)(\lambda^* - \lambda^t) + \frac{\psi''(\xi)}{2}(\lambda^* - \lambda^t)^2 - 1$$

where $\psi(\lambda^t) - 1 = \psi'(\lambda^t)(\lambda^t - \lambda^{t+1})$. This gives $\lambda^* - \lambda^{t+1} = \left|\frac{\psi''(\xi)}{2\psi'(\lambda^t)}\right|(\lambda^* - \lambda^t)^2$. Using Lemma 7.14,

$$\frac{\psi''(\xi)}{2\left|\psi'(\lambda^t)\right|} \leq \frac{(\lambda^*)^4\psi''(\lambda^*)}{2\xi^4\left|\psi'(\lambda^t)\right|} \leq \frac{(\lambda^*)^4\psi''(\lambda^*)}{2(\lambda^t)^4\left|\psi'(\lambda^*)\right|}\,.$$

For $\lambda \in \mathbb{R}_{>\mathbf{0}}$, $\frac{\mathrm{d}}{\mathrm{d}\lambda}\log(-\psi'(\lambda)) = \frac{\psi''(\lambda)}{\psi'(\lambda)}$. Moreover, for any $0 < x < y$ we have $x^3\left|\psi'(x)\right| \leq y^3\left|\psi'(y)\right|$ by (78), giving $\log(-\psi'(x)) - \log(-\psi'(y)) \leq 3(\log y - \log x)$. Then,

$$\begin{aligned}
\left|\frac{\psi''(\lambda)}{\psi'(\lambda)}\right| = -\frac{\psi''(\lambda)}{\psi'(\lambda)} &= \lim_{h\to 0}\frac{\log(-\psi'(\lambda)) - \log(-\psi'(\lambda+h))}{h} \\
&\leq \lim_{h\to 0}\frac{3(\log(\lambda+h) - \log\lambda)}{h} \\
&= 3\lim_{h\to 0}\frac{\log\left(1+\frac{h}{\lambda}\right)}{h} = \frac{3}{\lambda}\,.
\end{aligned}$$

Therefore, we arrive that

$$\frac{\lambda^* - \lambda^{t+1}}{\lambda^*} \leq \frac{(\lambda^*)^5\psi''(\lambda^*)}{2(\lambda^t)^4\left|\psi'(\lambda^*)\right|}\cdot\left(\frac{\lambda^*-\lambda^t}{\lambda^*}\right)^2 \leq \frac{(\lambda^*)^5}{2(\lambda^t)^4}\cdot\frac{3}{\lambda^*}\left(\frac{\lambda^*-\lambda^t}{\lambda^*}\right)^2 \leq \frac{3}{2}\left(\frac{\lambda^*}{\lambda^0}\right)^4\left(\frac{\lambda^*-\lambda^t}{\lambda^*}\right)^2.$$

Part (ii). Following (79), we use the bound $\frac{3}{2}x^4\left(1-\frac{1}{x}\right) < \frac{1}{2}$ for any $x \in [1, 1.1]$ to get $\frac{\lambda^*-\lambda^t}{\lambda^*} < \frac{1}{2^{2^t}}$ for any $t \geq 0$. Therefore, for $t \geq \log\log\left(\frac{4}{\delta}\right)$, we have $\frac{\lambda^*-\lambda^t}{\lambda^*} < \frac{\delta}{4}$ which gives $\frac{\lambda^*}{\lambda^t} < \frac{1}{1-\delta/4}$. By Lemma 7.3 again,

$$1 = \psi(\lambda^*) < \psi(\lambda^t) \leq \left(\frac{\lambda^*}{\lambda^t}\right)^2\psi(\lambda^*) < \frac{1}{(1-\delta/4)^2} < 1+\delta\,.$$

∎

As in [77], we safeguard Newton's method with an initial guess $\lambda^0 \in [\lambda_1, \lambda^*)$ that satisfies $\lambda^* \leq 1.1\lambda^0$ using a binary search on $[\lambda_1, \lambda_2]$. We give the following binary search subroutine for root finding on a bounded interval, which performs a geometric binary search without taking square roots.

Lemma 7.16. *BinarySearch$(f, \lambda_1, \lambda_2, \varepsilon)$ is correct and takes $O\left(\log\log\left(\frac{\lambda_2}{\lambda_1}\right) + \log\left(\frac{1}{\varepsilon}\right)\right)$ many operations.*

Proof. Since f is decreasing on $[\lambda_1, \lambda_2(1+\varepsilon)]$, $f(\lambda_1(1+\varepsilon)^{2^N}) \leq f(\lambda_2) \leq 0$. Then, by the construction of $\hat{\lambda}$, $f(\hat{\lambda}) > 0$, $\hat{\lambda} \geq \lambda_1$, and also $\hat{\lambda} < \lambda_2$ by the monotonicity of f. Also, $\hat{\lambda}$ satisfies $f(\hat{\lambda}(1+\varepsilon)) \leq 0$ by the termination condition.

Algorithm 5: BinarySearch$(f, \lambda_1, \lambda_2, \varepsilon)$

Input : $0 < \lambda_1 < \lambda_2$ with $f(\lambda_1) > 0 \geq f(\lambda_2)$, $\varepsilon \in (0, 1)$, f a decreasing function on $[\lambda_1, \lambda_2(1+\varepsilon)]$.
Output: $\hat{\lambda} \in [\lambda_1, \lambda_2]$ such that $f(\hat{\lambda}) > 0 \geq f(\hat{\lambda}(1+\varepsilon))$

$N \leftarrow 0$;
while $\lambda_1(1+\varepsilon)^{2^N} < \lambda_2$ **do** $N \leftarrow N + 1$;
$L \leftarrow 0$, $R \leftarrow 2^N$;
while $R > L + 1$ **do**
  $M \leftarrow \frac{L+R}{2}$;
  if $f(\lambda_1(1+\varepsilon)^M) > 0$ **then** $L \leftarrow M$;
  else $R \leftarrow M$;
return $\hat{\lambda} = \lambda_1(1+\varepsilon)^L$

It takes $O\left(\log\log\left(\lambda_2/\lambda_1\right) + \log\left(1/\varepsilon\right)\right)$ to find the desirable N and then the algorithm conducts a binary search in some subset of $\{0, 1, ..., 2^N\}$, which shows the running time.

For computational efficiency, we also need that L and R remain integer throughout. This follows since $R - L$ is divided by 2 in each iteration. The initial value is $R - L = 2^N$, and hence remains an integer power of 2 throughout. This also shows that BinarySearch$(f, \lambda_1, \lambda_2, \varepsilon)$ can be implemented in strongly polynomial time. □

Finally, we conclude this section by proving the correctness and running time of HybridNewton$(\lambda_1, \lambda_2, \delta)$:

Proof of Lemma 7.5. HybridNewton$(\lambda_1, \lambda_2, \delta)$ first calls BinarySearch$(\psi, \lambda_1, \lambda_2, 0.1)$ to find $\lambda^0 \in [\lambda_1, \lambda_2]$ with $\psi(\lambda^0) > 1 \geq \psi(1.1\lambda^0)$, taking $O(\log\log(\lambda_2/\lambda_1))$ many iterations. By the monotonicity of ψ, $\lambda^0 < \lambda^* \leq 1.1\lambda^0$. Then by Lemma 7.15, $\lambda^t \in [\lambda^0, \lambda^*)$ with $1 \leq \psi(\lambda^t) < 1 + \delta$ is returned by the subsequent Newton's method starting at λ^0 in $\log\log(4/\delta)$ many iterations.

In each iteration, HybridNewton$(\lambda_1, \lambda_2, \delta)$ at most computes $\psi(\lambda^t)$ by Lemma 7.4 and $\psi'(\lambda^t)$ by Lemma 7.14 for a fixed $\lambda^t \in [\lambda_1, \lambda_2]$, which amounts to $O(1)$ linear system solves of size m. As a result, we have the strongly polynomial computability. We can adopt Woodbury matrix identity to compute $\psi(\lambda^{t+1})$ and $\psi'(\lambda^{t+1})$ based on the previous iterate $\psi(\lambda^t)$ and $\psi'(\lambda^t)$ in HybridNewton$(\lambda_1, \lambda_2, \delta)$. One can verify that now we have $O(1)$ linear system solves of size $\min\{m, |I|, |J|\}$ in each iteration. This results in the stated running time in Lemma 7.5. □

7.5 General Trust Region Problems

In this section, we consider the general trust region problems in e.g. [15, 28, 54] and explain their relation to (TR-2) and the applicability of our methods.

Before that, let us mention that the standard ridge regression problem can be seen as a special case of (TR-2). Given a matrix $\mathbf{X} \in \mathbb{R}^{m\times k}$ and $b \in \mathbb{R}^m$, where the columns represent regressor variables and b the dependent variable, and the rows represent m observations. For some 'tuning parameter' $\lambda \geq 0$, the goal is to find a coefficient vector $\beta \in \mathbb{R}^{k+1}$, indexed from 0 to k, such that

$$\beta = \arg\min \sum_{i=1}^{m} \left(\beta_0 + \sum_{j=1}^{k} \beta_j \mathbf{X}_{i,j} - b_i\right)^2 + \lambda \sum_{j=1}^{k} \beta_i^2 \,.$$

This can be equivalently written as a trust region problem, with a trust region constraint $\sum_{j=1}^{k} \beta_i^2 \leq t^2$. These two forms are used interchangeably in the statistical literature, see e.g., [34, Section 3.4.1]. Our algorithm shows how to find the λ corresponding (up to a small error) to the given trust region constraint in strongly polynomial time.

The above can be easily encoded in the form (TR-2) with

$$\mathbf{B} = \begin{pmatrix} \mathbf{1}_m & \mathbf{X} & -\mathbf{I}_m \end{pmatrix} ,$$

where the first $k + 1$ coordinates represent β, and the last m coordinates the regression error. We set I as the coordinates corresponding to $\mathbf{X}$ and J as the coordinates in $\mathbf{I}_n$. Note that the first coordinate representing β_0 does not belong to either set; this column can be easily eliminated using row operations, and thus we obtain the form (TR-2).

We now turn to the general ℓ_2-trust region problem in the form

$$\begin{aligned} \min\ & \langle x, \mathbf{H}x\rangle + \langle c, x\rangle \\ & \|\mathbf{M}x - a\| \le 1\,. \end{aligned} \tag{80}$$

For $c \in \mathbb{R}^n$, $\mathbf{H} \in \mathbb{R}^{n\times n}$, $\mathbf{M} \in \mathbb{R}^{k\times n}$, $a \in \mathbb{R}^k$. In general, $\mathbf{H} \in \mathbb{R}^{n\times n}$ is not assumed to be positive semidefinite, and thus, the problem is not convex. Surprisingly, the problem is polynomial-time solvable even in this case, as shown by Ye [77]. There are also various numerical algorithms e.g., [28].

We show that under the following assumption, (80) is solvable in strongly polynomial time using our algorithm.

$$\inf_{x\in\mathbb{R}^n} \langle x, \mathbf{H}x\rangle + \langle c, x\rangle > -\infty\,, \tag{81}$$

that is, the unconstrained minimum value of the objective in (80) is bounded from below. This is equivalent to assuming $\mathbf{H} \succeq \mathbf{0}$ and $c \in \mathrm{Im}(\mathbf{H})$.

Let us consider LDLT factorization $\mathbf{H} = \mathbf{U}\mathbf{D}\mathbf{U}^\top$, where $\mathbf{D}$ a nonnegative diagonal matrix with diagonal entries d_i. This can be computed in $O(n^3)$ using Cholesky factorization. By the assumption $c \in \mathrm{Im}(\mathbf{H})$, it follows that $\left\langle c, \mathbf{U}^{-\top}_{\cdot,i}\right\rangle = 0$ whenever $d_i = 0$.

We define $y \in \mathbb{R}^{2n}$ with a partition $[2n] = I \cup J$ into two sets of n indices, such that

$$y_I = \mathbf{M}x - a\,, \quad y_J = \mathbf{U}^\top x + \frac{1}{2}\mathbf{D}^\dagger\mathbf{U}^{-1}c\,.$$

This substitution can be encoded in the linear system $\mathbf{B}y = b$, where

$$\mathbf{B} \coloneqq \begin{pmatrix} -\mathbf{I}_n & \mathbf{M}\mathbf{U}^{-\top}\end{pmatrix}\,, \quad b \coloneqq a + \frac{1}{2}\mathbf{M}\mathbf{U}^{-\top}\mathbf{D}^\dagger\mathbf{U}^{-1}c\,,$$

and the two blocks represent the sets I and J. The objective can be written as

$$\langle x, \mathbf{H}x\rangle + \langle c, x\rangle = \sum_{i\in J} d_i y_i^2 - \frac{1}{4}c^\top\mathbf{U}^{-\top}\mathbf{D}^\dagger\mathbf{U}^{-1}c\,.$$

After removing the constant term, (80) is equivalent to

$$\begin{aligned} \min \sum_{i\in J} & d_i y_i^2 \\ \|y_I\|^2 & \le 1 \\ \mathbf{B}y & = b\,. \end{aligned} \tag{82}$$

This form differs from (TR-2) because $\|y_J\|^2$ is replaced by weighted squares. Note that if we were able to compute their square roots, then weights could be encoded in the matrix $\mathbf{B}$. Without square root computations, we need to keep this more general form. However, the algorithm described in this section can be used essentially unchanged. We define $y(\lambda) = \arg\min_{\mathbf{B}y=b} \sum_{i\in J} d_i y_i^2 + \lambda\|y_I\|^2$, and $\psi(\lambda) = \|y_I(\lambda)\|^2$. We can compute $y(\lambda)$ by solving a linear system as before; and since the problem is equivalent to the setting where we rescale the coordinates in J by $\sqrt{d_i}$, all properties remain unchanged.

8 Solving the ℓ_∞-Trust Region Problem

In this section we provide a proof to Theorem 1.10 from the introduction by describing how to solve the program (TR-max), which we restate here for convenience.

$$\min \|y_J\|_\infty \quad \text{s.t.} \quad \mathbf{B}y = b\,,\ \|y_I\|_\infty \le 1\,. \tag{TR-max}$$

Our goal is to find a δ-optimal solution as in Definition 1.8. We will throughout assume that $b \in \mathrm{Im}(\mathbf{B})$ as otherwise infeasibility is trivial to prove through a projection in time $O(n^\omega)$. Throughout this section, the input $(\mathbf{B}, b, I, J)$ as well as the required accuracy $\delta > 0$ will be fixed. For convenience, we assume $\delta \le 1/n$. We let $\lambda^\star$ denote the optimum value of (TR-max).

Recall the definition of an approximate ℓ_∞-regression solver from Definition 2.24 and the subroutine $\textsc{LMFeas}(\mathbf{B}, b, \delta)$ from Theorem 2.25 implementing the solver in time $n^{\tilde\omega+o(1)}\log\left(\frac{1}{\delta}\right)$. We can use this to check feasibility of (TR-max), and also to make approximate queries on $\lambda^\star$. To state this subroutine, we first need to define the appropriate certificates.

Definition 8.1 ($\mathcal{U}^\varepsilon$-certificates and $\mathcal{L}$-certificates)**.** Let $\varepsilon \geq 0$, and $\lambda \in \mathbb{R}_{\geq \mathbf{0}} \cup \{\infty\}$. We say that $y \in \mathbb{R}^m$ is an $\mathcal{U}^\varepsilon$-certificate for λ if $\mathbf{B}y = b$, $\|y_J\|_\infty \leq (1+\varepsilon)\lambda$, $\|y_I\|_\infty \leq 1+\varepsilon$. We say that $z \in \mathbb{R}^m$ is an $\mathcal{L}$-certificate for λ if it is a Farkas certificate showing that $\mathbf{B}y = b$, $\|y_J\|_\infty \leq \lambda$, $\|y_I\|_\infty \leq 1$ is infeasible. Namely,

(i) $\|\mathbf{B}_I^\top z\|_1 < \langle b, z\rangle$ if $\lambda = 0$;

(ii) $\|\mathbf{B}_I^\top z\|_1 + \lambda\|\mathbf{B}_J^\top z\|_1 < \langle b, z\rangle$ if $0 < \lambda < \infty$;

(iii) $\mathbf{B}_J^\top z = \mathbf{0}_J$ and $\|\mathbf{B}_I^\top z\|_1 < \langle b, z\rangle$ if $\lambda = \infty$.

Further, let $\mathcal{U}^\varepsilon \subseteq \mathbb{R}_{\geq \mathbf{0}} \cup \{\infty\}$ be the set of λ values for which a $\mathcal{U}^\varepsilon$-certificate exists, and let $\mathcal{L} \subseteq \mathbb{R}_{\geq \mathbf{0}} \cup \{\infty\}$ for which an $\mathcal{L}$-certificate exists.

Clearly, $\mathcal{U}^\varepsilon \cup \mathcal{L} = \mathbb{R}_{\geq \mathbf{0}} \cup \{\infty\}$. Note that if (TR-max) is feasible, then $\mathcal{L} = [0, \lambda^\star)$, and $\mathcal{U}^\varepsilon = [\alpha, \infty)$ for some $\alpha \leq \lambda^\star$. However, these sets may overlap: $\alpha < \lambda^\star$ is possible.

Lemma 8.2. *Given* $(\mathbf{B}, b, I, J)$*, there exists a subroutine* LMGuess(λ, ε) *that, for given* $\lambda \in \mathbb{R}_{\geq \mathbf{0}} \cup \{\infty\}$*, either returns a* $\mathcal{U}^\varepsilon$*-certificate or an* $\mathcal{L}$*-certificate for* λ *in* $n^{\tilde{\omega}+o(1)} \log\left(\frac{1}{\varepsilon}\right)$.

Proof. If $\lambda = 0$, we simply call LMFeas$(\mathbf{B}_I, b, \varepsilon)$. If $0 < \lambda < \infty$, then let $u \in \mathbb{R}^n$ be defined by $u_I = \mathbf{1}_I$ and $u_J = \lambda\mathbf{1}_J$. Then, LMFeas$(\mathbf{B}\mathrm{diag}(u), b, \varepsilon)$ yields the desired outcome. For the feasibility case $\lambda = \infty$, we run LMFeas$(\mathbf{B}', b', \varepsilon)$ where $\mathbf{B}' = (\mathbf{I}_m - \mathbf{B}_J\mathbf{B}_J^\dagger)\mathbf{B}_I$ and $b' = (\mathbf{I}_m - \mathbf{B}_J\mathbf{B}_J^\dagger)b$. Note that $\mathbf{B}'y' = b'$ for $y' \in \mathbb{R}^I$ if and only if there exists $y'' \in \mathbb{R}^J$ such that $\mathbf{B}(y', y'') = b$; we can obtain such a solution as $y'' = \mathbf{B}_J^\dagger(b - \mathbf{B}_I y')$. □

Assume we have bounds $0 < \lambda_1 < \lambda_2$ with the guarantees $\lambda_1 \in \mathcal{L}$ and $\lambda_2 \in \mathcal{U}^{\delta/4}$. Then, we can use a binary search subroutine, the same way as in BinarySearch in Section 7.4 to find a δ-optimal solution to (TR-max) in $O(\log\log(\lambda_2/\lambda_1) + \log(1/\delta))$ calls to LMGuess. Namely, we find values $\lambda' \leq \lambda'' \leq (1+\delta/4)\lambda'$ such that $\lambda' \in \mathcal{L}$ and $\lambda'' \in \mathcal{U}^{\delta/4}$. We summarize this in the following lemma:

Lemma 8.3. *Given* $(\mathbf{B}, b, I, J)$*, there exists a subroutine* LM-Binary$(\lambda_1, \lambda_2, \delta)$ *that, for given* $0 < \lambda_1 < \lambda_2$ *such that* $\lambda_1 \in \mathcal{L}$ *and* $\lambda_2 \in \mathcal{U}^{\delta/4}$*, returns a* δ*-optimal solution to* (TR-max) *in time* $n^{\tilde{\omega}+o(1)} \log\left(\frac{1}{\delta}\right)\left(\log\log\left(\frac{\lambda_2}{\lambda_1}\right) + \log\left(\frac{1}{\delta}\right)\right)$.

Our ℓ_∞-Trust Region solver follows a similar principle as the ℓ_2-Trust Region solver in Section 7. We identify critical points based on an approximate singular value computation. Using binary search, we can narrow down the search interval to $[\lambda_1, \lambda_2]$ with $\lambda_1 \in \mathcal{L}$ and $\lambda_2 \in \mathcal{U}^{\delta/4}$ so that every critical value is either much larger than λ_2 or much smaller than λ_1. If λ_2/λ_1 is polynomially bounded, we call LM-Binary$(\lambda_1, \lambda_2, \delta)$. If $\lambda_2 \gg \lambda_1$, then we use the approximate feasible solution computed at λ_2 to further narrow down the search space to $[\hat{\lambda}_1, \hat{\lambda}_2] \subseteq [\lambda_1, \lambda_2]$ with $\hat{\lambda}_2/\hat{\lambda}_1 = O(n)$, and call LM-Binary$(\hat{\lambda}_1, \hat{\lambda}_2, \delta)$.

8.1 Critical points from the lifting operator

We obtain critical values from the *lifting operator* $\mathbf{L} \in \mathbb{R}^{I \times J}$ defined as

$$\mathbf{L} \coloneqq \mathbf{B}_I^\dagger \mathbf{B}_J \left(\mathbf{I}_J - \mathbf{B}_J^\top \left(\mathbf{I}_m - \mathbf{B}_I \mathbf{B}_I^\dagger\right)\left(\mathbf{B}_J^\top \left(\mathbf{I}_m - \mathbf{B}_I\mathbf{B}_I^\dagger\right)\right)^\dagger\right). \tag{83}$$

If for a vector $y \in \mathbb{R}^J$, there exists $y' \in \mathbb{R}^I$ such that $\mathbf{B}_J y = \mathbf{B}_I y'$, then $y' = \mathbf{L}y$ is the unique minimum norm such vector. To better understand this map, let us decompose $\mathbb{R}^J = W_1 \oplus W_2 \oplus W_3$, where

$$\begin{aligned}
W_1 &= \ker(\mathbf{B}_J)\,,\\
W_2 &= \{y \in \mathbb{R}^J : y \perp \ker(\mathbf{B}_J), \mathbf{B}_J y \in \mathrm{Im}(\mathbf{B}_I)\} = \{y \in \mathbb{R}^J : \mathbf{B}_J y \in \mathrm{Im}(\mathbf{B}_I), y \in \mathrm{Im}(\mathbf{B}_J^\top)\}\,,\\
W_3 &= \{y \in \mathbb{R}^J : y \perp \ker((\mathbf{I}_m - \mathbf{B}_I\mathbf{B}_I^\dagger)\mathbf{B}_J)\} = \{y \in \mathbb{R}^J : y \in \mathrm{Im}(\mathbf{B}_J^\top(\mathbf{I}_m - \mathbf{B}_I\mathbf{B}_I^\dagger))\}\,.
\end{aligned}$$

Note that $\ker(\mathbf{L}) = W_1 \oplus W_3$. Further, note that $\mathbf{I}_J - \mathbf{B}_J^\top\left(\mathbf{I}_m - \mathbf{B}_I\mathbf{B}_I^\dagger\right)\left(\mathbf{B}_J^\top\left(\mathbf{I}_m - \mathbf{B}_I\mathbf{B}_I^\dagger\right)\right)^\dagger$ is the projection onto $\ker((\mathbf{I}_m - \mathbf{B}_I\mathbf{B}_I^\dagger)\mathbf{B}_J)$ and therefore acts as identity on W_2, and therefore $\mathbf{L}w_2 = \mathbf{B}_I^\dagger\mathbf{B}_J w_2$ for any $w_2 \in W_2$. For $y = (y_I, y_J)$ with an orthogonal decomposition $y_J = w_1 + w_2 + w_3$, $w_i \in W_i$, $i = 1, 2, 3$, we have that for $y' \coloneqq \mathbf{L}y_J$ the equality $y' = \mathbf{L}w_2 = \mathbf{B}_I^+\mathbf{B}_J w_2$ and so $\mathbf{B}_I y' = \mathbf{B}_J w_2$ holds. As such $\mathbf{B}y = \mathbf{B}_I y_I + \mathbf{B}_J y_J = \mathbf{B}_I y_I + \mathbf{B}_J(w_2 + w_3) = \mathbf{B}_I(y_I + y') + \mathbf{B}_J w_3 = \mathbf{B}(y_I + y', w_3)$. The vector $(y_I + y', w_3)$ gives the largest possible reduction ℓ_2-norm on the coordinates in J while preserving $\mathbf{B}y$; w_3 is the part that cannot be eliminated.

For some factor $\varrho \geq 1$, we compute a ϱ-approximate partial SVD $\mathbf{U} \in \mathbb{R}^{\mathrm{rk}(\mathbf{L}) \times J}$ (see Definition 2.26) of $\mathbf{L}$ using Lemma 2.27 or Lemma 2.28. We then compute the approximate lengths $\sigma_i \in \|\mathbf{U}_{i,:}\|_2 [0.5, 2]$ according to Proposition 2.1. Let us assume these are ordered decreasingly as $\sigma_1 \geq \cdots \geq \sigma_{\mathrm{rk}(\mathbf{L})} > 0$. These will be the *critical points* in the algorithm.

8.2 The algorithm

The algorithm TRW-SOLVE($\mathbf{B}, b, I, J, \delta$) is shown as Algorithm 6. We start by calls to LMGUESS(∞, δ) to first determine approximate feasibility of (TR-max) and then LMGUESS($0, \delta$) to determine whether an approximately feasible solution with objective value 0 exists.

If we conclude that the problem (TR-max) is approximately feasible but the objective value is strictly positive, we proceed by computing a ϱ-approximate partial SVD $\mathbf{U} \in \mathbb{R}^{\mathrm{rk}(\mathbf{L}) \times J}$ of the lifting operator $\mathbf{L}$; let $u^{(i)} \in \mathbb{R}^J$ denote the i-th row of $\mathbf{U}$. Let $\ell = \mathrm{rk}(\mathbf{L})$ and $\sigma_1 \geq \ldots \geq \sigma_\ell > 0$ be 2-approximations of the $\|u^{(i)}\|_2$ norms, ordered decreasingly.

We define the breakpoints $\mathcal{P}^\star$ for the algorithm to include 0, ∞, and for each $k \in [\ell]$, the values $\delta/(4\varrho n \sigma_k)$ and $4\varrho n/(\delta \sigma_k)$. We then perform binary search on $\mathcal{P}^\star$, making calls to LMGUESS($\lambda, \delta/4$) to find two consecutive values $\lambda_1 < \lambda_2$ such that $\lambda_1 \in \mathcal{L}$ and $\lambda_2 \in \mathcal{U}^{\delta/4}$.

If $\lambda_2/\lambda_1 \leq 64\varrho^2 n^2/\delta^2$, then we can find a $\delta/4$-approximate optimal solution y using LM-BINARY($\lambda_1, \lambda_2, \delta/4$). If the gap λ_2/λ_1 is above this threshold, i.e., $\lambda_2/\lambda_1 > 64\varrho^2 n^2/\delta^2$, then let $k \in [\ell]$ be the smallest value such that $\sigma_k < 1/\lambda_2$. By construction,

$$\|u^{(j)}\|_2 \leq 2\sigma_k < \frac{\delta}{2\varrho n \lambda_2}, \quad \forall k \leq j \leq \ell, \quad \text{and} \quad \|u^{(j)}\|_2 \geq \frac{\sigma_{j-1}}{2} \geq \frac{\sigma_{k-1}}{2} \geq \frac{2\varrho n}{\delta \lambda_1}, \quad \forall 1 \leq j \leq k-1. \tag{84}$$

By definition of the ϱ-approximate SVD, this in particular means that

$$\mathbf{L}\frac{u^{(j)}}{\|u^{(j)}\|_2} \leq \varrho \|u^{(j)}\|_2 \leq \frac{\delta}{2n\lambda_2}, \quad \forall k \leq j \leq \ell, \quad \text{and} \quad \mathbf{L}\frac{u^{(j)}}{\|u^{(j)}\|_2} \geq \frac{\|u^{(j)}\|_2}{\varrho} \geq \frac{2n}{\delta\lambda_1}, \quad \forall 1 \leq j \leq k-1, \tag{85}$$

that is, the directions $u^{(j)}$ for $j \geq k$ are very cheap to lift, while the directions $u^{(j)}$ with $j < k$ are very expensive to lift. The vectors $\{u^{(j)} : j \in [\ell]\}$ form a basis of W_2. Let us further decompose

$$W_2 = H_1 \oplus H_2, \quad H_1 \coloneqq \mathrm{span}\{u_k : 1 \leq j \leq k-1\}, \quad H_2 \coloneqq \mathrm{span}\{u_k : k \leq j \leq \ell\},$$

that is, H_1 is the subspace of 'expensive to lift' and H_2 the subspace of 'cheap to lift' directions. We now consider the $\mathcal{U}^{\delta/4}$-certificate $\bar{y}$ for λ_2, and modify it in two steps. First, we obtain $\tilde{y}$ by replacing $\bar{y}_J$ with $\mathbf{B}_J^\dagger \mathbf{B}_J \bar{y}_J$, i.e., projecting out the part in $W_1 = \ker(\mathbf{B}_J)$. We next set

$$z \coloneqq \sum_{j=k}^{\ell} \frac{\langle u^{(j)}, \tilde{y}_J \rangle}{\|u^{(j)}\|_2^2} u^{(j)}, \tag{86}$$

as the orthogonal projection of $\tilde{y}_J$ onto H_2. The algorithm defines $\hat{\lambda} = \|\tilde{y}_J - z\|_\infty$, and finds the δ-optimal solution by LM-BINARY($\frac{1}{n}\hat{\lambda}, \hat{\lambda}, \delta$).

To justify this choice, Lemma 8.4 below shows that $\hat{y} = (\tilde{y}_I + \mathbf{L}z, \tilde{y}_J - z)$ is a $\mathcal{U}^\delta$-certificate for $\hat{\lambda}$. In Lemma 8.5 we then show that $\lambda^\star \geq \hat{\lambda}/n$.

8.3 Analysis

Lemma 8.4. *Let $\bar{y}$ be a $\mathcal{U}^{\delta/4}$-certificate for λ_2, and $\tilde{y} \coloneqq (\bar{y}_I, \mathbf{B}_J^\dagger \mathbf{B}_J \bar{y}_I)$. For z as in (86), the vector*

$$\hat{y} \coloneqq (\tilde{y}_I + \mathbf{L}z, \tilde{y}_J - z) \tag{87}$$

is a $\mathcal{U}^\delta$ certificate for $\hat{\lambda} = \|\tilde{y}_J - z\|_\infty$.

Proof. Clearly, $z \in W_2$; $\mathbf{B}\hat{y} = b$ follows by the definition of the lifting operator. Also, $\|\tilde{y}_J - z\|_\infty \leq (1+\delta)\hat{\lambda}$ holds by the definition of $\hat{\lambda}$. It remains to show that $\|\hat{y}_I\|_\infty \leq 1 + \delta$. Since $\tilde{y}_I = \bar{y}_I$, and $\|\bar{y}_I\|_\infty \leq 1 + \delta/4$, it suffices to show that $\|\mathbf{L}z\|_2 \leq 3\delta/4$.

Algorithm 6: TRW-SOLVE$(\mathbf{B}, b, I, J, \delta)$

Input : An instance of (TR-max) with constraint matrix $\mathbf{B} \in \mathbb{R}^{m\times n}$, $\mathrm{rk}(\mathbf{B}) = m$, $b \in \mathbb{R}^m$, a partition $I \cup J = [n]$, $\delta \in (0, 1/n)$ and a subroutine to compute ϱ-approximate partial SVD.

Output: an δ-optimal solution $y \in \mathbb{R}^n$ to (TR-max) or infeasibility conclusion.

1 Call LMGUESS(∞, δ) ;
2 **if** *it returns an $\mathcal{L}$-certificate for* ∞ **then return** Infeasible and Terminate;
3 Call LMGUESS$(0, \delta)$;
4 **if** *it returns an $\mathcal{U}^\delta$-certificate $y \in \mathbb{R}^n$ for* 0 **then return** y and Terminate;
5 `// Compute lifting operator and its partial SVD`
6 Compute lifting operator $\mathbf{L} \leftarrow \mathbf{B}_I^\dagger \mathbf{B}_J \left(\mathbf{I}_J - \mathbf{B}_J^\top \left(\mathbf{I}_m - \mathbf{B}_I\mathbf{B}_I^\dagger\right) \left(\mathbf{B}_J^\top \left(\mathbf{I}_m - \mathbf{B}_I\mathbf{B}_I^\dagger\right)\right)^\dagger\right)$;
7 Compute ϱ-approximate partial SVD $\mathbf{U} \in \mathbb{R}^{\mathrm{rk}(\mathbf{L})\times J}$ of $\mathbf{L}$ using Lemmas 2.27 and 2.28 ;
8 $\ell \leftarrow \mathrm{rk}(\mathbf{L})$; $u^{(i)} \leftarrow \mathbf{U}_{i,\star}$, $i \in [\ell]$;
9 Compute σ_i as 2-approximation of $\|u^{(i)}\|$ for $i \in [k]$ using Proposition 2.1 ;
10 Reorder as $\sigma_1 \geq \cdots \geq \sigma_\ell > 0$;
11 $\mathcal{P}^* \leftarrow \left\{\frac{\delta}{4\varrho n \sigma_k}, \frac{4\varrho n}{\delta \sigma_k} : k \in [\ell]\right\} \cup \{0, \infty\}$;
12 Use LMGUESS$(\lambda, \delta/4)$ in binary search in $\mathcal{P}^*$ to find consecutive values $\lambda_1 < \lambda_2$ that $\lambda_1 \in \mathcal{L}$ and $\lambda_2 \in \mathcal{U}^{\delta/4}$;
13 **if** $\lambda_2/\lambda_1 \leq 64\varrho^2 n^2/\delta^2$ **then** $y \leftarrow$ LM-BINARY$(\lambda_1, \lambda_2, \delta/4)$;
14 **else**
15 $\quad \bar{y} \leftarrow \mathcal{U}^{\delta/4}$ certificate for λ_2 ;
16 $\quad \tilde{y} \leftarrow (\bar{y}_I, \mathbf{B}_J^\dagger \mathbf{B}_J \bar{y}_J)$;
17 $\quad z \leftarrow \sum_{j : \sigma_j < 1/\lambda_2} \frac{\langle u^{(j)}, \tilde{y}_J\rangle}{\|u^{(j)}\|_2^2} u^{(j)}$;
18 $\quad \hat{\lambda} \leftarrow \|\tilde{y}_J - z\|_\infty$;
19 $\quad y \leftarrow$ LM-BINARY$(\frac{1}{n}\hat{\lambda}, \hat{\lambda}, \delta)$
20 **return** y

By definition,

$$\|z\|_2^2 = \sum_{j=k}^{\ell} \frac{\left\langle u^{(j)}, \tilde{y}_J\right\rangle^2}{\|u^{(j)}\|_2^2} \leq \|\tilde{y}_J\|_2^2 \leq \|\bar{y}_J\|_2^2 \leq n^2 \|\bar{y}_J\|_\infty^2 \leq n^2 (1 + \delta/4)^2 \lambda_2^2 \,. \tag{88}$$

Using the definition of $\mathbf{U}$ as ϱ-approximate partial SVD and that $\|u^{(j)}\| \leq \delta/(2n\varrho\lambda_2)$ for each $j \geq k$, as well as the above bound,

$$\begin{aligned} \|\mathbf{L}z\|_2^2 \leq \varrho^2 \|\mathbf{U}z\|_2^2 &= \varrho^2 \sum_{j=k}^{\ell} \left\langle u^{(j)}, z\right\rangle^2 \\ &\leq \varrho^2 \|u^{(k)}\|_2^2 \cdot \|z\|_2^2 \\ &\leq \varrho^2 \frac{\delta^2}{4n^2\varrho^2\lambda_2^2} \cdot n^2 (1+\delta/4)^2 \lambda_2^2 \\ &\leq \frac{1}{4}(1+\delta/4)^2 \delta^2 < \left(\frac{3\delta}{4}\right)^2 \,. \end{aligned} \tag{89}$$

□

Lemma 8.5. *For the value $\hat{\lambda}$ defined in line 18, $\hat{\lambda}/n \in \mathcal{L}$ holds.*

Proof. If $\hat{\lambda}/n \leq \lambda_1$, then the statement is immediate because $\lambda_1 \in \mathcal{L}$. Assume for the rest of the proof that $\hat{\lambda}/n \notin \mathcal{L}$. Thus, there exists a vector $\bar{w} \in \mathbb{R}^n$ such that $\mathbf{B}\bar{w} = b$, $\|\bar{w}_J\|_\infty \leq \hat{\lambda}/n$, $\|\bar{w}_I\|_\infty \leq 1$. Consider $\hat{y}$ as in (87); this is a $\mathcal{U}^\delta$-certificate for $\hat{\lambda}$ according to Lemma 8.4.

By construction, $\hat{y}_J \in H_1 \oplus W_3$. Similarly to Lemma 8.4, we can replace $\bar{w}$ by another vector $\hat{w}$ such that

$$\mathbf{B}\hat{w} = b \,, \quad \|\hat{w}_I\|_\infty \leq 1 + \delta \,, \quad \|\hat{w}_J\|_\infty \leq \hat{\lambda}/\sqrt{n} \,, \quad \hat{w}_J \in H_1 \oplus W_3 \,.$$

Indeed, we can first replace $\bar{w}$ by the projection $(\bar{w}_I, \mathbf{B}_J^\dagger \mathbf{B}_J \bar{w}_J)$ that removes the part in W_1, and then we project out the part in H_2. The bound $\|\hat{w}_I\|_\infty \le 1+\delta$ follows as in Lemma 8.4, using that $\hat{\lambda}/n \le \|\hat{y}_J\|_2/n \le \|\hat{y}_J\|_\infty/\sqrt{n} \le \lambda_2 \cdot (1+\delta/4)/\sqrt{n} \le \lambda_2$ as $\bar{y}$ is a $\mathcal{U}^{\delta/4}$-certificate for λ_2. Further, $\|\hat{w}_J\|_2 \le \|\bar{w}_J\|_2$ and $\|\bar{w}_J\|_\infty \le \hat{\lambda}/n$ imply $\|\hat{w}_J\|_\infty \le \hat{\lambda}/\sqrt{n}$. Let

$$s := \hat{w} - \hat{y}\,.$$

By the above, $s_J \in H_1 \oplus W_3$. Moreover, since $s \in \ker(\mathbf{B})$, it follows that $s_J \in H_1$. Thus,

$$s_J = \sum_{j=1}^{k-1} \frac{\langle u^{(j)}, s_J \rangle}{\|u^{(j)}\|_2^2} u^{(j)}\,.$$

Similarly to (89), we can bound

$$\begin{aligned}\|\mathbf{L}s_J\|_2^2 &\ge \frac{1}{\varrho^2}\|\mathbf{U}s_J\|_2^2 = \frac{1}{\varrho^2}\sum_{j=1}^{k-1}\left\langle u^{(j)}, s_J\right\rangle^2 \\ &\ge \frac{1}{\varrho^2}\|u^{(k-1)}\|_2^2 \cdot \|s_J\|_2^2 \\ &\ge \frac{4n^2\|s_J\|_2^2}{\delta^2\lambda_1^2}\,,\end{aligned} \tag{90}$$

using (84). We further have

$$\|s_J\|_2 = \|\hat{w}_J - \hat{y}_J\|_2 \ge \|\hat{y}_J\|_\infty - \|\hat{w}_J\|_\infty \ge \hat{\lambda}(1-1/\sqrt{n})\,.$$

By the definition of the lifting operator and using the above bounds, we have

$$\|s_I\|_2^2 = \|\mathbf{L}s_J\|_2^2 \ge \frac{4n^2}{\delta^2\lambda_1^2}\hat{\lambda}^2\left(1-\frac{1}{\sqrt{n}}\right)^2 \ge \left(\frac{3}{2}\right)^2 n^2\,,$$

using that $\hat{\lambda} > n\lambda_1$ by assumption. We now get a contradiction from

$$1+\delta \ge \|\hat{w}_I\|_\infty \ge \frac{\|\hat{w}_I\|_2}{\sqrt{n}} \ge \frac{\|s_I\|_2 - \|\hat{y}_I\|_2}{\sqrt{n}} > \sqrt{n}\,. \tag{91}$$

□

Theorem 1.10. *For $\delta \in (0,1)$, there exists an algorithm* TRW-SOLVE$(\mathbf{B}, b, I, J, \delta)$ *that finds a δ-optimal solution to* (TR-max) *or certifies infeasibility in strongly polynomial time. The running time of the algorithm is randomized $n^{\tilde{\omega}+o(1)}\log(1/\delta)$ or deterministic $n^{\tilde{\omega}+o(1)}\log(1/\delta) + O(n^3)$.*

Proof. Correctness follows from Lemma 8.4 and Lemma 8.5. In particular, whenever LM-BINARY$(\lambda', \lambda'', \delta')$ is called, $\lambda' \in \mathcal{L}$ and $\lambda'' \in \mathcal{U}^{\delta'}$. The running time bound follows by Lemma 8.3. □

A Applying the Robust Interior Point Method

Theorem 2.23. *Let $\theta \in (1/8, 1)$. Consider an instance of* (LP). *There exists a subroutine* PATHFOLLOW(x, s, ϱ) *that, given $(\bar{x}, \bar{s}) \in \mathcal{N}^{-\infty}(\theta)$ with $\overline{\mu}(\bar{x}, \bar{s}) = \bar{\mu}$ and $\varrho \in (0,1)$, outputs $(x^{\text{out}}, s^{\text{out}}) \in \mathcal{N}^{-\infty}(3/4)$ with $\overline{\mu}(x^{\text{out}}, s^{\text{out}}) \le \varrho\bar{\mu}$ in time $O\left(n^{\tilde{\omega}+o(1)}\log\left(\frac{1}{\varrho(1-\theta)}\right)\right)$.*

Proof. We consider the extended system with primal variables $(x^+, x^-, t) \in \mathbb{R}^{2n+1}_{\ge \mathbf{0}}$ and dual variables $(y, y_0, s^+, s^-, g) \in \mathbb{R}^{m+1} \times \mathbb{R}^{2n+1}_{\ge \mathbf{0}}$

$$\begin{aligned}\min\ &\langle \bar{s}, x^+\rangle + \langle \tilde{c}, x^-\rangle \\ \mathbf{A}x^+ - \mathbf{A}x^- &= b \\ \bar{s}^\top x^+ + t &= b_0 \\ x^+, x^-, t &\ge \mathbf{0}\,.\end{aligned} \qquad \begin{aligned}\max\ \langle b, y\rangle &- b_0 y_0 \\ \mathbf{A}^\top y - \bar{s}y_0 + s^+ &= \bar{s} \\ -\mathbf{A}^\top y + s^- &= \tilde{c} \\ -y_0 + g &= 0 \\ s^+, s^-, g &\ge \mathbf{0}\,,\end{aligned} \tag{92}$$

Here,

$$M := \frac{512n^2}{(1-\theta)^2}\,, \quad \tilde{c} := \frac{M^2\bar{\mu}}{M\bar{\mu}\bar{s}^{-1} - \bar{x}}\,, \quad b_0 := Mn\bar{\mu} + \frac{M^2\bar{\mu}}{M-1}\,, \tag{93}$$

We set as initial solution

$$x^+ := M\bar{\mu}\bar{s}^{-1}\,, \quad x^- := M\bar{\mu}\bar{s}^{-1} - \bar{x}\,, \quad t := \frac{M^2\bar{\mu}}{M-1}\,,$$
$$y := \mathbf{0}_m\,, \quad y_0 := g = M-1\,, \quad s^+ := M\bar{s}\,, \quad s^- := \tilde{c}\,.$$

The initial solution is exactly on the central path with normalized gap value $M^2\bar{\mu}$.

Using Theorem 2.22, we can find primal and dual solutions (x^+, x^-, t) and (y, y_0, s^+, s^-, y_0) to the system (92) in the neighborhood $\mathcal{N}^{-\infty}(1/8)$ at normalized gap value

$$\mu \le \frac{\varrho\bar{\mu}}{2n+1}\,.$$

in time $O\left(n^{\tilde{\omega}+o(1)}\log\left(\frac{1}{\varrho(1-\theta)}\right)\right)$.

We show that $x^{\mathrm{out}} := x^+ - x^-$ and $s^{\mathrm{out}} := s^+/(1+y_0)$ satisfy $(x^{\mathrm{out}}, s^{\mathrm{out}}) \in \mathcal{N}^{-\infty}(1/4)$ for the original (LP) with normalized gap value $\mu^{\mathrm{out}} \le \frac{2n+1}{n}\mu \le \varrho\bar{\mu}/n$.

First, note that $(x^{\mathrm{out}}, s^{\mathrm{out}})$ live in the correct subspaces as $\mathbf{A}x^{\mathrm{out}} = \mathbf{A}x^+ - \mathbf{A}x^- = b$ and $\mathbf{A}^\top(y/(1+y_0)) + s^{\mathrm{out}} = \bar{s}$.

Claim A.1. *$s^+ \le 2n\bar{s}/(1-\theta)$ and $y_0 \le 1/64$.*

Proof. Note that $(\bar{x}, \mathbf{0}_n, 0)$ and $(\mathbf{0}_m, 0, \bar{s}, \tilde{c}, 0)$ are primal and dual solutions to (92) with optimality gap $n\bar{\mu}$. Consequently, the gap between $(\bar{x}, \mathbf{0}_n, 0)$ and (y, y_0, s^+, s^-, y_0) is at most $n\bar{\mu} + (2n+1)\mu < 2n\bar{\mu}$. In particular,

$$\bar{x}_i s_i^+ \le 2n\bar{\mu} \quad \forall i \in [n]\,.$$

On the other hand, from $(\bar{x}, \bar{s}) \in \mathcal{N}^{-\infty}(\theta)$ we get

$$\bar{x}_i\bar{s}_i \ge (1-\theta)\bar{\mu} \quad \forall i \in [n]\,.$$

The first part of the claim follows as $s_i^+ \le 2n\bar{\mu}\bar{x}_i^{-1} \le 2n(1-\theta)^{-1}\bar{s}_i$.

For the second part, we can similarly argue that the gap between (x^+, x^-, t) and $(\mathbf{0}_n, 0, \bar{s}, \tilde{c}, 0)$ is at most $2n\bar{\mu}$, implying $\bar{s}^\top x^+ \le 2n\bar{\mu}$. From here, we get

$$t = b_0 - \bar{s}^\top x^+ > Mn\bar{\mu} - 2n\bar{\mu} > 64\bar{\mu}\,.$$

By $y_0 t \le (2n+1)\mu \le \bar{\mu}$ and the choice of μ, the second part of the claim follows. $\square$

From the dual constraints, we have $s^+ + s^- = \bar{s}(1+y_0) + \tilde{c}$. From the above claim, we can conclude that s_i^- is significantly larger than s_i^+, as

$$\begin{aligned}
\frac{s_i^-}{s_i^+} &= \frac{\bar{s}(1+y_0) + \tilde{c}_i - s_i^+}{s_i^+} \ge \frac{\tilde{c}_i - s_i^+}{s_i^+} \\
&\ge \frac{M^2\bar{\mu}}{M\bar{\mu}\bar{s}_i^{-1} - \bar{x}_i} \cdot \frac{1-\theta}{2n\bar{s}_i} - 1 \\
&\ge \frac{M^2\bar{\mu}}{M\bar{\mu}\bar{s}_i^{-1}} \cdot \frac{1-\theta}{2n\bar{s}_i} - 1 \\
&= \frac{M(1-\theta)}{2n} - 1 \\
&= \frac{512}{1-\theta} \cdot \frac{n}{2} - 1 \\
&\ge \frac{128n}{1-\theta}\,.
\end{aligned}$$

Consequently, x_i^+ is much larger than x_i^- as

$$\frac{x_i^-}{x_i^+} \le \frac{(2n+1)\mu}{s_i^-} \cdot \frac{s_i^+}{(1-1/8)\mu} = \frac{2n+1}{1-1/8}\frac{s_i^+}{s_i^-} \le \frac{2n+1}{1-1/8} \cdot \frac{1-\theta}{128n} \le \frac{1}{16}\,.$$

By definition of x^+ and s^+, we have that $x_i^+ s_i^+ \geq (1-1/8)\mu$ and $\sum_i x_i^+ s_i^+ \leq (2n+1)\mu$.

Thus, we get the upper bound

$$\sum_i x_i^{\text{out}} s_i^{\text{out}} \leq \sum_i x_i^+ \frac{s_i^+}{1+y_0} \leq \frac{(2n+1)\mu}{1} = (2n+1)\mu\,,$$

and consequently $\mu^{\text{out}} = \frac{1}{n}\sum_i x_i^{\text{out}} s_i^{\text{out}} \leq \frac{2n+1}{n}\mu \leq \varrho\frac{\bar{\mu}}{n}$.

Furthermore, note that $x_i^{\text{out}} = x_i^+ - x_i^- \geq x_i^+ - \frac{1}{16}x_i^+ = \frac{15}{16}x_i^+$, and $s_i^{\text{out}} = s_i^+/(1+y_0) \geq s_i^+/(1+\frac{1}{64})$, so we get the lower bound

$$x_i^{\text{out}} s_i^{\text{out}} \geq \frac{15}{16}x_i^+ \cdot \frac{s_i^+}{1+\frac{1}{64}} \geq \left(1-\frac{1}{8}\right)\frac{15}{16}\frac{1}{1+\frac{1}{64}}\mu > \left(1-\frac{1}{6}\right)\mu\,.$$

For the neighborhood parameter we therefore get the bound

$$\min_i \frac{x_i^{\text{out}} s_i^{\text{out}}}{n^{-1}\sum_i x_i^{\text{out}} s_i^{\text{out}}} \geq \left(1-\frac{1}{6}\right)\cdot\frac{n}{2n+1} \geq 1-\frac{3}{4}\,. \tag{94}$$

We can therefore conclude that $(x^{\text{out}}, s^{\text{out}}) \in \mathcal{N}^{-\infty}(3/4)$. □

Theorem 2.25. *There exists an approximate ℓ_∞-regression solver* $\text{LMFeas}(\mathbf{B}, b, \delta)$ *satisfying Definition 2.24 that runs in time* $O\left(n^{\tilde{\omega}+o(1)}\log\left(\frac{1}{\delta}\right)\right)$.

Proof. We start by computing $\bar{y} = \mathbf{B}^\dagger b$. If $\mathbf{B}\bar{y} \neq b$, then the linear system is infeasible. Otherwise, recall that $\bar{y} = \arg\min\{\|y\|^2 : \mathbf{B}y = b\}$. Thus, if $\|\bar{y}\|_2 > \sqrt{n}$, we may conclude that for any $y \in \mathbb{R}^n$ with $\mathbf{B}y = b$ we have $\|y\|_\infty \geq \|y\|_2/\sqrt{n} \geq \|\bar{y}\|_2/\sqrt{n} > 1$, and therefore the system $\mathbf{B}y = b$, $\|y\|_\infty \leq 1$ is infeasible.

This can be turned into a Farkas certificate of infeasibility for Definition 2.24. Recall that if $\text{rk}(\mathbf{B}) = m$, then $\mathbf{B}^\dagger = \mathbf{B}^\top(\mathbf{B}\mathbf{B}^\top)^{-1}$. Let $z \coloneqq (\mathbf{B}\mathbf{B}^\top)^{-1}b$. Then, $\mathbf{B}^\top z = \bar{y}$, and $\|\mathbf{B}^\top z\|_2^2 = \langle b, z\rangle$. On the other hand,

$$\|\mathbf{B}^\top z\|_1 \leq \sqrt{n}\|\mathbf{B}^\top z\|_2 < \|\mathbf{B}^\top z\|_2^2 = \langle b, z\rangle\ ,$$

as required.

In case that both $\mathbf{B}\bar{y} = b$ and $\|\bar{y}\|_2 \leq \sqrt{n}$, we consider the auxiliary LP

$$\begin{aligned} \min\ \ & \alpha \\ \mathbf{B}y &= b \\ -y + \alpha\mathbf{1}_n - w^+ &= -\mathbf{1}_n \\ y + \alpha\mathbf{1}_n - w^- &= -\mathbf{1}_n \\ w^+, w^-, \alpha &\geq 0\,. \end{aligned} \qquad \begin{aligned} \max\ \ & b^\top z - \mathbf{1}_n^\top(s^- + s^+) \\ \mathbf{B}^\top z - s^+ + s^- &= 0 \\ \mathbf{1}_n^\top s^+ + \mathbf{1}_n^\top s^- + \gamma &= 1 \\ s^+, s^-, \gamma &\geq 0\,. \end{aligned} \tag{95}$$

We initialize close to the central path with $\hat{y} = \bar{y}$, $\hat{\alpha} = 2n$, $\hat{w}^+ = (2n+1)\mathbf{1}_n - \bar{y}$, $\hat{w}^- = (2n+1)\mathbf{1}_n + \bar{y}$, $\hat{z} = 0$ $\hat{s}^+ = \hat{s}^- = \frac{1}{2n+1}\mathbf{1}_n$, $\hat{\gamma} = \frac{1}{2n+1}$. For this initialization we have that the normalized gap $\hat{\mu}$ is

$$\begin{aligned} \hat{\mu} &= \frac{1}{2n+1}\left(\hat{\alpha}\cdot\hat{\gamma} + \langle \hat{w}^+, \hat{s}^+\rangle + \langle \hat{w}^-, \hat{s}^-\rangle\right) = \frac{1}{2n+1}\left(\frac{2n}{2n+1} + \frac{(2n+1)\cdot n}{2n+1} + \frac{(2n+1)\cdot n}{2n+1}\right) \\ &= \frac{1}{2n+1}\left(\frac{2n}{2n+1} + 2n\right) < 1 \end{aligned}$$

and furthermore

$$\frac{\hat{\alpha}\cdot\hat{\gamma}}{\hat{\mu}} = \frac{1}{\frac{1}{2n+1}+1} \geq \frac{3}{4}$$

and for $i \in [n]$ we have

$$\frac{\hat{w}_i^+ \hat{s}_i^+}{\hat{\mu}} \geq \frac{2n+1-\|\bar{y}\|_\infty}{(2n+1)\hat{\mu}} \geq \frac{2n+1-\sqrt{n}}{(2n+1)\hat{\mu}} \geq \frac{2}{3}\frac{1}{\hat{\mu}} \geq \frac{2}{3}\,,$$

and analogously for $\hat{w}_i^- \hat{s}_i^-/\hat{\mu}$. Thus, the initialization is in the neighborhood $\mathcal{N}^{-\infty}(1/3)$.

We use Theorem 2.23 to find primal and dual solutions (y, w^+, w^-, α) and (z, s^+, s^-, γ) to the system (95) in the neighborhood $\mathcal{N}^{-\infty}(1/8)$ at normalized gap value $\bar{\mu} \leq \delta$ in time $O\left(n^{\tilde{\omega}+o(1)}\log\left(\frac{1}{\delta}\right)\right)$.

Note that the dual objective is upper bounded by $\langle b, z\rangle - \|\mathbf{B}^\top z\|_1$. Hence, if the dual objective value is positive, then z is a Farkas certificate of infeasibility of $\mathbf{B}y = b$, $\|y\|_\infty \leq 1$.

Otherwise, if the dual objective value is ≤ 0, then the primal objective value is $\alpha \leq \delta$. Thus, we have $\|y\|_\infty \leq 1 + \alpha \leq 1 + \delta$, and thus we can return the approximately feasible solution y. □

B Polarization in the wide neighborhood

In this Appendix, we prove Lemma 3.3, following the argument in [5]. We first introduce the following simple proposition which shows that if a straight line between two points in $\mathcal{N}^{-\infty}(\theta)$ is polarized, then changing the endpoints to other points in $\mathcal{N}^{-\infty}(\theta)$ with the same normalized duality gap, the straight lines still show ℓ_∞-polarization.

Proposition B.1. *For $\theta \in [0,1)$, let $z_0 = (x_0, s_0), z_1 = (x_1, s_1) \in \mathcal{N}^{-\infty}(\theta)$ with $\overline{\mu}(z_1) < \overline{\mu}(z_0)$. If the straight line $[z_1, z_0]$ is γ-polarized, then for any $z_0' \in \mathcal{N}^{-\infty}(\theta)$ with $\overline{\mu}(z_0') = \overline{\mu}(z_0)$ and any $z_1' \in \mathcal{N}^{-\infty}(\theta)$ with $\overline{\mu}(z_1') = \overline{\mu}(z_1)$, the straight line $[z_1', z_0']$ is $\frac{(1-\theta)^2\gamma}{4n^2}$-polarized.*

Proof. Let $\lambda \in [0,1]$. For each $i \in B$, we have $\frac{\lambda x_{1i} + (1-\lambda)x_{0i}}{x_{0i}} \geq \gamma$. Applying the bounds in Lemma 2.20, we obtain

$$\frac{\lambda x_{1i}' + (1-\lambda)x_{0i}'}{x_{0i}'} \geq \frac{1-\theta}{2n} \cdot \frac{\lambda x_{1i} + (1-\lambda)x_{0i}}{\frac{2n}{1-\theta}x_{0i}} \geq \frac{(1-\theta)^2\gamma}{4n^2}, \quad \text{for all } i \in B\,.$$

By an analogous derivation, we also have $\frac{\lambda s_{1i}' + (1-\lambda)s_{0i}'}{s_{0i}'} \geq \frac{(1-\theta)^2\gamma}{4n^2}$ for all $i \in N$. Hence, $[z_1', z_0']$ is $\frac{(1-\theta)^2\gamma}{4n^2}$-polarized. ☐

Lemma B.2 ([5, Lemma 3.7])**.** *For any $u, v > 0$,*

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

Lemma 3.3. *For $\theta \in [0,1)$, let $z_0 = (x_0, s_0), z_1 = (x_1, s_1) \in \mathcal{N}^{-\infty}(\theta)$ with $\overline{\mu}(z_1) < \overline{\mu}(z_0)$. If the straight line $[z_1, z_0] \subseteq \mathcal{N}^{-\infty}(\theta)$, then there exists a partition $B \cup N = [n]$ such that $[z_0, z_1]$ is $\frac{1-\theta}{4n}$-polarized with partition (B, N). Moreover, for any $z_0' \in \mathcal{N}^{-\infty}(\theta)$ with $\overline{\mu}(z_0') = \overline{\mu}(z_0)$ and any $z_1' \in \mathcal{N}^{-\infty}(\theta)$ with $\overline{\mu}(z_1') = \overline{\mu}(z_1)$, the straight lines $[z_1', z_0']$ are $\frac{(1-\theta)^3}{16n^3}$-polarized with (B, N).*

Proof. $\lambda z_1 + (1-\lambda)z_0 \in \mathcal{N}^{-\infty}(\theta)$ implies

$$\frac{(\lambda x_1 + (1-\lambda)x_0)(\lambda s_1 + (1-\lambda)s_0)}{\lambda x_1 s_1 + (1-\lambda)x_0 s_0} \geq \frac{(1-\theta)\overline{\mu}(\lambda z_1 + (1-\lambda)z_0)}{n(\lambda\mu_1 + (1-\lambda)\mu_0)}\mathbf{1} = \frac{1-\theta}{n}\mathbf{1}$$

Then, for all $i \in [n]$, applying Lemma B.2 with $u = x_{1i}/x_{0i}$ and $v = s_{1i}/s_{0i}$, we get

$$2\left(\frac{x_{1i}}{x_{0i}} + \frac{s_{1i}}{s_{0i}}\right) \geq \frac{(1-\lambda+\lambda x_{1i}/x_{0i})(1-\lambda+\lambda s_{1i}/s_{0i})}{1-\lambda+\lambda x_{1i}s_{1i}/x_{0i}s_{0i}} \geq \frac{1-\theta}{n}\,.$$

Let the polarized partition be

$$N := \left\{i \in [n] : \frac{x_{1i}}{x_{0i}} \geq \frac{s_{1i}}{s_{0i}}\right\}, \; B := [n] \setminus N\,. \tag{96}$$

For $i \in B$, we get

$$\frac{x_{1i}}{x_{0i}} \geq \frac{1}{2}\left(\frac{x_{1i}}{x_{0i}} + \frac{s_{1i}}{s_{0i}}\right) \geq \frac{1-\theta}{4n}$$

and hence for all $\lambda \in [0,1]$,

$$\frac{\lambda x_{1i} + (1-\lambda)x_{0i}}{x_{0i}} = 1 - \lambda + \lambda\frac{x_{1i}}{x_{0i}} \geq \min\left\{1, \frac{1-\theta}{4n}\right\} = \frac{1-\theta}{4n}\,.$$

By an analogous derivation, we get $\frac{\lambda s_{1i} + (1-\lambda)s_{0i}}{s_{0i}} \geq \frac{1-\theta}{4n}$ for $i \in N$. Hence, the straight line $[z_1, z_0]$ is $\frac{1-\theta}{4n}$-polarized. By Proposition B.1, the straight lines $[z_1', z_0']$ are $\frac{(1-\theta)^3}{16n^3}$-polarized.

☐

C Strongly Polynomial Eigenvalue Approximation

In this section we provide the strongly polynomial algorithms to compute approximate eigenvalues of positive semi-definite matrices as well as partial singular value decompositions of general matrices (see Definition 2.26 in Section 2.8).

Lemma C.1 (Pivoted Cholesky factorization)**.** *Let $\mathbf{M} \in \mathbb{R}^{n\times n}$ be a positive semi-definite matrix. One can in strongly polynomial time with $O(n^3)$ many arithmetic operations compute a Cholesky factorization*

$$\mathbf{M} = \mathbf{PLDL}^\top\mathbf{P}^\top,$$

where $\mathbf{P}$ is a permutation matrix, $\mathbf{D} = \operatorname{diag}(d_i)_{i\in[n]}$ and $\mathbf{L}$ is unit lower-triangular with all entries being bounded by 1 in absolute value. Furthermore, for the eigenvalues $(\lambda_i(\mathbf{M}))_{i\in[n]}$ of $\mathbf{M}$, we have

$$\frac{d_i}{n^2 2^{2n-2}} \leq \lambda_i(\mathbf{M}) \leq n^2 d_i\,, \quad \forall i \in [n]\,.$$

As a corollary of the above theorem, the trust region direction $\Delta z^{\mathrm{TR}} = (\Delta x^{\mathrm{TR}}, \Delta s^{\mathrm{TR}})$ used in TR2-IPM can be computed in strongly polynomial time, and hence, every iteration of TR2-IPM algorithm can be implemented in strongly polynomial time.

C.1 A deterministic pivoted Cholesky decomposition algorithm in cubic time

In this section, we give a strongly polynomial algorithm to compute multiplicative approximations to the eigenvalues of any positive semi-definite matrix and hence prove Lemma C.1. This algorithm is used as a subroutine in TR2-Solve($\mathbf{B}, b, I, J, \delta$) to compute the approximate critical points of ψ. In the second part of this section, we present a randomized algorithm given in [22] and explain how it can also be used to provide the approximate eigenvalues of a positive semi-definite matrix.

Lemma C.2 (Pivoted Cholesky via Schur Complements for PSD Matrices)**.** *Let $\mathbf{M} \in \mathbb{R}^{n\times n}$ be any positive semi-definite matrix. Then there exists a factorization*

$$\mathbf{M} = \mathbf{PLDL}^\top\mathbf{P}^\top,$$

where $\mathbf{P}$ is a permutation matrix, $\mathbf{D}$ is diagonal with nonnegative entries, and $\mathbf{L}$ is lower-triangular with unit diagonal entries and bounded off-diagonal elements:

$$\operatorname{diag}(\mathbf{L}) = 1 \quad \textit{and} \quad |\mathbf{L}_{ij}| \leq 1 \quad (\textit{for all } i > j).$$

This factorization can be computed in $O(n^3)$ time by selecting the largest diagonal entry as pivot at each stage and performing symmetric row/column interchanges.

Proof. We proceed by induction on the matrix dimension. The base case $n = 1$ is trivial. For the inductive step, assume the statement holds for any $(n-1)\times(n-1)$ positive semidefinite matrix.

Given an $n \times n$ positive semidefinite matrix $\widehat{\mathbf{M}}$, let $p \in [n]$ index its largest diagonal entry. Define $\mathbf{S} \in \mathbb{R}^{n\times n}$ as the permutation matrix that swaps the first and p-th rows/columns, and let $\mathbf{M} = \mathbf{S}\widehat{\mathbf{M}}\mathbf{S}^\top$. Now $\mathbf{M}_{1,1}$ is the largest diagonal entry of $\mathbf{M}$. If $\mathbf{M}_{1,1} = 0$, then $\mathbf{M} = \mathbf{0}_{n\times n}$ and the factorization is trivial. Otherwise, we can use the Schur complement to write:

$$\mathbf{M} = \begin{bmatrix} 1 & \mathbf{0}_{n-1}^\top \\ \mathbf{M}_{1,1}^{-1}\mathbf{M}_{[2:n],1} & \mathbf{I}_{n-1,n-1} \end{bmatrix} \begin{bmatrix} \mathbf{M}_{1,1} & \mathbf{0}_{n-1}^\top \\ \mathbf{0}_{n-1} & \mathbf{M}_{[2:n],[2:n]} - \mathbf{M}_{1,1}^{-1}\mathbf{M}_{[2:n],1}\mathbf{M}_{1,[2:n]} \end{bmatrix} \begin{bmatrix} 1 & \mathbf{M}_{1,1}^{-1}\mathbf{M}_{1,[2:n]} \\ \mathbf{0}_{n-1} & \mathbf{I}_{n-1,n-1} \end{bmatrix}. \tag{97}$$

Let $\widetilde{\mathbf{M}} = \mathbf{M}_{[2:n],[2:n]} - \mathbf{M}_{1,1}^{-1}\mathbf{M}_{[2:n],1}\mathbf{M}_{1,[2:n]}$ be the Schur complement. By standard properties, $\widetilde{\mathbf{M}}$ is positive semidefinite. Moreover, the positive semidefiniteness of $\mathbf{M}$ implies (e.g., by Sylvester's criterion) that principal minors are nonnegative and so for any $i > 1$:

$$0 \leq \det \begin{bmatrix} \mathbf{M}_{1,1} & \mathbf{M}_{1,i} \\ \mathbf{M}_{i,1} & \mathbf{M}_{i,i} \end{bmatrix} = \mathbf{M}_{1,1}\mathbf{M}_{i,i} - \mathbf{M}_{1,i}\mathbf{M}_{i,1} \leq \mathbf{M}_{1,1}^2 - \mathbf{M}_{i,1}^2\,, \tag{98}$$

which shows that $\mathbf{M}_{1,1} \geq |\mathbf{M}_{i,1}| = |\mathbf{M}_{1,i}|$.

By the inductive hypothesis, $\widetilde{\mathbf{M}}$ has a pivoted Cholesky decomposition $\widetilde{\mathbf{M}} = \widetilde{\mathbf{P}}\widetilde{\mathbf{L}}\widetilde{\mathbf{D}}\widetilde{\mathbf{L}}^\top\widetilde{\mathbf{P}}^\top$. Substituting this into our earlier factorization:

$$\mathbf{M} = \begin{bmatrix} 1 & \mathbf{0}_{n-1}^\top \\ \mathbf{M}_{1,1}^{-1}\mathbf{M}_{[2:n],1} & \mathbf{I}_{n-1,n-1} \end{bmatrix} \begin{bmatrix} 1 & \mathbf{0}_{n-1}^\top \\ \mathbf{0}_{n-1} & \widetilde{\mathbf{P}}\widetilde{\mathbf{L}} \end{bmatrix} \begin{bmatrix} \mathbf{M}_{1,1} & \\ & \widetilde{\mathbf{D}} \end{bmatrix} \begin{bmatrix} 1 & \mathbf{0}_{n-1}^\top \\ \mathbf{0}_{n-1} & \widetilde{\mathbf{L}}^\top\widetilde{\mathbf{P}}^\top \end{bmatrix} \begin{bmatrix} 1 & \mathbf{M}_{1,1}^{-1}\mathbf{M}_{1,[2:n]} \\ \mathbf{0}_{n-1} & \mathbf{I}_{n-1,n-1} \end{bmatrix}$$
$$= \begin{bmatrix} 1 & \mathbf{0}_{n-1}^\top \\ \mathbf{0}_{n-1} & \widetilde{\mathbf{P}} \end{bmatrix} \begin{bmatrix} 1 & \mathbf{0}_{n-1}^\top \\ \widetilde{\mathbf{P}}^\top\mathbf{M}_{1,1}^{-1}\mathbf{M}_{[2:n],1} & \widetilde{\mathbf{L}} \end{bmatrix} \begin{bmatrix} \mathbf{M}_{1,1} & \\ & \widetilde{\mathbf{D}} \end{bmatrix} \begin{bmatrix} 1 & \mathbf{M}_{1,1}^{-1}\mathbf{M}_{1,[2:n]}\widetilde{\mathbf{P}} \\ \mathbf{0}_{n-1} & \widetilde{\mathbf{L}}^\top \end{bmatrix} \begin{bmatrix} 1 & \mathbf{0}_{n-1}^\top \\ \mathbf{0}_{n-1} & \widetilde{\mathbf{P}}^\top \end{bmatrix}$$

Define the final matrices:

$$\mathbf{P} \coloneqq \mathbf{S}^\top \begin{bmatrix} 1 & \mathbf{0}_{n-1}^\top \\ \mathbf{0}_{n-1} & \widetilde{\mathbf{P}} \end{bmatrix}, \qquad \mathbf{L} = \begin{bmatrix} 1 & \mathbf{0}_{n-1}^\top \\ \widetilde{\mathbf{P}}^\top\mathbf{M}_{1,1}^{-1}\mathbf{M}_{1,[2:n]} & \widetilde{\mathbf{L}} \end{bmatrix} \mathbf{S}^\top, \qquad \mathbf{D} = \begin{bmatrix} \mathbf{M}_{1,1} & \\ & \widetilde{\mathbf{D}} \end{bmatrix}. \tag{99}$$

We observe that $\mathbf{P}$ is a permutation matrix, $\mathbf{D}$ is diagonal with nonnegative entries, and $\mathbf{L}$ is lower-triangular with unit diagonal entries and off-diagonal elements bounded in absolute value by 1. For the original matrix $\widehat{\mathbf{M}}$, we have:

$$\widehat{\mathbf{M}} = \mathbf{S}^\top\mathbf{M}\mathbf{S} = \mathbf{S}^\top \begin{bmatrix} 1 & \mathbf{0}_{n-1}^\top \\ \mathbf{0}_{n-1} & \widetilde{\mathbf{P}} \end{bmatrix} \mathbf{L}\mathbf{D}\mathbf{L}^\top \begin{bmatrix} 1 & \mathbf{0}_{n-1}^\top \\ \mathbf{0}_{n-1} & \widetilde{\mathbf{P}} \end{bmatrix}^\top \mathbf{S} = \mathbf{P}\mathbf{L}\mathbf{D}\mathbf{L}^\top\mathbf{P}^\top \tag{100}$$

Since all operations except the recursive call take $O(n^2)$ time, the total running time is $O(n^3)$. □

Hence, given $\mathbf{M} \succeq \mathbf{0}$, one can compute its pivoted Cholesky factorization $\mathbf{M} = \mathbf{P}\mathbf{L}\mathbf{D}\mathbf{L}^\top\mathbf{P}^\top$ as in Lemma C.2. We next show that the diagonal entries of $\mathbf{D}$ provide multiplicative estimates on the eigenvalues of $\mathbf{M}$. For this, we require the following two standard results.

Fact C.3 (Inverse Norm Bound of Triangular factor)**.** *Let $\mathbf{L}$ be an $n \times n$ lower-triangular matrix with* $\operatorname{diag}(\mathbf{L}) = 1$*. Then its inverse $\mathbf{L}^{-1}$ satisfies*

$$\|\mathbf{L}^{-1}\|_\infty \le 2^{n-1}\|\mathbf{L}\|_\infty^{n-1}.$$

In particular, if $|\mathbf{L}_{ij}| \le 1$ for all $i > j$ then $\|\mathbf{L}^{-1}\|_\infty \le 2^{n-1}$.

Proof. The absolute value of the entry (i,j) of $\mathbf{L}^{-1}$ is given by $|\operatorname{adj}(\mathbf{L})_{i,j}/\det(\mathbf{L})|$, where $\operatorname{adj}(\mathbf{L})$ is the adjugate matrix of $\mathbf{L}$. Clearly, $\det(\mathbf{L}) = 1$. Furthermore, $\operatorname{adj}(\mathbf{L})_{i,j}$ is the determinant of a $(n-1)\times(n-1)$ submatrix $\bar{\mathbf{L}}$ of $\mathbf{L}$. Since $\mathbf{L}$ is a lower triangular matrix, the submatrix $\bar{\mathbf{L}}$ is a lower Hessenberg matrix with $\|\bar{\mathbf{L}}\|_\infty \le \|\mathbf{L}\|_\infty$.

Using the Leibniz formula, we can bound:

$$|\det(\bar{\mathbf{L}})| = \left| \sum_{\tau \in S_{n-1}} \operatorname{sgn}(\tau) \prod_{k=1}^{n-1} \bar{\mathbf{L}}_{k,\tau(k)} \right| \le |S_{n-1}| \|\bar{\mathbf{L}}\|_\infty^{n-1} \le (n-1)!\ \|\mathbf{L}\|_\infty^{n-1},$$

where S_{n-1} is the permutation group on $n-1$ elements of size $(n-1)!$. However, due to the Hessenberg structure of $\bar{\mathbf{L}}$, the number of permutations $\tau \in S_{n-1}$ such that $\bar{\mathbf{L}}_{k,\tau(k)} \ne 0$ for all $k \in [n-1]$ is bounded by 2^{n-1}. Hence, $|\det(\bar{\mathbf{L}})| \le 2^{n-1}\|\mathbf{L}\|_\infty^{n-1}$. □

Fact C.4. *Let $\mathbf{M} \in \mathbb{R}^{n\times n}$ and $\mathbf{N} \in \mathbb{R}^{n\times n}$ be symmetric, and let $\mathbf{R} \in \mathbb{R}^{n\times n}$ be invertible. Let $\mathbf{M} = \mathbf{R}\mathbf{N}\mathbf{R}^\top$. Then, for any $i = 1,\ldots,n$, we have that*

$$\frac{1}{\|\mathbf{R}^{-1}\|^2}\sigma_i(\mathbf{N}) \le \sigma_i(\mathbf{M}) \le \|\mathbf{R}\|^2\sigma_i(\mathbf{N}).$$

Corollary C.5 (Pivoted Cholesky factorization approximates eigenvalues)**.** *Let $\mathbf{M}$ be a positive semidefinite matrix and let $\mathbf{P}\mathbf{L}\mathbf{D}\mathbf{L}^\top\mathbf{P}^\top$ be a pivoted Cholesky decomposition of $\mathbf{M}$ as in Lemma C.2. Then, for any $i = 1,\ldots,n$ we have that*

$$\frac{1}{2^{2(n-1)}}\sigma_i(\mathbf{D}) \le \sigma_i(\mathbf{M}) \le n^2\sigma_i(\mathbf{D}).$$

Proof. Note that $\mathbf{P}$ does not affect the spectrum of $\mathbf{M}$ and so by Fact C.4 we have that

$$\frac{1}{\|\mathbf{L}^{-1}\|^2}\sigma_i(\mathbf{D}) \le \sigma_i(\mathbf{M}) \le \|\mathbf{L}\|^2\sigma_i(\mathbf{D}).$$

By Fact C.3 we have that $\|\mathbf{L}^{-1}\|n\|\mathbf{L}^{-1}\|_\infty \leq n2^{n-1}$ and furthermore $\|\mathbf{L}\|_2 \leq n\|\mathbf{L}\|_\infty = n$. We conclude that

$$\frac{1}{n^2 2^{2n-2}}\sigma_i(\mathbf{D}) \leq \sigma_i(\mathbf{M}) \leq n^2\sigma_i(\mathbf{D}).$$

□

The proof of Lemma C.1 readily follows.

Proof of Lemma C.1. It is an immediate consequence of Lemma C.2 and Corollary C.5. □

C.2 A randomized algorithm in matrix multiplication time

In this section, we show that the computations in the algorithm by Diakonikolas, Tzamos, and Kane [22] can be modified to run in matrix multiplication time $\tilde{O}(n^\omega)$. Let us first verbatim state their result.

Proposition C.6 ([22, Proposition 4.1])**.** *Given an $n \times n$ PSD matrix $\mathbf{M}$, an accuracy parameter $\varepsilon > 0$ and a failure probability $\delta > 0$, there is an algorithm (Algorithm 7) that computes orthogonal vectors $q_1, \dots, q_n$ and scalars a_i, $i \in [n]$ such that the matrix $\widehat{\mathbf{M}} = \sum_{i=1}^n a_i q_i q_i^\top$ satisfies the following: for all $v \in \mathbb{R}^n$, it holds that*

$$|v^\top(\mathbf{M} - \widehat{\mathbf{M}})v| \leq \varepsilon(v^\top \mathbf{M} v)\,. \tag{101}$$

The algorithm performs $\mathrm{poly}(n/\varepsilon, \log(1/\delta))$ arithmetic operations on $\mathrm{poly}(n/\varepsilon, \log(1/\delta), b)$-bit numbers, where b is the bit complexity of the entries of $\mathbf{M}$.

We will only require the correctness statement in Proposition C.6. We need certain modifications compared to [22], because in our computational model, we cannot operate on bits. However, we show that the algorithm can be implemented in $\tilde{O}(n^\omega)$ in the real RAM model. In particular, we will now describe how the algorithm can be implemented in $\tilde{O}(n^\omega)$ time. First, note that the t-th power of $\mathbf{M}$ for $t = O(n^6/\varepsilon^2 \log(n/\delta))$ can be computed in $O(\log t)$ many matrix multiplications via repeated squaring, hence $\tilde{O}(n^\omega(\log(1/\varepsilon) + \log\log(1/\delta)))$ overall time. Furthermore, the size of the numbers is polynomially bounded in the input size. The other computational bottleneck is computing the vectors q_i in line 4. A naïve sequential algorithm would take $O(n^2)$ time per call, resembling the overall $O(n^3)$ complexity of the previous section, where pivoting forced us to compute the orthogonalization sequentially. Here, however, as proven by the correctness in Proposition C.6 no pivoting is needed. The procedure therefore corresponds to computing the QR decomposition of the matrix $\mathbf{M}^t\mathbf{A}$. This matrix has size $n \times n$, and so the QR decomposition can be computed in $\tilde{O}(n^\omega)$ time [18, 23, 72]. This concludes the analysis that the overall algorithm can be implemented in $\tilde{O}(n^\omega(\log(1/\varepsilon) + \log\log(1/\delta)))$ time. The guarantee in Proposition C.6 will provide us with $(1+\epsilon)$-approximations to the eigenvalues of the matrix $\mathbf{M}$. More precisely, the values $a_i = \frac{q_i^\top \mathbf{M} q_i}{q_i^\top q_i}$ approximate the eigenvalues of $\mathbf{M}$ within a factor of $1 \pm \varepsilon$. The proof is standard, nonetheless we provide it for completeness in the appendix (Lemma C.9). Based on this discussion, we arrive at the following theorem.

Theorem C.7. *Given an $n \times n$ PSD matrix $\mathbf{M}$, an accuracy parameter $\varepsilon > 0$ and a failure probability $\delta > 0$, there is an algorithm (Algorithm 7) that can be implemented in $\tilde{O}(n^\omega(\log(1/\epsilon) + \log(\log(1/\delta))))$ time that computes a matrix $\mathbf{Q} \in \mathbb{R}^{n\times n} = [q_1, \dots, q_n]$ with orthogonal columns such that for $a_i = q_i^\top \mathbf{M} q_i/(q_i^\top q_i)$ and $\widehat{\mathbf{M}} = \sum_i a_i q_i q_i^\top$ for all $v \in \mathbb{R}^n$, it holds that*

$$|v^\top(\mathbf{M} - \widehat{\mathbf{M}})v| \leq \varepsilon(v^\top \mathbf{M} v)\,. \tag{102}$$

Furthermore, we have that $(1-\varepsilon)\sigma_i(\mathbf{M}) \leq a_i \leq (1+\varepsilon)\sigma_i(\mathbf{M})$.

C.3 Strongly polynomial singular value approximation

Lemma C.8. *Given a matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ with $m \leq n$ one can in time $\widetilde{O}(n^3)$, deterministically compute a diagonal matrix $\mathbf{D}$ and a matrix $\mathbf{U}$ with orthogonal columns such that for all $x \in \mathbb{R}^n$ we have that $\sqrt{\mathbf{D}}\mathbf{Q}^\top$ is a ϱ-partial SVD of $\mathbf{A}$ for $\varrho = 2^{O(n)}$, i.e.,*

$$\frac{1}{\varrho}\|\sqrt{\mathbf{D}}\mathbf{Q}^\top x\|_2 \leq \|\mathbf{A}x\|_2 \leq \varrho\|\sqrt{\mathbf{D}}\mathbf{Q}^\top x\|_2\,. \tag{103}$$

Algorithm 7: Approximate Eigendecomposition

Input: An $n \times n$ PSD matrix $\mathbf{M}$, an accuracy parameter $\epsilon > 0$ and a failure probability $\delta > 0$.

Let $\mathbf{A}$ be a random $n \times n$ matrix where the entries are i.i.d. uniform samples from $\{1, 2, \ldots, N\}$, for N at least a sufficiently large constant multiple of n/δ.

Let $w_1, w_2, \ldots, w_n$ be the column vectors of $\mathbf{M}^t\mathbf{A}$, for t a sufficiently large constant multiple of $n^6/\epsilon^2 \log(n/\delta)$.

for $i = 1$ **to** n **do**

  Let q_i be the projection of w_i onto the orthogonal complement of $w_1, w_2, \ldots, w_{i-1}$.

  Let $a_i = 0$ if $q_i = 0$ and $a_i = q_i^\top \mathbf{M} q_i/(q_i \cdot q_i)$ otherwise.

end

return $\{a_i, q_i\}$

Proof. Compute the Cholesky decomposition $\mathbf{A}^\top\mathbf{A} = \mathbf{PLDL}^\top\mathbf{P}^\top$ as in Lemma C.2, where $\mathbf{P}$ a permutation matrix, $\mathbf{L}$ lower-triangular with unit diagonal and off-diagonal entries bounded in absolute value by 1, and $\mathbf{D}$ diagonal with nonnegative entries. Let us for the remainder of the proof w.l.o.g. assume that $\mathbf{P} = \mathbf{I}_n$. Let $(\mathbf{Q}, \mathbf{R})$ be the QR decomposition of $\mathbf{L}$, i.e., $\mathbf{L} = \mathbf{QR}$ with $\operatorname{diag}(\mathbf{R}) = \mathbf{1}_n$. We claim that $\mathbf{Q}$ has good conditioning: First, note that

$$\|\mathbf{Q}_{:,i}\|_2 \le \|\mathbf{L}_{:,i}\|_2 \le \sqrt{n} \tag{104}$$

Furthermore,

$$\|\mathbf{Q}_{:,i}\|_2 \ge \|\mathbf{Q}_{\le i,i}\|_2 = \|\mathbf{L}_{\le i,\le i}\mathbf{R}^{-1}_{\le i,i}\|_2 \ge \frac{1}{\|\mathbf{L}^{-1}_{\le i,\le i}\|_2}|[\mathbf{R}^{-1}]_{i,i}| \ge \frac{1}{\|\mathbf{L}^{-1}\|_2}|[\mathbf{R}^{-1}]_{i,i}| \ge \frac{1}{n\|\mathbf{L}^{-1}\|_\infty}|[\mathbf{R}^{-1}]_{i,i}| \ge \frac{1}{n2^n}\,, \tag{105}$$

where the first inequality uses the upper triangular structure of $\mathbf{R}^{-1}$, the third inequality uses the triangular structure of $\mathbf{L}$ and the last inequality uses that Fact C.3 and $[\mathbf{R}^{-1}]_{i,i} = 1$. From here, note that

$$\|\mathbf{R}\|_2 = \|\mathbf{Q}^{-1}\mathbf{L}\|_2 \le \|\mathbf{Q}^{-1}\|_2\|\mathbf{L}\|_2 \le n2^n \cdot n2^n = n^2 2^{2n}\,, \tag{106}$$

where we used (105) and the fact that $\|\mathbf{L}\|_2 \le n\|\mathbf{L}\|_\infty = n$. Also, $\|\mathbf{R}^{-1}\|_2 = \|\mathbf{L}^{-1}\mathbf{Q}\|_2 \le \|\mathbf{L}^{-1}\|_2\|\mathbf{Q}\|_2 \le 2^n \cdot n$. From here, note that for any vector $v \in \mathbb{R}^n$ we have that

$$\|\mathbf{A}\mathbf{Q}^{-\top}v\|_2 = \|\sqrt{\mathbf{D}}\mathbf{L}^\top\mathbf{Q}^{-\top}v\|_2 = \|\sqrt{\mathbf{D}}\mathbf{R}^\top v\|_2\,. \tag{107}$$

We will find lower and upper bounds on the last term. Let $w = \mathbf{R}^\top v$. Then, by upper triangularity of $\mathbf{R}$ and the aforementioned conditioning, we have for all $i \in [n]$ that

$$\|w_{\le i}\|_2 \le \|\mathbf{R}\|_2\|v_{\le i}\|_2 \le n^2 2^{2n}\|v_{\le i}\|_2 \tag{108}$$

Furthermore, by upper triangularity of $\mathbf{R}$ we have that

$$\begin{aligned}
\|w_{\le i}\|_2 &\ge \|[\mathbf{R}_{\le i,\le i}]^{-1}\|_2^{-1}\|v_{\le i}\|_2 \ge \|\mathbf{R}^{-1}\|_2^{-1}\|v_{\le i}\|_2 \ge n^{-2}\|\mathbf{R}^{-1}\|_\infty^{-1}\|v_{\le i}\|_2 \\
&\ge n^{-2}2^{-(n-1)}\|\mathbf{R}\|_\infty^{-(n-1)}\|v_{\le i}\|_2 && \text{(by Fact C.3)} \\
&\ge n^{-2}2^{-n}n^{-2}2^{-2n}\|v_{\le i}\|_2 && \text{(by (106))} \\
&= n^{-4}2^{-3n}\|v_{\le i}\|_2\,. && (109)
\end{aligned}$$

From here, using that the diagonal of $\mathbf{D}$ is non-increasing we get that

$$\begin{aligned}
\|\sqrt{\mathbf{D}}w\|_2 &\le \sqrt{n}\max_i \sqrt{d_i}\|w_{\le i}\|_2 \le \sqrt{n}\max_i \sqrt{d_i}\cdot n^2 2^{2n}\|v_{\le i}\|_2 \\
&\le n^{2.5}2^{2n}\max_i \|\sqrt{\mathbf{D}_{\le i}}v_{\le i}\|_2 \le n^{2.5}2^{2n}\|\sqrt{\mathbf{D}}v\|_2\,,
\end{aligned}$$

and

$$\|\sqrt{\mathbf{D}}v\|_2 \le \sqrt{n}\max_i \sqrt{d_i}|v_i| \le \sqrt{n}\max_i \sqrt{d_i}\|v_{\le i}\|_2 \le \sqrt{n}\max_i \sqrt{d_i}\cdot n^4 2^{3n}\|w_{\le i}\|_2 \le n^{4.5}2^{3n}\|\sqrt{\mathbf{D}}w\|_2\,,$$

We can therefore conclude that for all $v \in \mathbb{R}^n$ we have that

$$\frac{1}{n^{4.5}2^{3n}}\|\sqrt{\mathbf{D}}v\|_2 \leq \|\mathbf{A}\mathbf{Q}^{-\top}v\|_2 \leq n^{2.5}2^{2n}\|\sqrt{\mathbf{D}}v\|_2 \,, \tag{110}$$

or equivalently, for all $x \in \mathbb{R}^n$ we have that

$$\frac{1}{n^{4.5}2^{3n}}\|\sqrt{\mathbf{D}}\mathbf{Q}^\top x\|_2 \leq \|\mathbf{A}x\|_2 \leq n^{2.5}2^{2n}\|\sqrt{\mathbf{D}}\mathbf{Q}^\top x\|_2 \,, \tag{111}$$

which proves the lemma. □

Lemma 2.27 (Deterministic strongly polynomial partial SVD)**.** *There exists an $\tilde{O}(n^3 \max\{1, \log(1/\varepsilon)\})$ deterministic algorithm that given $m \times n$ matrix $\mathbf{A}$ with $m \leq n$ and $\varepsilon > 0$ computes a partial $(1+\varepsilon)$-SVD $\mathbf{U}$ of $\mathbf{A}$.*

Proof. Let $\mathbf{M} = [\mathbf{A}^\top \mathbf{A}]^p$ for some $p = O(n/\varepsilon)$ in time $\tilde{O}(n^\omega \log(p)) = \tilde{O}(n^\omega \log(\varepsilon^{-1}))$ and apply the algorithm from Lemma C.8 to $\mathbf{M}$ to obtain $\mathbf{D}$ and $\mathbf{Q}$ such that for some $\varrho = 2^{O(n)}$

$$\frac{1}{\varrho}\mathbf{Q}\mathbf{D}\mathbf{Q}^\top \preceq \mathbf{M}^2 \preceq \varrho\mathbf{Q}\mathbf{D}\mathbf{Q}^\top \,. \tag{112}$$

By operator monotonicity we therefore have that

$$\varrho^{-1/p}\left(\mathbf{Q}\mathbf{D}\mathbf{Q}^\top\right)^{1/p} \preceq \mathbf{M}^{2/p} \preceq \varrho^{1/p}\left(\mathbf{Q}\mathbf{D}\mathbf{Q}^\top\right)^{1/p} \,, \tag{113}$$

By orthogonality of the columns of $\mathbf{Q}$, we have that

$$(\mathbf{Q}\mathbf{D}\mathbf{Q}^\top)^{1/p} = \mathbf{Q}(\mathbf{Q}^\top\mathbf{Q})^{1/p-1}\mathbf{D}^{1/p}\mathbf{Q}^\top = \mathbf{Q}^{-\top}(\mathbf{Q}^\top\mathbf{Q})^{1/p}\mathbf{D}^{1/p}\mathbf{Q}^\top = \bar{\mathbf{Q}}(\mathbf{Q}^\top\mathbf{Q})^{1/p}\mathbf{D}^{1/p}\bar{\mathbf{Q}}^\top \,, \tag{114}$$

where $\bar{\mathbf{Q}}$ is the matrix with orthonormal columns obtained by normalizing the columns of $\mathbf{Q}$, so in particular $\bar{\mathbf{Q}}^\top\bar{\mathbf{Q}} = \mathbf{I}$ and $\bar{\mathbf{Q}}^{-\top} = \bar{\mathbf{Q}}$. Furthermore, $\mathbf{M}^{2/p} = \mathbf{A}^\top\mathbf{A}$. We can therefore conclude that for all $x \in \mathbb{R}^n$ we have that

$$\varrho^{-1/p}\|(\mathbf{Q}^\top\mathbf{Q})^{1/(2p)}\mathbf{D}^{1/(2p)}\bar{\mathbf{Q}}^\top x\|_2^2 \leq \|\mathbf{A}x\|_2^2 \leq \varrho^{1/p}\|(\mathbf{Q}^\top\mathbf{Q})^{1/(2p)}\mathbf{D}^{1/(2p)}\bar{\mathbf{Q}}^\top x\|_2^2 \,. \tag{115}$$

Note that in our computational model we cannot take roots. In particular, we cannot compute some of the matrices above exactly. However, we can access $\mathbf{Q}$ exactly. By computing $\hat{d}_i \approx \|\mathbf{A}\mathbf{Q}_{:,i}\|_2/\|\mathbf{Q}_{:,i}\|_2$ to high multiplicative accuracy $1+\varepsilon/4$ for $i \in [n]$ we can obtain the diagonal matrix $\hat{\mathbf{D}} = \operatorname{diag}(d_1, \ldots, d_n)$ such that for all $i \in [n]$ we have that

$$\frac{1}{1+\varepsilon}\|\hat{\mathbf{D}}\mathbf{Q}^\top x\|_2 \leq \varrho^{-1/p}\frac{1}{1+\frac{\varepsilon}{4}}\|\hat{\mathbf{D}}\mathbf{Q}^\top x\|_2 \leq \|\mathbf{A}x\|_2 \leq \varrho^{1/p}\left(1+\frac{\varepsilon}{4}\right)\|\hat{\mathbf{D}}\mathbf{Q}^\top x\|_2 \leq (1+\varepsilon)\|\hat{\mathbf{D}}\mathbf{Q}^\top x\|_2 \,, \tag{116}$$

where the first and last inequality follow from our sufficiently large choice of $p = O(n/\varepsilon)$.

□

Lemma 2.28 (Randomized strongly polynomial partial SVD)**.** *There exists an $\tilde{O}(n^\omega \max\{1, \log(1/\varepsilon)\})$ randomized algorithm that given $m \times n$ matrix $\mathbf{A}$ with $m \leq n$ and $\varepsilon > 0$ computes a partial $(1+\varepsilon)$-SVD $\mathbf{U}$ of $\mathbf{A}$.*

Proof. Apply Theorem C.7 to the matrix $\mathbf{A}^\top\mathbf{A}$ with precision $\varepsilon/4$ to obtain a diagonal matrix $\mathbf{D}$ and a matrix $\mathbf{Q}$ with orthogonal columns such that for all $v \in \mathbb{R}^n$ we have that

$$|v^\top(\mathbf{A}^\top\mathbf{A} - \mathbf{Q}\mathbf{D}\mathbf{Q}^\top)v| \leq \frac{\varepsilon}{4}(v^\top\mathbf{A}^\top\mathbf{A}v) \,. \tag{117}$$

The partial $1+\varepsilon$ SVD is given by $\sqrt{\mathbf{D}}\mathbf{Q}^\top$. However, our computational model does not allow us to take roots. We can still approximate these values by computing $\hat{d}_i \approx \|\mathbf{A}\mathbf{Q}_{:,i}\|_2/\|\mathbf{Q}_{:,i}\|_2$ for $i \in [n]$ up to multiplicative accuracy $1+\varepsilon/4$ in time $O(n \log\log(\varepsilon^{-1}))$ per column. For the matrix $\hat{\mathbf{D}} = \operatorname{diag}\left(\hat{d}_1, \ldots, \hat{d}_n\right)$ we then have that for all $x \in \mathbb{R}^n$ that

$$\frac{1}{1+\varepsilon}\|\hat{\mathbf{D}}\mathbf{Q}^\top x\|_2 \leq \frac{1}{1+\frac{\varepsilon}{4}}\left(1-\frac{\varepsilon}{4}\right)\|\hat{\mathbf{D}}\mathbf{Q}^\top x\|_2 \leq \|\mathbf{A}x\|_2 \left(1+\frac{\varepsilon}{4}\right)\left(1+\frac{\varepsilon}{4}\right)\|\hat{\mathbf{D}}\mathbf{Q}^\top x\|_2 \leq (1+\varepsilon)\|\hat{\mathbf{D}}\mathbf{Q}^\top x\|_2 \,, \tag{118}$$

as desired. □

The following lemma is a classical result. We state and prove it for self-containment.

Lemma C.9. *Let* $\mathbf{M} \in \mathbb{R}^{n\times n}$ *be a positive semidefinite (PSD) matrix. Let* $\widehat{\mathbf{M}} = \sum_{i=1}^n a_i q_i q_i^\top$ *with* $a_i \geq 0$ *for all* $i \in [n]$ *where the vectors* q_i *are orthogonal and furthermore*

$$\left|v^\top(\mathbf{M} - \widehat{\mathbf{M}})v\right| \leq \varepsilon\big(v^\top \mathbf{M} v\big) \quad \textit{for all } v \in \mathbb{R}^n,$$

where $0 < \varepsilon < 1$*. Then the eigenvalues of* $\widehat{\mathbf{M}}$*, namely* $a_i \|q_i\|^2$*, approximate the eigenvalues of* $\mathbf{M}$ *within a factor of* $1 \pm \varepsilon$*.*

Proof. Because $\mathbf{M}$ and $\widehat{\mathbf{M}}$ are PSD, for any $v \in \mathbb{R}^n$ we have

$$-\varepsilon v^\top \mathbf{M} v \leq v^\top(\widehat{\mathbf{M}} - \mathbf{M})v \leq \varepsilon v^\top \mathbf{M} v.$$

If furthermore $v \neq \mathbf{0}$, then dividing by $v^\top v > 0$ and rearranging gives the inequality

$$(1-\varepsilon)\frac{v^\top \mathbf{M}\, v}{v^\top v} \leq \frac{v^\top \widehat{\mathbf{M}} v}{v^\top v} \leq (1+\varepsilon)\frac{v^\top \mathbf{M} v}{v^\top v}.$$

In other words, the Rayleigh quotient of $\widehat{\mathbf{M}}$ for any vector v lies within a $1 \pm \varepsilon$ factor of the Rayleigh quotient of $\mathbf{M}$. Recall the Courant–Fischer (min–max) theorem for the ith largest eigenvalue of a symmetric PSD matrix A:

$$\lambda_i(A) = \max_{\dim(U)=n-i+1} \min_{\substack{v\in U \\ v\neq 0}} \frac{v^\top A v}{v^\top v}.$$

Applying this formula to $A = \mathbf{M}$ and to $A = \widehat{\mathbf{M}}$, and using the fact that every Rayleigh quotient of $\widehat{\mathbf{M}}$ is within $(1 \pm \varepsilon)$ times the corresponding Rayleigh quotient of $\mathbf{M}$, one obtains

$$(1-\varepsilon)\,\lambda_i(\mathbf{M}) \leq \lambda_i(\widehat{\mathbf{M}}) \leq (1+\varepsilon)\,\lambda_i(\mathbf{M}).$$

Therefore, each eigenvalue of $\widehat{\mathbf{M}}$ is a $1 \pm \varepsilon$ relative approximation to the corresponding eigenvalue of $\mathbf{M}$.

As $\widehat{\mathbf{M}} = \sum_{i=1}^n a_i q_i q_i^\top$, with the q_i are orthogonal, we have assuming w.l.o.g. the ordering $a_1 \|q_1\|_2^2 \geq \cdots \geq a_n \|q_n\|_2^2$ that the ith largest eigenvalue of $\widehat{\mathbf{M}}$ is $a_i \|q_i\|_2^2$. ☐

References

[1] MOSEK optimization toolbox for MATLAB. https://docs.mosek.com/latest/toolbox.pdf, 2025. Release 11.0.13.

[2] W. Ai. Neighborhood-following algorithms for linear programming. *Science in China Series A: Mathematics*, 47(6):812–820, 2004.

[3] W. Ai and S. Zhang. An $O(\sqrt{n}L)$ iteration primal-dual path-following method, based on wide neighborhoods and large updates, for monotone lcp. *SIAM Journal on Optimization*, 16(2):400–417, 2005.

[4] 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.

[5] X. Allamigeon, D. Dadush, G. Loho, B. Natura, and L. A. Végh. Interior point methods are not worse than simplex. *SIAM Journal on Computing*, page (available online), 2025.

[6] 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.

[7] K. M. Anstreicher. Volumetric path following algorithms for linear programming. *Mathematical programming*, 76(1):245–263, 1997.

[8] D. Armentano, C. Beltrán, P. Bürgisser, F. Cucker, and M. Shub. A stable, polynomial-time algorithm for the eigenpair problem. *Journal of the European Mathematical Society*, 20(6):1375–1437, 2018.

[9] J. Banks, J. Garza-Vargas, A. Kulkarni, and N. Srivastava. Pseudospectral shattering, the sign function, and diagonalization in nearly matrix multiplication time. *Foundations of Computational Mathematics*, 23(6):1959–2047, 2023.

[10] J.-Y. Cai. Computing Jordan normal forms exactly for commuting matrices in polynomial time. *International Journal of Foundations of Computer Science*, 5(03n04):293–302, 1994.

[11] T. F. Chan. Rank revealing qr factorizations. *Linear algebra and its applications*, 88:67–82, 1987.

[12] S. Chandrasekaran and I. C. Ipsen. On rank-revealing factorisations. *SIAM Journal on Matrix Analysis and Applications*, 15(2):592–622, 1994.

[13] M. B. Cohen, Y. T. Lee, and Z. Song. Solving linear programs in the current matrix multiplication time. *Journal of the ACM (JACM)*, 68(1):1–39, 2021.

[14] M. Colombo and J. Gondzio. Further development of multiple centrality correctors for interior point methods. *Computational Optimization and Applications*, 41(3):277–305, 2008.

[15] A. R. Conn, N. I. Gould, and P. L. Toint. *Trust region methods*. SIAM, 2000.

[16] 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.

[17] D. Dadush, Z. K. Koh, B. Natura, N. Olver, and L. A. Végh. A strongly polynomial algorithm for linear programs with at most two nonzero entries per row or column. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pages 1561–1572, 2024.

[18] J. Demmel, I. Dumitriu, and O. Holtz. Fast linear algebra is stable. *Numerische Mathematik*, 108(1):59–91, Oct. 2007.

[19] J. W. Demmel. *Applied numerical linear algebra*. SIAM, 1997.

[20] A. Deza, E. Nematollahi, R. Peyghami, and T. Terlaky. The central path visits all the vertices of the klee–minty cube. *Optimisation Methods and Software*, 21(5):851–865, 2006.

[21] A. Deza, E. Nematollahi, and T. Terlaky. How good are interior point methods? klee–minty cubes tighten iteration-complexity bounds. *Mathematical Programming*, 113(1):1–14, 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] E. Elmroth and F. G. Gustavson. Applying recursion to serial and parallel qr factorization leads to better performance. *IBM Journal of Research and Development*, 44(4):605–624, 2000.

[24] R. Fagin, A. Lotem, and M. Naor. Optimal aggregation algorithms for middleware. In *Proceedings of the twentieth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, pages 102–113, 2001.

[25] J. Gondzio. Multiple centrality corrections in a primal-dual method for linear programming. *Computational optimization and applications*, 6(2):137–156, 1996.

[26] J. Gondzio. Interior point methods 25 years later. *European Journal of Operational Research*, 218(3):587–601, 2012.

[27] C. C. Gonzaga. Path-following methods for linear programming. *SIAM review*, 34(2):167–224, 1992.

[28] N. I. Gould, S. Lucidi, M. Roma, and P. L. Toint. Solving the trust-region subproblem using the lanczos method. *SIAM Journal on Optimization*, 9(2):504–525, 1999.

[29] M. Grötschel, L. Lovász, and A. Schrijver. *Geometric algorithms and combinatorial optimization*, volume 2. Springer Science & Business Media, 1988.

[30] 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, 1996.

[31] O. Güler. Limiting behavior of weighted central paths in linear programming. *Mathematical Programming*, 65(1):347–363, 1994.

[32] B. Haeupler, R. Hladík, V. Rozhoň, R. E. Tarjan, and J. Tetěk. Universal optimality of Dijkstra via beyond-worst-case heaps. In *2024 IEEE 65th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 2099–2130. IEEE, 2024.

[33] B. Haeupler, R. Hladík, V. Rozhoň, R. E. Tarjan, and J. Tětek. Bidirectional dijkstra's algorithm is instance-optimal. In *2025 Symposium on Simplicity in Algorithms (SOSA)*, pages 202–215. SIAM, 2025.

[34] T. Hastie, R. Tibshirani, and J. Friedman. The elements of statistical learning: data mining, inference, and prediction, 2009.

[35] P.-F. Hung and Y. Ye. An asymptotical o(nl)-iteration path-following linear programming algorithm that uses wide neighborhoods. *SIAM Journal on Optimization*, 6(3):570–586, 1996.

[36] F. Jarre and M. Wechs. Extending mehrotra's corrector for linear programs. *Advanced Modeling and Optimization*, 1(2):38–60, 1999.

[37] S. Jiang, Z. Song, O. Weinstein, and H. Zhang. A faster algorithm for solving general LPs. In *53rd Annual ACM SIGACT Symposium on Theory of Computing (STOC)*, pages 823–832. ACM, 2021.

[38] 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.

[39] M. Kojima, S. Mizuno, and A. Yoshise. A primal-dual interior point algorithm for linear programming. In N. Megiddo, editor, *Progress in Mathematical Programming: Interior-Point and Related Methods*. Springer, 1989.

[40] 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.

[41] Y. T. Lee and A. Sidford. Path finding methods for linear programming: Solving linear programs in o (vrank) iterations and faster algorithms for maximum flow. In *2014 IEEE 55th Annual Symposium on Foundations of Computer Science*, pages 424–433. IEEE, 2014.

[42] Y. T. Lee and A. Sidford. Solving linear programs with sqrt (rank) linear system solves. *arXiv preprint arXiv:1910.08033*, 2019.

[43] Y. T. Lee and S. S. Vempala. Tutorial on the robust interior point method. *arXiv preprint arXiv:2108.04734*, 2021.

[44] A. Louis and S. S. Vempala. Accelerated Newton iteration for roots of black box polynomials. In *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 732–740. IEEE, 2016.

[45] N. Megiddo. Pathways to the optimal set in linear programming. In *Progress in mathematical programming: Interior-point and related methods*, pages 131–158. Springer, 1989.

[46] N. Megiddo, S. Mizuno, and T. Tsuchiya. A modified layered-step interior-point algorithm for linear programming. *Mathematical Programming*, 82(3):339–355, 1998.

[47] S. Mehrotra. On the implementation of a primal-dual interior point method. *SIAM Journal on optimization*, 2(4):575–601, 1992.

[48] S. Mehrotra and Z. Li. Convergence conditions and krylov subspace–based corrections for primal-dual interior-point method. *SIAM Journal on Optimization*, 15(3):635–653, 2005.

[49] S. Mizuno. A new polynomial time method for a linear complementarity problem. *Mathematical programming*, 56(1):31–43, 1992.

[50] S. Mizuno, M. J. Todd, and Y. Ye. On adaptive-step primal-dual interior-point algorithms for linear programming. *Mathematics of Operations Research*, 18(4):964–981, Nov. 1993.

[51] R. D. 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.

[52] 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.

[53] 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.

[54] J. J. Moré and D. C. Sorensen. Computing a trust region step. *SIAM Journal on scientific and statistical computing*, 4(3):553–572, 1983.

[55] Y. Nesterov and A. Nemirovskii. *Interior-point polynomial algorithms in convex programming*. SIAM, 1994.

[56] C.-T. Pan. On the existence and computation of rank-revealing lu factorizations. *Linear Algebra and its Applications*, 316(1-3):199–222, 2000.

[57] V. Y. Pan and Z. Q. Chen. The complexity of the matrix eigenproblem. In *Proceedings of the thirty-first annual ACM symposium on Theory of computing*, pages 507–516, 1999.

[58] J. Peng, T. Terlaky, and Y. Zhao. A predictor-corrector algorithm for linear optimization based on a specific self-regular proximity function. *SIAM Journal on Optimization*, 15(4):1105–1127, 2005.

[59] J. Renegar. A polynomial-time algorithm, based on Newton's method, for linear programming. *Mathematical Programming*, 40(1-3):59–93, 1988.

[60] J. Renegar. *A mathematical view of interior-point methods in convex optimization*. SIAM, 2001.

[61] M. Rojas, S. A. Santos, and D. C. Sorensen. A new matrix-free algorithm for the large-scale trust-region subproblem. *SIAM Journal on optimization*, 11(3):611–646, 2001.

[62] T. Roughgarden. *Beyond the worst-case analysis of algorithms*. Cambridge University Press, 2021.

[63] S. Smale. Newton's method estimates from data at one point. In *The merging of disciplines: New directions in pure, applied, and computational mathematics*, pages 185–196. Springer, 1986.

[64] G. Sonnevend, J. Stoer, and G. Zhao. On the complexity of following the central path of linear programs by linear extrapolation. *Methods of Operations Research*, 63:19–31, 1989.

[65] J. F. Sturm and S. Zhang. On a wide region of centers and primal-dual interior point algorithms for linear programming. *Mathematics of Operations Research*, 22(2):408–431, 1997.

[66] M. Tanneau, M. F. Anjos, and A. Lodi. Design and implementation of a modular interior-point solver for linear optimization. *Mathematical Programming Computation*, 13(3):509–551, 2021.

[67] É. Tardos. A strongly polynomial algorithm to solve combinatorial linear programs. *Operations Research*, pages 250–256, 1986.

[68] M. J. Todd. A lower bound on the number of iterations of an interior-point algorithm for linear programming. In *Numerical Analysis 1993*, pages 237–259. CRC Press, 1993.

[69] M. J. Todd and Y. Ye. A lower bound on the number of iterations of long-step primal-dual linear programming algorithms. *Annals of Operations Research*, 62(1):233–252, 1996.

[70] P. M. Vaidya. A new algorithm for minimizing convex functions over convex sets. *Mathematical programming*, 73(3):291–341, 1996.

[71] 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*, pages 259–278. SIAM, 2020.

[72] J. van den Brand. *Unifying Matrix Data Structures: Simplifying and Speeding up Iterative Algorithms*, page 1–13. Society for Industrial and Applied Mathematics, Jan. 2021.

[73] 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.

[74] A. Vladu. Breaking the barrier of self-concordant barriers: Faster interior point methods for M-matrices. In *Proceedings of the 57th Annual ACM Symposium on Theory of Computing (STOC)*, pages 2213–2224, 2025.

[75] J. H. Wilkinson. Global convergene of tridiagonal QR algorithm with origin shifts. *Linear Algebra and its Applications*, 1(3):409–420, 1968.

[76] V. V. Williams, Y. Xu, Z. Xu, and R. Zhou. New bounds for matrix multiplication: from alpha to omega. In *Proceedings of the 2024 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 3792–3835. SIAM, 2024.

[77] Y. Ye. A new complexity result on minimization of a quadratic function with a sphere constraint. In *Recent Advances in Global Optimization*, pages 19–31. Princeton University Press, 1992.

[78] Y. Ye. *Interior point algorithms: theory and analysis*. John Wiley & Sons, 2011.

[79] G. Zhao. On the relationship between the curvature integral and the complexity of path-following methods in linear programming. *SIAM Journal on Optimization*, 6(1):57–73, 1996.

[80] G. Zhao and J. Stoer. Estimating the complexity of a class of path-following methods for solving linear programs by curvature integrals. *Applied Mathematics and Optimization*, 27(1):85–103, 1993.