# Computational Guarantees for Restarted PDHG for LP based on “Limiting Error Ratios” and LP Sharpness

Zikai Xiong* Robert M. Freund†

April 29, 2024

**Abstract**

In recent years, there has been growing interest in solving linear optimization problems – or more simply “LP” – using first-order methods in order to avoid the costly matrix factorizations of traditional methods for huge-scale LP instances. The restarted primal-dual hybrid gradient method (PDHG) – together with some heuristic techniques – has emerged as a powerful tool for solving huge-scale LPs. However, the theoretical understanding of the restarted PDHG and the validation of various heuristic implementation techniques are still very limited. Existing complexity analyses have relied on the Hoffman constant of the LP KKT system, which is known to be overly conservative, difficult to compute (and hence difficult to empirically validate), and fails to offer insight into instance-specific characteristics of the LP problems. These limitations have limited the capability to discern which characteristics of LP instances lead to easy versus difficult LP instances from the perspective of difficulty of computation. With the goal of overcoming these limitations, in this paper we introduce and develop two purely geometry-based condition measures for LP instances: the “limiting error ratio” and the LP sharpness. We provide new computational guarantees for the restarted PDHG based on these two condition measures. For the limiting error ratio, we provide a computable upper bound and show its relationship with the data instance’s proximity to infeasibility under perturbation. For the LP sharpness, we prove its equivalence to the stability of the LP optimal solution set under perturbation of the objective function. We validate our computational guarantees in terms of these condition measures via specially constructed instances. Conversely, our computational guarantees validate the practical efficacy of certain heuristic techniques (row preconditioners and step-size tuning) that improve computational performance in practice. Finally, we present computational experiments on LP relaxations from the MIPLIB dataset that demonstrate the promise of various implementation strategies.

## 1 Introduction

The focus of this paper is on solving huge-scale instances of linear optimization problems – or more simply “LP”. LP problems abound across a wide variety of applications from manufacturing, transportation, service sciences, to computational science and engineering. Up until very recently, the most successful methods for solving LP problems have been simplex and pivoting methods [9] and interior-point methods [44]; these methods have been extensively studied and implemented in state-of-the-art commercial solvers [18]. In most cases, they are able to obtain high-accuracy

*MIT Operations Research Center, 77 Massachusetts Avenue, Cambridge, MA 02139, USA. zikai@mit.edu. Research supported by AFOSR Grant No. FA9550-22-1-0356.

†MIT Sloan School of Management, 77 Massachusetts Avenue, Cambridge, MA 02139, USA. rfreund@mit.edu. Research supported by AFOSR Grant No. FA9550-22-1-0356.

solution, but the success of these methods relies on repeatedly solving a linear system in each iteration. For LP instances of huge-scale, the matrix factorizations required for solving the linear systems can be prohibitively costly. Moreover, the matrix factorizations are often unable to exploit the natural sparsity of a given LP instance and can have prohibitively large memory requirements. In contrast, first-order methods (FOMs) – and in particular the primal-dual hybrid gradient method (PDHG) [8] – are emerging as an alternative for solving huge-scale LP problems because they do not require the repeated solution of linear equations nor do they impose large memory requirements, thus reducing per-iteration costs. The primary task within each iteration of a FOM for LP is the gradient computation, which typically only requires matrix-vector multiplications (and so can fully take advantage of the sparsity of the LP instance). Moreover, FOMs are more suitable for distributed and parallel computation, and can benefit from modern computational architectures that accelerate computation through distributed systems and graphics processing units (GPUs). And indeed this compatibility with modern hardware architectures underscores the growing importance of FOMs for solving larger-scale LP instances.

One of the better-known implementations of FOMs for solving LP is the solver PDLP [1], which is based on the primal-dual hybrid gradient method (PDHG) [8] to solve the saddlepoint formulation of LP. In the experiments reported in [1], PDLP was able to outperform the commercial solver Gurobi when the LP problem was large-scale. A recent GPU implementation of the PDLP further outperforms traditional algorithms implemented in the state-of-art commercial solvers on more LP instances [28]. Furthermore, [32] presents a distributed version of PDLP that is used to solve practical LP problems with 92 billion non-zeros in the constraint matrix – which is way beyond the capability of any simplex or interior-point method. PDLP is based on PDHG [8] – often referred to as the Chambolle-and-Pock method. PDHG is an operator-splitting method with alternating updates between the primal and dual variables. On top of running the base algorithm PDHG, schemes for restarting PDHG have also been proven in theory to help PDHG achieve faster linear convergence [3, 27] and this theory has yielded impressive speed-ups in practice as well. And in addition to using restarts, PDLP also utilizes various heuristic techniques such as presolving, row preconditioning, and step-size tuning [1].

Like many algorithms for solving LP, the theoretical guarantees for restarted PDHG are far more conservative than what is observed in practice. Indeed, the behavioral drivers of the performance of PDHG in practice have not been well understood. In [3] a global Hoffman constant of the KKT system is relied on as part of the proof of linear convergence of the restarted PDHG. This global Hoffman constant is often overly conservative, and so the gap between the theoretical performance of PDHG and its (impressive) practical performance has been rather large. For example, consider the following small LP instance

$$\mathrm{LP}_\gamma: \quad \min_{x,y\in\mathbb{R}} \ \cos(\gamma)\cdot x - \sin(\gamma)\cdot y \qquad \text{s.t.} \ \begin{cases} \sin(\gamma)\cdot x + \cos(\gamma)\cdot y = 1 \\ x \geq 0, \ y \geq 0 \ , \end{cases} \tag{1.1}$$

for $\gamma \in (0, \pi/2)$, whose optimal solution is $(x^\star, y^\star) = (1, 1/\cos(\gamma))$. We consider $(x, y)$ to be nearly-optimal if $\|x - x^\star, y - y^\star\|_1 \leq 10^{-10}$. As $\gamma \searrow 0$, Figure 1 shows that $\mathrm{LP}_\gamma$ becomes more ill-conditioned in terms of the Hoffman constant of the KKT system (orange line) and this is reflected in the iteration bound of [3] (red line), but this family of LP

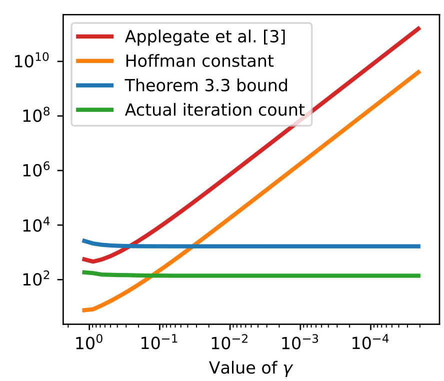


Figure 1: Values of the theoretical iteration bound of [3], Hoffman constant of the KKT system, actual iteration count, and theoretical bound in Theorem 3.3 for $\mathrm{LP}_\gamma$.

instances is consistently easy for the restarted PDHG to solve (green line). This points to the need to develop new condition measures for LP instances, and new theory using such condition measures, whose theoretical iteration bounds are much closer to computational practice. In this paper we propose two such condition measures – which we call LimitingERand LP sharpness, and associated computational theory that is much closer to computational practice, and the blue line in Figure 1 shows how our condition measures and theory translate into iteration bounds for this simple problem that are much closer to computational practice.

Another concurrent work that also use a Hoffman constant to study the convergence of PDHG for LP is [29], which studies regular PDHG (without restarts) applied to LP instances, and uncovers a two-phase behavior: the initial phase is characterized by sublinear convergence, followed by a second phase with linear convergence. The linear convergence of the latter phase upper-bounded using the Hoffman constant of a reduced linear system. The duration of their initial phase is still not well understood, and there are no upper-bound guarantees as of this writing.

In the context of these developments, in this paper we aim to address and answer the following general questions: what are the more natural (and more computable) "condition measures" or "condition numbers" that impact the actual and/or the theoretical performance of PDHG applied to solve LP? How are these condition measures related to existing other condition measures?, and how are these condition measures related to geometric characteristics of LP instances? And can we use answers to these questions to validate current heuristic enhancements of restarted PDHG and develop new heuristic enhancements of restarted PDHG?

## 1.1 Outline and contributions

We study LP in the standard form:

$$\min_{x\in\mathbb{R}^n} \; c^\top x \quad \text{s.t. } Ax = b,\; x \geq 0 \;, \tag{1.2}$$

where the constraint matrix $A \in \mathbb{R}^{m\times n}$, the right-hand side vector $b \in \mathbb{R}^m$, and the objective vector $c \in \mathbb{R}^n$. This paper introduces two geometry-based condition measures: limiting error ratio (LimitingER) and LP sharpness. Based on these condition two measures, we develop new computational guarantees for restarted PDHG.

In Section 3, we provide new computational guarantees for restarted PDHG based on our condition measures LimitingER and the LP sharpness.

The LimitingER is defined as the ratio of the distance from a point (that is in a neighborhood of optimal solutions) to the feasible set to the distance from this point to the inequality constraints. In Section 4 we show that the existence of a nicely interior point guarantees a favorable LimitingER. We also present a computable upper bound the LimitingER and we show a relationship between the LimitingER and the data-perturbation-based condition number of Renegar [40].

The LP sharpness measures how quickly the objective function grows away from the optimal solution set among all feasible points. We prove in Section 5 that the LP sharpness is equal to the the smallest relative perturbation of the objective function $c$ that yields a different optimal solution set that is nonempty and not a subset of the original optimal solution set. We also characterize the LP sharpness in terms of the edges emanating away from the optimal solution set.

In Section 6 we present computational experiments aimed to validate the extent to which our theoretical iteration bounds are "valid", by which we mean that the bounds are consistent with computational practice on specifically chosen families of test problems. We also develop and test two heuristic enhancements of restarted PDHG whose performance is evaluated on LP relaxations from the MIPLIB 2017 dataset.

### 1.2 Other related works for large-scale LPs

In addition to [1, 3, 29] already discussed, other studies have also analyzed the performance of PDHG and its variants for solving LP problems. [26] introduced a stochastic variant of PDHG for solving LPs. [2] studied how to use PDHG for detecting infeasible LP instances, and [27] demonstrated that the regular PDHG (without restarts) also achieves linear convergence on LP problems, but it is slower than its restarted version.

Several other FOMs have also been studied for solving huge-scale LP instances. The method ABIP proposed in [25] is an ADMM-based interior-point method that leverages the framework of the homogeneous self-dual interior-point method and employs ADMM to solve the inner log-barrier problems. [25] proves a sublinear rate of convergence for gradient iterations. In practice, [10] proposes some further enhancements to ABIP, including preconditioning, restarts, and parameter tuning based on ABIP. The SCS method proposed in [35, 34] utilizes ADMM to directly solve the homogeneous self-dual formulation for a general conic optimization problem. [23] studied the superlinear convergence of a semismooth Newton augmented Lagrangian method applied to LP instances.

### 1.3 Notation

For a matrix $A \in \mathbb{R}^{m\times n}$, let $\text{Null}(A) := \{x \in \mathbb{R}^n : Ax = 0\}$ denote the null space of $A$ and $\text{Im}(A) := \{Ax : x \in \mathbb{R}^n\}$ denote the image of $A$. For any set $\mathcal{X} \subset \mathbb{R}^n$, let $P_{\mathcal{X}} : \mathbb{R}^n \to \mathbb{R}^n$ denote the Euclidean projection onto $\mathcal{X}$, namely, $P_{\mathcal{X}}(x) := \arg\min_{\hat{x}\in\mathcal{X}} \|x - \hat{x}\|$. Unless otherwise specified, $\|\cdot\|$ denotes the Euclidean norm. For $M \in \mathbb{S}^n_+$, the set of symmetric positive-semi-definite matrices in $\mathbb{R}^{n\times n}$, we use $\|\cdot\|_M$ to denote the semi-norm $\|z\|_M := \sqrt{z^\top M z}$. For any $x \in \mathbb{R}^n$ and $\mathcal{X} \subset \mathbb{R}^n$, the Euclidean distance between $x$ and $\mathcal{X}$ is denoted by $\text{Dist}(x, \mathcal{X}) := \min_{\hat{x}\in\mathcal{X}} \|x - \hat{x}\|$ and the $M$-norm distance between $x$ and $\mathcal{X}$ is denoted by $\text{Dist}_M(x, \mathcal{X}) := \min_{\hat{x}\in\mathcal{X}} \|x-\hat{x}\|_M$. For simplicity of notation, we use $[n]$ to denote the set $\{1, 2, \ldots, n\}$. For $A \in \mathbb{R}^{n\times n}$, $A^\dagger$ denotes the Moore-Penrose inverse of $A$. For any matrix $A$, $\sigma^+_{\max}(A)$ and $\sigma^+_{\min}(A)$ denote the largest and smallest non-zero singular values of $A$. For an affine subset $V$, let $\vec{V}$ denote the associated linear subspace of $V$, namely $V = \vec{V} + v$ for every $v \in V$. Let $\mathbb{R}^n_+$ and $\mathbb{R}^n_{++}$ denote the nonnegative and strictly positive orthant in $\mathbb{R}^n$, respectively. Let $e$ denote the vector of ones, namely $e = (1, \ldots, 1)^\top$ whose dimension is dictated by context. For a vector $v \in \mathbb{R}^n$, $v^+$ and $v^-$ respectively denote the vector of positive parts and negative parts of $v$, i.e., the components of $v^+$ and $v^-$ are $(v^+)_i = \max\{v_i, 0\}$ and $(v^-)_i = \max\{-v_i, 0\}$ for $i \in [n]$. The operator norm $\|A\|$ of a matrix $A$ is defined as $\|A\| = \sup_{x\neq 0} \frac{\|Ax\|}{\|x\|}$. For a symmetric matrix $A$, $A \succeq 0$ means $A \in \mathbb{S}^n_+$. For a linear subspace $\vec{V} \subset \mathbb{R}^n$, $\vec{V}^\perp$ denotes the orthogonal complement of $\vec{V}$.

## 2 PDHG for Linear Programming

In this paper, we address the LP problem in its standard form (1.2). The feasible set of (1.2), denoted as $\mathcal{F}_p$, is the intersection of the affine subspace $V_p := \{x \in \mathbb{R}^n : Ax = b\}$ and the nonnegative orthant $\mathbb{R}^n_+$, $\mathcal{F}_p := V_p \cap \mathbb{R}^n_+$. The LP problem (1.2) can also be expressed as the saddlepoint problem:

$$\min_{x\in\mathbb{R}^n_+} \max_{y\in\mathbb{R}^m} L(x, y) := c^\top x + b^\top y - x^\top A^\top y \tag{2.1}$$

where $L(x, y)$ is the Lagrangian function. The corresponding linear programming dual problem of (1.2) is:

$$\max_{y\in\mathbb{R}^m,\ s\in\mathbb{R}^n} \quad b^\top y \quad \text{s.t. } A^\top y + s = c,\ s \geq 0\ . \tag{2.2}$$

We further make the following assumption about (1.2) and its dual problem (2.2).

**Assumption 1.** *We assume that the LP problem (1.2) has an optimal solution, and non-optimal feasible solutions exist for the duality-paired problems (1.2) and (2.2).*

Note that if (1.2) has a feasible solution, then there exists $x_0$ such that $b = Ax_0$. If the dual problem (2.2) also has a feasible solution, then for any dual feasible solutions $y_1$, $y_2$ satisfying $A^\top y_1 = A^\top y_2$, the corresponding objective values of $y_1$ and $y_2$ are equal, because

$$b^\top y_1 = x_0^\top A^\top y_1 = x_0^\top A^\top y_2 = b^\top y_2 \ .$$

And since the slack values for $y_1$ and $y_2$ are the same, namely $s := c - A^\top y_1 = c - A^\top y_2$, this implies that dual feasible solutions with identical slack values $s$ have identical objective values. For such a feasible slack value $s$ we can define $\hat{y} := \hat{y}(s) := (AA^\top)^\dagger A(c - s)$, then the objective of the dual problem (2.2) can be expressed as a linear function of $s$, namely $b^\top \hat{y} = b^\top (AA^\top)^\dagger A(c - s)$. Define $q := A^\top (AA^\top)^\dagger b$ ; then (2.2) is equivalent to the following (dual) problem on $s$:

$$\max_{s \in \mathbb{R}^n} \ q^\top (c - s) \quad \text{ s.t. } s \in c + \mathrm{Im}(A^\top), \ s \geq 0 \ . \tag{2.3}$$

Similar to the primal problem (1.2), the feasible set of (2.3) is also the intersection of an affine subspace $V_d := c + \mathrm{Im}(A^\top)$ and the nonnegative orthant $\mathbb{R}^n_+$. Let us denote the feasible set of (2.3) as $\mathcal{F}_d := V_d \cap \mathbb{R}^n_+$.

To summarize, we can rewrite the primal and dual problems in the following symmetric formats:

$$\begin{aligned} \text{(P)} \quad \mathcal{X}^\star := \arg\min_{x \in \mathbb{R}^n} \ & c^\top x \\ \text{s.t. } & x \in \mathcal{F}_p := V_p \cap \mathbb{R}^n_+ \\ & V_p := q + \mathrm{Null}(A) \end{aligned} \qquad \begin{aligned} \text{(D)} \quad \mathcal{S}^\star := \arg\max_{s \in \mathbb{R}^n} \ & q^\top (c - s) \\ \text{s.t. } & s \in \mathcal{F}_d := V_d \cap \mathbb{R}^n_+ \\ & V_d := c + \mathrm{Im}(A^\top) \end{aligned} \tag{2.4}$$

where $q := A^\top (AA^\top)^\dagger b$. (This re-formulation of the dual was, to the best of our knowledge, first proposed in [43].) Here the sets of primal and dual optima are $\mathcal{X}^\star$ and $\mathcal{S}^\star$, respectively, and we we use $\mathcal{Y}^\star$ to denote the corresponding optimal solutions $y$ associated with $\mathcal{S}^\star$. Let us denote by $\vec{V}_p$ and $\vec{V}_d$ the linear subspaces associated with the affine subspaces $V_p$ and $V_d$. Then $\vec{V}_p$ and $\vec{V}_d$ are orthogonal complements. The following fact collects some useful properties of the symmetric formulation (2.4):

**Fact 2.1.** *In the symmetric formulation (2.4), $\vec{V}_d$ is the orthogonal complement of $\vec{V}_p$, i.e., $\vec{V}_d = \vec{V}_p^\perp$. Furthermore, $P_{\vec{V}_p}(c) \in \vec{V}_p$ and $P_{\vec{V}_p}(c) = \arg\min_{v \in V_d} \|v\|$, and $q \in \vec{V}_d$ and $q = \arg\min_{v \in V_p} \|v\|$.*

Note that the duality gap in (2.4) is equal to $\mathrm{Gap}(x, s) := c^\top x - q^\top (c - s)$. Then a solution pair $(x, s)$ is optimal for (2.4) if and only if the following conditions are met:

- Primal feasibility: $\mathrm{Dist}(x, V_p) = 0$ and $\mathrm{Dist}(x, \mathbb{R}^n_+) = 0$,
- Dual feasibility: $\mathrm{Dist}(s, V_d) = 0$ and $\mathrm{Dist}(s, \mathbb{R}^n_+) = 0$, and
- Nonpositive duality gap: $\mathrm{Gap}(x, s) := c^\top x - q^\top (c - s) \leq 0$.

The optimal primal-dual solutions can then be directly written as

$$\mathcal{X}^\star \times \mathcal{S}^\star := \left\{ (x, s) \,\middle|\, x \in V_p, \ x \in \mathbb{R}^n_+, \ s \in V_d, \ s \in \mathbb{R}^n_+, \ \mathrm{Gap}(x, s) \leq 0 \right\} \ .$$

In our theoretical development we will measure the error of a non-optimal pair $(x, s)$ using the distance to optima, defined as:

$$\mathcal{E}_d(x, s) := \max\{\mathrm{Dist}(x, \mathcal{X}^\star), \mathrm{Dist}(s, \mathcal{S}^\star)\} \ . \tag{2.5}$$

(The distance to optima is not conveniently computable, and so in practice it is more typical to compute the relative error defined as $\mathcal{E}_r(x, y) := \frac{\|Ax^+ - b\|}{1+\|b\|} + \frac{\|(c - A^\top y)^-\|}{1+\|c\|} + \frac{|c^\top x^+ - b^\top y|}{1+|c^\top x^+|+|b^\top y|}$.)

For the remainder of this paper we will use $z \in \mathbb{R}^{m+n}$ to denote the pair $(x, y)$, which are the primal and dual variables for the duality-paired LP problems (1.2) and (2.2), respectively. We will freely use this notation with sub/superscripts and other modifications, so that, for example, $\bar{z}^{n,k} = (\bar{x}^{n,k}, \bar{y}^{n,k})$. In a similar way we will use $\mathcal{Z}^\star$ to denote the set of optimal solutions of the saddlepoint problem (2.1), which can be also be written as the cartesian product $\mathcal{X}^\star \times \mathcal{Y}^\star$.

### 2.1 PDHG for linear programming

The PDHG was introduced in [12, 38] to solve convex-concave saddlepoint problems. One PDHG update for the saddlepoint problem formulation of LP (2.1) is defined in Algorithm 1. As we use

**Algorithm 1:** One step of PDHG on $(x, y)$ for problem (2.1)

```
Function OnePDHG(x, y)
    x⁺ ← P_{ℝⁿ₊}(x − τ(c − Aᵀy));
    y⁺ ← y + σ(b − A(2x⁺ − x));
    return (x⁺, y⁺);
```

$z := (x, y) \in \mathbb{R}^{m+n}$ to denote the combined primal/dual iterates, then PDHG generates iterates as follows:

$$z^{k+1} \leftarrow \textsc{OnePDHG}(z^k) \quad \text{for } k = 0, 1, 2, \ldots .$$

The convergence guarantees for PDHG rely on the step-sizes $\tau$ and $\sigma$ being sufficiently small. In particular, if the following condition is satisfied:

$$M := \begin{pmatrix} \frac{1}{\tau} I_n & -A^\top \\ -A & \frac{1}{\sigma} I_m \end{pmatrix} \succeq 0 \ , \tag{2.6}$$

then PDHG's average iterates will converge to a saddlepoint of the problem (2.1) sublinearly [7]. The above requirement is equivalently written as:

$$\tau > 0,\ \sigma > 0, \ \text{ and } \tau\sigma \le \left(\frac{1}{\sigma_{max}^+(A)}\right)^2 \ , \tag{2.7}$$

where $\sigma_{max}^+(A)$ is the largest positive singular value of $A$.

Furthermore, the matrix $M$ defined in (2.6) turns out to be particularly useful in analyzing the convergence of PDHG through its induced inner product norm defined by $\|z\|_M := \sqrt{z^\top M z}$, which will be used extensively in the rest of this paper.

### 2.2 Sublinear convergence of the average iterates

For notational convenience let us define:

$$\lambda_{\max} := \sigma_{\max}^+(A)\,,\ \lambda_{\min} := \sigma_{\min}^+(A)\,,\ \text{ and } \kappa := \frac{\lambda_{\max}}{\lambda_{\min}} \ . \tag{2.8}$$

To measure the convergence of PDHG for LP, [3] introduces the "normalized duality gap," which will serve as an upper bound for evaluating error tolerances and distances to optimality. [3] also presents an algorithm to efficient approximate it.

**Definition 2.1** (Normalized duality gap, (4a) in [3]). *For any $z = (x, y) \in \mathbb{R}^n_+ \times \mathbb{R}^m$ and $r > 0$, define*

$$B(r; z) := \{\hat{z} := (\hat{x}, \hat{y}) : \hat{x} \ge 0 \text{ and } \|\hat{z} - z\|_M \le r\} \ .$$

*The normalized duality gap of the saddlepoint problem (2.1) is then defined as*

$$\rho(r; z) := \left(\frac{1}{r}\right) \max_{\hat{z} \in B(r;z)} \left[L(x, \hat{y}) - L(\hat{x}, y)\right] \ . \tag{2.9}$$

Note in Definition 2.1 that $B(r; z)$ is technically not a ball in the usual sense of the term, since the requirement that $\hat{x} \ge 0$ means that $B(r; z)$ is not necessarily symmetric relative to its center $z$. We have the following lemma (which is a variation of Lemma 3 of [3] but applies to distances instead of errors) that shows that the normalized duality gap provides an upper bound on the distances to feasibility and on the magnitude of the duality gap.

**Lemma 2.1.** *For any $r > 0$, $\bar{z} := (\bar{x}, \bar{y})$ such that $\bar{x} \ge 0$, and $\bar{s} := c - A^\top \bar{y}$, the normalized duality gap $\rho(r; \bar{z})$ provides the following bounds:*

1. *Primal near-feasibiltiy:* $\mathrm{Dist}(\bar{x}, V_p) \le \frac{1}{\sqrt{\sigma}\lambda_{\min}} \cdot \rho(r; \bar{z})$ *and* $\mathrm{Dist}(\bar{x}, \mathbb{R}^n_+) = 0$,
2. *Dual near-feasibility:* $\mathrm{Dist}(\bar{s}, V_d) = 0$ *and* $\mathrm{Dist}(\bar{s}, \mathbb{R}^n_+) \le \frac{1}{\sqrt{\tau}} \cdot \rho(r; \bar{z})$, *and*
3. *Duality gap:* $\mathrm{Gap}(\bar{x}, \bar{s}) \le \max\{r, \|\bar{z}\|_M\}\rho(r; \bar{z})$.

*Proof.* From the definition of $\rho(r; \cdot)$ we have:

$$L(\bar{x}, y) - L(x, \bar{y}) \le r\rho(r; \bar{z}) \quad \text{for any } z \in B(r; \bar{z}) \ . \tag{2.10}$$

We first examine primal near-feasibility. It holds trivially from the supposition that $\bar{x} \ge 0$ that $\mathrm{Dist}(\bar{x}, \mathbb{R}^n_+) = 0$. Let us now show the upper bound on $\mathrm{Dist}(\bar{x}, V_p)$. We assume that $\mathrm{Dist}(\bar{x}, V_p) > 0$ as otherwise the upper bound holds trivially. Let $\hat{x} = \arg\min_{x \in V_p} \|x - \bar{x}\|$ and hence $\mathrm{Dist}(\bar{x}, V_p) = \|\hat{x} - \bar{x}\|$. Note from the standard optimality conditions that $\hat{x} - \bar{x} \in \mathrm{Im}(A^\top)$ and hence there exists $w$ such that $\hat{x} - \bar{x} = A^\top w$ and also $w \in \mathrm{Im}(A)$. It further holds that $A^\top w \ne 0$, since $\mathrm{Dist}(\bar{x}, V_p) > 0$.

Define $y := \bar{y} + \sqrt{\sigma} r \cdot w/\|w\|$ and set $z := (\bar{x}, y)$, whereby $z \in B(r; \bar{z})$ and hence from (2.10) we have

$$r\rho(r; \bar{z}) \ge L(\bar{x}, y) - L(\bar{x}, \bar{y}) = (b - A\bar{x})^\top (y - \bar{y}) = (\hat{x} - \bar{x})^\top A^\top (y - \bar{y}) = w^\top A A^\top w \sqrt{\sigma} r/\|w\| \ .$$

It then follows that

$$\mathrm{Dist}(\bar{x}, V_p) = \|\hat{x} - \bar{x}\| = \|A^\top w\| \le \frac{\rho(r; \bar{z})}{\sqrt{\sigma}} \cdot \frac{\|w\|}{\|A^\top w\|} = \frac{\rho(r; \bar{z})}{\sqrt{\sigma}} \cdot \frac{\|w\|}{\|w\|_{AA^\top}} \le \frac{\rho(r; \bar{z})}{\sqrt{\sigma}\lambda_{\min}} \ ,$$

where the last inequality above follows since $\lambda_{\min} = \min_{v \in \mathrm{Im}(A)} \frac{\|v\|_{AA^\top}}{\|v\|}$. This proves item *1*.

Let us now examine dual near-infeasibility. Notice that by definition it holds that $\bar{s} \in V_d$. Define $x := \bar{x} + \sqrt{\tau} r \cdot (\bar{s})^- / \|(\bar{s})^-\|$ and set $z := (x, \bar{y})$, whereby $z \in B(r; \bar{z})$ and hence from (2.10) we have

$$r\rho(r; \bar{z}) \ge L(\bar{x}, \bar{y}) - L(x, \bar{y}) = (c - A^\top \bar{y})^\top (\bar{x} - x) = -\bar{s}^\top (\bar{s})^- \sqrt{\tau} r / \|(\bar{s})^-\| = \sqrt{\tau} r \|(\bar{s})^-\| \ ,$$

and hence $\mathrm{Dist}(\bar{s}, \mathbb{R}^n_+) = \|(\bar{s})^-\| \le \frac{1}{\sqrt{\tau}} \cdot \rho(r; \bar{z})$. This proves item *2*.

Lastly, we examine the duality gap $\mathrm{Gap}(\bar{x}, \bar{s}) = c^\top \bar{x} - b^\top \bar{y}$, and we consider two cases, namely $\bar{z} = 0$ and $\bar{z} \ne 0$. If $\bar{z} = 0$, then $\mathrm{Gap}(\bar{x}, \bar{s}) = c^\top \bar{x} - b^\top \bar{y} = 0$, which satisfies the duality gap bound trivially. If $\bar{z} \neq 0$, then define $z := \bar{z} - \min\{\frac{r}{\|\bar{z}\|_M}, 1\}\bar{z}$, which satisfies $\|z - \bar{z}\|_M \leq r$. Substituting this value of $z$ in (2.10) yields:

$$r\rho(r;\bar{z}) \geq L(\bar{x}, y) - L(x, \bar{y}) = \min\left\{\frac{r}{\|\bar{z}\|_M}, 1\right\}(c^\top \bar{x} - b^\top \bar{y}) \ , \tag{2.11}$$

which simplifies to

$$c^\top \bar{x} - b^\top \bar{y} \leq \max\{r, \|\bar{z}\|_M\}\rho(r;\bar{z}) \ . \tag{2.12}$$

This proves the desired bound in item *3*. □

Let the $k$-th iterate of PDHG be denoted as $z^k := (x^k, y^k)$ for $k = 0, 1, \ldots$, and let the average of the first $K$ iterates be denoted as $\bar{z}^K = (\bar{x}^K, \bar{y}^K) := \frac{1}{K}\sum_{i=1}^K (x^i, y^i)$ for $K \geq 1$. The following lemma concerns the sublinear convergence of the average iterates of PDHG for the saddlepoint LP formulation (2.1).

**Lemma 2.2** (Sublinear convergence of PDHG, Remark 2 in [8])**.** *Suppose that $\sigma, \tau$ satisfy (2.7). For all $K \geq 1$, and for all $x \geq 0$ and $y$ and $z = (x, y)$, it holds that*

$$L(\bar{x}^K, y) - L(x, \bar{y}^K) \leq \frac{\|z - z^0\|_M^2}{K} \ . \tag{2.13}$$

Furthermore, the iterates generated by PDHG satisfy the following desirable distance properties to the set of saddlepoints $\mathcal{Z}^\star$, as stated in the following lemma.

**Lemma 2.3** (Nonexpansive property)**.** *Suppose that $\sigma, \tau$ satisfy (2.7). For any saddlepoint $z^*$ of (2.1), and for all $k \geq 0$,*

$$\|z^{k+1} - z^\star\|_M \leq \|z^k - z^\star\|_M \ . \tag{2.14}$$

*Therefore under the assignment $z := z^k$ or $z := \bar{z}^k$ it holds that $\|z - z^\star\|_M \leq \|z^0 - z^\star\|_M$.*

Lemma 2.3 is essentially a restatement of Proposition 2 in [3]. The inequality (2.14), also known as the nonexpansive property, appears in many other operator splitting methods [24, 41, 3].

In [3] it was proved that the normalized duality gap converges to zero at the rate $O(1/K)$. Here we show the proof for completeness and also because it is so short. From the triangle inequality it holds that

$$\|z - z^0\|_M^2 \leq \left(\|z - \bar{z}^K\|_M + \|\bar{z}^K - z^0\|_M\right)^2 \ ,$$

which then implies via Lemma 2.2 that every $z \in B(\|\bar{z}^K - z^0\|_M; \bar{z}^K)$ satisfies

$$L(\bar{x}^K, y) - L(x, \bar{y}^K) \leq \frac{1}{K}\|z - z^0\|_M^2 \leq \frac{\left(\|z - \bar{z}^K\|_M + \|\bar{z}^K - z^0\|_M\right)^2}{K} \leq \frac{4}{K}\|\bar{z}^K - z^0\|_M^2 \ .$$

Therefore

$$\rho(\|\bar{z}^K - z^0\|_M; \bar{z}^K) \leq \frac{4\|\bar{z}^K - z^0\|_M}{K} \ .$$

For any $z^\star \in \mathcal{Z}^\star$, it follows from Lemma 2.3 that $\|\bar{z}^K - z^0\|_M \leq \|\bar{z}^K - z^\star\|_M + \|z^\star - z^0\|_M \leq 2\|z^\star - z^0\|_M$. We therefore have the following corollary.

**Corollary 2.4.** *Suppose that $\sigma, \tau$ satisfy (2.7). Then for all $K \geq 1$ it holds that*

$$\rho(\|\bar{z}^K - z^0\|_M; \bar{z}^K) \leq \frac{8\,\mathrm{Dist}_M(z^0, \mathcal{Z}^\star)}{K} \ . \tag{2.15}$$

The following theorem presents the (sublinear) $O(1/K)$ convergence of PDHG to primal/dual optimal solutions.

**Theorem 2.5.** *Suppose that $\sigma, \tau$ satisfy (2.7), and PDHG is initiated with $z^0 = (x^0, y^0) := (0, 0)$. Then for any $K \geq 1$ and $\bar{x}^K := \frac{1}{K}\sum_{i=1}^K x^i$ and $\bar{s}^K := \frac{1}{K}\sum_{i=1}^K (c - A^\top y^i)$, the following hold:*

1. *Primal near-feasibiltiy:* $\operatorname{Dist}(\bar{x}^K, V_p) \leq \frac{1}{\sqrt{\sigma}\lambda_{\min}} \cdot \frac{8 \operatorname{Dist}_M(0, \mathcal{Z}^\star)}{K}$ *and* $\operatorname{Dist}(\bar{x}^K, \mathbb{R}^n_+) = 0$ ,
2. *Dual near-feasibility:* $\operatorname{Dist}(\bar{s}^K, V_d) = 0$ *and* $\operatorname{Dist}(\bar{s}^K, \mathbb{R}^n_+) \leq \frac{1}{\sqrt{\tau}} \cdot \frac{8 \operatorname{Dist}_M(0, \mathcal{Z}^\star)}{K}$ *, and*
3. *Duality gap:* $\operatorname{Gap}(\bar{x}^K, \bar{s}^K) \leq \frac{16 \operatorname{Dist}_M(0, \mathcal{Z}^\star)^2}{K}$ .

**Remark 2.6.** *The quantity $\operatorname{Dist}_M(0, \mathcal{Z}^\star)$ in Theorem 2.5 is not so natural for measuring the quality of a candidate LP solution, even though it arises quite naturally in the analysis of PDHG (see [8]). At the end of this section, we present Proposition 2.8, which states that $\operatorname{Dist}_M(0, \mathcal{Z}^\star)$ can be replaced by $\frac{\sqrt{2}}{\sqrt{\tau}} \operatorname{Dist}(0, \mathcal{X}^\star) + \frac{\sqrt{2}}{\sqrt{\sigma}\lambda_{\min}} \operatorname{Dist}(c, \mathcal{S}^\star)$, and therefore the results in Theorem 2.5 can be restated in terms of Euclidean norms on the spaces of the primal and dual variables $x$ and $s$.*

Towards the proof of Theorem 2.5 we first present the following lemma.

**Lemma 2.7.** *Suppose $z^a$, $z^b$, and $z^c$ satisfy the nonexpansive properties: $\|z^b - z^\star\|_M \leq \|z^a - z^\star\|_M$ and $\|z^c - z^\star\|_M \leq \|z^a - z^\star\|_M$ for every $z^* \in \mathcal{Z}^*$. Then*

$$\max\{\|z^b - z^c\|_M, \|z^b\|_M\} \leq 2 \operatorname{Dist}_M(z^a, \mathcal{Z}^\star) + \|z^a\|_M \ . \tag{2.16}$$

*Proof.* For any $z^\star \in \mathcal{Z}^\star$ we have

$$\begin{array}{rl} & \max\{\|z^b - z^c\|_M, \|z^b\|_M\} \\ \leq & \max\{\|z^b - z^\star\|_M + \|z^c - z^\star\|_M, \|z^b - z^\star\|_M + \|z^\star\|_M\} \\ \leq & \max\{\|z^a - z^\star\|_M + \|z^a - z^\star\|_M, \|z^a - z^\star\|_M + \|z^\star\|_M\} \\ \leq & \max\{\|z^a - z^\star\|_M + \|z^a - z^\star\|_M, \|z^a - z^\star\|_M + \|z^a - z^\star\|_M + \|z^a\|_M\} \\ = & 2\|z^a - z^\star\|_M + \|z^a\|_M \ , \end{array} \tag{2.17}$$

where the second inequality uses the nonexpansive properties. Setting $z^\star := \arg\min_{z \in \mathcal{Z}^\star} \|z - z^a\|_M$ completes the proof. □

*Proof of Theorem 2.5.* The upper bounds for the primal near-feasibility and dual near-feasibility follow directly from Lemma 2.1 and Corollary 2.4. To prove the bound on the duality gap, let $r = \|\bar{z}^K - z^0\|_M$, and let $z^\star \in \mathcal{Z}^\star$. Then it follows from the duality gap bound in Lemma 2.1 that

$$\begin{aligned} \operatorname{Gap}(\bar{x}^K, \bar{s}^K) &\leq \max\{\|\bar{z}^K - z^0\|_M, \|\bar{z}^K\|_M\} \cdot \rho(\|\bar{z}^K - z^0\|_M; \bar{z}^K) \\ &\leq \|\bar{z}^K\|_M \cdot \frac{8\|z^0 - z^\star\|_M}{K} \ , \end{aligned}$$

where the second inequality above uses Corollary 2.4 and $z^0 = (0, 0)$. Now we apply Lemma 2.7 with $z^a = z^0 = (0, 0)$ and $z^b = z^c = \bar{z}^K$, which yields $\|\bar{z}^K\|_M \leq 2 \operatorname{Dist}_M(0, \mathcal{Z}^\star)$, and we obtain $\operatorname{Gap}(\bar{x}^K, \bar{s}^K) \leq \frac{16 \operatorname{Dist}_M(0, \mathcal{Z}^\star)^2}{K}$, which completes the proof. □

Note that while the product of the step-sizes $\tau\sigma$ is the relevant quantity in the conditions for convergence in (2.7), the ratio between $\tau$ and $\sigma$ can also play a significant role in balancing the three different near-optimality conditions in Theorem 2.5.

Similar to [3], we will develop and analyze a restart scheme for PDHG in Section 3, and we will show that this scheme exhibits linear convergence based on two different kinds of condition measures – one for the limiting error ratios, and the other for the LP sharpness. To pave the way for the analysis in Section 3, we will first discuss limiting error ratios in Section 4, followed by the LP sharpness in Section 5. We end this section with a short subsection on relevant norms.

### 2.3 On norms for PDHG and LP

Theorem 2.5 shows that the iterates of PDHG converge sublinearly to the feasibility sets of the primal and dual as well as the half-space of non-positive duality gaps, with bounds that use the $M$-norm $\|\cdot\|_M$. While the $M$-norm arises naturally in the analysis of PDHG, it has the disadvantage that its quadratic form is not separable in $x$ and $y$ and also it implicitly includes the step-sizes $\tau$ and $\sigma$ in its definition. Towards the goal of stating results in terms of the Euclidean norms on $x$ and $y$, we introduce the following $N$-norm on $(x,y) \in \mathbb{R}^{m+n}$, whose quadratic form is separable in $x$ and $y$. Define:

$$\|(x,y)\|_N := \sqrt{\frac{1}{\tau}\|x\|^2 + \frac{1}{\sigma}\|y\|^2} \ \text{ where } \ N := \begin{pmatrix} \frac{1}{\tau} I_n & \\ & \frac{1}{\sigma} I_m \end{pmatrix} .$$

In comparison with the $M$-norm, the $N$-norm offers advantages in both computation and analysis. Furthermore, when $\tau$ and $\sigma$ are sufficiently small, the $M$-norm and $N$-norm are equivalent up to well-specified constants related to $\tau$ and $\sigma$, as follows.

**Proposition 2.8.** *Suppose that $\sigma, \tau$ satisfy (2.7). Then for any $z = (x,y) \in \mathbb{R}^{m+n}$, it holds that*

$$\sqrt{1 - \sqrt{\tau\sigma}\lambda_{\max}} \cdot \|z\|_N \le \|z\|_M \le \sqrt{2}\|z\|_N \ . \tag{2.18}$$

*Furthermore, if $x \in \mathbb{R}^n_+$ and $s := c - A^\top y \in \mathbb{R}^n$ then*

$$\mathrm{Dist}_M(z, \mathcal{Z}^\star) \le \frac{\sqrt{2}}{\sqrt{\tau}} \mathrm{Dist}(x, \mathcal{X}^\star) + \frac{\sqrt{2}}{\sqrt{\sigma}\lambda_{\min}} \mathrm{Dist}(s, \mathcal{S}^\star) , \quad \textit{and} \tag{2.19}$$

$$\mathrm{Dist}_M(z, \mathcal{Z}^\star) \ge \sqrt{1 - \sqrt{\tau\sigma}\lambda_{\max}} \cdot \max\left\{ \frac{1}{\sqrt{\tau}} \mathrm{Dist}(x, \mathcal{X}^\star), \frac{1}{\sqrt{\sigma}\lambda_{\max}} \mathrm{Dist}(s, \mathcal{S}^\star) \right\} \ . \tag{2.20}$$

The proof of Proposition 2.8 is presented in Appendix B.

## 3 Computational Guarantees for PDHG with Restarts

The general PDHG algorithm typically attains only sublinear convergence for general convex-concave saddlepoint problems, as discussed in [8]. However, [3] has shown that using a well-designed restart strategy improves the convergence of PDHG in both practice and theory. Algorithm 2 describes a general restart scheme for PDHG, based mainly on [3].

---

**Algorithm 2:** PDHG with general Restart scheme

**Input:** Initial iterate $z^{0,0} := (x^{0,0}, y^{0,0})$, $n \leftarrow 0$ ;
**repeat**
  **initialize the inner loop:** inner loop counter $k \leftarrow 0$ ;
  **repeat**
    **conduct one step of PDHG:** $z^{n,k+1} \leftarrow \textsc{OnePDHG}(z^{n,k})$ ;
    **compute the average iterate in the inner loop.** $\bar{z}^{n,k+1} \leftarrow \frac{1}{k+1}\sum_{i=1}^{k+1} z^{n,i}$ ;
    $k \leftarrow k+1$ ;
  **until** *Some (verifiable) restart condition is satisfied by* $\bar{z}^{n,k}$ ;
  **restart the outer loop:** $z^{n+1,0} \leftarrow \bar{z}^{n,k}$, $n \leftarrow n+1$ ;
**until** *Either $z^{n,0}$ is a saddlepoint or $z^{n,0}$ satisfies some other convergence condition* ;
**Output:** $z^{n,0}$ $( = (x^{n,0}, y^{n,0}))$

---

Here we use the notation $z^{k+1} \leftarrow \textsc{OnePDHG}(z^k)$ to denote an iteration of PDHG as described in Algorithm 1. The double superscript on the variable $z^{n,k}$ indexes the outer iteration counter followed by the inner iteration counter, so that $z^{n,k}$ is the $k$-th inner iteration of the $n$-th outer loop. In order to implement Algorithm 2 it is necessary to specify a (verifiable) restart condition on the average iterate $\bar{z}^{n,k}$ in line **8** that is used to determine when to re-start PDHG. We will primarily consider Algorithm 2 using the following restart condition in line **8**:

$$\rho(\|\bar{z}^{n,k} - z^{n,0}\|_M; \bar{z}^{n,k}) \leq \beta \cdot \rho(\|z^{n,0} - z^{n-1,0}\|_M; z^{n,0}) \ , \tag{3.1}$$

for a specific value of $\beta \in (0, 1)$ (in fact we will use $\beta = 1/e$ where $e$ is the base of the natural logarithm). In this way (3.1) is nearly identical to the condition used in [3]. Note that condition (3.1) essentially states that the normalized duality gap shrinks by the factor $\beta$ between restart values $\bar{z}^{n,k}$ and $z^{n,0}$. Note also that one of the reasons for using condition (3.1) is that the normalized duality gap can be easily computed. We will present a computational guarantee for Algorithm 2 in Theorem 3.3 that shows linear convergence at a rate that depends critically on two different types of geometry-related condition measures. The first condition measure is the limiting error ratio and the second condition measure is the LP sharpness.

## 3.1 First condition measure: limiting error ratio (LimitingER)

We first introduce and discuss the limiting error ratio, or "LimitingER" for short. Consider the following generic LP problem:

$$\mathcal{U}^\star := \arg\min_{u \in \mathbb{R}^n} \ g^\top u \quad \text{s.t. } u \in \mathcal{F} := V \cap \mathbb{R}^n_+ \ , \tag{3.2}$$

of which the duality-paired LP problems in (2.4) are each instantiations. The feasible set $\mathcal{F}$ of (3.2) is specifically expressed as the intersection of an affine subspace $V$ and the nonnegative orthant $\mathbb{R}^n_+$. For this format we define the "error ratio" (ER) and the "limiting error ratio" (LimitingER) as follows:

**Definition 3.1** (Error ratios)**.** *Let $\mathcal{F}$ be expressed as the intersection of an affine subspace $V$ and the nonnegative orthant $\mathbb{R}^n_+$. For any $u \in V \setminus \mathcal{F}$ the error ratio (ER) of $\mathcal{F}$ at $u$ is defined as:*

$$\theta(u) := \frac{\text{Dist}(u, \mathcal{F})}{\text{Dist}(u, \mathbb{R}^n_+)} \ , \tag{3.3}$$

*and for $u \in \mathcal{F}$ we define $\theta(u) := 1$. The limiting error ratio (LimitingER) is the quantity $\theta^\star(\mathcal{U}^\star)$ defined as:*

$$\theta^\star(\mathcal{U}^\star) := \lim_{\varepsilon \to 0} \left( \sup_{u \in V, \, \text{Dist}(u, \mathcal{U}^\star) \leq \varepsilon} \theta(u) \right) \ , \tag{3.4}$$

*which is the supremum of $\theta(u)$ for all $u \in V$ approaching $\mathcal{U}^\star$.*

The measure $\theta(u)$ (and also $\theta^\star(\mathcal{U}^\star)$) is similar to error bounds proposed in the literature such as those in [36, 22]. We call $\theta(u)$ an *error ratio* because it is defined exclusively for $u$ in the affine subspace $V$ and it represents the ratio of two types of errors: the distance to the feasible region and the distance to the nonnegative orthant $\mathbb{R}^n_+$. Indeed, many papers have studied different formulations of global upper bounds for the error ratios of linear inequality systems from different perspectives, starting with the celebrated Hoffman bound [19], including [31, 5, 30, 16] among many others, see [36] for a comprehensive survey of relevant results.

Note that for $u \in \mathcal{F} \setminus \mathbb{R}^n_+$ it always holds that $\theta(u) \geq 1$ because the numerator in (3.3) is always at least as large as the denominator (since $\mathcal{F} \subset \mathbb{R}^n_+$). Note also that $\theta(u)$ is bounded above by a Hoffman bound $\mathcal{H}(K)$ on the linear inequality system $Ku = h, u \geq 0$ where $K, h$ are chosen so that $V = \{u : Ku = h\}$. Hence $\theta^\star(\mathcal{U}^\star) \leq \mathcal{H}(K)$ as well. However, $\mathcal{H}(K)$ is likely to be an excessively conservative bound on $\theta^\star(\mathcal{U}^\star)$, since (i) $\mathcal{H}(K)$ is a global bound whereas $\theta^\star(\mathcal{U}^\star)$ is defined locally, and (ii) $u$ must satisfy $u \in V$ in the definition of (3.4) and so has no error in the system $Ku = h$. In fact, we will show in Section 4 the following computable upper bound on $\theta^\star(\mathcal{U}^\star)$, which shows that $\theta^\star(\mathcal{U}^\star)$ cannot be too large if (i) the radius of the optimal solution set of (3.2) is not too large, and (ii) there is a feasible solution $u$ that is not not too close to the boundary of $\mathbb{R}^n_+$ and not too far from the optimal solution set. Such a solution is related to the concept of a "reliable solution" in [11].

**Proposition 3.1.** *Suppose $u_a \in \mathcal{U}^\star$ and there exists $R_a$ for which $\mathcal{U}^\star \subset \{u : \|u - u_a\| \leq R_a\}$, then it holds that $\theta^\star(\mathcal{U}^\star) \leq B^\star$ for the $B^\star$ defined as follows:*

$$B^\star := \inf_{r>0,\ u\in\mathbb{R}^n} \frac{R_a + \|u - u_a\|}{r} \quad \text{s.t. } u \in V,\ u \geq r \cdot e \ . \tag{3.5}$$

Proposition 3.1 above is the first assertion of Proposition 4.2 in Section 4, where we will also show that the optimization problem in (3.5) is easily transformed into a convex second-order cone optimization problem. We also show other properties of $\theta^\star(\mathcal{U}^\star)$ in Section 4, including its relation to the data-perturbation condition number of Renegar [40].

For simplicity of notation, we will use $\theta_p(x)$ and $\theta_d(s)$ to denote the ER for $\mathcal{F}_p$ and $\mathcal{F}_d$ as specified in (2.4). Additionally, we use $\theta_p^\star(\mathcal{X}^\star)$ and $\theta_d^\star(\mathcal{S}^\star)$ to denote the LimitingERs for the primal and dual problems specified in (2.4).

### 3.2 Second condition measure: LP sharpness

Once again we consider the generic LP problem (3.2), of which the duality-paired LP problems in (2.4) are each instantiations. We define the LP sharpness as follows:

**Definition 3.2** (LP sharpness)**.** *Let $f^*$ denote the optimal objective function value of the generic LP instance (3.2), and define $H^\star := \{u \in \mathbb{R}^n : g^\top u = f^\star\}$. The LP sharpness is defined as:*

$$\mu := \inf_{u\in\mathcal{F}\setminus\mathcal{U}^\star} \frac{\text{Dist}(u, V \cap H^\star)}{\text{Dist}(u,\ \mathcal{U}^\star)} \ . \tag{3.6}$$

The LP sharpness measures how quickly the objective function grows away from the optimal solution set $\mathcal{U}^\star$ among all feasible points. Note that $\mu \leq 1$ because $\mathcal{U}^\star \subset V \cap H^\star$ and so the numerator is at most as large as the denominator in (3.6). In the case of LP it is easy to see that $\mu > 0$, but in more general conic optimization this may not be the case.

Sharpness was first introduced by Polyak in [39] as a useful analytical tool in convex minimization. For example, sharpness plus some mild smoothness assumptions can lead to linear convergence of the subgradient descent method via the use of restarts, see [46]. [3] generalizes the sharpness concept from convex optimization to primal-dual saddlepoint problems by defining sharpness on the normalized duality gap. Here we apply the definition of sharpness directly (and naturally) to the LP optimization problem itself.

Note that the LP sharpness $\mu$ is invariant under positive re-scaling of $g$ or under any perturbation $\Delta g \in \vec{V}^\perp$ of $g$, because these changes do not affect either $V \cap H^\star$ or $\mathcal{U}^\star$ (even though they may alter $H^\star$). In Section 5 we will discuss other properties of LP sharpness, including its relationship to the stability of $\mathcal{U}^\star$ under perturbation of $g$, as well as computational issues in computing $\mu$ for a given LP instance.

**Remark 3.2.** *It is straightforward to compute the numerator of* (3.6)*, namely* $\mathrm{Dist}(u, V \cap H^\star) = \frac{|g^\top u - f^\star|}{\|P_{\vec{V}}(g)\|}$*, where* $\vec{V}$ *is the linear subspace associated to* $V$*.*

We will use $\mu_p$ and $\mu_d$ to denote the LP sharpness for the primal and dual problems as specified in (2.4), respectively.

### 3.3 Computational guarantees for PDHG based on LimitingER and LP sharpness

In this section we present and prove our main computational guarantee for PDHG using the simple restart condition defined as follows:

**Definition 3.3** ($\beta$-restart condition)**.** *For a given* $\beta \in (0, 1)$*, the iteration* $(n, k)$ *satisfies the* $\beta$*-restart condition if* $n \geq 1$ *and condition* (3.1) *is satisfied, or* $n = 0$ *and* $k = 1$*.*

In the following theorem we suppose for simplicity and ease of exposition that $c \in \vec{V}$. Also recall the definitions of $\lambda_{\max}$, $\lambda_{\min}$, and $\kappa$ from (2.8).

**Theorem 3.3.** *Suppose* $c \in \vec{V}$*, and that Assumption 1 holds, and that Algorithm 2 (PDHG with Restarts) is run starting from* $z^{0,0} = (x^{0,0}, y^{0,0}) = (0, 0)$ *using the* $\beta$*-restart condition with* $\beta := 1/e$*. Furthermore, let the step-sizes be chosen as follows:*

$$\tau = \frac{\|q\|}{2\kappa\|c\|} \quad \text{and} \quad \sigma = \frac{\|c\|}{2\|q\|\lambda_{\max}\lambda_{\min}} \ . \tag{3.7}$$

*Let* $T$ *be the total number of PDHG iterations that are run in order to obtain* $n$ *for which* $(x^{n,0}, s^{n,0})$ *satisfies* $\mathcal{E}_d(x^{n,0}, s^{n,0}) \leq \varepsilon$*. Then*

$$T \ \leq \ 9e \cdot \mathcal{L} \cdot \ln\left(16e \cdot \mathcal{L} \cdot \frac{\mathcal{E}_d(x^{0,0}, s^{0,0})}{\varepsilon} \cdot \left(1 + \kappa\frac{\|c\|}{\|q\|}\right)\left(1 + \frac{\|q\|}{\|c\|}\right)\right) + 1 \ , \tag{3.8}$$

*where* $\mathcal{L}$ *is defined as follows:*

$$\mathcal{L} := 8.5\kappa\left(\frac{1}{\mu_p} + \frac{1}{\mu_d}\right)\left(\theta_p^\star(\mathcal{X}^\star) + \theta_d^\star(\mathcal{S}^\star) + \frac{\mathrm{Dist}(0, \mathcal{X}^\star)}{\mathrm{Dist}(0, V_p)} + \frac{\mathrm{Dist}(c, \mathcal{S}^\star)}{\mathrm{Dist}(0, V_d)}\right) \ . \tag{3.9}$$

In the theorem we have supposed for simplicity/ease of exposition that $c \in \vec{V}$. Of course, if $c \notin \vec{V}$ one can project $c$ onto $\vec{V}$ if such a projection is not a computational burden (though it might be in the huge-dimensional regime). The theorem holds more generally when $c \notin \vec{V}$, but with substantially more notation and a more complicated expression for the upper bound.

Note that the prescribed step-sizes chosen in (3.7) are relatively easy to compute so long as estimates of the largest and smallest positive singular values of $A$ are easy to compute. In particular there is no need to compute either the LP sharpness or the LimitingER. The optimality tolerance criterion used in the theorem is $\mathcal{E}_d(x, s)$, which is the distance to optima of the primal variable $x$ and the dual slack variable $s$.

Let us now examine the components of the total iteration bound in (3.7) a bit closer. As mentioned above, the ratio of the initial error to the target error $\mathcal{E}_d(x^{0,0}, s^{0,0})/\varepsilon$ appears inside the logarithm term and reflects the global linear convergence property of the algorithm.

The quantity $\mathcal{L}$ appears both inside and outside of the logarithm term and itself involves $\mu_p$ and $\mu_d$ (the LP sharpness for both the primal and dual problems), $\theta_p^\star(\mathcal{X}^\star)$ and $\theta_d^\star(\mathcal{S}^\star)$ (the LimitingER for both the primal and dual problems), and $\frac{\mathrm{Dist}(0,\mathcal{X}^\star)}{\mathrm{Dist}(0,V_p)}$ and $\frac{\mathrm{Dist}(c,\mathcal{S}^\star)}{\mathrm{Dist}(0,V_d)}$. The numerator in $\frac{\mathrm{Dist}(0,\mathcal{X}^\star)}{\mathrm{Dist}(0,V_p)}$ is the norm of the least-norm primal optimal solution, which measures the stability of the dual problem under perturbation of $c$, and the denominator can be interpreted as a lower bound on the norm of any (and every) feasible or optimal solution of the primal, whereby the quotient is interpreted as a relative measure of stability. We argue similarly for $\frac{\mathrm{Dist}(c,\mathcal{S}^\star)}{\mathrm{Dist}(0,V_d)}$. Since the theorem assumes that $c \in \vec{V}_p$, it follows from Fact 2.1 that $\|c\| = \mathrm{Dist}(0, V_d)$ and hence $\mathrm{Dist}(c, \mathcal{S}^\star) \leq \|c\| + \mathrm{Dist}(0, \mathcal{S}^\star) \leq 2\,\mathrm{Dist}(0, \mathcal{S}^\star)$. Therefore $\frac{\mathrm{Dist}(c,\mathcal{S}^\star)}{\mathrm{Dist}(0,V_d)} \leq 2\frac{\mathrm{Dist}(0,\mathcal{S}^\star)}{\mathrm{Dist}(0,V_d)}$ and so is within a factor of 2 of the same measure of stability as for the primal. Notice that higher values of $\mu_p^{-1}$, $\mu_d^{-1}$, $\theta_p^\star(\mathcal{X}^\star)$, and/or $\theta_d^\star(\mathcal{S}^\star)$ result in a higher value of $\mathcal{L}$. Also, all four cross-terms between LimitingERs and the reciprocal of the LP sharpness are present in $\mathcal{L}$, as well as all four cross-terms between the relative distance to optima and the reciprocal of the LP sharpness. Last of all, we note that it follows from the definitions of LP sharpness and LimitingER that all terms in $\mathcal{L}$ are invariant under any positive scalar rescaling of the data $A$, $b$, or $c$.

All other constants that appear in (3.8) are naturally connected to the data of the problem, namely $\|c\|$, $\|q\|$, and $\kappa := \frac{\lambda_{\max}}{\lambda_{\min}}$, and it follows from Fact 2.1 that $\|c\|$ and $\|q\|$ can also be interpreted as the distances to $V_d$ and $V_p$.

**Remark 3.4.** *It is curious to note that instead of using the step-sizes given in* (3.7)*, the following choice of step-sizes:*

$$\tau = \frac{\mu_d\|q\|}{2\kappa\mu_p\|c\|} \quad \text{and} \quad \sigma = \frac{\mu_p\|c\|}{2\mu_d\|q\|\lambda_{\max}\lambda_{\min}} \tag{3.10}$$

*leads to a different bound on the total number of PDHG iterations $T$, namely:*

$$T \;\leq\; 9e \cdot \mathcal{L} \cdot \ln\left(16e \cdot \mathcal{L} \cdot \frac{\mathcal{E}_d(x^{0,0}, s^{0,0})}{\varepsilon} \cdot \left(1 + \kappa\frac{\mu_p\|c\|}{\mu_d\|q\|}\right)\left(1 + \frac{\mu_d\|q\|}{\mu_p\|c\|}\right)\right) + 1\;, \tag{3.11}$$

*using a structurally better value of the scalar $\mathcal{L}$, namely*

$$\mathcal{L} := 16\kappa\left(\frac{\theta_p^\star(\mathcal{X}^\star)}{\mu_p} + \frac{\theta_d^\star(\mathcal{S}^\star)}{\mu_d} + \frac{\mathrm{Dist}(c, \mathcal{S}^\star)}{\mu_p \cdot \mathrm{Dist}(0, V_d)} + \frac{\mathrm{Dist}(0, \mathcal{X}^\star)}{\mu_d \cdot \mathrm{Dist}(0, V_p)}\right)\;, \tag{3.12}$$

*which is potentially much smaller than* (3.9) *because there are fewer cross-terms involving $\mu_p$, $\mu_d$, $\theta_p^\star(\mathcal{X}^\star)$, and $\theta_d^\star(\mathcal{S}^\star)$. (This result will be proven later in this section.) However, the use of the step-sizes* (3.10) *requires knowledge of the LP sharpness constants $\mu_p$ and $\mu_d$ (or just their ratio), which are likely to be neither known nor easily computable.*

Towards the proof of Theorem 3.3 and Remark 3.4, we first prove the following lemma.

**Lemma 3.5.** *Suppose Algorithm 2 is run starting from $z^{0,0} = (x^{0,0}, y^{0,0}) = (0, 0)$ using the $\beta$-restart condition, and let the step-sizes $\sigma$ and $\tau$ satisfy* (2.7)*. Suppose also that there exists $\mathcal{L} \geq 1$ such that*

$$\mathrm{Dist}_M(z^{n,0}, \mathcal{Z}^\star) \;\leq\; \mathcal{L} \cdot \rho(\|z^{n,0} - z^{n-1,0}\|_M; z^{n,0}) \tag{3.13}$$

*for all $n \geq 1$. Let $T$ be the total number of PDHG iterations that are run in order to obtain $n$ for which $z^{n,0} = (x^{n,0}, s^{n,0})$ satisfies $\mathcal{E}_d(x^{n,0}, s^{n,0}) \leq \varepsilon$. Then*

$$T \;\leq\; \frac{9}{\beta\ln(1/\beta)} \cdot \mathcal{L} \cdot \ln\left(\tilde{c} \cdot \mathcal{L} \cdot \left(\frac{\mathcal{E}_d(x^{0,0}, s^{0,0})}{\varepsilon}\right)\right) + 1\;, \tag{3.14}$$

*where $\tilde{c} := \frac{8\sqrt{2}}{\beta\sqrt{1-\sqrt{\tau\sigma}\lambda_{\max}}}\left(\sqrt{\tau} + \sqrt{\sigma}\lambda_{\max}\right) \cdot \left(\frac{1}{\sqrt{\tau}} + \frac{1}{\sqrt{\sigma}\lambda_{\min}}\right)$.*

*Proof.* We first bound the number of inner iterations $k$ of PDHG between restarts in the outer loop. When $n = 0$ we have $k = 1$. For $n \geq 1$ we show that $k \leq 9\mathcal{L}/\beta$. To see this, note that it follows from Corollary 2.4 that

$$\rho(\|\bar{z}^{n,k} - z^{n,0}\|_M; \bar{z}^{n,k}) \leq \frac{8\,\mathrm{Dist}_M(z^{n,0}, \mathcal{Z}^\star)}{k} \ . \tag{3.15}$$

We may presume that $\rho(\|z^{n,0} - z^{n-1,0}\|_M; z^{n,0}) \neq 0$, for otherwise it follows from Lemma 2.1 that $z^{n,0} \in \mathcal{Z}^\star$, and Algorithm 2 would have terminated already in line **10**. Let us rewrite (3.15) as:

$$\frac{\rho(\|\bar{z}^{n,k} - z^{n,0}\|_M; \bar{z}^{n,k})}{\rho(\|z^{n,0} - z^{n-1,0}\|_M; z^{n,0})} \leq \frac{8}{k} \cdot \frac{\mathrm{Dist}_M(z^{n,0}, \mathcal{Z}^\star)}{\rho(\|z^{n,0} - z^{n-1,0}\|_M; z^{n,0})} \ . \tag{3.16}$$

It then follows from (3.16) and (3.13) that $k = \lceil 8\mathcal{L}/\beta \rceil$ suffices to ensure that condition (3.1) is satisfied. Since $\mathcal{L} \geq 1$ and $\beta \in (0, 1)$, such a $k$ is no larger than $9\mathcal{L}/\beta$.

Next we prove an upper bound on the number of outer iterations. When $n = 0$ PDHG restarts when $k = 1$, and it follows from Corollary 2.4 and inequality (2.19) that the initial normalized duality gap is upper bounded as follows:

$$\rho(\|\bar{z}^{0,1} - z^{0,0}\|_M; \bar{z}^{0,1}) \leq 8\,\mathrm{Dist}_M(z^{0,0}, \mathcal{Z}^\star) \leq 8 \left( \frac{\sqrt{2}}{\sqrt{\tau}} + \frac{\sqrt{2}}{\sqrt{\sigma}\lambda_{\min}} \right) \mathcal{E}_d(x^{0,0}, s^{0,0}) \ . \tag{3.17}$$

Now note from (2.20) that

$$\begin{aligned}
\mathrm{Dist}_M(z^{n,0}, \mathcal{Z}^\star) &\geq \gamma \cdot \max\left\{ \frac{\mathrm{Dist}(x^{n,0}, \mathcal{X}^\star)}{\sqrt{\tau}}, \frac{\mathrm{Dist}(s^{n,0}, \mathcal{S}^\star)}{\sqrt{\sigma}\lambda_{\max}} \right\} \\
&\geq \gamma \cdot \min\left\{ \frac{1}{\sqrt{\tau}}, \frac{1}{\sqrt{\sigma}\lambda_{\max}} \right\} \cdot \mathcal{E}_d(x^{n,0}, s^{n,0}) \ ,
\end{aligned}$$

where $\gamma := \sqrt{1 - \sqrt{\sigma\tau}\lambda_{\max}}$. Substituting this inequality back into (3.13) yields:

$$\mathcal{E}_d(x^{n,0}, s^{n,0}) \leq \frac{\mathcal{L}}{\gamma} \cdot \max\{\sqrt{\tau}, \sqrt{\sigma}\lambda_{\max}\} \cdot \rho(\|z^{n,0} - z^{n-1,0}\|_M; z^{n,0}) \ . \tag{3.18}$$

According to the restart condition, we have $\rho(\|z^{n,0} - z^{n-1,0}\|_M; z^{n,0}) \leq \beta \cdot \rho(\|z^{n-1,0} - z^{n-2,0}\|_M; z^{n-1,0})$ for each $n \geq 2$. And noting that $z^{1,0} = \bar{z}^{0,1}$, it follows that:

$$\begin{aligned}
\rho(\|z^{n,0} - z^{n-1,0}\|_M; z^{n,0}) \ &\leq \beta^{n-1} \cdot \rho(\|z^{1,0} - z^{0,0}\|_M; z^{1,0}) \\
&= \beta^{n-1} \cdot \rho(\|\bar{z}^{0,1} - z^{0,0}\|_M; \bar{z}^{0,1}) \leq 8\beta^{n-1} \left( \frac{\sqrt{2}}{\sqrt{\tau}} + \frac{\sqrt{2}}{\sqrt{\sigma}\lambda_{\min}} \right) \mathcal{E}_d(x^{0,0}, s^{0,0}) \ ,
\end{aligned} \tag{3.19}$$

where the second inequality uses (3.17). Combining (3.18) and (3.19) yields:

$$\begin{aligned}
\mathcal{E}_d(x^{n,0}, s^{n,0}) &\leq \frac{\mathcal{L}}{\gamma} \cdot \max\{\sqrt{\tau}, \sqrt{\sigma}\lambda_{\max}\} \cdot 8\beta^{n-1} \cdot \left( \frac{\sqrt{2}}{\sqrt{\tau}} + \frac{\sqrt{2}}{\sqrt{\sigma}\lambda_{\min}} \right) \mathcal{E}_d(x^{0,0}, s^{0,0}) \\
&\leq 8\beta^{n-1} \cdot \mathcal{L} \cdot \frac{\sqrt{2}}{\gamma} \cdot \left( \sqrt{\tau} + \sqrt{\sigma}\lambda_{\max} \right) \cdot \left( \frac{1}{\sqrt{\tau}} + \frac{1}{\sqrt{\sigma}\lambda_{\min}} \right) \cdot \mathcal{E}_d(x^{0,0}, s^{0,0}) \ .
\end{aligned} \tag{3.20}$$

Note that $\gamma = \sqrt{1-\sqrt{\sigma\tau}\lambda_{\max}}$, whereby (3.20) implies that for any $\varepsilon > 0$, $\mathcal{E}_d(x^{n,0}, s^{n,0}) \leq \varepsilon$ for all

$$n \geq \left\lceil \frac{\ln\left(8\mathcal{L} \cdot \frac{\sqrt{2}}{\sqrt{1-\sqrt{\sigma\tau}\lambda_{\max}}} \cdot (\sqrt{\tau} + \sqrt{\sigma}\lambda_{\max}) \cdot \left(\frac{1}{\sqrt{\tau}} + \frac{1}{\sqrt{\sigma}\lambda_{\min}}\right) \cdot \mathcal{E}_d(x^{0,0}, s^{0,0}) \cdot \varepsilon^{-1}\right)}{\ln(\beta^{-1})} \right\rceil + 1 \ , \quad (3.21)$$

which must be true when

$$n \geq \frac{\ln\left(8\mathcal{L} \cdot \frac{\sqrt{2}}{\sqrt{1-\sqrt{\sigma\tau}\lambda_{\max}}} \cdot (\sqrt{\tau} + \sqrt{\sigma}\lambda_{\max}) \cdot \left(\frac{1}{\sqrt{\tau}} + \frac{1}{\sqrt{\sigma}\lambda_{\min}}\right) \cdot \mathcal{E}_d(x^{0,0}, s^{0,0}) \cdot \varepsilon^{-1} \cdot \beta^{-1}\right)}{\ln(\beta^{-1})} + 1 \ . \quad (3.22)$$

The upper bound for the total number of iterations of PDHG now follows from (3.22) and noting that the inner loop at $n = 0$ uses $k = 1$ iteration of PDHG whereas for all $n \geq 1$ we have bounded the number of PDHG iterations by $k \leq 9\mathcal{L}/\beta$. □

**Remark 3.6.** *The dependency on $\beta$ in the bound on $T$ in Lemma 3.5 is mainly in the leading term $\frac{9\mathcal{L}}{\beta \ln(1/\beta)}$, and this term is minimized by setting $\beta = 1/e$ which results in the leading term being $9\mathcal{L}e$. Also, when $\beta = 1/e$ and the step-sizes are chosen so that $\sqrt{\sigma\tau} = 0.5/\lambda_{\max}$, then $\tilde{c} = 16e\left(1 + \frac{\sqrt{\tau}}{\sqrt{\sigma}\lambda_{\min}} + \frac{\sqrt{\sigma}\lambda_{\max}}{\sqrt{\tau}} + \kappa\right)$, which only involves the ratio $\tau/\sigma$ and the largest and the smallest singular values of $A$.*

Our next task is to show the existence of a value of $\mathcal{L}$ for which (3.13) holds. As part of this task we will need the following three very particular lemmas.

**Lemma 3.7.** *For any $\tilde{z} = (\tilde{x}, \tilde{y})$, let $z^\star \in \arg\min_{z \in \mathcal{Z}^\star} \|z - \tilde{z}\|_M$, and define for all $t \geq 0$*

$$\tilde{z}_t := \tilde{z} + t \cdot (\tilde{z} - z^\star) \quad (3.23)$$

*(whereby $\tilde{z}_0 = \tilde{z}$). Suppose that there exist nonnegative scalars $C_1$, $C_2$, $C_3$ such that the following inequality ($\mathcal{I}_t$) holds for $t = 0$:*

$$\mathrm{Dist}_M(\tilde{z}_t, \mathcal{Z}^\star) \leq C_1 \cdot \mathrm{Dist}(\tilde{x}_t, V_p) + C_2 \cdot \mathrm{Dist}(\tilde{s}_t, \mathbb{R}^n_+) + C_3 \cdot \max\{0, \mathrm{Gap}(\tilde{x}_t, \tilde{s}_t)\} \ . \quad (\mathcal{I}_t)$$

*Then inequality ($\mathcal{I}_t$) holds for all $t \geq 0$.*

*Proof.* Let $z^\star \in \arg\min_{z \in \mathcal{Z}^\star} \|z - \tilde{z}\|_M$ be given. Then because $\mathcal{Z}^\star$ is a convex set, it follows that $z^\star \in \arg\min_{z \in \mathcal{Z}^\star} \|z - \tilde{z}_t\|_M$ for all $t \geq 0$. Therefore $\mathrm{Dist}_M(\tilde{z}_t, \mathcal{Z}^\star) = \|z^\star - \tilde{z}_t\|_M = (1+t) \cdot \|\tilde{z} - z^\star\|_M = (1+t) \cdot \mathrm{Dist}_M(\tilde{z}, \mathcal{Z}^\star)$.

Regarding the terms on the right-hand side of ($\mathcal{I}_t$), for $t \geq 0$ we have $\mathrm{Dist}(\tilde{x}_t, V_p) = (1+t) \cdot \mathrm{Dist}(\tilde{x}, V_p)$ and $\max\{0, \mathrm{Gap}(\tilde{x}_t, \tilde{s}_t)\} = (1+t) \cdot \max\{0, \mathrm{Gap}(\tilde{x}, \tilde{s})\}$. It also holds that:

$$\mathrm{Dist}(\tilde{s}_t, \mathbb{R}^n_+) = \|(\tilde{s}_t)^-\| = \|(\tilde{s} + t \cdot (\tilde{s} - s^\star))^-\| \geq \|(1+t)(\tilde{s})^-\| = (1+t) \cdot \mathrm{Dist}(\tilde{s}, \mathbb{R}^n_+) \ ,$$

where the inequality above follows since $s^\star \geq 0$ and hence $(\tilde{s} + t \cdot (\tilde{s} - s^\star))^- \geq (1+t)(\tilde{s})^-$. Combining the above equalities and inequalities proves ($\mathcal{I}_t$) for all $t \geq 0$. □

**Lemma 3.8.** *Suppose that $\sigma$, $\tau$ satisfy (2.7). For any $x^0 \in \mathbb{R}^n_+$, $y^0 \in \mathbb{R}^m$ and $s^0 := c - A^\top y^0 \in \mathbb{R}^n$, then $z^0 := (x^0, y^0)$ satisfies:*

$$\mathrm{Dist}_M(z^0, \mathcal{Z}^\star) \leq C_1 \cdot \mathrm{Dist}(x^0, V_p) + C_2 \cdot \mathrm{Dist}(s^0, \mathbb{R}^n_+) + C_3 \cdot \max\{0, \mathrm{Gap}(x^0, s^0)\} \ , \quad (3.24)$$

*where*

$$
\begin{aligned}
C_1 &:= \left(\theta_p^\star(\mathcal{X}^\star)\|P_{\vec{V}_p}(c)\| + \|P_{\vec{V}_p^\perp}(c)\|\right)\cdot\left(\frac{3\sqrt{2}}{\sqrt{\tau}\mu_p\|P_{\vec{V}_p}(c)\|} + \frac{\sqrt{2}}{\sqrt{\sigma}\mu_d\|q\|\lambda_{\min}}\right)\\
C_2 &:= \theta_d^\star(\mathcal{S}^\star)\|q\|\cdot\left(\frac{2\sqrt{2}}{\sqrt{\sigma}\mu_d\|q\|\lambda_{\min}} + \frac{\sqrt{2}}{\sqrt{\tau}\mu_p\|P_{\vec{V}_p}(c)\|}\right)\\
C_3 &:= \left(\frac{\sqrt{2}}{\sqrt{\tau}\mu_p\|P_{\vec{V}_p}(c)\|} + \frac{\sqrt{2}}{\sqrt{\sigma}\mu_d\|q\|\lambda_{\min}}\right) \ .
\end{aligned}
\tag{3.25}
$$

*Proof.* We presume that $z^0 \notin \mathcal{Z}^\star$, for otherwise (3.24) follows trivially. Let $\tilde{x}^0$ be the projection of $x^0$ onto $V_p$, namely $\tilde{x}^0 = P_{V_p}(x^0)$. Then $\tilde{x}^0 - x^0$ is orthogonal to $\vec{V}_p$. Also let $\hat{x}$ and $\hat{s}$ be the projections of $\tilde{x}^0$ and $s^0$ onto $\mathcal{F}_p$ and $\mathcal{F}_d$, respectively, namely $\hat{x} := \arg\min_{x\in\mathcal{F}_p}\|x-\tilde{x}^0\|$ and $\hat{s} := \arg\min_{s\in\mathcal{F}_d}\|s-s^0\|$. Since $\hat{x}$ and $\hat{s}$ are feasible for $\mathcal{F}_p$ and $\mathcal{F}_d$, respectively, the duality gap $\mathrm{Gap}(\hat{x},\hat{s})$ is nonnegative. Furthermore, we have

$$
\begin{aligned}
\mathrm{Gap}(\hat{x},\hat{s}) &= c^\top\hat{x} - q^\top(c-\hat{s}) = \mathrm{Gap}(x^0,s^0) + c^\top(\hat{x}-x^0) + q^\top(\hat{s}-s^0)\\
&\le \mathrm{Gap}(x^0,s^0) + c^\top(\hat{x}-x^0) + \|q\|\cdot\mathrm{Dist}(s^0,\mathcal{F}_d) \ ,
\end{aligned}
\tag{3.26}
$$

and also

$$
\begin{aligned}
c^\top(\hat{x}-x^0) &= \left(P_{\vec{V}_p}(c) + P_{\vec{V}_p^\perp}(c)\right)^\top\left((\hat{x}-\tilde{x}^0) + (\tilde{x}^0-x^0)\right)\\
&= P_{\vec{V}_p}(c)^\top\left(\hat{x}-\tilde{x}^0\right) + P_{\vec{V}_p^\perp}(c)^\top\left(\tilde{x}^0-x^0\right)\\
&\le \|P_{\vec{V}_p}(c)\|\cdot\left\|\hat{x}-\tilde{x}^0\right\| + \|P_{\vec{V}_p^\perp}(c)\|\cdot\left\|\tilde{x}^0-x^0\right\|\\
&= \|P_{\vec{V}_p}(c)\|\cdot\mathrm{Dist}\left(\tilde{x}^0,\mathcal{F}_p\right) + \|P_{\vec{V}_p^\perp}(c)\|\cdot\mathrm{Dist}(x^0,V_p) \ ,
\end{aligned}
\tag{3.27}
$$

where the second equality above is due to $\hat{x}-\tilde{x}^0 \in \vec{V}_p$ (because both $\hat{x}$, $\tilde{x}^0 \in V_p$) and $\tilde{x}^0 - x^0 \in \vec{V}_p^\perp$. Substituting (3.27) into (3.26) then yields:

$$
\mathrm{Gap}(\hat{x},\hat{s}) \le \mathrm{Gap}(x^0,s^0) + \|P_{\vec{V}_p}(c)\|\cdot\mathrm{Dist}\left(\tilde{x}^0,\mathcal{F}_p\right) + \|P_{\vec{V}_p^\perp}(c)\|\cdot\mathrm{Dist}(x^0,V_p) + \|q\|\cdot\mathrm{Dist}(s^0,\mathcal{F}_d) \ .
\tag{3.28}
$$

Now we aim to replace the distance term involving $\tilde{x}^0$ in the right-hand side of (3.28) with a term involving $x^0$. From the definition of the ER $\theta_p(\cdot)$ we have:

$$
\begin{aligned}
\mathrm{Dist}(\tilde{x}^0,\mathcal{F}_p) &= \theta_p(\tilde{x}^0)\cdot\mathrm{Dist}(\tilde{x}^0,\mathbb{R}^n_+)\\
&\le \theta_p(\tilde{x}^0)\cdot\|\tilde{x}^0-x^0\| = \theta_p(\tilde{x}^0)\cdot\mathrm{Dist}(x^0,V_p) \ ,
\end{aligned}
\tag{3.29}
$$

where the inequality uses $x^0\in\mathbb{R}^n_+$. Note that $\mathrm{Dist}(x^0,\mathcal{F}_p) \le \mathrm{Dist}(\tilde{x}^0,\mathcal{F}_p) + \|x^0-\tilde{x}^0\| = \mathrm{Dist}(\tilde{x}^0,\mathcal{F}_p) + \mathrm{Dist}(x^0,V_p)$, so using (3.29) we obtain

$$
\mathrm{Dist}(x^0,\mathcal{F}_p) \le (\theta_p(\tilde{x}^0)+1)\cdot\mathrm{Dist}(x^0,V_p) \ .
\tag{3.30}
$$

Similarly, since $s^0\in V_d$, using the ER $\theta_d(\cdot)$ we have:

$$
\mathrm{Dist}(s^0,\mathcal{F}_d) \le \theta_d(s^0)\cdot\mathrm{Dist}(s^0,\mathbb{R}^n_+) \ .
\tag{3.31}
$$

Substituting (3.29) and (3.31) into (3.28) yields:

$$
\mathrm{Gap}(\hat{x},\hat{s}) \le \mathrm{Gap}(x^0,s^0) + \left(\|P_{\vec{V}_p}(c)\|\theta_p(\tilde{x}^0) + \|P_{\vec{V}_p^\perp}(c)\|\right)\cdot\mathrm{Dist}(x^0,V_p) + \|q\|\cdot\theta_d(s^0)\cdot\mathrm{Dist}(s^0,\mathbb{R}^n_+) \ .
\tag{3.32}
$$

Let us now use (3.32) to bound the distances to optima. Note that the duality gap $\mathrm{Gap}(\hat{x}, \hat{s})$ is an upper bound for both $c^\top \hat{x} - f^\star$ and $f^\star - q^\top(c - \hat{s})$. Then because $\hat{x} \in V_p$ and $\hat{s} \in V_d$ we have from Remark 3.2 that:

$$\begin{aligned} \mathrm{Dist}(\hat{x}, V_p \cap \{x : c^\top x = f^\star\}) &\le \frac{\mathrm{Gap}(\hat{x}, \hat{s})}{\|P_{\vec{V}_p}(c)\|} , \\ \mathrm{Dist}(\hat{s}, V_d \cap \{s : q^\top (c - s) = f^\star\}) &\le \frac{\mathrm{Gap}(\hat{x}, \hat{s})}{\|P_{\vec{V}_d}(q)\|} , \end{aligned} \tag{3.33}$$

and note that $\|P_{\vec{V}_d}(q)\| = \|q\|$ because $q \in \vec{V}_d$. Using the LP sharpness $\mu_p$ and $\mu_d$, we have:

$$\begin{aligned} \mathrm{Dist}(\hat{x}, \mathcal{X}^\star) &\le \frac{\mathrm{Dist}(\hat{x}, V_p \cap \{x : c^\top x = f^\star\})}{\mu_p} \le \frac{1}{\mu_p} \cdot \frac{\mathrm{Gap}(\hat{x}, \hat{s})}{\|P_{\vec{V}_p}(c)\|} , \\ \mathrm{Dist}(\hat{s}, \mathcal{S}^\star) &\le \frac{\mathrm{Dist}(\hat{s}, V_d \cap \{s : q^\top (c - s) = f^\star\})}{\mu_d} \le \frac{1}{\mu_d} \cdot \frac{\mathrm{Gap}(\hat{x}, \hat{s})}{\|q\|} . \end{aligned} \tag{3.34}$$

Now since $\mathrm{Dist}(x^0, \mathcal{X}^\star) \le \|x^0 - \hat{x}\| + \mathrm{Dist}(\hat{x}, \mathcal{X}^\star) = \mathrm{Dist}(x^0, \mathcal{F}_p) + \mathrm{Dist}(\hat{x}, \mathcal{X}^\star)$, using (3.30) and (3.34) implies that:

$$\mathrm{Dist}(x^0, \mathcal{X}^\star) \le (\theta_p(\tilde{x}^0) + 1) \cdot \mathrm{Dist}(x^0, V_p) + \frac{1}{\mu_p} \cdot \frac{\mathrm{Gap}(\hat{x}, \hat{s})}{\|P_{\vec{V}_p}(c)\|} . \tag{3.35}$$

Combining (3.32) and (3.35) we obtain:

$$\begin{aligned} \mathrm{Dist}(x^0, \mathcal{X}^\star) &\le (\theta_p(\tilde{x}^0) + 1) \cdot \mathrm{Dist}(x^0, V_p) \\ &\quad + \frac{\mathrm{Gap}(x^0, s^0) + \left( \|P_{\vec{V}_p}(c)\| \theta_p(\tilde{x}^0) + \|P_{\vec{V}_p^\perp}(c)\| \right) \cdot \mathrm{Dist}(x^0, V_p) + \|q\| \cdot \theta_d(s^0) \cdot \mathrm{Dist}(s^0, \mathbb{R}^n_+)}{\mu_p \|P_{\vec{V}_p}(c)\|} \\ &= \frac{\mathrm{Gap}(x^0, s^0)}{\mu_p \|P_{\vec{V}_p}(c)\|} + \left( \frac{\theta_p(\tilde{x}^0)}{\mu_p} + \frac{\|P_{\vec{V}_p^\perp}(c)\|}{\mu_p \|P_{\vec{V}_p}(c)\|} + \theta_p(\tilde{x}^0) + 1 \right) \cdot \mathrm{Dist}(x^0, V_p) + \frac{\|q\| \theta_d(s^0)}{\mu_p \|P_{\vec{V}_p}(c)\|} \cdot \mathrm{Dist}(s^0, \mathbb{R}^n_+) . \end{aligned} \tag{3.36}$$

Note that because $\mu_p \le 1$ and $\theta_p(\tilde{x}^0) \ge 1$, it follows that (3.36) can be relaxed to:

$$\mathrm{Dist}(x^0, \mathcal{X}^\star) \le \frac{\mathrm{Gap}(x^0, s^0)}{\mu_p \|P_{\vec{V}_p}(c)\|} + \left( \frac{3\theta_p(\tilde{x}^0)}{\mu_p} + \frac{\|P_{\vec{V}_p^\perp}(c)\|}{\mu_p \|P_{\vec{V}_p}(c)\|} \right) \cdot \mathrm{Dist}(x^0, V_p) + \frac{\|q\| \theta_d(s^0)}{\mu_p \|P_{\vec{V}_p}(c)\|} \cdot \mathrm{Dist}(s^0, \mathbb{R}^n_+) . \tag{3.37}$$

Using almost identical logic applied to $s^0$ instead of $x^0$, we obtain:

$$\mathrm{Dist}(s^0, \mathcal{S}^\star) \le \frac{\mathrm{Gap}(x^0, s^0)}{\mu_d \|q\|} + \frac{2\theta_d(s^0)}{\mu_d} \cdot \mathrm{Dist}(s^0, \mathbb{R}^n_+) + \frac{\|P_{\vec{V}_p}(c)\| \theta_p(\tilde{x}^0) + \|P_{\vec{V}_p^\perp}(c)\|}{\mu_d \|q\|} \cdot \mathrm{Dist}(x^0, V_p) . \tag{3.38}$$

Combining (3.37) with (3.38) and using the right-most inequality of (2.19), it follows that

$$\mathrm{Dist}_M(z^0, \mathcal{Z}^\star) \le \bar{C}_1(z^0) \cdot \mathrm{Dist}(x^0, V_p) + \bar{C}_2(z^0) \cdot \mathrm{Dist}(s^0, \mathbb{R}^n_+) + \bar{C}_3(z^0) \cdot \max\{0, \mathrm{Gap}(x^0, s^0)\} , \tag{3.39}$$

where

$$\begin{aligned}
\bar{C}_1(z^0) &:= \Big(\theta_p(P_{V_p}(x^0))\|P_{\vec{V}_p}(c)\| + \|P_{\vec{V}_p^\perp}(c)\|\Big)\cdot\left(\tfrac{3\sqrt{2}}{\sqrt{\tau}\mu_p\|P_{\vec{V}_p}(c)\|} + \tfrac{\sqrt{2}}{\sqrt{\sigma}\mu_d\|q\|\lambda_{\min}}\right) \\
\bar{C}_2(z^0) &:= \theta_d(s^0)\|q\|\cdot\left(\tfrac{2\sqrt{2}}{\sqrt{\sigma}\mu_d\|q\|\lambda_{\min}} + \tfrac{\sqrt{2}}{\sqrt{\tau}\mu_p\|P_{\vec{V}_p}(c)\|}\right) \\
\bar{C}_3(z^0) &:= C_3 \ ,
\end{aligned} \tag{3.40}$$

and notice in the definition $\bar{C}_1$ we have written $\theta_p(P_{V_p}(x^0))$ since in fact $\tilde{x}^0 := P_{V_p}(x^0)$. Now notice that (3.39) is nearly identical to (3.24), except that the constants $\bar{C}_1(z^0)$ and $\bar{C}_2(z^0)$ use $\theta_p(P_{V_p}(x^0))$ instead of $\theta_p^\star(\mathcal{X}^\star)$, and use $\theta_d(s^0)$ instead of $\theta_d^\star(\mathcal{S}^\star)$.

To finish the proof, let $z^\star \in \arg\min_{z\in\mathcal{Z}^\star}\|z - z^0\|_M$ be fixed, and define $z^\lambda := (1-\lambda)z^0 + \lambda z^\star$ for all $\lambda \in [0,1)$. Then (3.39) holds for $z^\lambda$, namely:

$$\mathrm{Dist}_M(z^\lambda,\mathcal{Z}^\star) \le \bar{C}_1(z^\lambda)\cdot\mathrm{Dist}(x^\lambda,V_p)+\bar{C}_2(z^\lambda)\cdot\mathrm{Dist}(s^\lambda,\mathbb{R}^n_+)+\bar{C}_3(z^\lambda)\cdot\max\{0,\mathrm{Gap}(x^\lambda,s^\lambda)\}\ , \tag{3.41}$$

since $z^\lambda$ satisfies the same hypotheses as $z^0$. And since $z^\star \in \arg\min_{z\in\mathcal{Z}^\star}\|z - z^\lambda\|_M$ we can invoke Lemma 3.7. It follows from Lemma 3.7 that for all $t \ge 0$ with $z_t^\lambda := z^\lambda + t(z^\lambda - z^\star)$ that

$$\mathrm{Dist}_M(z_t^\lambda,\mathcal{Z}^\star) \le \bar{C}_1(z^\lambda)\cdot\mathrm{Dist}(x_t^\lambda,V_p)+\bar{C}_2(z^\lambda)\cdot\mathrm{Dist}(s_t^\lambda,\mathbb{R}^n_+)+\bar{C}_3(z^\lambda)\cdot\max\{0,\mathrm{Gap}(x_t^\lambda,s_t^\lambda)\}\ . \tag{3.42}$$

Setting $t = \lambda/(1-\lambda)$ yields $z_t^\lambda = z^0$, whereby:

$$\mathrm{Dist}_M(z^0,\mathcal{Z}^\star) \le \bar{C}_1(z^\lambda)\cdot\mathrm{Dist}(x^0,V_p)+\bar{C}_2(z^\lambda)\cdot\mathrm{Dist}(s^0,\mathbb{R}^n_+)+\bar{C}_3(z^\lambda)\cdot\max\{0,\mathrm{Gap}(x^0,s^0)\}\ . \tag{3.43}$$

Now let $\lambda \to 1$, whereby $z^\lambda \to z^\star$, and so $\limsup_{\lambda\to 1}\theta_p(P_{V_p}(x^\lambda)) \le \theta_p^\star(\mathcal{X}^\star)$ and therefore $\limsup_{\lambda\to 1}\bar{C}_1(z^\lambda) \le C_1$. Similarly $\limsup_{\lambda\to 1}\theta_d(s^\lambda)) \le \theta_d^\star(\mathcal{S}^\star)$ and therefore $\limsup_{\lambda\to 1}\bar{C}_2(z^\lambda) \le C_2$. We therefore have from (3.43) that

$$\mathrm{Dist}_M(z^0,\mathcal{Z}^\star) \le C_1\cdot\mathrm{Dist}(x^0,V_p) + C_2\cdot\mathrm{Dist}(s^0,\mathbb{R}^n_+) + C_3\cdot\max\{0,\mathrm{Gap}(x^0,s^0)\}\ . \tag{3.44}$$

□

**Lemma 3.9.** *Suppose that the initial iterate of PDHG is $z^0 := (0,0)$, and let $z^b,\ z^c \in \mathbb{R}^n_+\times\mathbb{R}^m$ satisfy $z^b \ne z^c$ and the nonexpansive inequalities $\|z^b - z^\star\|_M \le \|z^0 - z^\star\|_M$ and $\|z^c - z^\star\|_M \le \|z^0 - z^\star\|_M$ for all $z^\star \in \mathcal{Z}^\star$. Then it holds that:*

$$\begin{aligned}
&\mathrm{Dist}_M(z^b,\mathcal{Z}^\star) \le \\
&\rho(\|z^b - z^c\|_M; z^b)\cdot\left(\begin{array}{l}
\left(\frac{3\sqrt{2}}{\sqrt{\sigma\tau}\lambda_{\min}\mu_p} + \frac{\sqrt{2}}{\sigma\lambda_{\min}^2\mu_d}\cdot\frac{\|P_{\vec{V}_p}(c)\|}{\|q\|}\right)\cdot\left(\theta_p^\star(\mathcal{X}^\star) + \frac{\|P_{\vec{V}_p^\perp}(c)\|}{\|P_{\vec{V}_p}(c)\|}\right) \\
+\left(\frac{2\sqrt{2}}{\sqrt{\sigma\tau}\lambda_{\min}\mu_d} + \frac{\sqrt{2}}{\tau\mu_p}\cdot\frac{\|q\|}{\|P_{\vec{V}_p}(c)\|}\right)\cdot\theta_d^\star(\mathcal{S}^\star) \\
+\left(\frac{4}{\sqrt{\tau}\mu_p\|P_{\vec{V}_p}(c)\|} + \frac{4}{\sqrt{\sigma}\mu_d\|q\|\lambda_{\min}}\right)\cdot\left(\frac{1}{\sqrt{\tau}}\mathrm{Dist}(0,\mathcal{X}^\star) + \frac{1}{\lambda_{\min}\sqrt{\sigma}}\mathrm{Dist}(c,\mathcal{S}^\star)\right)
\end{array}\right) .
\end{aligned} \tag{3.45}$$

*Proof.* The proof is a combination of Lemmas 2.1, 2.7, and 3.8. Setting $s^b = c - A^\top y^b$ it follows from Lemma 2.1 that

$$\begin{aligned}
&\mathrm{Dist}(x^b,V_p) \le \tfrac{1}{\sqrt{\sigma}\lambda_{\min}}\cdot\rho(\|z^b - z^c\|_M; z^b)\ , \\
&\mathrm{Dist}(s^b,\mathbb{R}^n_+) \le \tfrac{1}{\sqrt{\tau}}\cdot\rho(\|z^b - z^c\|_M; z^b)\ , \\
&\mathrm{Gap}(x^b,s^b) \le \max\{\|z^b - z^c\|_M, \|z^b\|_M\}\rho(\|z^b - z^c\|_M; z^b)\ .
\end{aligned} \tag{3.46}$$

Also, from Lemma 3.8 it follows that $\mathrm{Dist}_M(z^b, \mathcal{Z}^\star)$ can be bounded using the terms in the left-hand side of (3.46):

$$\mathrm{Dist}_M(z^b, \mathcal{Z}^\star) \le C_1 \cdot \mathrm{Dist}(x^b, V_p) + C_2 \cdot \mathrm{Dist}(s^b, \mathbb{R}^n_+) + C_3 \cdot \mathrm{Gap}(x^b, s^b) \ , \tag{3.47}$$

where $C_1$, $C_2$ and $C_3$ are the scalars defined in (3.25). Substituting (3.46) into (3.47) yields:

$$\mathrm{Dist}_M(z^b, \mathcal{Z}^\star) \le \left( \frac{C_1}{\sqrt{\sigma}\lambda_{\min}} + \frac{C_2}{\sqrt{\tau}} + C_3 \cdot \max\{\|z^b - z^c\|_M, \|z^b\|_M\} \right) \rho(\|z^b - z^c\|_M; z^b) \ . \tag{3.48}$$

From Lemma 2.7 it holds that

$$\begin{aligned} \max\{\|z^b - z^c\|_M, \|z^b\|_M\} &\le 2\,\mathrm{Dist}_M(z^0, \mathcal{Z}^\star) + \|z^0\|_M = 2\,\mathrm{Dist}_M(0, \mathcal{Z}^\star) \\ &\le \frac{2\sqrt{2}}{\sqrt{\tau}} \mathrm{Dist}(0, \mathcal{X}^\star) + \frac{2\sqrt{2}}{\sqrt{\sigma}\lambda_{\min}} \mathrm{Dist}(c, \mathcal{S}^\star) \ , \end{aligned}$$

where the second inequality uses (2.19). Substituting this inequality into (3.48) yields

$$\mathrm{Dist}_M(z^b, \mathcal{Z}^\star) \le \left( \frac{C_1}{\sqrt{\sigma}\lambda_{\min}} + \frac{C_2}{\sqrt{\tau}} + \frac{2\sqrt{2}C_3}{\sqrt{\tau}} \mathrm{Dist}(0, \mathcal{X}^\star) + \frac{2\sqrt{2}C_3}{\lambda_{\min}\sqrt{\sigma}} \mathrm{Dist}(c, \mathcal{S}^\star) \right) \rho(\|z^b - z^c\|_M; z^b) \ . \tag{3.49}$$

The proof is completed by substituting the values of $C_1, C_2, C_3$ defined in (3.25) into (3.49), which then yields (3.45). □

We now prove Theorem 3.3 and Remark 3.4.

*Proof of Theorem 3.3.* For any $n \ge 1$ it holds that $z^{n,0} = \bar{z}^{n-1,K} = \frac{1}{K}\sum_{i=1}^K z^{n-1,i}$ where $K$ is the total number of inner iterations of PDHG run in the outer loop iteration $n-1$, whose initial iterate value is $z^{n-1,0}$. It then follows from the nonexpansive properties of PDHG from Lemma 2.3 that $\|z^{n,0} - z^\star\|_M \le \|z^{n-1,0} - z^\star\|_M$ for each $z^\star \in \mathcal{Z}^\star$. By telescoping inequalities we then have

$$\|z^{n,0} - z^\star\|_M \le \|z^{n-1,0} - z^\star\|_M \le \cdots \le \|z^{0,0} - z^\star\|_M$$

for each $z^\star \in \mathcal{Z}^\star$. Now notice that $z^0 := z^{0,0}$, $z^b := z^{n,0}$, and $z^c := z^{n-1,0}$ satisfy the hypotheses of Lemma 3.9, whereby it holds that

$$\begin{aligned} &\mathrm{Dist}_M(z^{n,0}, \mathcal{Z}^\star) \le \\ &\rho(\|z^{n,0} - z^{n-1,0}\|_M; z^{n,0}) \cdot \left( \begin{array}{l} \left( \frac{3\sqrt{2}}{\sqrt{\sigma\tau}\lambda_{\min}\mu_p} + \frac{\sqrt{2}}{\sigma\lambda_{\min}^2\mu_d} \cdot \frac{\|P_{\vec{V}_p}(c)\|}{\|q\|} \right) \cdot \left( \theta_p^\star(\mathcal{X}^\star) + \frac{\|P_{\vec{V}_p^\perp}(c)\|}{\|P_{\vec{V}_p}(c)\|} \right) \\ + \left( \frac{2\sqrt{2}}{\sqrt{\sigma\tau}\lambda_{\min}\mu_d} + \frac{\sqrt{2}}{\tau\mu_p} \cdot \frac{\|q\|}{\|P_{\vec{V}_p}(c)\|} \right) \cdot \theta_d^\star(\mathcal{S}^\star) \\ + \left( \frac{4}{\sqrt{\tau}\mu_p\|P_{\vec{V}_p}(c)\|} + \frac{4}{\sqrt{\sigma}\mu_d\|q\|\lambda_{\min}} \right) \cdot \left( \frac{1}{\sqrt{\tau}} \mathrm{Dist}(0, \mathcal{X}^\star) + \frac{1}{\lambda_{\min}\sqrt{\sigma}} \mathrm{Dist}(c, \mathcal{S}^\star) \right) \end{array} \right) . \end{aligned} \tag{3.50}$$

From the supposition of Theorem 3.3 we have $c \in \vec{V}_p$ and therefore $\|P_{\vec{V}_p}(c)\| = \|c\|$, $\|P_{\vec{V}_p^\perp}(c)\| = 0$, and $\|c\| = \mathrm{Dist}(0, V_d)$. Also by construction we have $\|q\| = \mathrm{Dist}(0, V_p)$. Substituting in the step-sizes (3.7) and using the above norm equalities in (3.50) yields:

$$\begin{aligned} &\mathrm{Dist}_M(z^{n,0}, \mathcal{Z}^\star) \le \rho(\|z^{n,0} - z^{n-1,0}\|_M; z^{n,0}) \cdot \\ &6\sqrt{2}\kappa \left( \frac{1}{\mu_p} + \frac{\frac{1}{3}}{\mu_d} \right) \theta_p^\star(\mathcal{X}^\star) + 4\sqrt{2}\kappa \left( \frac{\frac{1}{2}}{\mu_p} + \frac{1}{\mu_d} \right) \theta_d^\star(\mathcal{S}^\star) + 8\kappa \left( \frac{1}{\mu_p} + \frac{1}{\mu_d} \right) \left[ \frac{\mathrm{Dist}(0, \mathcal{X}^\star)}{\mathrm{Dist}(0, V_p)} + \frac{\mathrm{Dist}(c, \mathcal{S}^\star)}{\mathrm{Dist}(0, V_d)} \right] . \end{aligned} \tag{3.51}$$

Now notice that the value of $\mathcal{L}$ specified in (3.9) is at least as large as the second line in (3.51), whereby it holds that

$$\mathrm{Dist}_M(z^{n,0}, \mathcal{Z}^\star) \leq \mathcal{L} \cdot \rho(\|z^{n,0} - z^{n-1,0}\|_M; z^{n,0})$$

for the value of $\mathcal{L}$ specified in (3.9). Therefore condition (3.13) of Lemma 3.5 is satisfied, and it follows from Lemma 3.5 that $T$ satisfies (3.14) with the value of $\tilde{c}$ specified in the statement of the lemma, namely:

$$T \ \leq \ \frac{9}{\beta \ln(1/\beta)} \cdot \mathcal{L} \cdot \ln\left(\tilde{c} \cdot \mathcal{L} \cdot \left(\frac{\mathcal{E}_d(x^{0,0}, s^{0,0})}{\varepsilon}\right)\right) + 1 \ . \tag{3.52}$$

Substituting in the step-sizes (3.7) and $\beta = 1/e$ into the value of $\tilde{c}$ in Lemma 3.5 we find that:

$$\tilde{c} = 16e \cdot \left(1 + \kappa\frac{\|c\|}{\|q\|}\right)\left(1 + \frac{\|q\|}{\|c\|}\right) \ ,$$

and using $\beta = 1/e$ and the value of $\tilde{c}$ above in (3.52) we finally arrive at:

$$T \ \leq \ 9e \cdot \mathcal{L} \cdot \ln\left(16e \cdot \mathcal{L} \cdot \left(\frac{\mathcal{E}_d(x^{0,0}, s^{0,0})}{\varepsilon}\right)\left(1 + \kappa\frac{\|c\|}{\|q\|}\right)\left(1 + \frac{\|q\|}{\|c\|}\right)\right) + 1 \ , \tag{3.53}$$

which completes the proof of the theorem. □

We now prove Remark 3.4 using essentially the identical logic as above.

*Proof of Remark 3.4.* Substituting in the step-sizes (3.10) and using the norm equalities $\|P_{\vec{V}_p}(c)\| = \|c\|$, $\|P_{\vec{V}_p^\perp}(c)\| = 0$, $\|q\| = \mathrm{Dist}(0, V_p)$, and $\|c\| = \mathrm{Dist}(0, V_d)$ into (3.50) yields:

$$\begin{aligned} &\mathrm{Dist}_M(z^{n,0}, \mathcal{Z}^\star) \\ &\quad\leq \ \rho(\|z^{n,0} - z^{n-1,0}\|_M; z^{n,0}) \cdot \left[2\sqrt{2}\kappa\left(\frac{4\theta_p^\star(\mathcal{X}^\star) + 4\sqrt{2}\cdot\frac{\mathrm{Dist}(c,\mathcal{S}^\star)}{\mathrm{Dist}(0,V_d)}}{\mu_p} + \frac{3\theta_d^\star(\mathcal{S}^\star) + 4\sqrt{2}\cdot\frac{\mathrm{Dist}(0,\mathcal{X}^\star)}{\mathrm{Dist}(0,V_p)}}{\mu_d}\right)\right] \ . \end{aligned} \tag{3.54}$$

Now notice that the value of $\mathcal{L}$ specified in (3.12) is at least as large as the square-bracketed term in (3.54), whereby it holds that

$$\mathrm{Dist}_M(z^{n,0}, \mathcal{Z}^\star) \leq \mathcal{L} \cdot \rho(\|z^{n,0} - z^{n-1,0}\|_M; z^{n,0})$$

for the value of $\mathcal{L}$ specified in (3.12). Therefore condition (3.13) of Lemma 3.5 is satisfied, and it follows from Lemma 3.5 that $T$ satisfies (3.14) with the value of $\tilde{c}$ specified in the statement of the lemma, namely:

$$T \ \leq \ \frac{9}{\beta \ln(1/\beta)} \cdot \mathcal{L} \cdot \ln\left(\tilde{c} \cdot \mathcal{L} \cdot \left(\frac{\mathcal{E}_d(x^{0,0}, s^{0,0})}{\varepsilon}\right)\right) + 1 \ . \tag{3.55}$$

Substituting in the step-sizes (3.10) and $\beta = 1/e$ into the value of $\tilde{c}$ in Lemma 3.5 we find that:

$$\tilde{c} = 16e \cdot \left(1 + \kappa\frac{\mu_p\|c\|}{\mu_d\|q\|}\right)\left(1 + \frac{\mu_d\|q\|}{\mu_p\|c\|}\right) \ ,$$

and using $\beta = 1/e$ we finally arrive at:

$$T \ \leq \ 9e \cdot \mathcal{L} \cdot \ln\left(16e \cdot \mathcal{L} \cdot \left(\frac{\mathcal{E}_d(x^{0,0}, s^{0,0})}{\varepsilon}\right) \cdot \left(1 + \kappa\frac{\mu_p\|c\|}{\mu_d\|q\|}\right)\left(1 + \frac{\mu_d\|q\|}{\mu_p\|c\|}\right)\right) + 1 \ , \tag{3.56}$$

which completes the proof of the remark. □

**Remark 3.10.** *Both Theorem 3.3 and Remark 3.4 assume that* $c \in \vec{V}_p$*, or equivalently that* $\|P_{\vec{V}_p^\perp}(c)\| = 0$*, for simplicity and ease of exposition. Unpacking the proofs above, we see that* $\|P_{\vec{V}_p^\perp}(c)\|$ *only affects the proof in the bound* (3.50)*. Suppose that* $\|P_{\vec{V}_p^\perp}(c)\| \neq 0$ *but is not too large, say,* $\|P_{\vec{V}_p^\perp}(c)\| \leq \theta_p^\star(\mathcal{X}^\star)\|P_{\vec{V}_p}(c)\|$*. Then the right-hand side of* (3.50)*, and hence the value of* $\mathcal{L}$ *in Lemma 3.5, will only increase by a constant factor of at most* 2*.*

# 4 Properties of the limiting error ratio $\theta^\star(\mathcal{U}^\star)$ (LimitingER)

In this section we present some relevant properties of the limiting error ratio (LimitingER) $\theta^\star(\mathcal{U}^\star)$. In Theorem 4.1 we characterize an upper bound on $\theta^\star(\mathcal{U}^\star)$ that is connected to the notion of a "nicely interior" point in a convex set – which itself is critical to the complexity of separation-oracle methods [14]. We also present in Proposition 4.2 a convex optimization problem (actually a conic optimization problem with one second-order cone constraint) whose solution provides an upper bound on $\theta^\star(\mathcal{U}^\star)$, thus showing that computing a bound on $\theta^\star(\mathcal{U}^\star)$ is computationally tractable. Last of all, in Theorem 4.4 we show that the error ratio $\theta(u)$ is upper-bounded by a simple quantity involving the data-perturbation condition number DistInfeas$(\cdot)$ of Renegar [40].

Our setup once again is the generic LP problem (3.2) where $V$ is an affine subspace. We will assume in this subsection that $\mathcal{U}^\star$ is nonempty, and recall the definition of the limiting error ratio (LimitingER) $\theta^\star(\mathcal{U}^\star)$ in (3.4). Let $\mathcal{F}_{++}$ denote the strictly feasible solutions of (3.2), namely $\mathcal{F}_{++} := V \cap \mathbb{R}^n_{++}$. We do not necessarily assume that $\mathcal{F}_{++} \neq \emptyset$.

## 4.1 Upper bound on $\theta^\star(\mathcal{U}^\star)$ based on "nicely interior" feasible solutions

The following theorem presents an upper bound on the limiting error ratio $\theta^\star(\mathcal{U}^\star)$ using the existence of a "nicely interior" point in $\mathcal{F}_{++}$. (In the theorem we use the convention that the infimum over an empty set is $+\infty$.)

**Theorem 4.1.** *For the general LP problem* (3.2)*, suppose that the optimal solution set* $\mathcal{U}^\star$ *is nonempty and bounded. Then*

$$\theta^\star(\mathcal{U}^\star) \leq \sup_{u^\star \in \mathcal{U}^\star} \inf_{u_{\text{int}} \in \mathcal{F}_{++}} \frac{\|u^\star - u_{\text{int}}\|}{\min_i (u_{\text{int}})_i} \ . \tag{4.1}$$

This theorem states that if every optimal solution $u^\star$ has a nicely interior point near to it – in the sense that there exists $u_{\text{int}} \in \mathcal{F}_{++}$ that is simultaneously close to $u^\star$ and far from the boundary of the nonnegative orthant $\mathbb{R}^n_+$, then the LimitingER value $\theta^\star(\mathcal{U}^\star)$ will not be excessively large. In the case when $\mathcal{U}^\star$ is a singleton, then (4.1) simplifies to finding a single nicely interior point that balances the distance from the optimal solution (in the numerator above) with the distance to the boundary of $\mathbb{R}^n_+$ (in the denominator above). We will prove Theorem 4.1 at the end of this section. (Note that the concept of a nicely interior point is quite similar to that of a "reliable solution" in [11], see also [13] for connections to Renegar's data-perturbation condition number DistInfeas$(\cdot)$ [40].)

## 4.2 A computable upper bound for the limiting error ratio $\theta^\star(\mathcal{U}^\star)$

Here we show how, in principle, we can use the upper bound in Theorem 4.1 to construct a computationally tractable convex optimization problem that computes an upper bound on the LimitingER $\theta^\star(\mathcal{U}^\star)$.

We first suppose that we can compute, without too much extra computational effort, a ball that contains the optimal solution set $\mathcal{U}^\star$. That is, we suppose we can compute a point $u_a \in \mathcal{U}^\star$ and a radius value $R_a$ such that $\mathcal{U}^\star \subset B(u_a, R_a)$, so that every optimal solution is within a distance $R_a$ from the optimal solution $u_a$. If $\mathcal{U}^\star$ is a singleton, then $R_a = 0$ trivially. If $\mathcal{U}^\star$ is not a singleton, then one choice of $u_a$ is the analytic center of $\mathcal{U}^\star$ (see Sonnevend [42], also [33]), from which one can then easily construct a bounding ellipsoid $\mathcal{E}^{\text{out}}$ that contains $\mathcal{U}^\star$ and then examine the eigenstructure $\mathcal{E}^{\text{out}}$ to compute a suitable value of $R_a$.

**Proposition 4.2.** *Suppose $u_a \in \mathcal{U}^\star$ and there exists $R_a$ for which $\mathcal{U}^\star \subset \{u : \|u - u_a\| \le R_a\}$, then it holds that $\theta^\star(\mathcal{U}^\star) \le B^\star$ for $B^\star$ defined as follows:*

$$B^\star := \inf_{r>0,\ u\in\mathbb{R}^n} \frac{R_a + \|u - u_a\|}{r} \quad \text{s.t. } u \in V,\ u \ge r\cdot e \ . \tag{4.2}$$

*Furthermore, if $V$ is conveyed as $V = \{u \in \mathbb{R}^n : Ku = h\}$, then*

$$B^\star := \min_{v\in\mathbb{R}^n,\ \alpha\in\mathbb{R}} R_a\alpha + \|v - u_a\alpha\| \quad \text{s.t. } Kv = h\alpha,\ v \ge e,\ \alpha \ge 0 \ . \tag{4.3}$$

*Proof.* From Theorem 4.1 we have:

$$\begin{aligned}
\theta^\star(\mathcal{U}^\star) &\le \sup_{u^\star\in\mathcal{U}^\star} \inf_{u_{\text{int}}\in\mathcal{F}_{++}} \frac{\|u^\star - u_{\text{int}}\|}{\min_i (u_{\text{int}})_i} \le \sup_{u\in B(u_a,R_a)} \inf_{u_{\text{int}}\in\mathcal{F}_{++}} \frac{\|u - u_{\text{int}}\|}{\min_i (u_{\text{int}})_i} \\
&\le \sup_{u\in B(u_a,R_a)} \inf_{u_{\text{int}}\in\mathcal{F}_{++}} \frac{\|u - u_a\| + \|u_a - u_{\text{int}}\|}{\min_i (u_{\text{int}})_i} \le \inf_{u_{\text{int}}\in\mathcal{F}_{++}} \frac{R_a + \|u_a - u_{\text{int}}\|}{\min_i (u_{\text{int}})_i} \\
&= \inf_{r>0,\ u\in\mathbb{R}^n} \frac{R_a + \|u - u_a\|}{r} \quad \text{s.t. } u \in V,\ u \ge r\cdot e \ ,
\end{aligned} \tag{4.4}$$

and notice that the final right-hand side is precisely $B^\star$, which proves (4.2). Next notice that (4.2) and (4.3) are equivalent via the elementary projective transformations $u = v/\alpha$ and $(v, \alpha) = (u/r, 1/r)$ if we add the additional constraint $\alpha > 0$ to (4.3). However, since we are only interested in the optimal objective value of (4.2) and (4.3), solving the (4.3) yields the same optimal objective value as (4.2). ∎

Note that the formulation of the upper bound in (4.3) is a convex optimization problem of essentially the same size as that of the original LP problem (3.2), and its only non-linear component is the norm term $\|v - u_a\alpha\|$ in the objective function. This can easily be handled by a single second-order cone constraint, or can be replaced with an $\ell_1$ or $\ell_\infty$ norm which then can be converted to a pure LP problem.

Before proving Theorem 4.1, we first introduce a more general result about $\theta(u)$. Suppose $u \in V \setminus \mathcal{F}$ and $u_{\text{int}} \in \mathcal{F}_{++}$, then the line segment from $u$ to $u_{\text{int}}$ will contain a unique point that lies on the boundary of $\mathcal{F}$, and let us denote this point by $\mathcal{F}(u; u_{\text{int}})$. More formally we have

$$\mathcal{F}(u; u_{\text{int}}) := \arg\min_{\tilde u} \{\|u - \tilde u\| : \tilde u \in \mathcal{F} \ ,\ \tilde u := \lambda u_{\text{int}} + (1-\lambda)u \text{ for some } \lambda \in \mathbb{R}\} \ . \tag{4.5}$$

The following lemma will be used in our proof of Theorem 4.1.

**Lemma 4.3.** *For the general LP presentation (3.2), suppose that Assumption 1 holds. Then for $u \in V \setminus \mathcal{F}$ and $u_{\text{int}} \in \mathcal{F}_{++}$, it holds that*

$$\theta(u) \le \frac{\|\mathcal{F}(u; u_{\text{int}}) - u_{\text{int}}\|}{\min_i (u_{\text{int}})_i} \le \frac{\|u - u_{\text{int}}\|}{\min_i (u_{\text{int}})_i}, \tag{4.6}$$

*where $\mathcal{F}(u; u_{\text{int}})$ is given by (4.5).*

*Proof.* Suppose that $u_{\text{int}}$ is any given strictly feasible point in $\mathcal{F}_{++}$. Let $r := \min_i (u_{\text{int}})_i$, and so $r > 0$. In the line segment connecting $u_{\text{int}}$ and $u$, let $v := \mathcal{F}(u; u_{\text{int}})$ defined in (4.5). Then because $r > 0$ and $u \in V \setminus \mathcal{F}$, there exists $\lambda \in (0, 1)$ for which $v = \lambda u_{\text{int}} + (1 - \lambda)u$. Also we have $v \in \partial \mathbb{R}^n_+$ and there exists $i \in [n]$ such that $v_i = 0$, whereby $0 = v_i = \lambda (u_{\text{int}})_i + (1 - \lambda) u_i$ and $u_i < 0$ and so:

$$\frac{(u_{\text{int}})_i}{|u_i|} = \frac{1-\lambda}{\lambda} = \frac{\|v - u_{\text{int}}\|}{\|v - u\|} \ . \tag{4.7}$$

And since $(u_{\text{int}})_i \geq r$ it follows that

$$\frac{\|v - u\|}{|u_i|} = \frac{\|v - u_{\text{int}}\|}{(u_{\text{int}})_i} \leq \frac{\|v - u_{\text{int}}\|}{r} \ . \tag{4.8}$$

On the left-most term of (4.8) it follows from $u_i < 0$ that

$$\frac{\|v - u\|}{|u_i|} \geq \frac{\|v - u\|}{\|(u)^-\|} = \frac{\|v - u\|}{\text{Dist}(u, \mathbb{R}^n_+)} \ . \tag{4.9}$$

Combining (4.8) and (4.9) yields

$$\frac{\|u - \mathcal{F}(u; u_{\text{int}})\|}{\text{Dist}(u, \mathbb{R}^n_+)} = \frac{\|u - v\|}{\text{Dist}(u, \mathbb{R}^n_+)} \leq \frac{\|v - u_{\text{int}}\|}{(u_{\text{int}})_i} \leq \frac{\|v - u_{\text{int}}\|}{r} = \frac{\|\mathcal{F}(u; u_{\text{int}}) - u_{\text{int}}\|}{\min_i (u_{\text{int}})_i} \ . \tag{4.10}$$

Noting that the numerator of the right-most term of (4.10) is bounded by

$$\|\mathcal{F}(u; u_{\text{int}}) - u_{\text{int}}\| \ \leq \|\mathcal{F}(u; u_{\text{int}}) - u_{\text{int}}\| + \|\mathcal{F}(u; u_{\text{int}}) - u\| = \|u - u_{\text{int}}\| \ , \tag{4.11}$$

and the left-most term (4.10) satisfies $\frac{\|u - \mathcal{F}(u; u_{\text{int}})\|}{\text{Dist}(u, \mathbb{R}^n_+)} \geq \theta(u)$, we therefore have $\theta(u) \leq \frac{\|\mathcal{F}(u; u_{\text{int}}) - u_{\text{int}}\|}{\min_i (u_{\text{int}})_i}$. Last of all, since $\|u - u_{\text{int}}\| \geq \|\mathcal{F}(u; u_{\text{int}}) - u_{\text{int}}\|$, $\theta(u)$ is also further upper bounded by $\frac{\|u - u_{\text{int}}\|}{\min_i (u_{\text{int}})_i}$, which completes the proof. □

Lemma 4.3 shows that the error ratio $\theta(u)$ is upper-bounded by the ratio of the distance from $u$ to $u_{\text{int}}$ to the distance of $u_{\text{int}}$ to the boundary of the nonnegative orthant. We now use Lemma 4.3 to prove Theorem 4.1.

*Proof of Theorem 4.1.* If $\mathcal{F}_{++} = \emptyset$, then the right-hand side of (4.1) is equal to $+\infty$ so (4.1) is trivially true. We therefore consider the case when $\mathcal{F} \neq \emptyset$. For any optimal solution $u^\star \in \mathcal{U}^\star$ and a given associated strictly feasible solution $u_{\text{int}} \in \mathcal{F}_{++}$, let the $\{u^k\}_{k=1}^\infty$ be a sequence in $V$ that converges to $u^\star$, whereby from Lemma 4.3 it holds that

$$\theta(u^k) \leq \frac{\|u^k - u_{\text{int}}\|}{\min_i (u_{\text{int}})_i} \leq \frac{\|u^\star - u_{\text{int}}\| + \|u^\star - u^k\|}{\min_i (u_{\text{int}})_i} \ .$$

Taking the limit as $k \to \infty$ on both sides, and noting that $\lim_{k\to\infty} \|u^\star - u^k\| = 0$, it thus follows that $\limsup_{k\to\infty} \theta(u^k) \leq \frac{\|u^\star - u_{\text{int}}\|}{\min_i (u_{\text{int}})_i}$. And since $u_{\text{int}}$ is any strictly feasible point, taking the infimum over all such $u_{\text{int}}$ yields

$$\limsup_{k\to\infty} \theta(u^k) \leq \inf_{u_{\text{int}} \in \mathcal{F}_{++}} \frac{\|u^\star - u_{\text{int}}\|}{\min_i (u_{\text{int}})_i} \ . \tag{4.12}$$

We now seek to prove:

$$\theta^\star(\mathcal{U}^\star) := \lim_{\varepsilon \to 0} \sup_{u \in V,\, \text{Dist}(u, \mathcal{U}^\star) \leq \varepsilon} \theta(u) \leq \sup_{u^\star \in \mathcal{U}^\star} \inf_{u_{\text{int}} \in \mathcal{F}_{++}} \frac{\|u^\star - u_{\text{int}}\|}{\min_i (u_{\text{int}})_i} \ . \tag{4.13}$$

If this were false, there would exist $\delta > 0$ and a sequence $\{\bar{u}^k\}_{k=1}^{\infty}$ in $V$ such that $\mathrm{Dist}(\bar{u}^k, \mathcal{U}^\star) \le 1/k$ and $\theta(\bar{u}^k) \ge \delta + \sup_{u^\star \in \mathcal{U}^\star} \inf_{u_{\mathrm{int}} \in \mathcal{F}_{++}} \frac{\|u^\star - u_{\mathrm{int}}\|}{\min_i (u_{\mathrm{int}})_i}$. Note that the points in the sequence $\{\bar{u}^k\}_{k=1}^{\infty}$ all lie in the compact set $\{u : \mathrm{Dist}(u, \mathcal{U}^\star) \le 1\}$ as $\mathcal{U}^\star$ is convex, closed and bounded. Therefore there exists a subsequence of $\{\bar{u}^k\}_{k=1}^{\infty}$ that converges to a limit point in $\mathcal{U}^\star$. This violates (4.12), and so provides a contradiction, whereby (4.13) is true, thus completing the proof. □

Finally notice that Proposition 3.1 is a restatement of (4.2), which was proved as part of the proof of Proposition 4.2.

### 4.3 Relationship between the LimitingER and the distance to infeasibility

We have established the relationship between $\theta^\star(\mathcal{U}^\star)$ and the geometric properties of the feasible sets. In this subsection, we demonstrate that this relationship also extends to the distance of the data from infeasibility. The concept of distance to infeasibility was initially utilized to assess the complexity of LP [40]. Previous research, such as [30, 20], has also studied the connections between global upper bounds of error bounds and the existence of perturbations. Here we show it also holds for the ER $\theta(u)$ and the LimitingER $\theta^\star(\mathcal{U}^\star)$.

Specifically, suppose that the affine subspace $V$ is given by $\{u \in \mathbb{R}^n : Ku = h\}$ for an $m \times n$ real matrix $K$ and a vector $h$ in $\mathbb{R}^m$. Let $\mathrm{SOLN}(K, h)$ denote the feasible set corresponding to $(K, h)$, namely $\mathrm{SOLN}(K, h) := \{u \in \mathbb{R}^n : Ku = h, u \ge 0\}$. We suppose that $\mathrm{SOLN}(K, h) \ne \emptyset$, in which case the "distance to infeasibility" of the data $(K, h)$ is defined as follows:

$$\mathrm{DistInfeas}(K, h) := \inf \left\{ \|\Delta K\| + \|\Delta h\| : \mathrm{SOLN}(K + \Delta K, h + \Delta h) = \emptyset \right\} \ ,$$

see [40]. Now we have the following general theorem about the relationship between $\theta(u)$ and the distance to infeasibility.

**Theorem 4.4.** *Suppose that $\mathcal{F}$ is nonempty for the general LP formulation (3.2). Then for every $u \in V \setminus \mathcal{F}$, it holds that*

$$\theta(u) \le \frac{\|K\|(1 + \|u\|)}{\mathrm{DistInfeas}(K, h)} \ . \tag{4.14}$$

This theorem shows that the larger the distance to infeasibility is (namely the larger the least data perturbation to infeasibility), then the smaller the error ratio $\theta(u)$ must be. The inequality (4.14) looks similar in spirit to Theorem 1.1 part (1) of Renegar [40], even though the setup and context is structurally different from that considered here.

The idea of the proof is to construct, for each $u \in V \setminus \mathcal{F}$, a suitable perturbation $(\Delta K, \Delta h)$ of $(K, h)$ for which $\mathrm{DistInfeas}(K + \Delta K, h + \Delta h) = 0$ and $\theta(u) \le \frac{\|K\|(1+\|u\|)}{\|\Delta K\| + \|\Delta h\|}$. Before presenting the formal proof of the theorem, we establish several key properties of points $u \in V \setminus \mathcal{F}$.

Let $\bar{u} \in V \setminus \mathcal{F}$ be fixed and given, and let $\hat{u}$ be the projection of $\bar{u}$ onto $\mathcal{F}$, denoted as $\hat{u} := P_{\mathcal{F}}(\bar{u})$. Then $\hat{u}$ solves the following convex quadratic program:

$$\min_{u \in \mathbb{R}^n} \ \tfrac{1}{2}\|u - \bar{u}\|^2 \ , \quad \mathrm{s.t.} \ Ku = h, \ u \ge 0 \ , \tag{4.15}$$

whereby there exist multipliers $\hat{y}$ and $\hat{s}$ that together with $\hat{u}$ satisfy the KKT optimality conditions:

$$K\hat{u} = h, \ \hat{u} \ge 0, \ \hat{u} - \bar{u} = K^\top \hat{y} + \hat{s}, \ \hat{s} \ge 0, \ \hat{u}^\top \hat{s} = 0 \ . \tag{4.16}$$

Note that since $\bar{u} \in V \setminus \mathcal{F}$, then $K\bar{u} = K\hat{u} = h$. In addition, the following proposition holds for $(\hat{u}, \hat{s}, \hat{y})$:

**Proposition 4.5.** *For any $u \in \mathbb{R}^n_+$ it holds that $-\|\hat{u}-\bar{u}\|^2 = \hat{s}^\top \bar{u} < \hat{s}^\top \hat{u} = 0 \le \hat{s}^\top u$.*

*Proof.* This first equality follows from (4.16) since $\|\hat{u}-\bar{u}\|^2 = (K^\top y+\hat{s})^\top(\hat{u}-\bar{u})$, $K(\hat{u}-\bar{u})=0$, and $\hat{u}^\top\hat{s}=0$, whereby $\|\hat{u}-\bar{u}\|^2 = -\hat{s}^\top\bar{u}$. The first inequality follows trivially since $\hat{s}^\top\bar{u} = -\|\hat{u}-\hat{u}\|^2 < 0$ and $\hat{u}^\top\hat{s}=0$. And the last inequality follows since $u \ge 0$ and $\hat{s}\ge 0$. □

Let $H$ be the hyperplane defined as $H := \{u : \hat{s}^\top u = 0\}$. Proposition 4.5 implies that $H$ separates $\bar{u}$ and $\mathbb{R}^n_+$, and $\hat{u} \in H$. We denote the projection of $\bar{u}$ onto $H$ as $\check{u}$, namely $\check{u} = P_H(\bar{u})$ which has closed form $\check{u} = \bar{u} - \frac{\hat{s}^\top\bar{u}}{\|\hat{s}\|^2}\cdot\hat{s}$. For simplicity of exposition we use $a$ to denote $\check{u}-\bar{u}$ and use $b$ to denote $\hat{u}-\bar{u}$, namely

$$a := \check{u}-\bar{u} = -\frac{\hat{s}^\top\bar{u}}{\|\hat{s}\|^2}\cdot\hat{s}\ , \quad b := \hat{u}-\bar{u} = K^\top\hat{y}+\hat{s}\ . \tag{4.17}$$

**Proposition 4.6.** *For $a$ and $b$ defined in (4.17) it holds that*

1. $a = \frac{\|b\|^2}{\|\hat{s}\|^2}\cdot\hat{s}$ ,
2. $\|b\| \ge \|a\| > 0$ , *and*
3. $a - \frac{\|a\|^2}{\|b\|^2}\cdot b = K^\top w$, *where* $w := -\frac{\|b\|^2}{\|\hat{s}\|^2}\cdot\hat{y}$ .

*Proof.* To prove item 1, note from Proposition 4.5 that $-\|b\|^2 = -\|\hat{u}-\bar{u}\|^2 = \hat{s}^\top\bar{u}$, whereby $a = -\frac{\hat{s}^\top\bar{u}}{\|\hat{s}\|^2}\cdot\hat{s} = \frac{\|b\|^2}{\|\hat{s}\|^2}\cdot\hat{s}$.

To prove item 2, notice that since $Kb = (K\hat{u}-K\bar{u}) = h-h = 0$, it follows that $\|\hat{s}\|^2 = \|b-K^\top\hat{y}\|^2 = \|b\|^2 + \hat{y}^\top KK^\top\hat{y} \ge \|b\|^2$. And since we have from item 1 that $\|a\| = \frac{\|b\|^2}{\|\hat{s}\|}$, this implies $\|a\|/\|b\| = \|b\|/\|\hat{s}\| \le 1$ which proves the first inequality in item 2. To prove that $\|a\|>0$, note that we cannot have $\|b\|=0$ since $\bar{u}\in V\setminus\mathcal{F}$ and $\hat{u}\in\mathcal{F}$, and we cannot have $\hat{s}=0$, for otherwise Proposition 4.5 would also imply that $b = \hat{u}-\bar{u} = 0$. Therefore from item 1 we have $\|a\|>0$.

To prove item 3, we first show that $\|a\|^2 = b^\top a$. From item 1 and the definition of $b$, we have $a^\top b = \frac{\|b\|^2}{\|\hat{s}\|^2}\cdot\hat{s}^\top(\hat{s}+K^\top\hat{y})$. Since $K(\hat{s}+K^\top\hat{y}) = K\hat{u}-K\bar{u} = h-h = 0$, it follows that

$$\hat{s}^\top(\hat{s}+K^\top\hat{y}) = \|\hat{s}\|^2 + \hat{s}^\top K^\top\hat{y} + \hat{y}^\top K(\hat{s}+K^\top\hat{y}) = \|\hat{s}+K^\top\hat{y}\|^2 = \|b\|^2\ , \tag{4.18}$$

and therefore $a^\top b = \frac{\|b\|^4}{\|\hat{s}\|^2} = \|a\|^2$ (from item 1). This proves $\|a\|^2 = b^\top a$ and then we have

$$\begin{aligned} a - \frac{\|a\|^2}{\|b\|^2}\cdot b = a - \frac{b^\top a}{\|b\|^2}\cdot b &= \frac{\|b\|^2}{\|\hat{s}\|^2}\cdot\hat{s} - \frac{\|b\|^2\cdot b^\top\hat{s}}{\|\hat{s}\|^2\|b\|^2}\cdot b = \frac{\|b\|^2}{\|\hat{s}\|^2}\cdot\hat{s} - \frac{\|b\|^2\cdot(\hat{s}+K^\top\hat{y})^\top\hat{s}}{\|\hat{s}\|^2\|b\|^2}\cdot(\hat{s}+K^\top\hat{y}) \\ &= \frac{\|b\|^2}{\|\hat{s}\|^2}\cdot\hat{s}\cdot\left(1-\frac{(\hat{s}+K^\top\hat{y})^\top\hat{s}}{\|b\|^2}\right) - \frac{(\hat{s}+K^\top\hat{y})^\top\hat{s}}{\|\hat{s}\|^2}\cdot K^\top\hat{y}\ . \end{aligned} \tag{4.19}$$

Here, the second equality follows from item 1, and the third equality uses $b = \hat{s}+K^\top\hat{y}$. Substituting (4.18) into (4.19) yields $a - \frac{\|a\|^2}{\|b\|^2}\cdot b = -\frac{\|b\|^2}{\|\hat{s}\|^2}\cdot K^\top\hat{y} = K^\top w$, thus proving item 3. □

With the above propositions established, we now prove Theorem 4.4.

*Proof of Theorem 4.4.* Let $\bar{u}\in V\setminus\mathcal{F}$ be given. We will use all of the notation developed earlier in this subsection, including $\hat{u} = P_{\mathcal{F}}(\bar{u})$, the KKT multipliers $(\hat{y},\hat{s})$, $H := \{u : \hat{s}^\top u = 0\}$, and $\check{u} = P_H(\bar{u}) = \bar{u} - \frac{\hat{s}^\top\bar{u}}{\|\hat{s}\|^2}\cdot\hat{s}$. Additionally, let $a$ and $b$ be as given in (4.17) and $w = -\frac{\|b\|^2}{\|\hat{s}\|^2}\cdot\hat{y}$ as specified in Proposition 4.6.

We first examine the case when $\hat{y} \neq 0$, and hence $w \neq 0$. Let us consider the following perturbations of $K$ and $h$:

$$\Delta K := \frac{\|a\|^2}{\|w\|^2\|b\|^2} \cdot w(b-a)^\top, \quad \Delta h := \Delta K \bar{u} - \varepsilon w \ , \tag{4.20}$$

where $\varepsilon > 0$ is a small positive scalar. Then:

$$\begin{aligned}(K+\Delta K)^\top w = K^\top w + \Delta K^\top w &= a - \frac{\|a\|^2}{\|b\|^2} \cdot b + \frac{\|a\|^2}{\|w\|^2\|b\|^2} \cdot (b-a) \cdot w^\top w \\ &= \left(1 - \frac{\|a\|^2}{\|b\|^2}\right) a = \left(1 - \frac{\|a\|^2}{\|b\|^2}\right) \cdot \frac{\|b\|^2}{\|\hat{s}\|} \cdot \hat{s} \geq 0 \ ,\end{aligned} \tag{4.21}$$

where the second and the fourth equalities are due to Proposition 4.6, and the final inequality is also due to Proposition 4.6. Furthermore, we have

$$w^\top(h + \Delta h) = w^\top(K+\Delta K)\bar{u} - \varepsilon\|w\|^2 = \left(1 - \frac{\|a\|^2}{\|b\|^2}\right) \cdot \frac{\|b\|^2}{\|\hat{s}\|} \cdot \hat{s}^\top \bar{u} - \varepsilon\|w\|^2 < 0 \ , \tag{4.22}$$

where the strict inequality follows since $\hat{s}^\top \bar{u} < 0$ from Proposition 4.5, $\|a\| \leq \|b\|$ from Proposition 4.6, and $\|w\| > 0$ by supposition for this case. Examining (4.21) and (4.22) yields $(K+\Delta K)^\top w \geq 0$ and $w^\top(h+\Delta h) < 0$, which implies via Farkas' lemma that $\operatorname{SOLN}(K+\Delta K, h+\Delta h) = \emptyset$, and hence $\operatorname{DistInfeas}(K,h) \leq \|\Delta K\| + \|\Delta h\|$.

Let us now bound the size of $\|\Delta K\|$ and $\|\Delta h\|$. From (4.20) we have

$$\|\Delta K\| \leq \frac{\|a\|^2}{\|w\|\|b\|^2} \cdot \|b-a\| \ . \tag{4.23}$$

Since $a - \frac{\|a\|^2}{\|b\|^2} \cdot b = K^\top w$, we also have $\|K\| \geq \left\| a - \frac{\|a\|^2}{\|b\|^2} \cdot b \right\| \cdot \frac{1}{\|w\|}$. Therefore

$$\|K\| \geq \left\| \|a\| \cdot \frac{a}{\|a\|} - \frac{\|a\|^2}{\|b\|} \cdot \frac{b}{\|b\|} \right\| \cdot \frac{1}{\|w\|} = \left\| \frac{\|a\|^2}{\|b\|} \cdot \frac{a}{\|a\|} - \|a\| \cdot \frac{b}{\|b\|} \right\| \cdot \frac{1}{\|w\|} = \frac{\|a\|}{\|b\|} \cdot \|b-a\| \cdot \frac{1}{\|w\|} \ , \tag{4.24}$$

where the first equality above follows from squaring both sides and rearranging terms using Proposition 4.6. Combining (4.23) and (4.24) yields $\frac{\|\Delta K\|}{\|K\|} \leq \frac{\|a\|}{\|b\|} = \frac{\|\check{u}-\bar{u}\|}{\|\hat{u}-\bar{u}\|} = \frac{\operatorname{Dist}(\bar{u},H)}{\operatorname{Dist}(\bar{u},\mathcal{F})}$. From Proposition 4.5, because $H$ separates $\hat{u}$ and $\mathbb{R}^n_+$, it follows that $\operatorname{Dist}(\bar{u},H) \leq \operatorname{Dist}(\bar{u},\mathbb{R}^n_+)$, whereby $\frac{\|\Delta K\|}{\|K\|} \leq \frac{\operatorname{Dist}(\bar{u},\mathbb{R}^n_+)}{\operatorname{Dist}(\bar{u},\mathcal{F})} = \frac{1}{\theta(\bar{u})}$. Moreover, since $\Delta h := \Delta K \bar{u} - \varepsilon w$, it follows that $\|\Delta h\| \leq \|\Delta K\| \cdot \|\bar{u}\| + \varepsilon\|w\| \leq \frac{\|\bar{u}\|}{\theta(\bar{u})}\|K\| + \varepsilon\|w\|$. Finally, we can add the inequalities $\theta(\bar{u})\|\Delta K\| \leq \|K\|$ and $\theta(\bar{u})\|\Delta h\| \leq \|u\|\|K\| + \theta(\bar{u})\varepsilon\|w\|$ which yields after rearranging $\theta(u) \leq \frac{\|K\|(1+\|u\|+\varepsilon\cdot\theta(\bar{u})\|w\|/\|K\|)}{\|\Delta K\|+\|\Delta h\|}$. And since $\operatorname{DistInfeas}(K,h) \leq \|\Delta K\| + \|\Delta h\|$ we have $\theta(u) \leq \frac{\|K\|(1+\|u\|+\varepsilon\cdot\theta(\bar{u})\|w\|/\|K\|)}{\operatorname{DistInfeas}(K,h)}$. Taking the limit as $\varepsilon \to 0$ then proves the result in the case when $\hat{y} \neq 0$.

Next we consider the case when $\hat{y} = 0$. It follows from (4.16) that $\hat{u} - \bar{u} = \hat{s}$. Let $I := \{i : \hat{u}_i > 0\}$ and $J := [n] \setminus I$. Then we have $\hat{s}_I = 0$ and $\hat{u}_I = \bar{u}_I$, and $\hat{s}_J = -\bar{u}_J$. This implies that $\hat{u} = \bar{u}^+ = P_{\mathbb{R}^n_+}(\bar{u})$, and hence $\operatorname{Dist}(\bar{u},\mathcal{F}) = \|\hat{u}-\bar{u}\| = \|\hat{s}\| = \operatorname{Dist}(\bar{u},\mathbb{R}^n_+)$, and hence $\theta(\bar{u}) = \operatorname{Dist}(\bar{u},\mathcal{F})/\operatorname{Dist}(\bar{u},\mathbb{R}^n_+) = 1$. Now let $\Delta K = -K$, and for any $\varepsilon > 0$ let $\Delta h$ be any vector satisfying $\|\Delta h\| \leq \varepsilon$ and $h + \Delta h \neq 0$. Then $(K+\Delta K, h+\Delta h) = (0, h+\Delta h)$ whereby $\operatorname{SOLN}(K+\Delta K, h+\Delta h) = \emptyset$. Therefore $\operatorname{DistInfeas}(K,h) \leq \|\Delta K\| + \|\Delta h\| \leq \|K\| + \varepsilon$ for all $\varepsilon > 0$, and thus $\operatorname{DistInfeas}(K,h) \leq \|K\|$. Finally, we have in this case that $\theta(\bar{u}) = 1 \leq \frac{\|K\|}{\operatorname{DistInfeas}(K,h)} \leq \frac{\|K\|(1+\|\bar{u}\|)}{\operatorname{DistInfeas}(K,h)}$, which completes the proof. □

We also have the following relationship between $\theta^\star(\mathcal{U}^\star)$ and the distance to infeasibility.

**Corollary 4.7.** *Suppose that the general LP formulation* (3.2) *has an optimal solution. If* $\mathrm{DistInfeas}(K,h) > 0$*, then it holds that*

$$\theta^\star(\mathcal{U}^\star) \le \frac{1 + \max_{u\in\mathcal{U}^\star}\|u\|}{\mathrm{DistInfeas}(K,h)} \ . \tag{4.25}$$

*Proof.* By the definition of $\theta^\star(\mathcal{U}^\star)$ in (3.4) and Theorem 4.4, we have

$$\theta^\star(\mathcal{U}^\star) \ = \lim_{\varepsilon\to 0}\left(\sup_{u\in V,\,\mathrm{Dist}(u,\mathcal{U}^\star)\le\varepsilon} \theta(u)\right) \le \lim_{\varepsilon\to 0}\left(\sup_{u\in V,\,\mathrm{Dist}(u,\mathcal{U}^\star)\le\varepsilon}\left(\frac{1+\|u\|}{\mathrm{DistInfeas}(K,h)}\right)\right) \ , \tag{4.26}$$

which is exactly (4.25). $\square$

Both Theorem 4.4 and Corollary 4.7 imply that the farther the feasible set $\mathcal{F}$ is from infeasibility (namely, the larger $\mathrm{DistInfeas}(K,h)$ is), then the smaller $\theta(u)$ and $\theta^\star(\mathcal{U}^\star)$ must be.

# 5 LP Sharpness, stability under perturbation, and computing $\mu$

In this section we present a characterization of the LP sharpness $\mu$ in terms of the least relative perturbation $\Delta c$ of the objective function vector $c$ that yields a different optimal solution set that is nonempty and not a subset of the original solution set. We also present a characterization of the LP sharpness via polyhedral geometry, with implications for computing the LP sharpness $\mu$.

## 5.1 Characterization of LP sharpness and the stability of optimal solutions under perturbation

Let $\mathrm{OPT}(g,\mathcal{F})$ denote the set of optimal solutions of the general LP problem (3.2), namely

$$\mathrm{OPT}(g,\mathcal{F}) := \arg\min_{u\in\mathcal{F}} \ g^\top u \ .$$

Also let $C_{\mathcal{U}^\star}$ denote the recession cone of the optimal solution set $\mathcal{U}^\star$, and let $C^*_{\mathcal{U}^\star}$ denote its (positive) dual cone, namely $C^*_{\mathcal{U}^\star} = \{w : w^\top u \ge 0 \text{ for all } u \in C_{\mathcal{U}^\star}\}$. Then for any $\Delta g \in \vec{V}^\perp$, $\mathrm{OPT}(g,\mathcal{F}) = \mathrm{OPT}(g+\Delta g,\mathcal{F})$. Additionally, if $\Delta g \notin C^*_{\mathcal{U}^\star}$, then $\mathrm{OPT}(g+\Delta g,\mathcal{F}) = \emptyset$ (since there exists $u \in C_{\mathcal{U}^\star}$ for which $g^\top u < 0$ and hence the resulting LP instance has unbounded objective value).

Intuition suggests that the LP sharpness value $\mu$ should be related to objective function perturbations that alter the set of optimal solutions. Indeed, we have the following theorem that characterizes this relationship completely, namely $\mu$ is the smallest relative perturbation $\Delta g$ of $g$ that yields a different optimal solution set that is nonempty and not a subset of the original optimal solution set. More precisely we have:

**Theorem 5.1.** *Consider the general LP problem* (3.2) *under Assumption 1, and let* $\mu$ *be the LP sharpness of* (3.2)*. Then*

$$\mu = \inf_{\Delta g}\left\{\frac{\|P_{\vec{V}}(\Delta g)\|}{\|P_{\vec{V}}(g)\|} : \mathrm{OPT}(g+\Delta g,\mathcal{F}) \neq \emptyset \ \textit{ and } \ \mathrm{OPT}(g+\Delta g,\mathcal{F}) \not\subset \mathrm{OPT}(g,\mathcal{F})\right\} \ . \tag{5.1}$$

Note that under Assumption 1 we must have $\|P_{\vec{V}}(g)\| > 0$ as otherwise all feasible solutions would be optimal (which violates Assumption 1). The proof of Theorem 5.1 is in two parts, where each part proves an inequality version of (5.1) in one of the two possible directions of the inequality. The following lemma proves the "$\le$" version of (5.1).

**Lemma 5.2.** *Consider the general LP problem* (3.2) *under Assumption 1, and let $\mu$ be the LP sharpness of* (3.2)*. Then*

$$\mu \le \inf_{\Delta g} \left\{ \frac{\|P_{\vec{V}}(\Delta g)\|}{\|P_{\vec{V}}(g)\|} : \mathrm{OPT}(g+\Delta g, \mathcal{F}) \neq \emptyset \ \ \textit{and} \ \mathrm{OPT}(g+\Delta g, \mathcal{F}) \not\subset \mathrm{OPT}(g, \mathcal{F}) \right\} \ . \tag{5.2}$$

*Proof.* Let $\Delta g$ satisfy $\mathrm{OPT}(g+\Delta g, \mathcal{F}) \neq \emptyset$ and $\mathrm{OPT}(g+\Delta g, \mathcal{F}) \not\subset \mathrm{OPT}(g, \mathcal{F})$, and let $\bar{u} \in \mathrm{OPT}(g+\Delta g, \mathcal{F}) \setminus \mathrm{OPT}(g, \mathcal{F})$. Denote the optimal objective value hyperplane by $H^\star := \{u : g^\top u = f^\star\}$. Let $\check{u} := P_{V\cap H^\star}(\bar{u}) = \bar{u} - \frac{g^\top \bar{u} - f^\star}{\|P_{\vec{V}}(g)\|^2} \cdot P_{\vec{V}}(g)$ and $\hat{u} := P_{\mathcal{U}^\star}(\bar{u})$. For simplicity, we use the notation $a := \check{u} - \bar{u}$ and $b := \hat{u} - \bar{u}$. From Definition 3.2 we have:

$$\mu \le \frac{\mathrm{Dist}(\bar{u}, V \cap H^\star)}{\mathrm{Dist}(\bar{u}, \mathcal{U}^\star)} = \frac{\|\check{u} - \bar{u}\|}{\|\hat{u} - \bar{u}\|} = \frac{\|a\|}{\|b\|} \ . \tag{5.3}$$

Our goal then is to prove that

$$\frac{\|a\|}{\|b\|} \le \frac{\|P_{\vec{V}}(\Delta g)\|}{\|P_{\vec{V}}(g)\|} \ , \tag{5.4}$$

and combining (5.4) with (5.3) will yield the proof.

Because $\check{u} \in H^\star$ and $\hat{u} \in H^\star$, we have $g^\top \check{u} = g^\top \hat{u}$, which implies that $g^\top a = g^\top b$. Furthermore, since $\bar{u} \in \mathrm{OPT}(g+\Delta g, \mathcal{F})$, we have $(g+\Delta g)^\top \bar{u} \le (g+\Delta g)^\top \hat{u}$, which implies that $(g+\Delta g)^\top b \ge 0$. Substituting $g^\top a = g^\top b$ into $(g+\Delta g)^\top b \ge 0$, we obtain $\Delta g^\top b \ge -g^\top a$. It follows directly from Assumption 1 that $\|P_{\vec{V}}(g)\| > 0$, and dividing both sides of this last inequality by $\|P_{\vec{V}}(g)\|$ yields

$$\left( \frac{\Delta g}{\|P_{\vec{V}}(g)\|} \right)^\top b \ge -\left( \frac{g}{\|P_{\vec{V}}(g)\|} \right)^\top a \ . \tag{5.5}$$

Regarding the right-hand side of (5.5), note that $\check{u} = u - \frac{g^\top \bar{u} - f^\star}{\|P_{\vec{V}}(g)\|^2} \cdot P_{\vec{V}}(g)$ and $a = -\frac{g^\top \bar{u} - f^\star}{\|P_{\vec{V}}(g)\|^2} \cdot P_{\vec{V}}(g)$, whereby:

$$-\left( \frac{g}{\|P_{\vec{V}}(g)\|} \right)^\top a = -\left( \frac{P_{\vec{V}}(g)}{\|P_{\vec{V}}(g)\|} \right)^\top a = \|a\| \ . \tag{5.6}$$

Regarding the left-hand side of (5.5), since $b = \hat{u} - \bar{u} \in \vec{V}$, it follows that $(\Delta g)^\top b = (P_{\vec{V}}(\Delta g))^\top b \le \|P_{\vec{V}}(\Delta g)\| \|b\|$. Substituting this inequality and (5.6) back into (5.5) yields (5.4), which as noted earlier combines with (5.3) to complete the proof. □

Before proving the "$\ge$" direction, we first establish a simple proposition. For a convex set $\mathcal{S}$ let $C_{\mathcal{S}}$ denote the recession cone of $S$, and let $C^*_{\mathcal{S}}$ denote the corresponding (positive) dual cone.

**Proposition 5.3.** *Let $\bar{u} \in \mathcal{F} \setminus \mathcal{U}^\star$, and let $\hat{u} := P_{\mathcal{U}^\star}(\bar{u})$, then*

- $(\hat{u} - \bar{u})^\top (u^\star - \hat{u}) \ge 0$ *for any $u^\star \in \mathcal{U}^\star$, and*
- $\hat{u} - \bar{u} \in C^*_{\mathcal{U}^\star}$ .

*Proof.* The first assertion follows directly from the optimality conditions for the projection of $\bar{u}$ onto $\mathcal{U}^\star$. For the second assertion, observe that for any $v \in C_{\mathcal{U}^\star}$ and any $u \in \mathcal{U}^\star$ we have $u + \lambda v \in \mathcal{U}^\star$ for all $\lambda \ge 0$, whereby it follows from the first assertion that $(\hat{u} - \bar{u})^\top (u + \lambda v - \hat{u}) \ge 0$ for all $\lambda \ge 0$ and hence $(\hat{u} - \bar{u})^\top v \ge 0$. Since $v$ is an arbitrary point in $C_{\mathcal{U}^\star}$ it holds that $\hat{u} - \bar{u} \in C^*_{\mathcal{U}^\star}$ . □

**Lemma 5.4.** *Consider the general LP problem* (3.2) *under Assumption 1, and let $\mu$ be the LP sharpness of* (3.2)*. Then*

$$\mu \ge \inf_{\Delta g} \left\{ \frac{\|P_{\vec{V}}(\Delta g)\|}{\|P_{\vec{V}}(g)\|} : \mathrm{OPT}(g+\Delta g, \mathcal{F}) \neq \emptyset \ \ \textit{and} \ \mathrm{OPT}(g+\Delta g, \mathcal{F}) \not\subset \mathrm{OPT}(g, \mathcal{F}) \right\} \ . \tag{5.7}$$

*Proof.* Note that by setting $\Delta g := -g$ that $\mathrm{OPT}(g + \Delta g, \mathcal{F}) = \mathcal{F} \not\subset \mathrm{OPT}(g, \mathcal{F}) = \mathcal{U}^\star$ under Assumption 1, and therefore the right-hand side of (5.7) is at most 1. Recall from the definition of LP sharpness that $\mu \le 1$. Therefore in the special case when $\mu = 1$ then (5.7) holds trivially.

Let us therefore consider the case $\mu < 1$. For any given $\bar{u} \in \mathcal{F} \setminus \mathcal{U}^\star$, we will construct a perturbation $\Delta g$ for which $\mathrm{OPT}(g + \Delta g, \mathcal{F}) \neq \emptyset$ and $\mathrm{OPT}(g + \Delta g, \mathcal{F}) \not\subset \mathrm{OPT}(g, \mathcal{F})$, and

$$\frac{\|P_{\vec{V}}(\Delta g)\|}{\|P_{\vec{V}}(g)\|} \le \frac{\mathrm{Dist}(\bar{u}, V \cap H^\star)}{\mathrm{Dist}(\bar{u}, \mathcal{U}^\star)} \ , \tag{5.8}$$

which then implies (5.7). We proceed as follows. Let $\bar{u} \in \mathcal{F} \backslash \mathcal{U}^\star$ be given, and define $\check{u} := P_{V \cap H^\star}(\bar{u}) = \bar{u} - \frac{g^\top \bar{u} - f^\star}{\|P_{\vec{V}}(g)\|^2} \cdot P_{\vec{V}}(g)$ and $\hat{u} := P_{\mathcal{U}^\star}(u)$. Similar to the notation used in the proof of Lemma 5.2, let $a := \check{u} - \bar{u}$ and $b := \hat{u} - \bar{u}$.

To construct the perturbation $\Delta g$ we first define

$$\bar{b} := \frac{-g^\top b}{\|b\|^2} \cdot b \tag{5.9}$$

and construct the perturbation $\Delta g := t \cdot \bar{b}$, where $t$ is the optimal objective value of the following optimization problem:

$$t \ := \ \max_{\tau} \ \tau \qquad \text{s. t.} \ \ \hat{u} \in \mathrm{OPT}(g + \tau \bar{b}, \mathcal{F}\} \ . \tag{5.10}$$

We aim to show that $t \in (0, 1]$. Towards the proof of this inclusion, let $\mathcal{E}1$ be the set of extreme points of $\mathcal{F}$ that are in $\mathcal{U}^\star$, and let $\mathcal{E}2$ be the set of extreme points of $\mathcal{F}$ that are not in $\mathcal{U}^\star$. Similarly let $\mathcal{R}1$ be the set of extreme rays of $\mathcal{F}$ that are also extreme rays of $\mathcal{U}^\star$, and let $\mathcal{R}2$ be the set of extreme rays of $\mathcal{F}$ that are not also extreme rays of $\mathcal{U}^\star$. Note that $\mathcal{E}1$ and $\mathcal{E}2$ are finite sets, and $\mathcal{R}1$ and $\mathcal{R}2$ are also finite sets. We can therefore re-write (5.10) as:

$$\begin{aligned}
\mathrm{OP}: \qquad t \ := \ \max_{\tau} \quad & \tau && && \text{(5.11)}\\
\text{s. t.} \quad & \tau \cdot \bar{b}^\top (\hat{u} - v^i) && \le -g^\top (\hat{u} - v^i) && \text{for each } v^i \in \mathcal{E}1 \quad \text{(5.12)}\\
& \tau \cdot \bar{b}^\top (\hat{u} - v^i) && \le -g^\top (\hat{u} - v^i) && \text{for each } v^i \in \mathcal{E}2 \quad \text{(5.13)}\\
& \tau \cdot \bar{b}^\top r^i && \ge -g^\top r^i && \text{for each } r^i \in \mathcal{R}1 \quad \text{(5.14)}\\
& \tau \cdot \bar{b}^\top r^i && \ge -g^\top r^i && \text{for each } r^i \in \mathcal{R}2 \quad \text{(5.15)}
\end{aligned}$$

First observe that $-g^\top b = -g^\top (\hat{u} - \bar{u}) > 0$, whereby $\bar{b}$ is a positive scaling of $b$. Also notice that $\tau = 0$ is feasible for OP, because $\bar{u} \in \mathrm{OPT}(g, \mathcal{F})$. It implies that the right-hand sides of (5.12) and (5.13) are nonnegative and the right-hand sides of (5.14) and (5.15) are nonpositive. Next we show that $t \le 1$. To see this, note that if $\tau > 1$ then

$$(g + \tau \cdot \bar{b})^\top (\hat{u} - \bar{u}) = (g + \tau \cdot \bar{b})^\top b = (g + \bar{b})^\top b + (\tau - 1)\bar{b}^\top b = (\tau - 1) \cdot (-g^\top b) > 0 \ ,$$

whereby $\hat{u} \notin \mathrm{OPT}(g + \tau \bar{b}, \mathcal{F})$ and thus $\tau$ is not feasible for (5.10).

It thus remains to show that $t > 0$, which we will demonstrate by examining the constraints of OP. Examining the constraints (5.12), when $v^i \in \mathcal{E}1$ the corresponding right-hand side is equal to 0 while $\bar{b}^\top (\hat{u} - v^i) \le 0$ (from Proposition 5.3), so these constraints are satisfied for all $\tau \ge 0$. Examining the constraints (5.13), when $v^i \in \mathcal{E}2$ the corresponding right-hand side is strictly positive so (5.13) is satisfied for all sufficiently small $\tau > 0$. Examining the constraints (5.14), when $r^i \in \mathcal{R}1$ the corresponding right-hand side is equal to 0 while $\bar{b}^\top r^i \ge 0$ (from Proposition 5.3), so these constraints are satisfied for all $\tau \ge 0$. And examining (5.15), when $r^i \in \mathcal{R}2$ the corresponding right-hand side is strictly negative, so (5.15) is satisfied for all sufficiently small $\tau > 0$. Therefore there exists $\tau > 0$ that satisfies all of the constraints of OP, and therefore $t > 0$.

Now let us show that $\mathrm{OPT}(g + \Delta g, \mathcal{F}) \neq \emptyset$ and $\mathrm{OPT}(g + \Delta g, \mathcal{F}) \not\subset \mathrm{OPT}(g, \mathcal{F})$. It follows from (5.10) that $(g + \Delta g)^\top(\hat{u} - u) \leq 0$ for any $u \in \mathcal{F}$ and therefore $\hat{u} \in \mathrm{OPT}(g + \Delta g, \mathcal{F})$ and therefore $\mathrm{OPT}(g + \Delta g, \mathcal{F}) \neq \emptyset$. Now notice that when $\tau$ is optimal for (5.10) (and its equivalent formulation OP), there exists either $v^i \in \mathcal{E}2$ for which the corresponding constraint in (5.13) is active, or $r^i \in \mathcal{R}2$ for which the corresponding constraint in (5.15) is active (or both). In the former case, $v^i \in \mathrm{OPT}(g + \Delta g, \mathcal{F})$ and in the latter case $\hat{u} + r^i \in \mathrm{OPT}(g + \Delta g, \mathcal{F})$. And in either case, we have $\mathrm{OPT}(g + \Delta g, \mathcal{F}) \not\subset \mathrm{OPT}(g, \mathcal{F})$.

Last of all, because $g^\top a = g^\top b$, $a \in \vec{V}$, and $\bar{b} \in \vec{V}$, we have

$$\|P_{\vec{V}}(\Delta g)\| = t \cdot \|\bar{b}\| = t \cdot \frac{-g^\top b}{\|b\|} \leq \frac{-g^\top b}{\|b\|} = \frac{-g^\top a}{\|b\|} = \frac{-P_{\vec{V}}(g)^\top a}{\|b\|} \leq \frac{\|a\|}{\|b\|} \cdot \|P_{\vec{V}}(g)\| = \frac{\mathrm{Dist}(\bar{u}, V \cap H^\star)}{\mathrm{Dist}(\bar{u}, \mathcal{U}^\star)} \cdot \|P_{\vec{V}}(g)\| \ .$$

This shows (5.8) and completes the proof. □

Last of all, Theorem 5.1 follows by combining Lemmas 5.2 and 5.4.

## 5.2 Polyhedral geometry of LP sharpness, and computation of $\mu$

In this subsection we present a characterization of the LP sharpness $\mu$ via polyhedral geometry, with implications for computing $\mu$. Before we go into details, we first convey our general results as follows. If the optimal solution set is a singleton, namely $\mathcal{U}^\star = \{u^\star\}$, then we will show that the LP sharpness $\mu$ is the smallest sharpness along all of the edges of $\mathcal{F}$ emanating from $u^\star$. One implication of this result is that if the LP instance is primal and dual nondegenerate, then the dual nondegeneracy implies that $\mathcal{U}^\star$ is a singleton, and the primal nondegeneracy implies that there are exactly $n - m$ edges emanating from $\mathcal{U}^\star$, and hence it will be very easy to compute the LP sharpness. The more general result that we will show is that LP sharpness is the smallest sharpness along all edges of $\mathcal{F}$ that intersect $\mathcal{U}^\star$ but are not subsets of $\mathcal{U}^\star$. In the absence of nondegeneracy there can be exponentially many such edges, and so computing the LP sharpness for either a primal or dual degenerate instance is not a tractable problem in general.

We now develop these results more formally. For any $u \in \mathcal{F} \setminus \mathcal{U}^\star$, let $u^\star = \arg\min_{v^\star \in \mathcal{U}^\star} \|v^\star - u\|$ be the projection of $u$ onto $\mathcal{U}^\star$, and we define the sharpness of the point $u$ to be:

$$G(u) := \frac{\mathrm{Dist}(u, V \cap H^\star)}{\mathrm{Dist}(u, \mathcal{U}^\star)} = \frac{\frac{g^\top(u - u^*)}{\|P_{\vec{V}}(g)\|}}{\|u - u^\star\|}$$

where in the above expression $u^\star := \arg\min_{v^\star \in \mathcal{U}^\star} \|v^\star - u\|$ is the projection of $u$ onto $\mathcal{U}^\star$. With this notation the LP sharpness (3.6) is $\mu = \inf_{u \in \mathcal{F} \setminus \mathcal{U}^\star} G(u)$. Next let us recall some notation about convex polyhedra, see [17]. An edge of a polyhedron is a 1-dimensional face of the polyhedron. And since $\mathcal{F}$ is a polyhedron it follows that $\mathcal{F}$ will have a finite number of edges. Furthermore, every edge will either be (i) a line segment joining two different vertices $v^1 \neq v^2$ of $\mathcal{F}$ which we denote by $\boldsymbol{e} = [v^1, v^2]$, or (ii) a half-line points $v + \theta r$ for all $\theta \geq 0$, where $v$ is a vertex of $\mathcal{F}$ and $r$ is an extreme ray of $\mathcal{F}$, and which we denote by $\boldsymbol{f} = [v; r]$. We will be concerned with the subset of edges of $\mathcal{F}$ which have one endpoint in $\mathcal{U}^\star$ but are not subsets of $\mathcal{U}^\star$, which we call edges emanating away from $\mathcal{U}^\star$ and which we denote as $\mathcal{M}$, and whose formal definition is:

$$\begin{aligned} \mathcal{M} := \mathcal{M}_1 \cup \mathcal{M}_2 \qquad \text{where} \quad & \mathcal{M}_1 := \{\boldsymbol{e} = [v^1, v^2] : \boldsymbol{e} \text{ is an edge of } \mathcal{F},\ v^1 \in \mathcal{U}^*, v^2 \notin \mathcal{U}^\star\} \\ \text{and} \quad & \mathcal{M}_2 := \{\boldsymbol{f} = [v, r] : \boldsymbol{f} \text{ is an edge of } \mathcal{F},\ v \in \mathcal{U}^*, g^\top r = 1\}\ . \end{aligned}$$

The theorem below shows that the LP sharpness can be characterized using the following two functions:

$$R_1(\boldsymbol{e}) := G(\tilde{u}) = \frac{\mathrm{Dist}(\tilde{u}, V \cap H^\star)}{\mathrm{Dist}(\tilde{u}, \mathcal{U}^\star)} \text{ for all edges } \boldsymbol{e} = [u^\star, \tilde{u}] \in \mathcal{M}_1 \text{ , and}$$

$$R_2(\boldsymbol{f}; \bar{\varepsilon}) := G(u^\star + \bar{\varepsilon} \cdot r) = \frac{\mathrm{Dist}(u^\star + \bar{\varepsilon} \cdot r, V \cap H^\star)}{\mathrm{Dist}(u^\star + \bar{\varepsilon} \cdot r, \mathcal{U}^\star)} \text{ for all edges } \boldsymbol{f} = [u^\star; r] \in \mathcal{M}_2 \text{ and all } \bar{\varepsilon} > 0 \ .$$

Because $\mathcal{M}$ is a finite set, we can write $\mathcal{M} = \{\boldsymbol{e}^i : i = 1, 2, \ldots, m_1\} \cup \{\boldsymbol{f}^j : j = 1, 2, \ldots, m_2\}$ for some integers $m_1, m_2$.

**Theorem 5.5.** *For any given $\bar{\varepsilon} > 0$, the LP sharpness is characterized as follows:*

$$\mu = \min\left\{R_1(\boldsymbol{e}^1), R_1(\boldsymbol{e}^2), \ldots, R_1(\boldsymbol{e}^{m_1}), R_2(\boldsymbol{f}^1; \bar{\varepsilon}), R_2(\boldsymbol{f}^2; \bar{\varepsilon}), \ldots, R_2(\boldsymbol{f}^{m_2}; \bar{\varepsilon})\right\} \ . \tag{5.16}$$

Before proving Theorem 5.5 we discuss its implications for computing the LP sharpness measure $\mu$. The computation of $R_1(\boldsymbol{e})$ or $R_2(\boldsymbol{f}; \bar{\varepsilon})$ for a given edge $\boldsymbol{e}$ or $\boldsymbol{f}$ requires computing a simple algebraic projection for the numerator of $R_i$ and requires solving a convex quadratic program associated with the projection onto $\mathcal{U}^\star$ for the denominator of $R_i$. Note that for a given basis of an optimal solution $u^\star$, there are at most $n$ edges emanating from $u^\star$ and all of them have closed-form formulation [6]. Therefore if all optimal bases have been enumerated then the computation of $\mu$ via Theorem 5.5 is itself fairly straightforward. The key issue therefore is the enumeration of all optimal bases. For a general polyhedron [21] proves that enumerating all vertices is NP-hard if the polyhedron is unbounded. However, in some cases enumerating all optimal bases is easy. For example, when there is exactly one optimal basis (which occurs if the LP instance is primal and dual nondegenerate), then solving the LP instance yields the unique optimal basis. Furthermore, when $\mathcal{U}^\star$ is bounded and all optimal bases are primal non-degenerate, [4] provides an algorithm for enumerating all optimal bases in $O(n^2 v)$ time complexity, where $v$ is the number of optimal bases. Hence when the number of optimal bases is small, enumerating all of them remains tractable and so computing $\mu$ is also tractable.

Our proof of Theorem 5.5 relies on the following two elementary propositions. For any $\varepsilon \geq 0$ define $S_\varepsilon$ to be the level set whose objective function value is exactly $\varepsilon$ larger than the optimal objective value, namely $S_\varepsilon := \mathcal{F} \cap \{v : g^\top v = f^\star + \varepsilon\}$.

**Proposition 5.6.** *If $\varepsilon_2 \geq \varepsilon_1 > 0$ and $S_{\varepsilon_i} \neq \emptyset$ for $i = 1, 2$, then $\inf_{u \in S_{\varepsilon_2}} G(u) \geq \inf_{u \in S_{\varepsilon_1}} G(u)$.*

*Proof.* For any given $u \in S_{\varepsilon_2}$, let $\hat{u} := P_{\mathcal{U}^\star}(u)$ and $t := \varepsilon_1/\varepsilon_2$. Then for $u_t := tu + (1-t)\hat{u}$ it holds that $u_t \in S_{\varepsilon_1}$. Also notice that $\mathrm{Dist}(u_t, \mathcal{U}^\star) = \|u_t - \hat{u}\| = t \cdot \|u - \hat{u}\| = t \cdot \mathrm{Dist}(u, \mathcal{U}^\star)$ and $\mathrm{Dist}(u_t, V \cap H^\star) = (1-t) \cdot \mathrm{Dist}(\hat{u}, V \cap H^\star) + t \cdot \mathrm{Dist}(u, V \cap H^\star) = t \cdot \mathrm{Dist}(u, V \cap H^\star)$. Therefore $G(u_t) = \frac{t \cdot \mathrm{Dist}(u, V \cap H^\star)}{t \cdot \mathrm{Dist}(u, \mathcal{U}^\star)} = G(u)$. Since this equality holds for all $u \in S_{\varepsilon_2}$, it follows that $\inf_{u \in S_{\varepsilon_2}} G(u) \geq \inf_{u \in S_{\varepsilon_1}} G(u)$, which proves the proposition. □

**Proposition 5.7.** *Let $u^\star \in \mathcal{U}^\star$ and $v \in \vec{V}$ satisfy $g^\top v > 0$. If $t_2 \geq t_1 > 0$ and $u^\star + t_i \cdot v \in \mathcal{F}$ for $i = 1, 2$, then $G(u^\star + t_1 \cdot v) \geq G(u^\star + t_2 \cdot v)$.*

*Proof.* For $t \geq 0$ define $g_1(t) := \mathrm{Dist}(u^\star + t \cdot v,\ V \cap H^\star)$ and $g_2(t) := \mathrm{Dist}(u^\star + t \cdot v,\ \mathcal{U}^\star)$. Then for $t > 0$ we have $G(u^\star + t \cdot v) = g_1(t)/g_2(t)$. Notice that $g_1(t) = \frac{t \cdot g^\top v}{\|P_{\vec{V}}(g)\|\|v\|}$ which is a nonnegative increasing linear function of $t$ with $g_1(0) = 0$. Also notice that $g_2(t)$ is convex and nonnegative for $t \geq 0$, and $g_2(0) = 0$, whereby $g_2(t)$ is a monotonically increasing nonnegative convex function for $t \geq 0$. Therefore $g_2(t)/g_1(t)$ is monotonically increasing on $t > 0$, whereby $G(t) = g_1(t)/g_2(t)$ is monotonically decreasing on $t > 0$, which proves the lemma. □

*Proof of Theorem 5.5.* First notice from the definition of $\mu$ that

$$\mu \le \min\left\{R_1(\boldsymbol{e}^1), R_1(\boldsymbol{e}^2), \ldots, R_1(\boldsymbol{e}^{m_1}), R_2(\boldsymbol{f}^1;\bar\varepsilon), R_2(\boldsymbol{f}^2;\bar\varepsilon), \ldots, R_2(\boldsymbol{f}^{m_2};\bar\varepsilon)\right\} \ . \tag{5.17}$$

For $\varepsilon > 0$ let us consider the level set $S_\varepsilon$ and suppose that $S_\varepsilon \neq \emptyset$, and let $\mathcal{ES}_\varepsilon$ denote the extreme points of $S_\varepsilon$. Then we claim that

$$\begin{aligned}\mu = \inf_{u\in\mathcal{F}\setminus\mathcal{U}^\star} G(u) &= \inf_{\varepsilon>0}\left(\inf_{u\in S_\varepsilon} G(u)\right) = \lim_{\varepsilon\to 0}\left(\inf_{u\in S_\varepsilon} G(u)\right)\\ &= \lim_{\varepsilon\to 0}\left(\frac{\frac{\varepsilon}{\|P_{\vec V}(g)\|}}{\sup_{u\in S_\varepsilon}\operatorname{Dist}(u,\mathcal{U}^\star)}\right) = \lim_{\varepsilon\to 0}\left(\frac{\frac{\varepsilon}{\|P_{\vec V}(g)\|}}{\max_{v^i\in\mathcal{ES}_\varepsilon}\operatorname{Dist}(v^i,\mathcal{U}^\star)}\right) \ .\end{aligned}$$

Here the first equality is the definition of $\mu$, the second equality is a restatement of the first expression, and the third equality is due to the monotonicity property of the sharpness function on the level sets $S_\varepsilon$ from Proposition 5.6. To prove the fourth equality, observe that the numerator of $G(u)$ is $\operatorname{Dist}(u, V\cap H^\star)$ which equals the constant $\frac{\varepsilon}{\|P_{\vec V}(g)\|}$ for $u\in S_\varepsilon$, and the denominator of $G(u)$ is $\operatorname{Dist}(u,\mathcal{U}^\star)$. For the fifth equality, observe that $\operatorname{Dist}(\cdot,\mathcal{U}^\star)$ is convex in $u$ and bounded from above and below on $S_\varepsilon$, and so attains its maximum at an extreme point of $S_\varepsilon$.

Next notice that since $\mathcal{F}$ is a polyhedron and $S_\varepsilon$ is a level set of $\mathcal{F}$, there exists $\bar\varepsilon > 0$ such that for all $\varepsilon\in(0,\bar\varepsilon)$ the extreme points of $S_\varepsilon$ all lie in the edges of $\mathcal{F}$ emanating away from $\mathcal{U}^\star$, namely $\mathcal{M}\cap S_\varepsilon = \mathcal{ES}_\varepsilon$. Therefore using the definition of $\mathcal{M} = \mathcal{M}_1\cup\mathcal{M}_2$ we have:

$$\mu = \lim_{\varepsilon\to 0}\left(\frac{\frac{\varepsilon}{\|P_{\vec V}(g)\|}}{\max_{v^i\in\mathcal{ES}_\varepsilon}\operatorname{Dist}(v^i,\mathcal{U}^\star)}\right) = \min\left\{\min_{\boldsymbol{e}\in\mathcal{M}_1}\left(\inf_{u\in\boldsymbol{e}:g^\top u\le f^\star+\bar\varepsilon} G(u)\right),\ \min_{\boldsymbol{f}\in\mathcal{M}_2}\left(\inf_{u\in\boldsymbol{f}:g^\top u\le f^\star+\bar\varepsilon} G(u)\right)\right\} \ . \tag{5.18}$$

It follows from Proposition 5.7 that $G(u)$ is decreasing on any edge emanating away from $\mathcal{U}^\star$. Thus for $\boldsymbol{e} = [v^1, v^2]\in\mathcal{M}_1$ with $v^1\in\mathcal{U}^\star$ and $v^2\notin\mathcal{U}^\star$ we have $\inf_{u\in\boldsymbol{e}:g^\top u\le f^\star+\bar\varepsilon} G(u) \ge G(u^2) = R_1(\boldsymbol{e})$, and similarly for $\boldsymbol{f} = [v;r]\in\mathcal{M}_2$ with $v\in\mathcal{U}^\star$ and $r$ being an extreme ray of $\mathcal{F}$ we have $\inf_{u\in\boldsymbol{f}:g^\top u\le f^\star+\bar\varepsilon} G(u) = G(v+\bar\varepsilon r) = R_2(\boldsymbol{f},\bar\varepsilon)$. Substituting these inequalities back into (5.18) yields

$$\mu \ge \min\left\{R_1(\boldsymbol{e}^1), R_1(\boldsymbol{e}^2), \ldots, R_1(\boldsymbol{e}^{m_1}), R_2(\boldsymbol{f}^1;\bar\varepsilon), R_2(\boldsymbol{f}^2;\bar\varepsilon), \ldots, R_2(\boldsymbol{f}^{m_2};\bar\varepsilon)\right\} \ ,$$

which combined with (5.17) yields the proof. □

# 6 Numerical Experiments

Here we present results of numerical experiments designed to test how consistent our theoretical bounds are with computational practice, as well as to demonstrate the value of various heuristics on practical computation, based on our theoretical results. All computation was conducted on the MIT Engaging Cluster, and each experiment used a 2.4 GHz 14 Core CPU and 32G RAM, with CentOS version 7. All experiments were implemented in Julia 1.8.5.

## 6.1 Simple validation experiments

We conducted five simple experiments to test the extent to which the iteration bounds in Theorem 3.3 are "valid", by which we mean that the bounds are directionally consistent with computational practice on a specifically chosen family of test problems.

**Experiment 1: Sensitivity to the Hoffman constant of the KKT system.** The bounds in Theorem 3.3 are based essentially on three condition measures: LP sharpness, LimitingER, and (relative) distance to optima. This is in contrast to the analysis of Applegate et al. [3] whose bounds are mostly based on the Hoffman constant of the KKT system of the LP instance. To test the sensitivity of Algorithm 2 to the Hoffman constant of the KKT system, we created the following family of LP instances in standard form (1.2) with $m = 1$, $n = 3$, and data $(A^1, b^1, c^1)$ parameterized by $\gamma \in (0, 1]$ as follows: $A^1_\gamma := \left[\frac{\sin(\gamma)}{\sqrt{2}}, \cos(\gamma), \frac{\sin(\gamma)}{\sqrt{2}}\right]$, $b^1 := 1$, $c^1_\gamma := \left[\frac{\cos(\gamma)}{\sqrt{2}}, -\sin(\gamma), \frac{\cos(\gamma)}{\sqrt{2}}\right]^\top$. This family of problems was designed to have the following properties: $\|c\| = 1$, $Ac = 0$, $\|q\| = 1$, with uniform values of LP sharpness values and LimitingER for both the primal and dual problems, but with increasing values of the Hoffman constant of the KKT system as $\gamma \searrow 0$.

The first three columns of the first row of Figure 2 show the LP sharpness values (computed using the methodology in Section 5.2), LimitingER values (computed using the upper bound methodology in Section 4.2), and relative distances to optima for this simple family of problems, which are all constant over the range of $\gamma \in (0, 1]$. In the fourth column we report the actual iterations of Algorithm 2 (to obtain a solution whose Euclidean distance to the optimum is at most $10^{-10}$), and the iteration bound of Theorem 3.3 as well as the iteration bound of Applegate et al. [3]. (Since these two bounds are based on linear convergence rates, we report the constant outside of the logarithmic term for simplicity, and we computed the Hoffman constant for the KKT system using the algorithm and code from [37].) Notice that the bound of Applegate et al. [3] grows exponentially in $\ln(1/\gamma)$ while the actual number of iterations and the bound of Theorem 3.3 are constant over $\gamma \in (0, 1]$. This simple example validates the absence of the Hoffman constant from the bound in Theorem 3.3, and shows for this simple family that the actual number of iterations of Algorithm 2 is constant as suggested by Theorem 3.3.

**Experiment 2: Sensitivity to LimitingER.** This simple experiment is designed to test the sensitivity of Algorithm 2 to the LimitingER. Similar in approach to Experiment 1, we created the family: $A^2_\gamma := \left[\frac{\cos(\gamma)}{\sqrt{2}}, \sin(\gamma), \frac{\cos(\gamma)}{\sqrt{2}}\right]$, $b^2 = 1$, $c^2_\gamma := \left[\frac{\sin(\gamma)}{\sqrt{2}}, -\cos(\gamma), \frac{\sin(\gamma)}{\sqrt{2}}\right]^\top$, where again $\|c\| = 1$, $Ac = 0$, $\|q\| = 1$, with constant values of LP sharpness for $\gamma \in (0, 1]$, but now the LimitingER value increases exponentially as $\gamma \searrow 0$. The second row of Figure 2 shows our results. In this family of instances the LP sharpness values are constant even as the LimitingER grows. Notice that the relative distance to optima also grows similar to the LimitingER; this must occur since the relative distances to optima are lower-bounded by the LimitingER , see [45]. The fourth column shows that for the smaller values of $\gamma$ that the bound in Theorem 3.3 follows a similar pattern – including the slope in the log-log plot – as the actual iterations of Algorithm 2.

**Experiment 3: Sensitivity to LP Sharpness.** This simple experiment is designed to test the sensitivity of restarted-PDHG to LP sharpness. We created the family: $A^3_\gamma := \left[\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right]$, $b^3 = 1$, $c^3_\gamma := \cos(\gamma) \cdot \left[\frac{-1}{\sqrt{6}}, \frac{-1}{\sqrt{6}}, \frac{2}{\sqrt{6}}\right]^\top + \sin(\gamma) \cdot \left[\frac{-1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0\right]^\top$. Similar to the previous experiments we have $\|c\| = 1$, $Ac = 0$, $\|q\| = 1$, with constant values of the LimitingER and the relative distances to optima for $\gamma \in (0, 1]$, but now the primal LP sharpness $\mu_p$ value decreases exponentially as $\gamma \searrow 0$. The third row of Figure 2 shows our results. Similar in spirit to Experiment 2, the fourth column shows that for the smaller values of $\gamma$ that the bound in Theorem 3.3 follows a similar pattern – including the slope of the log-log plot – as the actual iterations of Algorithm 2.

**Experiment 4: Sensitivity to simultaneous changes in LP sharpness and LimitingER.** We created the family: $A^4_\gamma := \left[\sin(\gamma), \frac{\cos(\gamma)}{\sqrt{2}}, -\frac{\cos(\gamma)}{\sqrt{2}}\right]$, $b^4 = 1$, $c^4_\gamma := \left[0, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right]^\top$, in which the primal LP sharpness $\mu_p$ decreases exponentially and the dual LimitingER $\theta^\star(\mathcal{S}^\star)$ increases exponentially as $\gamma \searrow 0$, see the fourth row of Figure 2 for the computational values. Examining the fourth column of this row, we see the multiplicative effect of these two condition measures both on the theoretical bounds of Theorem 3.3 as well as a doubling of the slope of the log-log plot of actual iteration counts, which aligns well with the theoretical results.

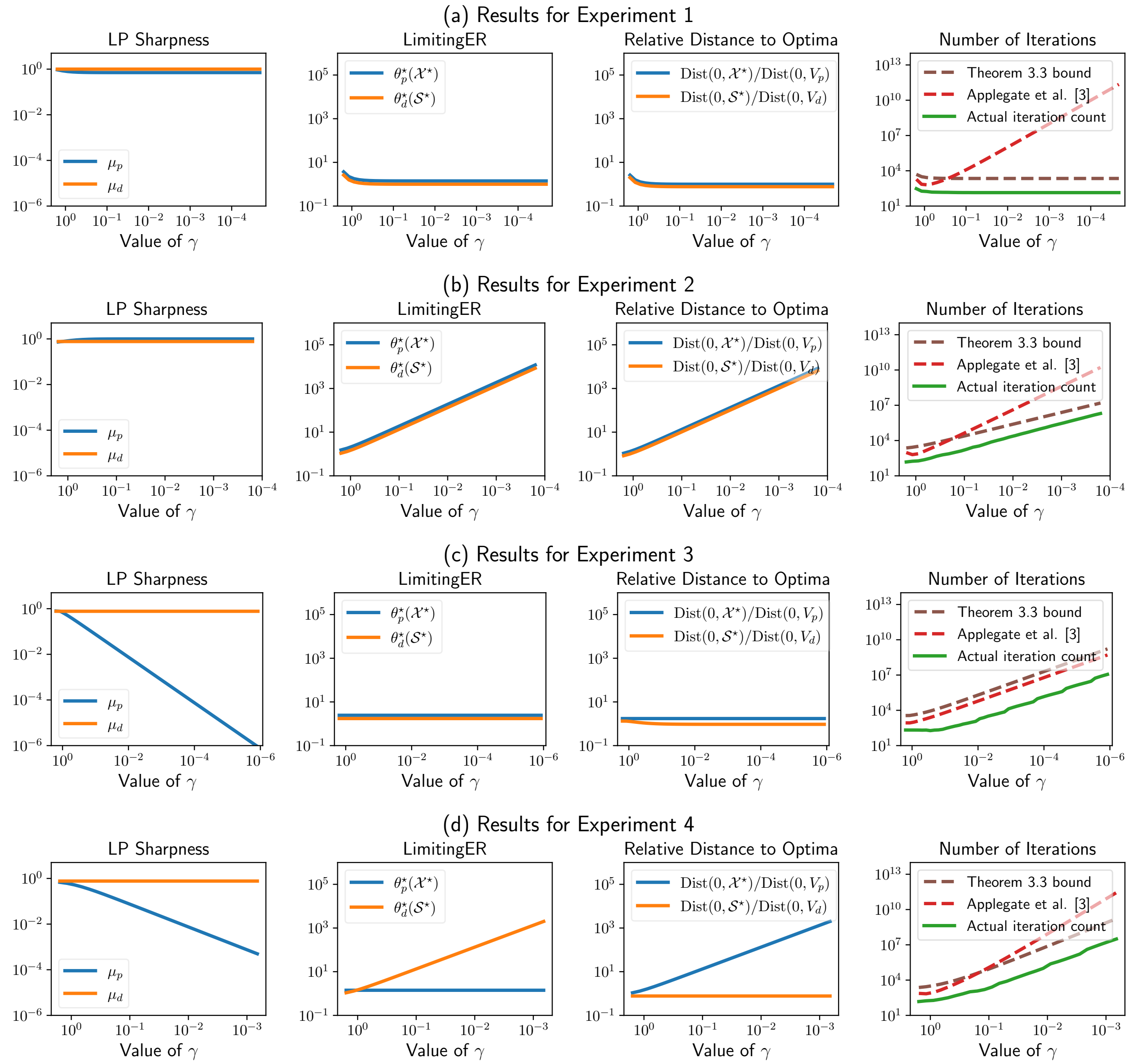


Figure 2: Values of LP sharpness, LimitingER, Relative Distance to Optima, theoretical iteration upper bound of Theorem 3.3, actual iteration count, and previous theoretical iteration upper bound of Applegate et al. [3] for the four simple validation experiments.

**Experiment 5: Effect of the step-size rule based on LP sharpness.** Remark 3.4 presented a step-size rule (3.7) based on knowledge of the LP sharpness measures $\mu_p$ and $\mu_d$ that leads to a structurally superior complexity bound for Algorithm 2. (However, this rule is impractical since the LP sharpness measures are neither known nor easily computable in practice.) In this experiment we test the utility of this rule using the simple family of LP instances $(A_\gamma^4,\, b^4,\, c_\gamma^4)$ described in Experiment 4, where for this simple family we know the LP sharpness measures. Figure 3 shows the theoretical upper bounds and the actual iteration numbers of Algorithm 2 with standard step-sizes (Theorem 3.3) and the step-sizes of Remark 3.4 on for this family of LP instances. The figure shows that this step-size rule reduces the actual number of iterations in line with the theory.

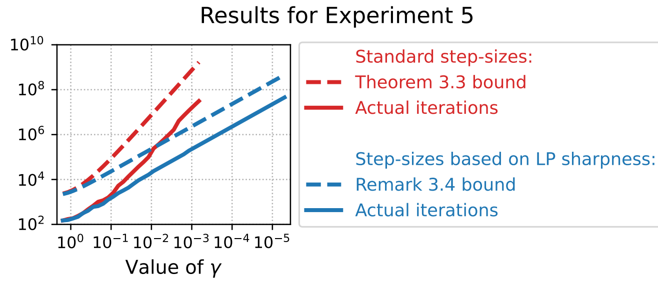


Figure 3: Theoretical upper bounds and the actual iteration numbers of Algorithm 2 with standard step-sizes (Theorem 3.3), and theoretical upper bounds and the actual iteration numbers of Algorithm 2 with the LP sharpness-based step-sizes of Remark 3.4, for the family of LP instances described by $(A^4_\gamma,\, b^4,\, c^4_\gamma)$ .

## 6.2 Computational evaluation of two theory-based heuristics on the MIPLIB 2017 dataset

In this subsection we introduce two heuristics that are inspired by our theoretical guarantees, and are designed to improve the practical performance of Algorithm 2. The first heuristic involves the choice of step-sizes $\tau$ and $\sigma$ for Algorithm 2. It was observed in [3] that even while keeping the product of the primal and dual step-sizes $\tau$ and $\sigma$ constant, that heuristically modifying the ratio $\tau/\sigma$ had the potential to improve the computational performance of restarted PDHG. Theoretical justification for that observation can be seen in the computational bounds for Algorithm 2 in Theorem 3.3 using different step-sizes $\tau$ for the primal and $\sigma$ for the dual in (3.7), and Remark 3.4 shows – at least in concept – how the complexity bound can be structurally improved by appropriately varying the ratio $\tau/\sigma$ while keeping the product constant, namely $\tau\sigma = 1/(4\lambda_{\max}^2)$. In the spirit of "learning from experience", our first heuristic is essentially an adaptation of the methodology in [3] to learn a reasonably good step-size ratio, and works as follows. We consider five possible choices of step-sizes, namely $(\tau,\sigma) = (40^\ell/2\lambda_{\max}, 40^{-\ell}/2\lambda_{\max})$ for $\ell = -1, -\frac{1}{2}, 0, \frac{1}{2}, 1$. For each of these step-size pairs we run Algorithm 2 for $5,000$ iterations from the same initial point $(x^{0,0}, y^{0,0}) = (0,0)$, and then choose which of the five step-sizes to use based on the smallest relative error $\mathcal{E}_r(x,y) := \frac{\|Ax^+ - b\|}{1+\|b\|} + \frac{\|(c-A^\top y)^-\|}{1+\|c\|} + \frac{|c^\top x^+ - b^\top y|}{1+|c^\top x^+|+|b^\top y|}$. The heuristic essentially spends $20,000$ iterations exploring/testing for a better step-size ratio. (We note that the relative error $\mathcal{E}_r(x,y)$ is upper-bounded by the distance to optima $\mathcal{E}_r d(x,s)$ (2.5), see Remark A.1.)

The second heuristic is also motivated by the computational bound in Theorem 3.3 where we observe in (3.8) that the bound grows at least linearly in the condition number $\kappa$ of the matrix $A$ (recall the definition of $\kappa$ in (2.8)). The heuristic is to compute and apply a (full-rank) row-preconditioner $D \in \mathbb{R}^{m\times m}$ to the equality constraints $Ax = b$ to yield the equivalent system $DAx = Db$ for which the condition number $\kappa' := \kappa(DA) := \frac{\lambda^+_{\max}(DA)}{\lambda^+_{\min}(DA}$ is reduced. Notice that for any such $D$ that the preconditioned LP instance

$$\min_{x\in\mathbb{R}^n} \; c^\top x \quad \text{s.t. } DAx = Db, \; x \geq 0 \tag{6.1}$$

and its dual problem have the identical duality-paired symmetric format (2.4) as the original LP instance; and so the LP sharpness, the LimitingER, and the relative distance to optima are unchanged by the preconditioner. Indeed the only quantity in the iteration bound (3.8) that is changed is the matrix condition number $\kappa(DA)$. In our second heuristic we work with the "complete" pre-conditioner $D := (AA^\top)^{-1/2}$, for which $\kappa' = \kappa(DA) = 1$, which requires one (potentially expensive) matrix factorization. Other first-order methods for LP, such as [25, 35], also compute and use a single matrix factorization throughout all iterations. (When the problem is very large and computing even one matrix factorization is not tractable, [1] proposed to use a diagonal preconditioner $D$, but there were no theoretical guarantees.)

We tested the usefulness of the two heuristics using the LP relaxations of the MIPLIB 2017 dataset [15], which is a collection of mixed-integer programs from real applications. We took the LP relaxations of the problems in the dataset and converted them to standard form so that Algorithm 2 can be directly applied. We ran Algorithm 2 to compare the following choice of heuristic strategies for step-sizes and preconditioners:

- **Simple Step-size**: this is the simple step-size rule originally used in the proofs in Applegate et al. [3], namely $\tau = \sigma = 1/(2\lambda_{\max})$,
- **Learned Step-size**: use an extra $20,000$ iterations to heuristically learn the best of five step-sizes as described above,
- **Preconditioner**: apply the preconditioner $D = (AA^\top)^{-1/2}$ as described above, and
- **Learned Step-size+Preconditioner**: apply both of the above heuristics.

Figure 4 illustrates the individual effects of the two heuristics on three representative problems, namely `nu120-pr9`, `n2seq36f` and `n3705`. The horizontal axis is the number of iterations and the vertical axis is the relative error $\mathcal{E}_r(x,y) := \frac{\|Ax^+ - b\|}{1+\|b\|} + \frac{\|(c - A^\top y)^-\|}{1+\|c\|} + \frac{|c^\top x^+ - b^\top y|}{1+|c^\top x^+|+|b^\top y|}$ computed using the original data of the LP instance for consistency. (The rather chaotic pattern of the early iterations of the Learned Step-size heuristic is due to the fact that the first $25,000$ iterations are used to test five different step-sizes.) For most of the LP instances in the MIPLIB 2017 we observed that the Learned Step-size heuristic enables much faster linear convergence, though `n3705` is an exception to this observation. We also observed that the preconditioner improves convergence significantly across all problems.

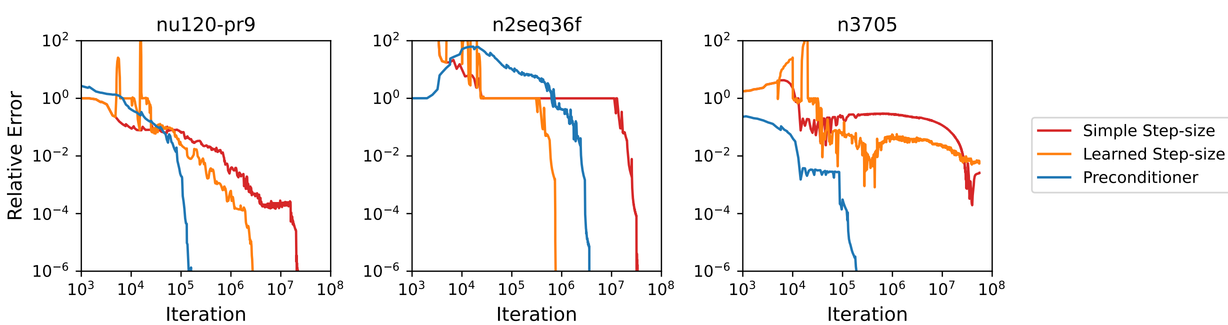


Figure 4: Performance of Algorithm 2 using two heuristic strategies, on problems `nu120-pr9`, `n2seq36f` and `n3705`.

Last of all, we tested all four combinations of heuristics on a large subset of the MIPLIB 2017 dataset, namely all LP relaxation problems in which $mn \leq 10^9$, of which there are 574 such problems in total. For this evaluation we consider an LP instance to be "solved" if Algorithm 2 computes a solution $(x, y)$ for which $\mathcal{E}_r(x, y) \leq 10^{-4}$. Figure 5 shows the fraction of solved problems (of the 574 instances) on the horizontal axis, and the maximum iterations (leftmost plot) and the maximum runtime (rightmost plot). Notice that these two techniques both help the restarted PDHG solve more problems in shorter time. Among the two heuristics, the preconditioner plays a prominent role in reducing the number of iterations, and also in reducing runtimes. Moreover, applying both heuristics is also valuable. Finally, it bears mentioning that if the problem is so large that the cost of working a matrix factorization is prohibitive, it is still possible to apply diagonal preconditioners to potentially improve the value of $\kappa$, see [1].

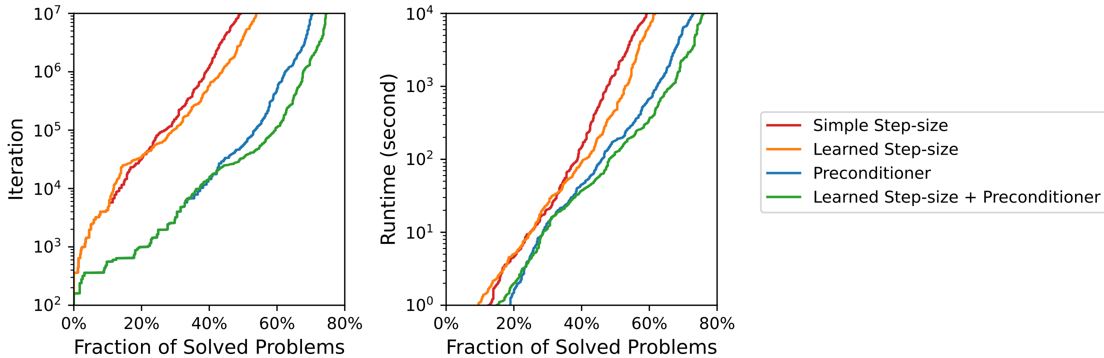


Figure 5: Performance of Algorithm 2 combined with heuristic strategies, on the 574 LP relaxation instances from the MIPLIB 2017 dataset.

The above experiments show that these two heuristics – which are motivated by our theoretical results – have clear potential to improve the practical performance of Algorithm 2, which also highlights the value of the theoretical understanding in the development of practical improvements in solution methods.

# A From the Distance to Optima to the Relative Error

Here we show that the relative error is upper-bounded by the distance to optima up to a scalar factor. For the pair $(x, y)$ and $s := c - A^\top y$, defining $\mathcal{E}_r(x, s) := \mathcal{E}_r(x, y)$, it follows that $\mathcal{E}_r(x, s) = \frac{\|Ax^+ - b\|}{1+\|b\|} + \frac{\|s^-\|}{1+\|c\|} + \frac{|c^\top x^+ - q^\top(c-s)|}{1+|c^\top x^+|+|q^\top(c-s)|}$, in which $q := A^\top(AA^\top)^\dagger b$.

**Remark A.1** (Relative error bounded by distance to optima)**.** *There exists a scalar constant $\bar{c}$ depending only on the data instance $(A, b, c)$, such that for any solution pair $(x, s)$, the relative error of $(x, s)$ is upper bounded by the distance to optima by a factor of $\bar{c}$, namely $\mathcal{E}_r(x, s) \leq \bar{c} \cdot \mathcal{E}_d(x, s)$ for any $(x, s)$. One such value of $\bar{c}$ is $\bar{c} = c_0 := \frac{2\|A\|}{1+\|b\|} + 2\|c\| + \|q\| + 1$.*

*Proof.* The error $\mathcal{E}_r(x, s)$ is comprised of three parts, namely $\frac{\|Ax^+ - b\|}{1+\|b\|}$, $\frac{\|s^-\|}{1+\|c\|}$ and $\frac{|c^\top x^+ - q^\top(c-s)|}{1+|c^\top x^+|+|q^\top(c-s)|}$. For any $x^\star \in \mathcal{X}^\star$ and $s^\star \in \mathcal{S}^\star$, because $\|x - x^\star\| \geq \|x - x^+\|$, we have

$$\frac{\|Ax^+ - b\|}{1 + \|b\|} \leq \frac{\|Ax - Ax^+\| + \|Ax - Ax^\star\|}{1 + \|b\|} \leq \frac{\|A\| \cdot (\|x - x^\star\| + \|x - x^+\|)}{1 + \|b\|} \leq \frac{2\|A\| \cdot \|x - x^\star\|}{1 + \|b\|}$$

and

$$\begin{aligned} &\frac{|c^\top x^+ - q^\top(c-s)|}{1 + |c^\top x^+| + |q^\top(c-s)|} \leq |c^\top x^+ - q^\top(c-s)| = |c^\top x^+ - c^\top x^\star - q^\top(c-s) + q^\top(c-s^\star)| \\ &\leq \|c\| \cdot (\|x - x^\star\| + \|x - x^+\|) + \|q\| \cdot \|s - s^\star\| \leq 2\|c\| \cdot \|x - x^\star\| + \|q\| \cdot \|s - s^\star\| \ . \end{aligned}$$

Similarly, because $\|s - s^+\| \leq \|s - s^\star\|$, we have $\frac{\|s^-\|}{1+\|c\|} = \frac{\|s-s^+\|}{1+\|c\|} \leq \frac{\|s-s^\star\|}{1+\|c\|} \leq \|s - s^\star\|$. Now let $x^\star := \arg\min_{\hat{x}\in\mathcal{X}^\star} \|x - \hat{x}\|$ and $s^\star := \arg\min_{\hat{s}\in\mathcal{S}^\star} \|s - \hat{s}\|$, then combining the above three inequalities implies that $\mathcal{E}_r(x, s) \leq \bar{c} \cdot \mathcal{E}_d(x, s)$. □

## B Proof of Proposition 2.8

We first prove the following elementary inequality:

**Proposition B.1.** *For any $y$ and $s = c - A^\top y$, it holds that* $\mathrm{Dist}(y, \mathcal{Y}^\star) \le \mathrm{Dist}(s, \mathcal{S}^\star) \cdot \frac{1}{\lambda_{\min}}$.

*Proof.* First observe that:

$$\mathrm{Dist}(s, \mathcal{S}^\star) = \mathrm{Dist}(c - A^\top y, \mathcal{S}^\star) = \mathrm{Dist}(c - A^\top y, c - A^\top(\mathcal{Y}^\star)) = \mathrm{Dist}(A^\top y, A^\top(\mathcal{Y}^\star)) = \mathrm{Dist}_{AA^\top}(y, \mathcal{Y}^\star) .$$

Let $AA^\top = PD^2P^\top$ denote the thin eigendecomposition of $AA^\top$, so that $P^\top P = I$ and $D$ is the diagonal matrix of positive singular values of $A$, whereby $D_{ii} \ge \min_j D_{jj} = \lambda_{\min}$ for each $i$. Now let $y^\star$ solve the shortest distance problem from $y$ to $\mathcal{Y}^\star$ in the norm $\|\cdot\|_{AA^\top}$, hence $y^* \in \mathcal{Y}^*$ and $\mathrm{Dist}_{AA^\top}(y, \mathcal{Y}^\star) = \|y - y^\star\|_{AA^\top}$, and let us write $y - y^\star = u + v$ where $u \in \mathrm{Im}(A)$ and $v \in \mathrm{Null}(A^\top)$. Then setting $\tilde{y} = y^\star + v$ and noting that $\tilde{y} \in \mathcal{Y}^\star$, we have:

$$\mathrm{Dist}_{AA^\top}(y, \mathcal{Y}^\star) \le \|y - \tilde{y}\|_{AA^\top} = \|u\|_{AA^\top} \ . \tag{B.1}$$

Next notice that since $u \in \mathrm{Im}(A) = \mathrm{Im}(AA^\top)$, there exists $\pi$ for which $u = AA^\top\pi$, and define $\lambda = D^2P^\top\pi$. It then follows that $u = P\lambda$, $\lambda = P^\top u$, and $\|u\| = \|\lambda\|$. We therefore have:

$$\begin{aligned} \mathrm{Dist}_{AA^\top}(y, \mathcal{Y}^\star)^2 &= (u+v)^\top AA^\top (u+v) \\ &= u^\top AA^\top u = \lambda^\top P^\top P D^2 P^\top P \lambda = \lambda^\top D^2 \lambda \ge \lambda_{\min}^2 \|\lambda\|^2 \ , \end{aligned} \tag{B.2}$$

and hence $\mathrm{Dist}(s, \mathcal{S}^\star) = \mathrm{Dist}_{AA^\top}(y, \mathcal{Y}^\star) \ge \lambda_{\min}\|\lambda\| = \lambda_{\min}\|u\| \ge \lambda_{\min}\,\mathrm{Dist}(y, \mathcal{Y}^\star)$, where the second inequality uses (B.1). Rearranging completes the proof. □

*Proof of Proposition 2.8.* We first prove (2.18). For a given $\alpha > 0$, the statement "$\|z\|_M \ge \alpha\|z\|_N$ for any $z$" is equivalent to

$$\|z\|_M^2 - \alpha^2\|z\|_N^2 = z^\top Q_\alpha z \ge 0 \ , \quad \text{where } Q_\alpha := \begin{pmatrix} \frac{1-\alpha^2}{\tau} I_n & -A^\top \\ -A & \frac{1-\alpha^2}{\sigma} I_m \end{pmatrix} \ ,$$

and hence $\|z\|_M \ge \alpha\|z\|_N$ for any $z$ if and only if $Q_\alpha \succeq 0$. A Schur complement argument then establishes that $Q_\alpha \succeq 0$ if and only if $(1-\alpha^2)^2/\sigma\tau \ge \lambda_{\max}^2$, which rearranges to $\alpha \le \sqrt{1 - \sqrt{\tau\sigma}\lambda_{\max}}$ (where the right-hand side is well defined due to (2.7)). This establishes the first inequality in (2.18). For the second inequality, note that the statement "$\|z\|_M \le \sqrt{2}\|z\|_N$ for any $z$" holds if and only $1/(\tau\sigma) \ge \lambda_{\max}^2$, which is satisfied due to (2.7)), completing the proof of (2.18).

Let us now prove (2.19). We have

$$\begin{aligned} \mathrm{Dist}_M(z, \mathcal{Z}^\star) &= \min_{\tilde{z} \in \mathcal{X}^\star \times \mathcal{Y}^\star} \|z - \tilde{z}\|_M \le \sqrt{2} \cdot \min_{\tilde{z} \in \mathcal{X}^\star \times \mathcal{Y}^\star} \|z - \tilde{z}\|_N \\ &\le \frac{\sqrt{2}}{\sqrt{\tau}} \mathrm{Dist}(x, \mathcal{X}^\star) + \frac{\sqrt{2}}{\sqrt{\sigma}} \mathrm{Dist}(y, \mathcal{Y}^\star) \le \frac{\sqrt{2}}{\sqrt{\tau}} \mathrm{Dist}(x, \mathcal{X}^\star) + \frac{\sqrt{2}}{\sqrt{\sigma}\lambda_{\min}} \mathrm{Dist}(s, \mathcal{S}^\star) \ , \end{aligned} \tag{B.3}$$

where the first inequality utilities (2.18) and the third inequality uses Proposition B.1.

We next prove (2.20). Again using (2.18), we have $\mathrm{Dist}_M(z, \mathcal{Z}^\star) \ge \sqrt{1 - \sqrt{\tau\sigma}\lambda_{\max}} \cdot \mathrm{Dist}_N(z, \mathcal{Z}^\star)$, and it also holds that $\mathrm{Dist}_N(z, \mathcal{Z}^\star) \ge \max \cdot \left\{ \frac{1}{\sqrt{\tau}} \mathrm{Dist}(x, \mathcal{X}^\star), \frac{1}{\sqrt{\sigma}} \mathrm{Dist}(y, \mathcal{Y}^\star) \right\}$. Furthermore, $\mathrm{Dist}(s, \mathcal{S}^\star) = \mathrm{Dist}(A^\top y, A^\top(\mathcal{Y}^\star)) \le \|A\| \cdot \mathrm{Dist}(y, \mathcal{Y}^\star)$ and $\|A\| = \lambda_{\max}$, which combined with the above two inequalities yields the proof of (2.20).

□

### Acknowledgements

The authors are grateful to Kim-Chuan Toh, Haihao Lu, Jinwen Yang, and Guanghao Ye for discussions and input concerning the ideas in this paper and related research ideas.


## References


[1] D. Applegate, M. Díaz, O. Hinder, H. Lu, M. Lubin, B. O'Donoghue, and W. Schudy. Practical large-scale linear programming using primal-dual hybrid gradient. In *Advances in Neural Information Processing Systems*, volume 34, pages 20243–20257, 2021.

[2] D. Applegate, M. Díaz, H. Lu, and M. Lubin. Infeasibility detection with primal-dual hybrid gradient for large-scale linear programming. *arXiv preprint arXiv:2102.04592*, 2021.

[3] D. Applegate, O. Hinder, H. Lu, and M. Lubin. Faster first-order primal-dual methods for linear programming using restarts and sharpness. *Mathematical Programming*, 201(1-2):133–184, 2023.

[4] D. Avis and K. Fukuda. A pivoting algorithm for convex hulls and vertex enumeration of arrangements and polyhedra. *Discrete & Computational Geometry*, 8(3):295–313, 1992.

[5] C. Bergthaller and I. Singer. The distance to a polyhedron. *Linear Algebra and its Applications*, 169:111–129, 1992.

[6] D. Bertsimas and J. N. Tsitsiklis. *Introduction to Linear Optimization.* Athena Scientific, Belmont, MA, 1997.

[7] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40:120–145, 2011.

[8] A. Chambolle and T. Pock. On the ergodic convergence rates of a first-order primal–dual algorithm. *Mathematical Programming*, 159(1-2):253–287, 2016.

[9] G. Dantzig. *Linear Programming and Extensions.* Princeton University Press, 1963.

[10] Q. Deng, Q. Feng, W. Gao, D. Ge, B. Jiang, Y. Jiang, J. Liu, T. Liu, C. Xue, Y. Ye, and C. Zhang. New developments of ADMM-based interior point methods for linear programming and conic programming. *arXiv preprint arXiv:2209.01793*, 2022.

[11] M. Epelman and R. M. Freund. Condition number complexity of an elementary algorithm for computing a reliable solution of a conic linear system. *Mathematical Programming*, 88(3):451–485, 2000.

[12] E. Esser, X. Zhang, and T. F. Chan. A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science. *SIAM Journal on Imaging Sciences*, 3(4):1015–1046, 2010.

[13] R. Freund and J. Vera. Some characterizations and properties of the "distance to ill-posedness" and the condition measure of a conic linear system. *Mathematical Programming*, 86:225–260, 1999.

[14] R. M. Freund and J. Vera. Equivalence of convex problem geometry and computational complexity in the separation oracle model. *Mathematics of Operations Research*, 34, 2009.

[15] A. Gleixner, G. Hendel, G. Gamrath, T. Achterberg, M. Bastubbe, T. Berthold, P. Christophel, K. Jarck, T. Koch, J. Linderoth, M. Lübbecke, H. D. Mittelmann, D. Ozyurt, T. K. Ralphs, D. Salvagnin, and Y. Shinano. MIPLIB 2017: data-driven compilation of the 6th mixed-integer programming library. *Mathematical Programming Computation*, 13(3):443–490, 2021.

[16] J.-L. Goffin. The relaxation method for solving systems of linear inequalities. *Mathematics of Operations Research*, 5(3):388–414, 1980.

[17] B. Grünbaum. *Convex Polytopes*. Wiley, New York, 1967.

[18] Gurobi Optimization, LLC. Gurobi optimizer reference manual, 2023.

[19] A. J. Hoffman. On approximate solutions of systems of linear inequalities. *Journal of Research of the National Bureau of Standards*, 49(4), 1952.

[20] H. Hu. Perturbation analysis of global error bounds for systems of linear inequalities. *Mathematical Programming*, 88:277–284, 2000.

[21] L. Khachiyan, E. Boros, K. Borys, V. Gurvich, and K. Elbassioni. Generating all vertices of a polyhedron is hard. *Discrete & Computational Geometry*, 39:174–190, 2008.

[22] A. S. Lewis and J.-S. Pang. Error bounds for convex inequality systems. In *Generalized Convexity, Generalized Monotonicity: Recent Results*, pages 75–110. Springer, 1998.

[23] X. Li, D. Sun, and K.-C. Toh. An asymptotically superlinearly convergent semismooth Newton augmented Lagrangian method for linear programming. *SIAM Journal on Optimization*, 30(3):2410–2440, 2020.

[24] J. Liang, J. Fadili, and G. Peyré. Convergence rates with inexact non-expansive operators. *Mathematical Programming*, 159:403–434, 2016.

[25] T. Lin, S. Ma, Y. Ye, and S. Zhang. An ADMM-based interior-point method for large-scale linear programming. *Optimization Methods and Software*, 36(2-3):389–424, 2021.

[26] H. Lu and J. Yang. Nearly optimal linear convergence of stochastic primal-dual methods for linear programming. *arXiv preprint arXiv:2111.05530*, 2021.

[27] H. Lu and J. Yang. On the infimal sub-differential size of primal-dual hybrid gradient method. *arXiv preprint arXiv:2206.12061*, 2022.

[28] H. Lu and J. Yang. cuPDLP. jl: A GPU implementation of restarted primal-dual hybrid gradient for linear programming in Julia. *arXiv preprint arXiv:2311.12180*, 2023.

[29] H. Lu and J. Yang. On the geometry and refined rate of primal-dual hybrid gradient for linear programming. *arXiv preprint arXiv:2307.03664*, 2023.

[30] Z.-Q. Luo and P. Tseng. Perturbation analysis of a condition number for linear systems. *SIAM Journal on Matrix Analysis and Applications*, 15(2):636–660, 1994.

[31] O. L. Mangasarian. A condition number for linear inequalities and linear programs. In *Proceedings of 6th Symposium über Operations Research, Universität Augsburg*, pages 3–15, 1981.

[32] V. Mirrokni. Google Research, 2022 & beyond: Algorithmic advances. 2023. https://ai.googleblog.com/2023/02/google-research-2022-beyond-algorithmic.html.

[33] Y. Nesterov and A. Nemirovskii. *Interior-point polynomial algorithms in convex programming*. SIAM, 1994.

[34] B. O'Donoghue. Operator splitting for a homogeneous embedding of the linear complementarity problem. *SIAM Journal on Optimization*, 31(3):1999–2023, 2021.

[35] B. O'donoghue, E. Chu, N. Parikh, and S. Boyd. Conic optimization via operator splitting and homogeneous self-dual embedding. *Journal of Optimization Theory and Applications*, 169:1042–1068, 2016.

[36] J.-S. Pang. Error bounds in mathematical programming. *Mathematical Programming*, 79(1-3):299–332, 1997.

[37] J. Pena, J. Vera, and L. Zuluaga. An algorithm to compute the Hoffman constant of a system of linear constraints. *arXiv preprint arXiv:1804.08418*, 2018.

[38] T. Pock, D. Cremers, H. Bischof, and A. Chambolle. An algorithm for minimizing the Mumford–Shah functional. In *IEEE 12th International Conference on Computer Vision*, pages 1133–1140. IEEE, 2009.

[39] B. T. Polyak. Sharp minima. In *Proceedings of the IIASA Workshop on Generalized Lagrangians and Their Applications, Laxenburg, Austria. Institute of Control Sciences Lecture Notes, Moscow*, 1979.

[40] J. Renegar. Some perturbation theory for linear programming. *Mathematical Programming*, 65(1-3):73–91, 1994.

[41] E. K. Ryu and W. Yin. *Large-Scale Convex Optimization: Algorithms & Analyses via Monotone Operators*. Cambridge University Press, 2022.

[42] G. Sonnevend. An 'analytic' center for polyhedrons and new classes of global algorithms for linear (smooth, convex) optimization. Technical report, Department of Numerical Analysis, Institute of Mathematics, Eötvös University, 1088, Budapest, Muzeum Körut 6-8, 1985. Preprint.

[43] M. J. Todd and Y. Ye. A centered projective algorithm for linear programming. *Mathematics of Operations Research*, 15(3):508–529, 1990.

[44] S. J. Wright. *Primal-Dual Interior-Point Methods*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, 1997.

[45] Z. Xiong and R. M. Freund. On the relation between LP sharpness and limiting error ratio and its complexity implications for restarted PDHG. *arXiv preprint arXiv:2312.13773*, 2023.

[46] T. Yang and Q. Lin. RSG: Beating subgradient method without smoothness and strong convexity. *Journal of Machine Learning Research*, 19(1):236–268, 2018.