On the Relation Between LP Sharpness and Limiting Error Ratio and Complexity Implications for Restarted PDHG

Zikai Xiong* Robert M. Freund†

December 2023

Abstract

There has been a recent surge in development of first-order methods (FOMs) for solving huge-scale linear programming (LP) problems. The attractiveness of FOMs for LP stems in part from the fact that they avoid costly matrix factorization computation. However, the efficiency of FOMs is significantly influenced – both in theory and in practice – by certain instance-specific LP condition measures. Xiong and Freund recently showed that the performance of the restarted primal-dual hybrid gradient method (PDHG) is predominantly determined by two specific condition measures: LP sharpness and Limiting Error Ratio. In this paper we examine the relationship between these two measures, particularly in the case when the optimal solution is unique (which is generic – at least in theory), and we present an upper bound on the Limiting Error Ratio involving the reciprocal of the LP sharpness. This shows that in LP instances where there is a dual nondegenerate optimal solution, the computational complexity of restarted PDHG can be characterized solely in terms of LP sharpness and the distance to optimal solutions, and simplifies the theoretical complexity upper bound of restarted PDHG for these instances.

1 Introduction

In this work, we study linear programming (LP) problems in the following form:

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & c^\top x \\ \text{s.t.} \quad & x \in V_p,\ x \geq 0 \end{aligned} \tag{1.1}$$

where V_p is an affine subspace in $\mathbb{R}^n$. For common standard-form problems we have $V_p = \{x \in \mathbb{R}^n : Ax = b\}$ for a given matrix $A \in \mathbb{R}^{m\times n}$ and a right-hand side vector $b \in \mathbb{R}^m$ [2]. Without loss of generality we assume that $c \in \vec{V}_p$ where $\vec{V}_p$ is the linear subspace associated with V_p. This is without loss of generality since we can replace c with its projection onto $\vec{V}_p$ and maintain the same optimal solution set and optimal objective function value. We therefore include this condition in our set of assumptions about (1.1) as follows:

Assumption 1. *The LP problem* (1.1) *has an optimal solution, the objective vector c is in the linear subspace $\vec{V}$, and non-optimal feasible solutions exist for* (1.1) *and its dual problem.*

*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.

Under Assumption 1, the primal and dual problems can be written in the following symmetric format:

$$
\begin{aligned}
&\text{(P)} \qquad \min_{x\in\mathbb{R}^n}\ c^\top x && \text{(D)} \qquad \max_{s\in\mathbb{R}^n}\ -q^\top s \\
&\qquad\quad \text{s.t. } x \in \mathcal{F}_p := V_p \cap \mathbb{R}^n_+ && \qquad\quad \text{s.t. } s \in \mathcal{F}_d := V_d \cap \mathbb{R}^n_+ \\
&\qquad\qquad\qquad V_p := q + \vec{V}_p && \qquad\qquad\qquad V_d := c + \vec{V}_d
\end{aligned}
\tag{PD}
$$

where $\vec{V}_d$ is the linear subspace associated with the affine subspace V_d, and $\vec{V}_p$ and $\vec{V}_d$ are orthogonal complements, namely,

$$
\mathbb{R}^n = \vec{V}_p + \vec{V}_d \quad \text{and} \quad \vec{V}_p \perp \vec{V}_d \ , \tag{1.2}
$$

and

$$
c \in \vec{V}_p \text{ and } c = \arg\min_{c_0\in V_d} \|c_0\|; \quad \text{also} \quad q \in \vec{V}_d \text{ and } q = \arg\min_{q_0\in V_p} \|q\| \ . \tag{1.3}
$$

This symmetric reformulation was first proposed in [10]. We will denote the primal and dual optimal solutions of (PD) as $\mathcal{X}^\star$ and $\mathcal{S}^\star$, respectively. The duality gap is

$$
\mathrm{Gap}(x,s) := c^\top x + q^\top s \ ,
$$

and the primal-dual optimal solution set is

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

Based on the symmetric primal-dual formulation (PD), [12] introduce two condition measures: LP sharpness and Limiting Error Ratio, and provide new computational guarantees for the restarted Primal-Dual Hybrid Gradient method (PDHG) involving these condition measures. Here we address the question of the relationship between LP sharpness and Limiting Error Ratioand whether the overall complexity proven in [12] can be simplified by replacing one these condition measures with an upper bound involving the other condition measure. We will demonstrate the following results for the case when the optimal solution of (1.1) is unique:

1. The primal Limiting Error Ratio is upper-bounded by the reciprocal of the primal LP sharpness multiplied by the relative distance to the dual optimal solutions. Similarly, the dual Limiting Error Ratio is upper-bounded by the reciprocal of the dual LP sharpness multiplied by the relative distance to the primal optimal solutions. However, we show that the reciprocal of the LP sharpness is not upper-bounded by the Limiting Error Ratio; this is shown by constructing a specific class of LP instances.

2. The overall worst-case complexity of restarted PDHG can be expressed using only the LP sharpness and the relative distance to optimal solutions. Similar to the results in [12], the choice of the primal and dual step-sizes in PDHG can affect occurrence of certain cross-terms in the complexity bounds.

We note that various condition numbers (measures) have been introduced for algorithms for LP with the aim of studying/explaining the theoretical and/or practical performance of algorithms for LP, including $\kappa(A)$ (the ratio of largest to smallest singular value of A [3]), the Hoffman constant [6], $\chi(A)$ and $\bar{\chi}(A)$ [9, 8], and Renegar's distance to ill-posedness [7]. There are also a variety of geometry-focused condition numbers involving characteristics of level-sets [5], symmetry measures of convex bodies [1, 4], etc. For the use of these and other condition measures in the context of LP see [11, 9, 4] among others.

Notation. Throughout this paper, unless explicitly stated otherwise, $\|\cdot\|$ denotes the Euclidean norm, and we use $\|\cdot\|_1$ to denote the ℓ_1 norm. For any $x \in \mathbb{R}^n$ and $\mathcal{X} \subset \mathbb{R}^n$, we denote the Euclidean distance from x to the set $\mathcal{X}$ as $\mathrm{Dist}(x, \mathcal{X}) := \min_{\hat{x} \in \mathcal{X}} \|x - \hat{x}\|$. The diameter of $\mathcal{X}$ is denoted as $\mathrm{Diam}(\mathcal{X}) := \max_{x, \hat{x} \in \mathcal{X}} \|x - \hat{x}\|$. For any matrix A, $\sigma^+_{\max}(A)$ and $\sigma^+_{\min}(A)$ denote the largest and smallest non-zero singular values of A, respectively. For an affine subspace V, let $\vec{V}$ denote the associated linear subspace of V, whereby $V = \vec{V} + v$ holds for every $v \in V$. We use $\mathbb{R}^n_+$ and $\mathbb{R}^n_{++}$ to denote the nonnegative orthant and strictly positive orthant in $\mathbb{R}^n$, respectively. For a linear subspace $\vec{V} \subset \mathbb{R}^n$, $\vec{V}^\perp$ denotes the orthogonal complement of $\vec{V}$. We use e to denote the all-ones vector in $\mathbb{R}^n$, namely $e = (1, \dots, 1)^\top$. We use $\|\cdot\|_1$ to denote the ℓ_1 norm.

Organization. This rest of this paper is structured as follows. In Section 2 we recall the definitions of the two condition measures introduced in [12]. In Section 3 we present our main result about the relationship between these two condition measures. In Section 4 we present an simplified version of the worst-case complexity analysis for restarted PDHG in [12] using the results from Section 3.

2 Two Geometry-based Condition Measures for LP

To study the computational guarantees for the restarted Primal-Dual Hybrid Gradient method (PDHG) , [12] has introduced two condition measures for LP problems (1.1) that together play an important role in the performance of restarted PDHG both in theory and in practice. The two condition measures under consideration are the Limiting Error Ratio and LP sharpness. The first condition measure, Limiting Error Ratio, is defined as follows, using the primal problem (1.1) as an illustrative case:

Definition 2.1 (Limiting Error Ratio, Definition 3.1 of [12]). *For any $x \in V_p \setminus \mathcal{F}_p$, the error ratio of $\mathcal{F}_p$ at x is defined as:*

$$\theta(x) := \frac{\mathrm{Dist}(x, \mathcal{F}_p)}{\mathrm{Dist}(x, \mathbb{R}^n_+)} \ , \tag{2.1}$$

and for any $x \in \mathcal{F}_p$ we define $\theta(x) := 1$. The Limiting Error Ratio is then defined as:

$$\theta^\star_p := \lim_{\epsilon \to 0} \left(\sup_{x \in V_p, \mathrm{Dist}(x, \mathcal{X}^\star) \le \epsilon} \theta(x) \right) . \tag{2.2}$$

And of course we can similarly define Limiting Error Ratio for the dual problem in (PD). We will use $\theta^\star_p$ and $\theta^\star_d$ to denote the Limiting Error Ratio for the primal and dual problems in (PD).

The second condition measure, LP sharpness, is defined as follows, where again we use the primal problem (1.1) as the illustrative case:

Definition 2.2 (LP sharpness, Definition 3.2 of [12]). *Let $f^\star$ denote the optimal objective function value of the problem (1.1), and define the hyperplane of the optimal objective value to be $H^\star := \{x \in \mathbb{R}^n : c^\top x = f^\star\}$. The LP sharpness is defined as follows:*

$$\mu := \inf_{x \in \mathcal{F}_p \setminus \mathcal{X}^\star} \frac{\mathrm{Dist}(x, V_p \cap H^\star)}{\mathrm{Dist}(x, \mathcal{X}^\star)} \ . \tag{2.3}$$

Again we can similarly define LP sharpness for the dual problem in (PD). For simplicity of notation we will use μ_p and μ_d to denote the LP sharpness of the primal and dual problems in (PD). Under Assumption 1 the numerator of (2.3) has a closed form:

$$\mathrm{Dist}(x, V_p \cap H^\star) = \frac{c^\top x - f^\star}{\|c\|} = \frac{c^\top x - f^\star}{\mathrm{Dist}(0, V_d)} \ ,$$

where the second equality uses (1.3). With this closed form, the following remark presents an alternative expression for the LP sharpness.

Remark 2.1. *Suppose that Assumption 1 holds. Then*

$$\mu_p := \inf_{x\in\mathcal{F}_p\setminus\mathcal{X}^\star} \frac{c^\top x - f^\star}{\|c\|\cdot \mathrm{Dist}(x,\mathcal{X}^\star)} = \frac{1}{\mathrm{Dist}(0,V_d)}\cdot \inf_{x\in\mathcal{F}_p\setminus\mathcal{X}^\star} \frac{c^\top x - f^\star}{\mathrm{Dist}(x,\mathcal{X}^\star)} \ , \tag{2.4}$$

and

$$\mu_d := \inf_{s\in\mathcal{F}_d\setminus\mathcal{S}^\star} \frac{q^\top s + f^\star}{\|q\|\cdot \mathrm{Dist}(s,\mathcal{S}^\star)} = \frac{1}{\mathrm{Dist}(0,V_p)}\cdot \inf_{s\in\mathcal{F}_d\setminus\mathcal{S}^\star} \frac{q^\top s + f^\star}{\mathrm{Dist}(s,\mathcal{S}^\star)} \ . \tag{2.5}$$

[12] uses the two measures Limiting Error Ratio and LP sharpness to establish an overall complexity bound for restarted PDHG for LP problems (1.1). In the next section we will demonstrate that Limiting Error Ratio is upper-bounded by the reciprocal of LP sharpness under the assumption that (1.1) has a unique optimal solution.

3 Relation Between Limiting Error Ratio and LP Sharpness

In this section we present our main result that establishes an upper bound on Limiting Error Ratio using the reciprocal of LP sharpness when the LP problem has a unique optimal solution.

Theorem 3.1. *For the primal LP problem (1.1), suppose that Assumption 1 holds and that the optimal solution is unique. Then the following inequality holds:*

$$\theta_p^\star \le \frac{1}{\mu_p}\cdot\sqrt{n}\cdot\left(\frac{\mathrm{Dist}(0,\mathcal{S}^\star)}{\mathrm{Dist}(0,V_d)} + \frac{\mathrm{Diam}(\mathcal{S}^\star)}{\mathrm{Dist}(0,V_d)}\right) \ . \tag{3.1}$$

Examining the two right-most fractions in the right-hand side of (3.1) we can interpret $\frac{\mathrm{Dist}(0,\mathcal{S}^\star)}{\mathrm{Dist}(0,V_d)}$ as the relative distance (from 0) to the dual optima set, and we can interpret $\frac{\mathrm{Diam}(\mathcal{S}^\star)}{\mathrm{Dist}(0,V_d)}$ as the relative diameter of the dual optimal set, because both of these quantities are (re-)scaled by the distance to the corresponding affine subspace V_d. Also observe that the sum of the distance to optima and the diameter is both lower-bounded and upper-bounded by the maximum-norm dual optimal solution, namely

$$\max_{s^\star\in\mathcal{S}}\|s^\star\| \le \mathrm{Dist}(0,\mathcal{S}^\star) + \mathrm{Diam}(\mathcal{S}^\star) \ ,$$

and also

$$\max_{s^\star\in\mathcal{S}}\|s^\star\| \ge \mathrm{Dist}(0,\mathcal{S}^\star)$$

and

$$\mathrm{Diam}(\mathcal{S}^\star) = \max_{u,v\in\mathcal{S}^\star}\|u-v\| \le 2\cdot\max_{s^\star\in\mathcal{S}}\|s^\star\| \ ,$$

which together imply that

$$\max_{s^\star\in\mathcal{S}}\|s^\star\| \le \mathrm{Dist}(0,\mathcal{S}^\star) + \mathrm{Diam}(\mathcal{S}^\star) \le 3\max_{s^\star\in\mathcal{S}}\|s^\star\| \ .$$

Overall, the above theorem guarantees that when the primal problem has a unique optimal solution, the value of the primal Limiting Error Ratio $\theta_p^\star$ cannot be large, provided that the reciprocal of the primal sharpness μ_p and the (relative) maximum norm of dual optimal solutions are both small. And of course a corresponding result also holds for the dual problem.

Corollary 3.2. *For the dual problem in* (PD)*, suppose that Assumption 1 holds and that the optimal dual solution is unique. Then the following inequality holds:*

$$\theta_d^\star \le \frac{1}{\mu_d} \cdot \sqrt{n} \cdot \left(\frac{\mathrm{Dist}(0, \mathcal{X}^\star)}{\mathrm{Dist}(0, V_p)} + \frac{\mathrm{Diam}(\mathcal{X}^\star)}{\mathrm{Dist}(0, V_p)} \right) \ . \tag{3.2}$$

Towards the proof of Theorem 3.1 we first review two useful results from [12] The first result is Proposition 3.1 in [12] which we re-state as follows:

Proposition 3.3 (Proposition 3.1 in [12])**.** *Let $x^\star \in \mathcal{X}^\star$ and suppose that $\mathcal{X}^\star \subset \{x : \|x - x^\star\| \le R\}$ for some $R > 0$. Then it holds that $\theta_p^\star \le B^\star$, where $B^\star$ is defined as follows:*

$$\begin{aligned} B^\star := \inf_{r>0, \bar{x}\in\mathbb{R}^n} \quad & \frac{R + \|\bar{x} - x^\star\|}{r} \\ s.t. \quad & \bar{x} \in V_p, \ \bar{x} \ge r \cdot e \ . \end{aligned} \tag{3.3}$$

This proposition states that the Limiting Error Ratio cannot be excessively large when the LP problem has a strictly feasible solution that is neither too large nor too close to the boundary of $\mathbb{R}^n_+$. The second result is Theorem 2.1 of [5], which presents a geometric relationship between the primal objective function level sets and the dual objective function level sets. To introduce this result we define two quantities:

$$\begin{aligned} r_\delta := \max_x \quad & \min_i \{x_i\} \\ \text{s. t.} \quad & x \in \mathcal{F}_p \\ & c^\top x \le f^\star + \delta \end{aligned} \tag{3.4}$$

for any $\delta \ge 0$, and

$$\begin{aligned} R_\varepsilon := \max_s \quad & \|s\|_1 \\ \text{s. t.} \quad & s \in \mathcal{F}_d \\ & -q^\top s \ge f^\star - \varepsilon \end{aligned} \tag{3.5}$$

for any $\varepsilon \ge 0$. The quantity r_δ is the positivity of the most positive x in the primal objective function δ-optimal level set; or equivalently as the distance to the boundary of the nonnegative orthant of point in the δ-optimal level set that is farthest from the boundary. The quantity R_ε can be interpreted as the norm of the maximum-norm point s in the dual ε-optimal level set. Using r_δ in (3.4) and R_ε in (3.5), we re-state Theorem 2.1 of [5] as follows:

Lemma 3.4 (Theorem 2.1 in [5])**.** *Suppose that the optimal objective value $f^\star$ of* (1.1) *is finite. If R_ε is positive and finite, then*

$$\min\{\varepsilon, \delta\} \le R_\varepsilon \cdot r_\delta \le \varepsilon + \delta \ . \tag{3.6}$$

Otherwise, $R_\varepsilon = 0$ if and only if $r_\delta = +\infty$, and $R_\varepsilon = +\infty$ if and only if $r_\delta = 0$.

This result states, for instance, that the quantities r_δ and R_ε are inversely related to within a factor of 2 in the case when $\delta = \varepsilon$. We will use this result in our proof of Theorem 3.1, which we now provide.

Proof of Theorem 3.1. Note that the optimal primal solution is unique, so we will denote this optimal solution as $x^\star$. Then according to Proposition 3.3, it holds for $\theta_p^\star$ that

$$\theta_p^\star \le \inf_{x_{int}\in(\mathcal{F}_p)_{++}} \frac{\|x^\star - x_{int}\|}{\min_i |(x_{int})_i|} \ . \tag{3.7}$$

For any $\delta > 0$, let $\bar{x}$ be

$$\begin{aligned} \bar{x} \in \arg\max_{x} \quad & \min_{i}\{x_i\} \\ \text{s.t.} \quad & x \in \mathcal{F}_p \\ & c^\top x \le f^\star + \delta \end{aligned} \tag{3.8}$$

and then because of (3.7) it holds that

$$\theta_p^\star \le \frac{\|x^\star - \bar{x}\|}{r_\delta} \ . \tag{3.9}$$

Now, due to Lemma 3.4,

$$\frac{1}{r_\delta} \le \frac{R_\delta}{\delta} \ . \tag{3.10}$$

Note that for any $p \in \mathbb{R}^n$, there exists the relation between norm $\|\cdot\|_1$ and the Euclidean norm $\|\cdot\|$ that $\|p\|_1 \le \sqrt{n}\|p\|$, so

$$\begin{aligned} R_\delta \ = \ \max_{s} \quad & \|s\|_1 & \le \ \max_{s} \quad & \sqrt{n}\cdot\|s\| \\ \text{s.t.} \quad & s \in \mathcal{F}_d & \text{s.t.} \quad & s \in \mathcal{F}_d \\ & -q^\top s \ge f^\star - \varepsilon & & -q^\top s \ge f^\star - \varepsilon \end{aligned}$$

which further implies

$$\begin{aligned} R_\delta &\le \sqrt{n}\cdot\left(\max_{s\in\mathcal{F}_d,\ -q^\top s \ge f^\star-\delta} \operatorname{Dist}(s,\mathcal{S}^\star) + \operatorname{Dist}(0,\mathcal{S}^\star) + \operatorname{Diam}(\mathcal{S}^\star)\right), \\ &\le \sqrt{n}\cdot\left(\frac{\delta}{\|q\|\cdot\mu_d} + \operatorname{Dist}(0,\mathcal{S}^\star) + \operatorname{Diam}(\mathcal{S}^\star)\right) \ . \end{aligned}$$

Here the second inequality is due to the definition of the LP sharpness μ_d. And furthermore, $\|q\| = \operatorname{Dist}(0, V_p)$ since (1.3). Now substitute the above inequality back to (3.10) and then (3.9) becomes

$$\theta_p^\star \le \|x^\star - \bar{x}\| \cdot \frac{\sqrt{n}}{\delta}\cdot\left(\frac{\delta}{\operatorname{Dist}(0,V_p)\cdot\mu_d} + \operatorname{Dist}(0,\mathcal{S}^\star) + \operatorname{Diam}(\mathcal{S}^\star)\right) \ . \tag{3.11}$$

Similarly, since $c^\top\bar{x} - c^\top x^\star \le \delta$, according to the definition of the LP sharpness μ_p, it holds that:

$$\|x^\star - \bar{x}\| = \operatorname{Dist}(\bar{x},\mathcal{X}^\star) \le \frac{\delta}{\|P_{\vec{V_p}}(c)\|\cdot\mu_p} = \frac{\delta}{\operatorname{Dist}(0,V_d)\cdot\mu_p} \ . \tag{3.12}$$

Substituting (3.12) back to (3.11) yields:

$$\theta_p^\star \le \frac{1}{\operatorname{Dist}(0,V_d)\cdot\mu_p}\cdot\sqrt{n}\cdot\left(\frac{\delta}{\operatorname{Dist}(0,V_p)\cdot\mu_d} + \operatorname{Dist}(0,\mathcal{S}^\star) + \operatorname{Diam}(\mathcal{S}^\star)\right) \ . \tag{3.13}$$

Note that the δ in the above inequality can be any positive scalar, so

$$\theta_p^\star \le \frac{\sqrt{n}}{\mu_p}\cdot\left(\frac{\operatorname{Dist}(0,\mathcal{S}^\star)}{\operatorname{Dist}(0,V_d)} + \frac{\operatorname{Diam}(\mathcal{S}^\star)}{\operatorname{Dist}(0,V_d)}\right) \ .$$

This is exactly (3.1). Now we have completed the proof. □

3.1 The reciprocal of LP sharpness cannot be upper-bounded by Limiting Error Ratio

In Theorem 3.1 we established an upper bound on the Limiting Error Ratio using the reciprocal of LP sharpness. In this subsection we provide an example where the reciprocal of LP sharpness is huge but the Limiting Error Ratio and the distance to optimal solutions are both small. Consider the family of LP problems parameterized by $\gamma \in [0, \pi/2)$ as follows:

$$\begin{aligned} \text{LP}_\gamma : \quad & \min_{x\in\mathbb{R}^3} \ c_\gamma^\top x \\ & \text{s.t.} \ \frac{\sqrt{3}}{3}\cdot\sum_{i=1}^{3} x_i = 1 \ , \ x \geq 0 \ , \end{aligned}$$

where c_γ is defined as follows:

$$c_\gamma := \cos(\gamma)\cdot\begin{pmatrix} \frac{-1}{\sqrt{6}} \\ \frac{-1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \end{pmatrix} + \sin(\gamma)\cdot\begin{pmatrix} \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix} \ . \tag{3.14}$$

For all $\gamma \in (0, \pi/2)$ the unique primal optimal solution is $x^\star = (\sqrt{3}, 0, 0)$. However, when $\gamma = 0$ all points on the line segment connecting $(\sqrt{3}, 0, 0)$ and $(0, \sqrt{3}, 0)$ are optimal solutions. Basically, as $\gamma \searrow 0$ the primal sharpness $\mu_p \searrow 0$ as well. We compute the Limiting Error Ratio and LP sharpness for $\gamma \searrow 0$, which are shown in Figure 1. We see from Figure 1 that μ_p decreases linearly in $O(\gamma)$ while μ_d, $\theta_p^\star$, $\theta_d^\star$, and the primal and dual relative distances to optima all remain constant. This example shows that the reciprocal of LP sharpness cannot be upper-bounded as a function of Limiting Error Ratio and the relative distance to optima.

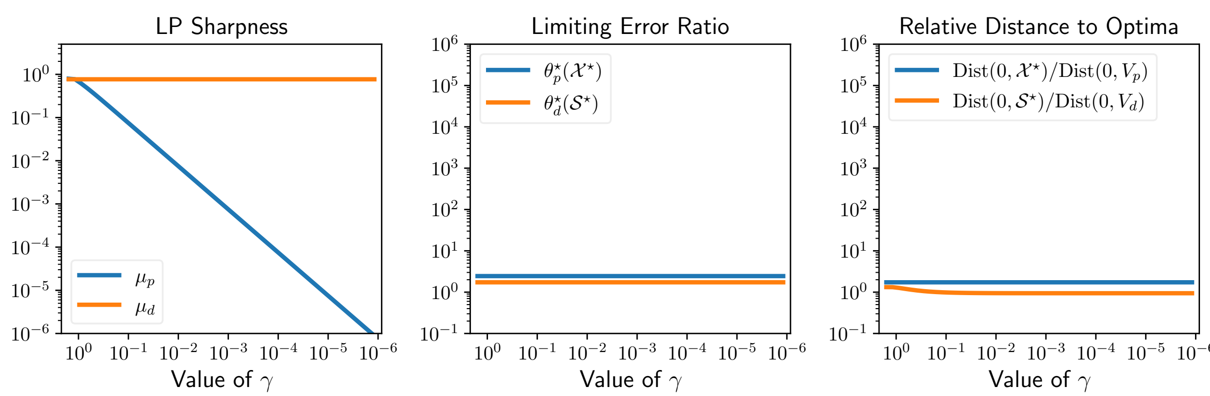

Figure 1: Condition measures of LP_γ.

4 Complexity Implication in the Restarted PDHG

In this section we show how Theorem 3.1 can be leveraged to yield a simplified complexity bound for restarted PDHG for LP. We first recall the complexity results in [12]. Let us assume that the affine subspace V_p in (1.1) is given by $\{x : Ax = b\}$ for a specific $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$. We define the following quantities:

$$\lambda_{\max} := \sigma_{\max}^{+}(A)\,,\ \lambda_{\min} := \sigma_{\min}^{+}(A)\,,\ \text{and } \kappa := \frac{\lambda_{\max}}{\lambda_{\min}}\ , \tag{4.1}$$

and the measure of error of a primal-dual pair (x, s) is the "distance to optima" error of (x, s) given by:

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

The following is a re-statement of Theorem 3.3 of [12] whose result is for the "standard" step-sizes based on knowledge of $\lambda_{\min}$, $\lambda_{\max}$, $\|c\|$ and $\|q\|$ as follows.

Lemma 4.1 (Theorem 3.3 of [12])**.** *Suppose that Assumption 1 holds, and that the restarted PDHG (Algorithm 2 in [12]) is run starting from $z^{0,0} = (x^{0,0}, y^{0,0}) = (0, 0)$ using the β-restart condition with $\beta := 1/e$. Furthermore, let the step-sizes be chosen as follows:*

$$\tau = \frac{\|q\|}{2\kappa\|c\|} \quad \textit{and} \quad \sigma = \frac{\|c\|}{2\|q\|\lambda_{\max}\lambda_{\min}}\ . \tag{4.2}$$

Let T be the total number of PDHG iterations that are run in order to obtain a solution (x, s) that satisfies $\mathcal{E}_d(x, s) \le \varepsilon$. Then

$$T \ \le\ 9e \cdot \mathcal{L} \cdot \ln\left(16e \cdot \mathcal{L} \cdot \frac{\mathcal{E}_d(x^{0,0}, s^{0,0})}{\varepsilon} \cdot \left(1 + \kappa\frac{\|c\|}{\|q\|}\right)\left(1 + \frac{\|q\|}{\|c\|}\right)\right) + 1\ , \tag{4.3}$$

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

$$\mathcal{L} := 8.5\kappa\left(\frac{1}{\mu_p} + \frac{1}{\mu_d}\right)\left(\theta_p^\star + \theta_d^\star + \frac{\mathrm{Dist}(0, \mathcal{X}^\star)}{\mathrm{Dist}(0, V_p)} + \frac{\mathrm{Dist}(c, \mathcal{S}^\star)}{\mathrm{Dist}(0, V_d)}\right)\ . \tag{4.4}$$

In addition to the "standard" step-sizes described in (4.2), [12] also develops potentially better step-sizes that involve knowledge of LP sharpness, which we will call the "optimized" step-sizes. The complexity of using the "optimized" step-sizes is as follows.

Lemma 4.2 (Remark 3.4 in [12])**.** *The following choice of step-sizes:*

$$\tau = \frac{\mu_d\|q\|}{2\kappa\mu_p\|c\|} \quad \textit{and} \quad \sigma = \frac{\mu_p\|c\|}{2\mu_d\|q\|\lambda_{\max}\lambda_{\min}} \tag{4.5}$$

leads to an alternative bound on the total number of PDHG iterations T:

$$T \ \le\ 9e \cdot \mathcal{L} \cdot \ln\left(16e \cdot \mathcal{L} \cdot \frac{\mathcal{E}_d(x^{0,0}, s^{0,0})}{\varepsilon} \cdot \left(1 + \kappa\frac{\mu_p\|c\|}{\mu_d\|q\|}\right)\left(1 + \frac{\mu_d\|q\|}{\mu_p\|c\|}\right)\right) + 1\ , \tag{4.6}$$

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

$$\mathcal{L} := 16\kappa\left(\frac{\theta_p^\star}{\mu_p} + \frac{\theta_d^\star}{\mu_d} + \frac{\mathrm{Dist}(c, \mathcal{S}^\star)}{\mu_p \cdot \mathrm{Dist}(0, V_d)} + \frac{\mathrm{Dist}(0, \mathcal{X}^\star)}{\mu_d \cdot \mathrm{Dist}(0, V_p)}\right)\ . \tag{4.7}$$

It can be observed when considering the terms outside the logarithm, that the number of iterations of restarted PDHG is $O(\mathcal{L})$, where $\mathcal{L}$ is defined in (4.4) or (4.7) depending on the step-sizes used. However, $\mathcal{L}$ in (4.4) and (4.7) involves LP sharpness, Limiting Error Ratio, and the relative distance to optima. According to Theorem 3.1 and Corollary 3.2, we can conclude that if both the primal and dual optimal solutions are unique, then:

$$\theta_p^\star \le \frac{\sqrt{n}}{\mu_p} \cdot \frac{\mathrm{Dist}(0, \mathcal{S}^\star)}{\mathrm{Dist}(0, V_d)} \quad \text{and} \quad \theta_d^\star \le \frac{\sqrt{n}}{\mu_d} \cdot \frac{\mathrm{Dist}(0, \mathcal{X}^\star)}{\mathrm{Dist}(0, V_p)} \ . \tag{4.8}$$

This allows us to provide an alternative expression for $\mathcal{L}$ as follows.

Corollary 4.3. *When Assumption 1 holds, and both the primal and dual problems in* (PD) *have unique optimal solutions, then the expression for* $\mathcal{L}$ *in* (4.4) *can be replaced by:*

$$\mathcal{L} := 8.5\kappa(\sqrt{n}+1)\left(\frac{1}{\mu_p}+\frac{1}{\mu_d}\right)\left(\frac{1}{\mu_p}\cdot\frac{\mathrm{Dist}(0, \mathcal{S}^\star)}{\mathrm{Dist}(0, V_d)}+\frac{1}{\mu_d}\cdot\frac{\mathrm{Dist}(0, \mathcal{X}^\star)}{\mathrm{Dist}(0, V_p)}\right) ,$$

and the expression for $\mathcal{L}$ *in* (4.7) *can be replaced by*

$$\mathcal{L} := 16\kappa(\sqrt{n}+1)\left(\frac{1}{\mu_p^2}\cdot\frac{\mathrm{Dist}(0, \mathcal{S}^\star)}{\mathrm{Dist}(0, V_d)}+\frac{1}{\mu_d^2}\cdot\frac{\mathrm{Dist}(0, \mathcal{X}^\star)}{\mathrm{Dist}(0, V_p)}\right) \ .$$

This allows for a worst-case complexity analysis of restarted PDHG without the direct use of Limiting Error Ratio.

References

[1] A. Belloni and R. M. Freund. On the symmetry function of a convex set. *Mathematical Programming*, 111(1-2):57–93, 2008.

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

[3] S. P. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[4] M. Epelman and R. M. Freund. A new condition measure, preconditioners, and relations between different measures of conditioning for conic linear systems. *SIAM Journal on Optimization*, 12(3):627–655, 2002.

[5] R. M. Freund. On the primal-dual geometry of level sets in linear and conic optimization. *SIAM Journal on Optimization*, 13(4):1004–1013, 2003.

[6] A. J. Hoffman. On approximate solutions of systems of linear inequalities. In *Selected Papers Of Alan J Hoffman: With Commentary*, pages 174–176. World Scientific, 2003.

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

[8] G. W. Stewart. On scaled projections and pseudoinverses. *Linear Algebra and its Applications*, 112:189–193, 1989.

[9] M. J. Todd. A Dantzig-Wolfe-like variant of Karmarkar's interior-point linear programming algorithm. *Operations Research*, 38(6):1006–1018, 1990.

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

[11] S. A. Vavasis and Y. Ye. A primal-dual interior point method whose running time depends only on the constraint matrix. *Mathematical Programming*, 74(1):79–120, 1996.

[12] Z. Xiong and R. Freund. Compuatational guarantees for restarted PDHG for LP based on "limiting error ratios" and LP sharpness. *arXiv preprint arXiv:2312.14774*, 2023.