Accessible Theoretical Complexity of the Restarted Primal-Dual Hybrid Gradient Method for Linear Programs with Unique Optima

Zikai Xiong
MIT Operations Research Center, 77 Massachusetts Avenue, Cambridge, MA 02139, USA, zikai@mit.edu.

Abstract. The restarted primal-dual hybrid gradient method (rPDHG) has recently emerged as an important tool for solving large-scale linear programs (LPs). For LPs with unique optima, we present an iteration bound of $\tilde{O}\left(\kappa\Phi \cdot \ln\left(\frac{\|w^\star\|}{\varepsilon}\right)\right)$, where ε is the target tolerance, κ is the standard matrix condition number, $\|w^\star\|$ is the norm of the optimal solution, and Φ is a geometric condition number of the LP sublevel sets. This iteration bound is "accessible" in the sense that computing it is no more difficult than computing the optimal solution itself. Indeed, we present a closed-form and tractably computable expression for Φ. This enables an analysis of the "two-stage performance" of rPDHG: we show that the first stage identifies the optimal basis in $\tilde{O}(\kappa\Phi)$ iterations, and the second stage computes an ε-optimal solution in $O\left(\|B^{-1}\|\|A\| \cdot \ln\left(\frac{\xi}{\varepsilon}\right)\right)$ additional iterations, where A is the constraint matrix, B is the optimal basis and ξ is the smallest nonzero in the optimal solution. Furthermore, computational tests mostly confirm the tightness of our iterations bounds. We also show a reciprocal relation between the iteration bound and three equivalent types of condition measures: (i) stability under data perturbation, (ii) proximity to multiple optima, and (iii) the LP sharpness of the instance. Finally, we analyze an "optimized" primal-dual reweighting which offers some intuition concerning the step-size heuristics used in practice.

Funding: This research was supported by AFOSR Grant No. FA9550-22-1-0356.
Key words: Linear optimization, First-order method, Convergence guarantees, Condition number

1. Introduction Linear Program (LP) has been a cornerstone of optimization since the 1950s, with far-reaching applications across diverse fields, including economics (see, e.g., Greene [25]), transportation (see, e.g., Charnes and Cooper [13]), manufacturing (see, e.g., Bowman [11], Hanssmann and Hess [27]), computer science (see, e.g., Cormen et al. [15]), and medicine (see, e.g., Wagner et al. [54]) among many others (see, e.g., Dantzig [16]). LP algorithms have also been extensively researched in the past several decades. Almost all LP algorithms to date are based on either simplex/pivoting methods and/or interior-point methods (IPMs). These classic methods form the foundation of modern solvers due to their reliability and robustness in providing high-quality solutions. However, both of them require repeatedly solving linear equation systems at each iteration using matrix factorizations, whose cost grows superlinearly in the size of the instance (as measured in the dimensions of problem and/or the number of nonzeros in the data). Consequently, as problem size increases, these methods become computationally impractical. Furthermore, matrix factorizations cannot efficiently leverage modern computational architectures, such as parallel computing on graphics processing units (GPUs). For these reasons, suitable first-order methods (FOMs) are emerging as attractive solution algorithms because they are "matrix-free," meaning they require no or perhaps only very few matrix factorizations, while their primary computational cost lies just in computing matrix-vector products when computing gradients and related quantities. Hence, FOMs are inherently more suitable for exploiting data sparsity, and for parallel computing using GPUs, and their iteration cost typically scales just linearly in the size of the instance.

The restarted primal-dual hybrid gradient method (rPDHG) has emerged as a particularly successful FOM for solving LPs. It directly addresses the saddlepoint formulation of LP (see Applegate et al. [6]), automatically detects infeasibility (see Applegate et al. [5]), and has natural extensions to conic linear programs in Xiong and Freund [58] and convex quadratic programs in Lu and Yang [39] and Huang et al. [30]. This algorithm has led to various implementations on CPUs (PDLP by Applegate et al. [4]) and GPUs (cuPDLP by Lu and Yang [38] and cuPDLP-C by Lu et al. [43]), equipped with several effective heuristics. Notably, the performance of the GPU implementations has surpassed classic algorithms (simplex methods and IPMs) on a significant number of problem instances as shown by Lu and Yang [38] and Lu et al. [43]. The strong practical performance has also sparked considerable industrial interest from mathematical optimization software companies. To date, rPDHG has been integrated into the state-of-art commercial solvers COPT 7.1 (see Ge et al. [24]) and Xpress 9.4 (see Biele and Gade [9]) as a new base algorithm for LP, complementing simplex methods and IPMs. It has also been added to Google OR-Tools (see Applegate et al. [4]), HiGHS (see Ge et al. [24]), and NVIDIA cuOpt (see Fender [21]). With rPDHG, many problems previously considered too large-scale for classic algorithms are now solvable; for instance, it is reported by Mirrokni [45] that a distributed version of rPDHG has been used to solve practical LP instances with more than 9.2×10^{10} nonzero components in the constraint matrix, a scale far beyond the capabilities of traditional methods. Another example is a representative large-scale benchmark instance called zib03. This instance, which took 16.5 hours to solve in 2021 as reported by Koch et al. [31], can now be solved in 15 minutes using rPDHG by Lu et al. [43]. Furthermore, it has recently activated the potentials of large-scale linear programming in real applications, including making targeted marketing policies by Lu et al. [36], solving large-scale integer programming instances by De Rosa and Khajavirad [17], and optimizing data center network traffic engineering by Lu and Applegate [35], the latter of which has been deployed in Google's production environment.

Despite the strong performance of rPDHG on many LP instances, certain aspects of its practical behavior remain poorly understood. Indeed, rPDHG sometimes performs poorly, even for some very small LP instances. Additionally, minor data perturbation of some easily solveable instances can lead to instances with substantially increased computational cost. Also, it has been observed that rPDHG often exhibits a "two-stage performance" phenomenon in which the second stage exhibits much faster local convergence, but this phenomenon has not been adequately explained or otherwise addressed by suitable theory.

To better understand the underlying behavior of rPDHG, it is important to have theory that is in synch with practical performance. However, many aspects of the existing theory cannot be adequately evaluated for practical relevance due to the difficulty of actually computing the quantities in the theoretical computational bounds. Applegate et al. [6] establish the linear convergence rate of rPDHG using the global Hoffman constant of the matrix K of the KKT system corresponding to the LP instance. Roughly speaking, the Hoffman constant is equal to the reciprocal of the smallest nonzero singular values of the submatrices of K, of which there are exponentially many (see Pena et al. [48]). While intuition suggests that the Hoffman constant is itself an overly conservative quantity in the computational complexity, we do not know this from experience on any non-trivial LPs since the Hoffman constant is not computable in reasonable time. Xiong and Freund [56] provide a tighter computational guarantee for rPDHG using two natural properties of the LP problem: LP sharpness and the "limiting error ratio." Furthermore, for LPs with extremely poor sharpness and the broader family of conic LPs, Xiong and Freund [58] provide computational guarantees for rPDHG based on three geometric measures of the primal-dual (sub)level set geometry. In addition, Lu and Yang [40] study the vanilla primal-dual hybrid gradient method (PDHG) using a trajectory-based analysis approach, and shows the two-stage performance of PDHG based on the Hoffman constant of a smaller linear system. However, despite these studies, none provides an iteration bound that is reasonably easy to compute, and so we cannot ascertain the extent to which any of these iteration bounds align with computational practice. To compute the iteration bounds, all existing works require prohibitively expensive operations, such as directly computing Hoffman constants (e.g., Applegate et al. [6], Lu and Yang [40]), solving multiple additional optimization problems (e.g., Lu and Yang [40], Xiong and Freund [56, 58]), or running a first-order method beforehand to obtain the solution trajectory (e.g., Lu and Yang [40]).

Due to the absence of an iteration bound that can be practically evaluated, we do not know the extent to which existing theoretical bounds align with computational practice. This makes it more difficult to do mathematical analysis to possibly improve the practical behavior of rPDHG either. Furthermore, the lack of a reasonably convenient expression of the iteration bound also hinders a deeper understanding of rPDHG's performance on specific families of LP instances, impedes theoretical validation of some effective practical heuristics, and potentially hampers the development of further practical enhancements.

This paper aims to make progress on the above issues by posing and trying to answer the following questions:

- Is there an *accessible* iteration bound of rPDHG that is easily computable?, we will say that an iteration bound is *accessible* if, roughly speaking, the quantities in the bound depend only on some intrinsic

properties of the LP instance whose computation is not more costly than solving the LP instance itself. (For example, the norm of the optimal solution is accessible, but the Hoffman constant is not accessible.)
- If we have an accessible iteration bound, can we use the bound to provide deeper insights into the practical performance of rPDHG? – particularly regarding the two-stage performance phenomenon, the sensitivity to minor data perturbations, and to possibly improve current heuristic components of practical implementations?

This paper focuses on LP instances with unique optimal solutions (of the primal and dual) and hence unique optimal bases, and proves an accessible iteration bound that is easily computable and indeed has a closed-form expression. A similar unique optimum assumption is often made in studies of large-scale LPs (e.g., Liu et al. [34] and Xiong and Freund [57]) and other optimization problems, such as semidefinite programs (e.g., Alizadeh et al. [1]) and general convex optimization (e.g., Drusvyatskiy and Lewis [19]). The unique optimum assumption is actually looser than the nondegeneracy assumption, because once the primal and dual optimal basic feasible solutions are nondegenerate, then the primal and dual solutions are unique and nondegenerate, which means the LP instance has one unique optimal basis. This nondegeneracy assumption is also used by almost all classic optimization textbooks, such as Bertsimas and Tsitsiklis [8], to simplify analysis and convey insights. We acknowledge the unique optimum assumption does not hold for a very large number of LP instances occurring in practice due to special problem structures (network substructure in particular) in very many real applications. But the property of unique optima for LP is generic (i.e., it holds almost surely) under most models of randomly generated LP instances. Here we use our accessible new iteration bound to provide valuable insights into the performance of rPDHG and may potentially inspire future research generalizing this work to the broader family of LP instances.

1.1. Outline and contributions In this paper, we consider the following standard form LPs:

$$\min_{x\in\mathbb{R}^n} \quad c^\top x \qquad \text{s.t. } Ax = b \ , \ x \geq 0 \ . \tag{1.1}$$

where $A \in \mathbb{R}^{m\times n}$ is the constraint matrix, $b \in \mathbb{R}^m$ is the right-hand side vector, and $c \in \mathbb{R}^n$ is the objective vector. The corresponding dual problem is:

$$\max_{y\in\mathbb{R}^m, s\in\mathbb{R}^n} \quad b^\top y \qquad \text{s.t. } A^\top y + s = c \ , \ s \geq 0 \ . \tag{1.2}$$

We will assume the optimal basis is unique, denoted by B, and let $x^\star$, $y^\star$ and $s^\star$ denote the optimal solutions. In Section 2, we revisit its saddlepoint formulation and its symmetric reformulation on the space of x and s. We also review rPDHG for solving LPs.

Section 3 presents the main result of the paper: an accessible iteration bound of rPDHG that has a closed-form expression. The bound takes the form

$$O\left(\kappa\Phi \cdot \ln\left(\kappa\Phi\frac{\|(x^\star, s^\star)\|}{\varepsilon}\right)\right) \ ,$$

where ε is the target tolerance, κ is the standard matrix condition number, and Φ is a geometric condition number of the LP sublevel sets that admits a closed-form expression. This new bound is actually proven equivalent to a bound in Xiong and Freund [58] (under the unique optimum assumption) but has a closed-form expression. Furthermore, Φ has an even simpler upper bound:

$$\Phi \leq \frac{\|x^\star + s^\star\|_1}{\min_{1\leq i\leq n}\left\{x_i^\star + s_i^\star\right\}} \cdot \left\|B^{-1}A\right\|_2 \ .$$

In Section 4, using the established accessible iteration bound, we provide a mathematical analysis of rPDHG's "two-stage performance." Specifically, we show that Stage I achieves finite-time optimal basis identification in

$$O\left(\kappa\Phi \cdot \ln\left(\kappa\Phi\right)\right)$$

iterations, and Stage II exhibits a faster local convergence rate and computes an ε-optimal solution in

$$O\left(\|B^{-1}\|\|A\| \cdot \ln\left(\frac{\min_{1\le i\le n}\{x_i^\star + s_i^\star\}}{\varepsilon}\right)\right)$$

additional iterations. The iteration bound of Stage II is independent of Φ and may thus be significantly lower than that of Stage I. This provides at least a partial explanation for the "two-stage performance" in theory.

In Section 5, using the expression of the new iteration bound, we study the relation between the iteration bound of rPDHG and three equivalent types of condition measures: (i) stability under data perturbations, (ii) proximity to multiple optima, and (iii) the LP sharpness of the instance. Specifically, we show that

$$\Phi = \frac{\|x^\star\|_1 + \|s^\star\|_1}{\min\{\zeta_p, \zeta_d\}} ,$$

where ζ_p and ζ_d denote the stability measures for the primal and dual problems. This relationship yields a new and tighter computational guarantee of rPDHG, and also quantifies the impact of tiny data perturbations on the convergence rate.

In Section 6, since the new iteration bounds can now be easily computed, we confirm their tightness via computational tests on LP instances. As predicted by the new iteration bounds, experiments show that tiny perturbations may indeed significantly alter Φ and the overall convergence rates. Additionally, the new iterations bounds are also confirmed matching the practice as our experiments show $\kappa\Phi$ and $\|B^{-1}\|\|A\|$ indeed play important roles in the global linear convergence rates and the two-stage performance. Since our new iteration bounds are proven equivalent to a bound in Xiong and Freund [58], the latter is also confirmed to match the practical behavior by our experiments.

In Section 7, thanks to the accessible iteration bound expression, we demonstrate that the reweighting that equalizes the ℓ_1-norms of the primal and dual optimal solutions can approximately minimize Φ and the overall iteration bound. This finding provides some intuition concerning the very effective heuristic of balancing the "primal weights."

1.2. Other related works In addition to the previously discussed papers, several other studies have analyzed the performance of PDHG and its variants. Hinder [28] and Lu and Yang [41] present instance-independent worst-case complexity bounds of rPDHG on totally-unimodular LPs and optimal transport problems. Lu and Yang [37] show that the last iterate of the vanilla PDHG without restarts also exhibits a linear convergence rate, dependent on the global Hoffman constant of the KKT system matrix. A recent concurrent work Lu and Yang [42] propose a new restart scheme for PDHG by restarting from the Halpern iterate instead of the average iterate. They prove an accelerated refined complexity bound compared to that of the vanilla PDHG proven in Lu and Yang [40]. This new bound is still based on the Hoffman constant of the reduced KKT system and employs a trajectory-based analysis approach.

There has been increasing interest in developing FOMs for LPs. Xiong and Freund [58] propose to use central-path Hessian-based rescaling to accelerate rPDHG, and Li et al. [32] design a learning-to-optimize method to emulate PDHG for solving LPs. Beyond PDHG, several other FOMs have been studied recently. Lin et al. [33] and Deng et al. [18] develop ABIP (and ABIP+), 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. O'donoghue et al. [47] and O'Donoghue [46] develop SCS, applying ADMM directly to the homogeneous self-dual formulation for general CLP problems. Basu et al. [7] utilize accelerated gradient descent to solve a smoothed dual form of LP. Wang et al. [55] use overparametrized neural networks to solve entropically regularized LPs. Very recently, Hough and Vavasis [29] use a Frank-Wolfe method to address the saddlepoint problem formulation, and Chen et al. [14] implement a Halpern Peaceman-Rachford method with semi-proximal terms to solve LPs.

1.3. Notation In this paper, we use $[n]$ as shorthand for $\{1,2,\ldots,n\}$. For a matrix $A\in\mathbb{R}^{m\times n}$, $\mathrm{Null}(A):=\{x\in\mathbb{R}^n : Ax=0\}$ denotes the null space of A and $\mathrm{Im}(A):=\{Ax : x\in\mathbb{R}^n\}$ denotes the image of A. For any $i\in[m]$ and $j\in[n]$, $A_{\cdot,i}$ and $A_{j,\cdot}$ denote the corresponding i-th column and j-th row of A, respectively. For any subset Θ of $[n]$, A_Θ denotes the submatrix of A formed by the columns indexed by Θ. We use $\|A\|_{\alpha,\beta}$ to denote the operator norm, i.e., $\|A\|_{\alpha,\beta}:=\sup_{x\neq 0}\frac{\|Ax\|_\alpha}{\|x\|_\beta}$. Specifically, $\|A\|_{2,\infty}=\max_{1\le i\le m}\left\|A_{i,\cdot}\right\|$ and $\|A\|_{1,2}=\max_{j\in[n]}\left\|A_{\cdot,j}\right\|$. We let $\|\cdot\|_M$ denote the inner product "norm" induced by M, namely, $\|z\|_M:=\sqrt{z^\top Mz}$. Unless otherwise specified, for a vector v, $\|v\|$ denotes the Euclidean norm, and for a matrix A, $\|A\|$ denotes $\|A\|_{2,2}$, the spectral norm of A. For any set $\mathcal{X}\subset\mathbb{R}^n$, $P_\mathcal{X}:\mathbb{R}^n\to\mathbb{R}^n$ denotes the Euclidean projection onto $\mathcal{X}$, namely, $P_\mathcal{X}(x):=\arg\min_{\hat{x}\in\mathcal{X}}\|x-\hat{x}\|$. For any $x\in\mathbb{R}^n$ and set $\mathcal{X}\subset\mathbb{R}^n$, the Euclidean distance between x and $\mathcal{X}$ is denoted by $\mathrm{Dist}(x,\mathcal{X}):=\min_{\hat{x}\in\mathcal{X}}\|x-\hat{x}\|$ and the M-norm distance between x and $\mathcal{X}$ is denoted by $\mathrm{Dist}_M(x,\mathcal{X}):=\min_{\hat{x}\in\mathcal{X}}\|x-\hat{x}\|_M$. For $x\in\mathbb{R}^n$, we use x^+ to denote the positive part of x. For any affine subspace V, we use $\vec{V}$ to denote the associated linear subspace corresponding to V. For any linear subspace $\vec{S}$ in $\mathbb{R}^n$, we use $\vec{S}^\perp$ to denote the corresponding complementary linear subspace of $\vec{S}$. In this paper, we use $O(\cdot)$ to hide factors of only absolute constants.

2. Preliminaries and Background Throughout this paper, we consider the primal problem (1.1) and its dual problem (1.2). For simplicity of analysis, we consistently assume that the rows of A are linearly independent and that (1.1) has at least one optimal solution. A primal-dual solution pair x and (y,s) is optimal if and only if they are feasible and the duality gap is zero, i.e.,

$$\mathrm{Gap}(x,y):=c^\top x-b^\top y=0\ . \tag{2.1}$$

Furthermore, (1.1) and (1.2) are equivalent to the following saddlepoint problem:

$$\min_{x\in\mathbb{R}^n_+}\max_{y\in\mathbb{R}^m}\quad L(x,y):=c^\top x+b^\top y-(Ax)^\top y\ . \tag{2.2}$$

The optimal solutions $x^\star$ and $(y^\star,s^\star)$ of (1.1) and (1.2), are also the saddle point of (2.2), and *vice versa*.

2.1. Symmetric formulation of LP Given that (1.2) includes the constraint $A^\top y+s=c$, it follows that for any feasible (y,s), $y=(AA^\top)^{-1}A(c-s)$. Let us define $q:=A^\top\left(AA^\top\right)^{-1}b$; then the objective function of y is equivalent to an objective function of s, i.e., $b^\top y=q^\top(c-s)$, and (1.2) is thus equivalent to the following (dual) problem on s :

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

We denote the duality gap for the pair (x,s) as $\mathrm{Gap}(x,s):=c^\top x-q^\top(c-s)$, which is equivalent to $\mathrm{Gap}(x,y)$ when $A^\top y+s=c$.

Note that the feasible set of the primal problem (1.1) is the intersection of the affine subspace $V_p:=q+\mathrm{Null}(A)$ and the nonnegative orthant. Similarly, the feasible set of (2.3) is the intersection of $V_d:=c+\mathrm{Im}\left(A^\top\right)$ and the nonnegative orthant. We can thus rewrite the primal and dual problems in the following symmetric form:

$$\begin{aligned}&\min_{x\in\mathbb{R}^n}\ c^\top x &&\max_{s\in\mathbb{R}^n}\ q^\top(c-s)\\ &\quad\text{s.t. } x\in\mathcal{F}_p:=V_p\cap\mathbb{R}^n_+ &&\quad\text{s.t. } s\in\mathcal{F}_d:=V_d\cap\mathbb{R}^n_+\end{aligned} \tag{2.4}$$

Let $\vec{V}_p$ and $\vec{V}_d$ denote the linear subspaces associated with the affine subspaces V_p and V_d, respectively. These subspaces are orthogonal complements. We then use $\mathcal{X}^\star$ and $\mathcal{S}^\star$ to denote the optimal solutions for the primal and the dual problem, respectively. Notably, any change in c within the space of $\vec{V}_p$ does not affect $\mathcal{X}^\star,\mathcal{S}^\star,V_p,V_d,\mathcal{F}_p$, or $\mathcal{F}_d$. Without loss of generality, we may sometimes assume that c is in $\mathrm{Null}(A)$, which can be achieved by replacing c with $P_{\vec{V}_p}(c)$ beforehand. This leads to the following symmetric properties for (1.1) and (2.3).

Fact 2.1. Suppose that $Ac=0$. Then $\vec{V}_d$ is the orthogonal complement of $\vec{V}_p$, i.e., $\vec{V}_d=\vec{V}_p^\perp$. Furthermore, $q^\top c=0$, and the objective function of (2.3) is equal to $-q^\top s$, and $\mathrm{Gap}(x,s)$ is equal to $c^\top x+q^\top s$. Additionally, $c\in\vec{V}_p$ and $c=\arg\min_{v\in V_d}\|v\|$, and $q\in\vec{V}_d$ and $q=\arg\min_{v\in V_p}\|v\|$.

This re-formulation of the dual was, to the best of our knowledge, first introduced in Todd and Ye [52]. We will use the notation $\mathcal{W}^\star$ to denote the primal-dual pairs, i.e.,

$$\mathcal{W}^\star := \mathcal{X}^\star \times \mathcal{S}^\star = \left\{(x^\star, s^\star) : x^\star \in \mathcal{X}^\star, s^\star \in \mathcal{S}^\star\right\} \ . \tag{2.5}$$

Our focus is on computing ε-optimal solutions, which are essentially solutions sufficiently close to $\mathcal{W}^\star$, as defined below.

Definition 2.1 (ε-optimal solution). A solution w is said to be ε-optimal if the Euclidean distance between w and $\mathcal{W}^\star$ is less than ε, i.e.,

$$\mathrm{Dist}(w, \mathcal{W}^\star) \le \varepsilon \ .$$

2.2. Restarted Primal-dual hybrid gradient method (rPDHG) The vanilla primal-dual hybrid gradient method (abbreviated as PDHG) was introduced by Esser et al. [20], Pock et al. [49] to solve general convex-concave saddlepoint problems, of which (2.2) is a specific subclass. For LP problems, let z denote the primal-dual pair (x, y), and then iteration of PDHG, denoted by $z^{k+1} = (x^{k+1}, y^{k+1}) \leftarrow \textsc{OnePDHG}(z^k)$, is defined as follows:

$$\begin{cases} x^{k+1} \leftarrow \left(x^k - \tau\left(c - A^\top y^k\right)\right)^+ \\ y^{k+1} \leftarrow y^k + \sigma\left(b - A\left(2x^{k+1} - x^k\right)\right) \end{cases} \tag{2.6}$$

where τ and σ are the primal and dual step-sizes, respectively.

Algorithm 1 presents the general restart scheme for PDHG. We refer to this algorithm as "rPDHG," short for restarted-PDHG.

Algorithm 1: rPDHG: restarted-PDHG

1 **Input:** Initial iterate $z^{0,0} := (x^{0,0}, y^{0,0}) = (0, 0)$, $n \leftarrow 0$, step-size τ, σ, and $\beta \in (0, 1)$;
2 **repeat**
3   **initialize the inner loop:** inner loop counter $k \leftarrow 0$;
4   **repeat**
5   **conduct one step of PDHG:** $z^{n,k+1} \leftarrow \textsc{OnePDHG}(z^{n,k})$;
6   **compute the average iterate in the inner loop.** $\bar{z}^{n,k+1} \leftarrow \frac{1}{k+1}\sum_{i=1}^{k+1} z^{n,i}$;
7         $k \leftarrow k + 1$;
8   **until** *satisfying the β-restart condition*;
9   **restart the outer loop:** $z^{n+1,0} \leftarrow \bar{z}^{n,k}$, $n \leftarrow n + 1$;
10 **until** *Either $z^{n,0}$ is a saddlepoint or $z^{n,0}$ satisfies some other convergence condition* ;
11 **Output:** $z^{n,0}$ ($= (x^{n,0}, y^{n,0})$)

Line 5 of Algorithm 1 is an iteration of the vanilla PDHG as described in (2.6). For each iterate $z^{n,k} = (x^{n,k}, y^{n,k})$, we define $s^{n,k} := c - A^\top y^{n,k}$ and $\bar{s}^{n,k} := c - A^\top \bar{y}^{n,k}$. The double superscript 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. Line 8 of Algorithm 1 specifies an easily verifiable restart condition proposed by Applegate et al. [6] and also used by Xiong and Freund [56, 58] and the practical implementation by Applegate et al. [4]. We will discuss it in more detail when using them.

The primary computational effort of Algorithm 1 is the OnePDHG in Line 5, which involves two matrix-vector products. In contrast to traditional methods such as simplex and interior-point methods, rPDHG does not require any matrix factorizations. It is worth noting that the step-sizes τ and σ need to be sufficiently small to ensure convergence. In particular, if $M := \begin{pmatrix} \frac{1}{\tau} I_n & -A^\top \\ -A & \frac{1}{\sigma} I_m \end{pmatrix}$ is positive semi-definite, then Chambolle and Pock [12] prove rPDHG's iterates will converge to a saddlepoint of (2.2). The above requirement can be equivalently expressed as:

$$\tau > 0, \ \sigma > 0, \ \text{ and } \ \tau\sigma \le \frac{1}{\|A\|^2} \ . \tag{2.7}$$

Furthermore, the matrix M turns out to be particularly useful in analyzing the convergence of rPDHG through its induced inner product norm defined as $\|z\|_M := \sqrt{z^\top M z}$. This norm will be extensively employed throughout the remainder of this paper.

2.3. LPs with unique optima This paper focuses particularly on LPs with unique optima, the problems satisfying the following assumption:

ASSUMPTION 2.1. *The linear optimization problem* (1.1) *has a unique optimal solution* $x^\star$*, and the dual problem* (1.2) *has a unique optimal solution* $(y^\star, s^\star)$*, i.e.,* $\mathcal{X}^\star = \{x^\star\}$*,* $\mathcal{Y}^\star = \{y^\star\}$ *and* $\mathcal{S}^\star = \{s^\star\}$*.*

When $\mathcal{S}^\star$ is a singleton, actually $\mathcal{Y}^\star$ is a singleton if and only if the rows of the constraint matrix A are linearly independent. This assumption is equivalent to having a unique optimal basis, and is also equivalent to the case that the primal and dual optimal basic feasible solutions are nondegenerate. Actually, in theory "almost all" LP instances have unique optima, as randomly generated instances are known to be nondegenerate almost surely (see Borgwardt [10]). The unique optimum assumption and the stronger nondegeneracy assumption are often used in large-scale linear programming (see, e.g., Liu et al. [34], Xiong and Freund [57]), semidefinite programs (see, e.g., Alizadeh et al. [1]), and general convex optimization (see, e.g., Drusvyatskiy and Lewis [19]). But in practice, due to special structures of real problems, this assumption does not always hold.

Under Assumption 2.1, the primal-dual pair of optimal solutions, $x^\star$ and $(y^\star, s^\star)$, are optimal basic feasible solutions, corresponding to the optimal basis $\Theta := \{i \in [n] : x_i^\star > 0\}$. Let $\bar{\Theta}$ denote the complement of Θ, i.e., $\bar{\Theta} := [n] \setminus \Theta$. Due to strict complementary slackness, $\bar{\Theta} = \{i \in [n] : s^\star > 0\}$. As $x^\star$ is an optimal basic feasible solution, there are exactly m components in Θ and $n - m$ components in $\bar{\Theta}$.

Since the algorithm is invariant under permutation of the variables, for simplicity of notations in this paper we assume that the optimal basis is $\{1, 2, \ldots, m\}$ and use B and N to denote the submatrices A_Θ and $A_{\bar{\Theta}}$, respectively. In other words,

$$\Theta = [m] = \{1, 2, \ldots, m\}, \quad \bar{\Theta} = [n] \setminus [m] = \{m+1, m+2, \ldots, n\} \quad \text{and} \quad A = \begin{pmatrix} B & N \end{pmatrix} . \tag{2.8}$$

With the above Θ and $\bar{\Theta}$, the indices of the nonzero components of $x^\star$ are exactly $[m]$, and the indices of the nonzero components of $s^\star$ are exactly $[n] \setminus [m]$.

Later in the paper we will frequently use the following quantities of the matrix A:

$$\lambda_{\max} := \sigma_{\max}^+(A), \; \lambda_{\min} := \sigma_{\min}^+(A), \; \kappa := \frac{\lambda_{\max}}{\lambda_{\min}} \tag{2.9}$$

where $\sigma_{\max}^+(A)$ and $\sigma_{\min}^+(A)$ denote the largest and the smallest nonzero singular values of A, respectively. And κ is often referred to as the matrix condition number of A.

3. Closed-form Complexity Bound of rPDHG This section presents the main result of the paper: an iteration bound of the global linear convergence that has a closed-form expression. First of all, we define the following quantity Φ:

$$\Phi := \left(\|x^\star\|_1 + \|s^\star\|_1\right) \cdot \max\left\{ \max_{1 \le j \le n-m} \frac{\sqrt{\left\|(B^{-1}N)_{\cdot,j}\right\|^2 + 1}}{s_{m+j}^\star}, \; \max_{1 \le i \le m} \frac{\sqrt{\left\|(B^{-1}N)_{i,\cdot}\right\|^2 + 1}}{x_i^\star} \right\} . \tag{3.1}$$

Notably, it leads to the following iteration bound of rPDHG.

THEOREM 3.1. *Suppose Assumption 2.1 holds and* $Ac = 0$*. When running Algorithm 1 (rPDHG) with* $\tau = \frac{1}{2\kappa}$*,* $\sigma = \frac{1}{2\lambda_{\max}\lambda_{\min}}$ *and* $\beta := 1/e$ *to solve the LP, the total number of* ONEPDHG *iterations required to compute an* ε*-optimal solution is at most*

$$O\left(\kappa\Phi \cdot \ln\left(\kappa\Phi \cdot \frac{\|w^\star\|}{\varepsilon}\right)\right) . \tag{3.2}$$

This new computational guarantee for rPDHG is an accessible iteration bound, as it has a closed-form expression that can be easily computed once the optimal solution has been identified. Examining the definition of Φ in (3.1), $B^{-1}N$ is essentially the tabular of the tabular simplex method at the optimal basis B. Overall, Φ is in closed form of the optimal solution/basis, making it very amenable to computation. Given the optimal basis, the matrix $B^{-1}N$ can also be easily computed via one matrix factorization followed by one matrix multiplication. Overall, computing Φ is almost as easy as solving the LP itself. In addition, κ can be computed by one singular value decomposition of A. Consequently, the bound (3.2) in Theorem 3.1 is accessible because computing it does not require solving any additional optimization problems beyond the original LP.

Among κ and Φ, κ is a standard definition and easy to compute and analyze. It is solely determined by the matrix A and is independent of the problem's geometry. Conversely, although Φ is defined in terms of the matrix A and the optimal solutions, it is equivalent to an intrinsic measure of the geometry detailed later in Section 3.1. Indeed, Φ is not affected by any parameters of Algorithm 1. In addition, replacing the constraint $Ax = b$ with any preconditioned constraint $DAx = Db$ does not change Φ either.

Furthermore, Φ has the following simplified upper bound:

Proposition 3.1. *The following inequality holds for Φ:*

$$\Phi \leq \frac{\|x^\star + s^\star\|_1}{\min_{1\leq i\leq n}\left\{x_i^\star + s_i^\star\right\}} \cdot \|B^{-1}A\|_2 \ .$$

Due to complementary slackness, all components of $x^\star + s^\star$ are strictly positive, and $\min_{1\leq i\leq n}\left\{x_i^\star + s_i^\star\right\}$ represents the minimum nonzero component of $x^\star$ and $s^\star$. This upper bound is the product of two factors: (i) $\frac{\|x^\star+s^\star\|_1}{\min_{1\leq i\leq n}\left\{x_i^\star+s_i^\star\right\}}$, the ratio between the ℓ_1-norm and the smallest nonzero of the optimal solution, and (ii) $\|B^{-1}A\|_2$, the spectral norm of $B^{-1}A$. For readers familiar with simplex methods, $B^{-1}A$ is the simplex tabular at the optimal basis B. Its proof directly computes the relaxation of Φ; we defer it to Appendix A.

It should be noted that Φ is also relevant to condition numbers of other methods, beyond its connection to the tableau in simplex methods. Firstly, $\min_{1\leq i\leq n}\left\{x_i^\star + s_i^\star\right\}$ and the ratio $\frac{\|w^\star\|}{\min_{1\leq i\leq n}\left\{x_i^\star+s_i^\star\right\}}$ appear in classic complexity analyses of interior-point methods, including the convergence behavior (e.g., Güler and Ye [26]), finite convergence to optimal solutions (e.g., Ye [60]), and identification of the optimal face (e.g, Mehrotra and Ye [44]). Additionally, Lu and Yang [40] demonstrate that PDHG (without restarts) exhibits faster local linear convergence within a neighborhood whose size relates to $\min_{1\leq i\leq n}\left\{x_i^\star + s_i^\star\right\}$. Furthermore, $\frac{\|w^\star\|}{\min_{1\leq i\leq n}\left\{x_i^\star+s_i^\star\right\}}$ also appears in finite termination analysis of interior-point methods by Anstreicher et al. [3], Potra [50], in the form that is multiplied by certain norms of $B^{-1}N$. Later in Section 4, we will show that Φ also plays an important role in rPDHG's finite time identification of the optimal basis. Notably, while these condition numbers typically appear inside logarithmic terms in the complexity of interior-point methods, rPDHG's complexity is linear with respect to Φ. This suggests that Φ has more profound implications for the complexity and practical convergence rates of rPDHG compared to interior-point methods. Beyond interior-point methods, the upper bound $\frac{\|x^\star\|_1}{\min_{1\leq i\leq m} x_i^\star}$ also plays a crucial role in the complexity analysis of simplex and policy-iteration methods for discounted Markov decision problems (see Ye [62]).

The rest of this section presents the proof of Theorem 3.1. Section 3.1 recalls the sublevel set condition numbers defined by Xiong and Freund [58] and their roles in rPDHG. Furthermore, Section 3.1 shows a key lemma of the equivalence relationship between Φ and the sublevel set condition numbers, which helps prove Theorem 3.1. After that, Section 3.2 proves the key lemma of the equivalence relationship.

3.1. Sublevel-set condition numbers and the proof of Theorem 3.1 Recall that $\mathcal{F}_p = V_p \cap \mathbb{R}^n_+$ and $\mathcal{F}_d = V_d \cap \mathbb{R}^n_+$ denote the feasible sets of the primal and dual problems, respectively. Let $\mathcal{F} := \mathcal{F}_p \times \mathcal{F}_d$ represent the primal-dual feasible set of the solution pair (x, s). The optimal solution can then be characterized as

$$\mathcal{W}^\star := \mathcal{F} \cap \{w = (x, s) \in \mathbb{R}^{2n} : \mathrm{Gap}(w) = 0\} \ ,$$

the set of feasible solutions with zero duality gap. The δ-sublevel set is similarly characterized as the feasible solutions whose duality gap is less than or equal to δ, formally defined as follows:

Definition 3.1 (δ-sublevel set $\mathcal{W}_\delta$). For $\delta \geq 0$, the δ-sublevel set $\mathcal{W}_\delta$ is defined as:

$$\mathcal{W}_\delta := \mathcal{F} \cap \{w = (x,s) \in \mathbb{R}^{2n} : \mathrm{Gap}(w) \leq \delta\} \ . \tag{3.3}$$

Based on $\mathcal{W}_\delta$, Xiong and Freund [58] introduce the following two geometric condition numbers: the diameter D_δ and the conic radius r_δ.

Definition 3.2 (Condition numbers of $\mathcal{W}_\delta$). For $\delta \geq 0$, the diameter of $\mathcal{W}_\delta$ is defined as

$$D_\delta := \max_{u,v\in\mathcal{W}_\delta} \|u - v\| \ . \tag{3.4}$$

And the conic radius of $\mathcal{W}_\delta$ is the optimal objective value of the optimization problem

$$r_\delta := \left(\max_{w\in\mathcal{W}_\delta, r\geq 0} r \qquad \text{s.t.} \ \left\{\hat{w} : \|\hat{w} - w\| \leq r\right\} \subseteq \mathbb{R}^{2n}_+\right) \ , \tag{3.5}$$

which is also equal to $\left(\max_{w\in\mathcal{W}_\delta} \mathrm{Dist}(w, \partial\mathbb{R}^{2n}_+)\right)$ and $\left(\max_{w\in\mathcal{W}_\delta} \min_{i\in[2n]} w_i\right)$.

These condition numbers play a crucial role in the iteration bound of rPDHG as follows:

Lemma 3.1. *Under the same settings of the LP instance and Algorithm 1 with Theorem 3.1, the total number of OnePDHG iterations T required to obtain the first outer iteration N such that $w^{N,0} = (x^{n,0}, c - A^\top y^{n,0})$ is ε-optimal is bounded above by*

$$T \leq 380\kappa \left(\liminf_{\delta\searrow 0} \frac{D_\delta}{r_\delta}\right) \left[\ln\left(380\kappa\left(\liminf_{\delta\searrow 0}\frac{D_\delta}{r_\delta}\right)\right) + \ln\left(\frac{\|w^\star\|}{\varepsilon}\right)\right] \ . \tag{3.6}$$

This lemma is a simple extension of Theorem 3.5 of Xiong and Freund [58], and its complete proof is deferred to Appendix A. For simplicity of notations, the rest of the paper uses $\hat{\Phi}$ to denote $\left(\liminf_{\delta\searrow 0} \frac{D_\delta}{r_\delta}\right)$:

$$\hat{\Phi} := \liminf_{\delta\searrow 0} \frac{D_\delta}{r_\delta} \ . \tag{3.7}$$

Lemma 3.1 implies that the linear convergence rate is mostly determined by κ, a condition number of the constraint matrix, and $\hat{\Phi}$, a condition number of the sublevel set. When $\mathcal{W}^\star$ is a singleton, the sublevel set $\mathcal{W}_\delta$ is always inside the tangent cone to $\mathcal{F}$ at $w^\star$. Although looking similar, $\hat{\Phi}$ is not equivalent to the "width" of the tangent cone at $w^\star$ (see a formal definition in Freund and Vera [23]). The latter is an inherent property of the tangent cone, but the former is also influenced by the direction of (c, q).

Actually, we have the following critical lemma, which states the equivalence between Φ and $\hat{\Phi}$:

Lemma 3.2. *The geometric condition number $\hat{\Phi}$ and Φ are equivalent up to a constant factor of* 2*, i.e.,*

$$\Phi \leq \hat{\Phi} \leq 2\Phi \ . \tag{3.8}$$

Because of Lemma 3.2, all previous discussions for Φ also apply to $\hat{\Phi}$. In addition, we can now directly prove Theorem 3.1.

Proof of Theorem 3.1. Directly applying Lemma 3.2 in the iteration bound of Lemma 3.1 yields the desired iteration bound (3.2). □

Furthermore, since Φ is actually equivalent to $\hat{\Phi}$ up to a constant, the new accessible iteration bound of Theorem 3.1 is also equivalent to the iteration bound of Lemma 3.1 (an iteration bound of Xiong and Freund [58]) up to a constant. Xiong and Freund [58] point out that proper central path based Hessian rescaling can improve $\hat{\Phi}$ to at most $2n$, so Lemma 3.2 indicates that this rescaling can also improve Φ to at most $2n$.

We now prove Lemma 3.2 in Section 3.2.

3.2. Proof of the equivalence between Φ and $\hat{\Phi}$ The sublevel set can be equivalently regarded as the "primal sublevel set" for an artificial LP problem whose variables contain both the primal and dual variables. Section 3.2.1 extends the definitions of the sublevel set to the primal space only and demonstrates how to compute its condition numbers approximately. Subsequently, Section 3.2.2 illustrates how to approximate the sublevel-set condition numbers by treating them as primal sublevel-set condition numbers of an artificial problem, thereby proving Lemma 3.2.

3.2.1. Condition numbers of the primal sublevel set Recall that the primal feasible set is $\mathcal{F}_p := V_p \cap \mathbb{R}^n_+$, the intersection of the nonnegative orthant $\mathbb{R}^n_+$ and the affine subspace of the linear equality constraints. We define the objective error of x as $\mathrm{E}^{\mathrm{p}}_{\mathrm{obj}}(x) := c^\top x - f^\star$, where $f^\star$ is the optimal objective $c^\top x^\star$ for an optimal $x^\star$ of (1.1). The optimal primal solution is the feasible solution with zero objective error. The primal δ-sublevel set is then defined as:

$$\mathcal{X}_\delta := \mathcal{F}_p \cap \left\{ x \in \mathbb{R}^n : \mathrm{E}^{\mathrm{p}}_{\mathrm{obj}}(x) := c^\top x - f^\star \le \delta \right\} \ , \tag{3.9}$$

the sets of feasible primal solutions whose objective error does not exceed δ. Analogous to Definition 3.2, we define the diameter and conic radius of $\mathcal{X}_\delta$ as

$$D^p_\delta := \max_{u,v \in \mathcal{X}_\delta} \|u - v\| \quad \text{and} \quad r^p_\delta := \max_{x \in \mathcal{X}_\delta} \mathrm{Dist}(x, \mathbb{R}^n_+) \ . \tag{3.10}$$

Now we show the representation of $\mathcal{X}_\delta$ and how to compute D^p_δ and r^p_δ.

Convex hull representation of the primal sublevel set $\mathcal{X}_\delta$. Under Assumption 2.1, the optimal primal and dual solutions are unique and nondegenerate, corresponding to a unique optimal basis. Each edge emanating from $\mathcal{X}^\star = \{x^\star\}$ connects to a basic feasible solution of an adjacent basis. There are exactly $n - m$ entering basic variables. Let the corresponding directions of these edges be given by the vectors $u^1, u^2, \dots, u^{n-m} \in \mathbb{R}^n$. Since $\Theta = [m]$ and $\bar{\Theta} = [n] \setminus [m]$, these vectors can be computed as follows:

$$u^j_{[m]} := -B^{-1} N_{\cdot,j}, \ \ u^j_{m+j} := 1, \ \ \text{and} \ \ u^j_k := 0 \text{ for all } k \notin [m] \text{ and } k \neq m + j \tag{3.11}$$

for each $j \in [n - m]$. Therefore, the $n - m$ edges are as follows:

$$\mathcal{E}^j := \left\{ x^\star + \theta \cdot u^j : \theta \ge 0, \ x^\star + \theta \cdot u^j \in \mathcal{F}_p \right\} \quad \text{for each } j \in [n - m] \ . \tag{3.12}$$

If $u^j \ge 0$, then $\mathcal{E}^j$ is an extreme ray. Otherwise, $\mathcal{E}^j$ connects to an adjacent basic feasible solution of $x^\star$. Based on these edges, when δ is sufficiently small so that it is no larger than the extreme points' best nonzero objective error $\bar{\delta}_p$ (which is always strictly positive for LP) defined by

$$\bar{\delta}_p := \begin{cases} \min\{\mathrm{E}^{\mathrm{p}}_{\mathrm{obj}}(x) : x \in \mathrm{EP}_{\mathcal{F}_p} \setminus \mathcal{X}^\star\} & \text{if } \mathrm{EP}_{\mathcal{F}_p} \setminus \mathcal{X}^\star \neq \emptyset \\ +\infty & \text{if } \mathrm{EP}_{\mathcal{F}_p} \setminus \mathcal{X}^\star = \emptyset \end{cases} \tag{3.13}$$

then $\mathcal{X}_\delta$ can be represented as the convex hull of these edges. Here we use $\mathrm{EP}_{\mathcal{F}_p}$ to denote the set of extreme points of $\mathcal{F}_p$.

Lemma 3.3. *Suppose that Assumption 2.1 holds and $\delta \in (0, \infty)$ lies in $(0, \bar{\delta}_p]$. Then $\mathcal{X}_\delta$ is represented by the following convex hull formulation:*

$$\mathcal{X}_\delta = \mathrm{Conv}\left(\{x^\star\} \cup \{x^j : j \in [n - m]\} \right) \tag{3.14}$$

where

$$x^j := x^\star + \frac{\delta}{s^\star_{m+j}} \cdot u^j \textit{ for each } j \in [n - m] \ . \tag{3.15}$$

Proof. Because $\delta \in (0, \bar{\delta}_p]$, the halfspace $\{x : c^\top x < f^\star + \bar{\delta}\}$ contains no other basic feasible solutions of $\mathcal{F}_p$ and intersect no other edges except the $n-m$ edges emanating from $\mathcal{X}^\star$, namely $\mathcal{E}^1, \mathcal{E}^2, \dots, \mathcal{E}^{n-m}$. Therefore, in addition to $x^\star$, the other extreme points of $\mathcal{X}_\delta$ are the intersection points of the hyperplane $\{x : c^\top x = f^\star + \delta\}$ and the edges $\mathcal{E}^1, \mathcal{E}^2, \dots, \mathcal{E}^{n-m}$. These intersection points all exist because $\delta \in (0, \bar{\delta}_p] \cap (0, \infty)$. Moreover, they are precisely the $\{x^j : j \in [n-m]\}$ defined in (3.15), as the objective errors $\mathrm{E}^{\mathrm{p}}_{\mathrm{obj}}(x^j)$ all equal δ (which is because $\mathrm{E}^{\mathrm{p}}_{\mathrm{obj}}(x^j) = (x^j)^\top s^\star = x^j_{m+j} s^\star_{m+j} = \frac{\delta}{s^\star_{m+j}} \cdot u^j_{m+j} \cdot s^\star_{m+j} = \delta u^j_{m+j} = \delta$). Therefore, $\mathcal{X}_\delta$ is indeed the convex hull of the $n-m+1$ points in $\{x^\star\}$ and $\{x^j : j \in [n-m]\}$. $\square$

Computing the diameter and conic radius of $\mathcal{X}_\delta$. We now provide an approximation of D^p_δ for sufficiently small δ.

LEMMA 3.4. *Suppose that Assumption 2.1 holds and $\delta \in (0, \bar{\delta}_p]$. Then we have*

$$\delta \cdot \max_{1 \le j \le n-m} \frac{\sqrt{\left\|B^{-1} N_{\cdot,j}\right\|^2 + 1}}{s^\star_{m+j}} \le D^p_\delta \le 2\delta \cdot \max_{1 \le j \le n-m} \frac{\sqrt{\left\|B^{-1} N_{\cdot,j}\right\|^2 + 1}}{s^\star_{m+j}} . \tag{3.16}$$

Proof. The diameter of a polyhedron is the maximum distance between any two extreme points of the polyhedron. Thus, $D^p_\delta \le \max_{i,j \in [n-m]} \|x^i - x^\star\| + \|x^j - x^\star\| \le 2 \cdot \max_{1 \le j \le n-m} \|x^j - x^\star\| = 2\delta \cdot \max_{1 \le j \le n-m} \frac{\|u^j\|}{s^\star_{m+j}}$, where the last equality uses Lemma 3.3. Conversely, $D^p_\delta \ge \max_{1 \le j \le n-m} \|x^j - x^\star\|$, which equals $\delta \cdot \max_{1 \le j \le n-m} \frac{\|u^j\|}{s^\star_{m+j}}$ (by Lemma 3.3). Finally, note from (3.11) that $\|u^j\| = \sqrt{\left\|B^{-1} N_{\cdot,j}\right\|^2 + 1}$ so the proof is completed. $\square$

Next, we show how to exactly compute r_δ when δ is sufficiently small. Specifically, we study δ small enough so that for $\{x^j : 1 \le j \le n-m\}$ defined in (3.15):

$$\min_{1 \le k \le m} x^j_k \ge x^j_{m+j} \text{ for all } j \in [n-m] . \tag{3.17}$$

In other words, δ is small enough such that x^j_{m+j} is one of the smallest nonzeros of x^j for all $j \in [n-m]$. Note that x^j is continuous with respect to the value of δ, and when $\delta = 0$, all the $n-m$ inequalities of (3.17) hold strictly. Therefore, there must exist a nonempty neighborhood of 0 in which all values of δ satisfy (3.17).

LEMMA 3.5. *Suppose that Assumption 2.1 holds and δ is sufficiently small so that $\delta \le \bar{\delta}_p$ and* (3.17) *holds. Then we have:*

$$r^p_\delta = \frac{\delta}{\|s^\star\|_1} . \tag{3.18}$$

Proof. By definition, $r^p_\delta = \max_{x \in \mathcal{X}_\delta} \mathrm{Dist}(x, \partial \mathbb{R}^n_+) = \max_{x \in \mathcal{X}_\delta} \min_{1 \le l \le n} x_l$. Since $\delta \le \bar{\delta}_p$, using the convex hull formulation of $\mathcal{X}_\delta$ presented in Lemma 3.3, r^p_δ can be equivalently written as:

$$r^p_\delta = \max_{\substack{\lambda \in \mathbb{R}^{n-m+1}_+ \\ \sum_{j=1}^{n-m+1} \lambda_j = 1}} \min_{1 \le l \le n} x(\lambda)_l \quad \text{in which} \quad x(\lambda) := \lambda_{n-m+1} \cdot x^\star + \sum_{j=1}^{n-m} \lambda_j \cdot x^j . \tag{3.19}$$

Due to the above definition of $x(\lambda)$ and the definition of x^j in (3.15), for each $m+j$ in $\bar{\Theta} = [n] \setminus [m]$, the component $x(\lambda)_{m+j}$ is given by $\lambda_j \cdot x^j_{m+j}$.

We now claim that a smallest component of $x(\lambda)$ is of an index in $[n] \setminus [m]$. On the one hand,

$$\min_{1 \le i \le m} x(\lambda)_i \overset{(3.19)}{\ge} \lambda_{n-m+1} \cdot \min_{1 \le i \le m} x^\star_i + \sum_{j=1}^{n-m} \lambda_j \cdot \min_{1 \le i \le m} x^j_i \ge 0 + \sum_{j=1}^{n-m} \lambda_j \cdot x^j_{m+j} . \tag{3.20}$$

where the last inequality uses (3.17) because δ is sufficiently small. On the other hand, because $x(\lambda)_{m+j} = \lambda_j \cdot x^j_{m+j}$,

$$\sum_{j=1}^{n-m} \lambda_j \cdot x^j_{m+j} = \sum_{j=1}^{n-m} x(\lambda)_{m+j} \geq \min_{1\leq j\leq n-m} x(\lambda)_{m+j} \; . \tag{3.21}$$

Overall, for this small δ, (3.20) and (3.21) ensure that a smallest component of $x(\lambda)$ is of an index in $[n] \setminus [m]$.

Consequently, when computing r^p_δ we only need to consider the components in $[n] \setminus [m]$.

$$r^p_\delta \overset{(3.19)}{=} \max_{\substack{\lambda\in\mathbb{R}^{n-m+1}_+ \\ \sum_{j=1}^{n-m+1}\lambda_j=1}} \min_{1\leq l\leq n} x(\lambda)_l = \max_{\substack{\lambda\in\mathbb{R}^{n-m+1}_+ \\ \sum_{j=1}^{n-m+1}\lambda_j=1}} \min_{m+1\leq l\leq n} x(\lambda)_l = \max_{\substack{\lambda\in\mathbb{R}^{n-m+1}_+ \\ \lambda_{n-m+1}=0 \\ \sum_{j=1}^{n-m}\lambda_j=1}} \min_{m+1\leq l\leq n} x(\lambda)_l \tag{3.22}$$

where the last equality follows from $x^\star_{[n]\setminus[m]} = 0$, which implies that λ_{n-m+1} in an optimal λ for r^p_δ must be 0. The value of r^p_δ in (3.22) is then equal to the optimal objective of

$$\begin{pmatrix} \max\limits_{\lambda\in\mathbb{R}^{n-m}} & \min\limits_{1\leq j\leq n-m} \lambda_j \cdot x^j_{m+j} \\ \text{s.t.} & \sum\limits_{1\leq j\leq n-m} \lambda_j = 1,\ \lambda\geq 0 \end{pmatrix} = \begin{pmatrix} \max\limits_{\lambda\in\mathbb{R}^{n-m}} & \min\limits_{1\leq j\leq n-m} \lambda_j \cdot \dfrac{\delta}{s^\star_{m+j}} \\ \text{s.t.} & \sum\limits_{1\leq j\leq n-m} \lambda_j = 1,\ \lambda\geq 0 \end{pmatrix} \tag{3.23}$$

where the equality uses $x^j_{m+j} = \frac{\delta}{s_{m+j}}$ according to (3.15) and (3.11). Finally, the optimal solution $\lambda^\star$ of (3.23) is given by $\lambda^\star_j = \frac{s^\star_{m+j}}{\sum_{k=1}^{n-m} s^\star_{m+k}} = \frac{s^\star_{m+j}}{\|s^\star\|_1}$ for each j, and the optimal objective is equal to $\frac{\delta}{\|s^\star\|_1}$. This establishes (3.18) and completes the proof. □

It is noteworthy that if we similarly define the dual sublevel set $\mathcal{S}_\delta$ and then Freund [22, Theorem 3.2.] shows $\delta \leq r^p_\delta \cdot \max_{s\in\mathcal{S}_\delta} \|s\|_1 \leq 2\delta$. A direct application of this result yields $\frac{1}{\|s^\star\|_1} \leq \lim_{\delta\to 0} \frac{r^p_\delta}{\delta} \leq \frac{2}{\|s^\star\|_1}$. However, Lemma 3.5 provides a slightly stronger result by precisely computing r^p_δ.

3.2.2. Proof of Lemma 3.2 In this subsection, we prove Lemma 3.2. We begin by demonstrating that the sublevel set $\mathcal{W}_\delta$ can be regarded as a primal sublevel set of an artificial problem. Using the results of Section 3.2.1, we then show how to approximate D_δ and compute r_δ, which subsequently allows us to approximate $\hat{\Phi}$ and prove Lemma 3.2.

Problem (2.3) can be transformed into the subsequent standard-form problem

$$\max_{s\in\mathbb{R}^n} \; q^\top(c-s) \quad \text{ s.t. } Qs = Qc,\ s\geq 0 \tag{3.24}$$

for any $Q \in \mathbb{R}^{(n-m)\times n}$ whose rows are linearly independent and orthogonal to the rows of A so that $\mathrm{Null}(Q) = \mathrm{Im}(A^\top)$. This equivalence holds because $\mathrm{Im}(A^\top) + c$ in (2.3) is identical to $\{s : Qs = Qc\}$ in (3.24). For problem (3.24), the optimal basis is $\bar{\Theta} = [n] \setminus [m]$, the complement of Θ. Although multiple choices of Q exist, the matrix $Q^{-1}_{\bar{\Theta}} Q_\Theta$ is always equal to $-(B^{-1}N)^\top$.

Lemma 3.6. *Suppose that Assumption 2.1 holds. The matrix $Q^{-1}_{\bar{\Theta}} Q_\Theta$ is equal to $-(B^{-1}N)^\top$.*

Proof. Given that $\mathrm{Null}(Q) = \mathrm{Im}(A^\top)$, we have $QA^\top = 0$, i.e.,

$$0 = QA^\top = \begin{pmatrix} Q_\Theta & Q_{\bar{\Theta}} \end{pmatrix} \begin{pmatrix} A^\top_\Theta \\ A^\top_{\bar{\Theta}} \end{pmatrix} = Q_\Theta A^\top_\Theta + Q_{\bar{\Theta}} A^\top_{\bar{\Theta}} = Q_\Theta B^\top + Q_{\bar{\Theta}} N^\top \; .$$

Since the optimal bases B and $Q_{\bar{\Theta}}$ are of full rank, we obtain $Q^{-1}_{\bar{\Theta}} Q_\Theta = -N^\top (B^\top)^{-1} = -(B^{-1}N)^\top$. □

Overall, the primal problem (1.1) and the dual problem (3.24) can be combined and reformulated as an equivalent standard-form LP problem in the product space of x and s:

$$\min_{w=(x,s)\in\mathbb{R}^{2n}} \begin{pmatrix} c \\ q \end{pmatrix}^\top w \quad \text{s.t.} \begin{pmatrix} A & 0 \\ 0 & Q \end{pmatrix} w = \begin{pmatrix} b \\ Qc \end{pmatrix},\ w \geq 0\ . \tag{3.25}$$

The above (3.25) is also in standard form, and satisfies Assumption 2.1. Furthermore, the duality gap $\mathrm{Gap}(x,s)$ of any (x,s) is the same as the objective error for (3.25) because

$$\mathrm{Gap}(x,s) = c^\top x - q^\top(c-s) = \left(\begin{smallmatrix} c \\ q \end{smallmatrix}\right)^\top w - q^\top c = \left(\begin{smallmatrix} c \\ q \end{smallmatrix}\right)^\top (w - w^\star) \tag{3.26}$$

where the last equality follows from $0 = \mathrm{Gap}(w^\star) = c^\top x^\star - q^\top(c - s^\star) = \left(\begin{smallmatrix} c \\ q \end{smallmatrix}\right)^\top w^\star - q^\top c$. The right-hand side of (3.26) is the objective error $\mathrm{E_{obj}}(w)$ of w, defined by $\mathrm{E_{obj}}(w) := \left(\begin{smallmatrix} c \\ q \end{smallmatrix}\right)^\top (w - w^\star)$. Consequently, (3.26) implies that the δ-sublevel set $\mathcal{W}_\delta$ is identical to the primal δ-sublevel set of (3.25). Therefore, utilizing the results in Section (3.2.1), we can directly approximately compute D_δ and r_δ, by treating them as the condition numbers of the primal sublevel set of (3.25).

Lemma 3.7. *Suppose that Assumption 2.1 holds. There exists a positive $\bar\delta$ such that for any $0 \leq \delta \leq \bar\delta$, it holds that*

$$\hat D_\delta \leq D_\delta \leq 2\hat D_\delta \quad \textit{and} \quad r_\delta = \frac{\delta}{\|x^\star\|_1 + \|s^\star\|_1}\ , \tag{3.27}$$

where

$$\hat D_\delta := \delta \cdot \max\left\{ \max_{1\leq j\leq n-m} \frac{\sqrt{\left\|(B^{-1}N)_{\cdot,j}\right\|^2+1}}{s^\star_{m+j}},\ \max_{1\leq i\leq m} \frac{\sqrt{\left\|(B^{-1}N)_{i,\cdot}\right\|^2+1}}{x^\star_i} \right\}\ . \tag{3.28}$$

Proof. Let H denote the constraint matrix of (3.24) for simplicity of notations. We use Ω to represent the indices of the optimal basis of (3.24), which is $\Theta \cup (n+\bar\Theta)$. Here $n+\bar\Theta$ denotes $\{n+j : j\in\bar\Theta\}$. Similarly, the complementary set is $\bar\Omega = \bar\Theta \cup (n+\Theta)$. We use $\Omega(i)$ to denote the i-th smallest index component of Ω and use $\bar\Omega(j)$ to denote the j-th smallest index component of $\bar\Omega$. Both Ω and $\bar\Omega$ contain exactly n components, and the optimal basis H_Ω is given by $\left(\begin{smallmatrix} A_\Theta & 0 \\ 0 & Q_{\bar\Theta} \end{smallmatrix}\right)$, or equivalently $\left(\begin{smallmatrix} B & 0 \\ 0 & Q_{\bar\Theta} \end{smallmatrix}\right)$. By an approach similar to that used in Section 2.1 for deriving the dual problem, the dual problem of (3.25) is symmetric with (3.25) and also has a unique optimal solution $\tilde w^\star$ equal to $(s^\star, x^\star)$.

We now prove the first half of (3.27) using Lemma 3.4. The term $\sqrt{\left\|B^{-1}N_{\cdot,j}\right\|^2+1}$ in (3.16) is $\sqrt{\left\|H_\Omega^{-1}H_{\cdot,\bar\Omega(j)}\right\|^2+1}$ for problem (3.25). And $s^\star_{\bar\Theta}$ in (3.16) is $\tilde w_{\bar\Omega}$ in problem (3.25). Therefore, Lemma 3.4 implies

$$\bar D_\delta \leq D_\delta \leq 2\bar D_\delta, \quad \text{where } \bar D_\delta := \delta\cdot\max_{j\in[n]} \frac{\sqrt{\left\|H_\Omega^{-1}H_{\cdot,\bar\Omega(j)}\right\|^2+1}}{\tilde w_{\bar\Omega(j)}}\ . \tag{3.29}$$

To compute the value of $\bar D_\delta$, we consider two cases based on the structure of $\bar\Omega = \bar\Theta\cup(n+\Theta)$. When $\bar\Omega(j)\in\bar\Theta$, we have $\tilde w^\star_{\bar\Omega(j)} = s^\star_{\bar\Omega(j)}$, and $H_\Omega^{-1}H_{\cdot,\bar\Omega(j)} = H_\Omega^{-1}\left(\begin{smallmatrix} A_{\cdot,\bar\Omega(j)} \\ 0 \end{smallmatrix}\right) = \left(\begin{smallmatrix} B^{-1}A_{\cdot,\bar\Omega(j)} \\ 0 \end{smallmatrix}\right)$. When $\bar\Omega(j)\in n+\Theta$, we have $\tilde w^\star_{\bar\Omega(j)} = x^\star_{\bar\Omega(j)-n}$, and $H_\Omega^{-1}H_{\cdot,\bar\Omega(j)} = H_\Omega^{-1}\left(\begin{smallmatrix} 0 \\ Q_{\cdot,\bar\Omega(j)-n} \end{smallmatrix}\right) = \left(\begin{smallmatrix} 0 \\ Q_{\bar\Theta}^{-1}Q_{\cdot,\bar\Omega(j)-n} \end{smallmatrix}\right) = \left(\begin{smallmatrix} 0 \\ -((B^{-1}N)^\top)_{\cdot,\bar\Omega(j)-n} \end{smallmatrix}\right)$, where the last inequality uses Lemma 3.6. Therefore,

$$\begin{aligned}
\bar D_\delta &= \delta\cdot\max\left\{ \max_{\bar\Omega(j)\in\bar\Theta} \frac{\sqrt{\left\|\left(\begin{smallmatrix} B^{-1}A_{\cdot,\bar\Omega(j)} \\ 0 \end{smallmatrix}\right)\right\|^2+1}}{s^\star_{\bar\Omega(j)}},\ \max_{\bar\Omega(j)\in n+\Theta} \frac{\sqrt{\left\|\left(\begin{smallmatrix} 0 \\ -((B^{-1}N)^\top)_{\cdot,\bar\Omega(j)-n} \end{smallmatrix}\right)\right\|^2+1}}{x^\star_{\bar\Omega(j)-n}} \right\} \\
&\overset{(2.8)}{=} \delta\cdot\max\left\{ \max_{1\leq j\leq n-m} \frac{\sqrt{\left\|(B^{-1}N)_{\cdot,j}\right\|^2+1}}{s^\star_{m+j}},\ \max_{1\leq i\leq m} \frac{\sqrt{\left\|(B^{-1}N)_{i,\cdot}\right\|^2+1}}{x^\star_i} \right\} = \hat D_\delta\ .
\end{aligned}$$

Finally, substituting $\bar{D}_\delta = \hat{D}_\delta$ back to (3.29) proves the first half of (3.27).

As for the second half of (3.27), note that $\tilde{w}^\star = (s^\star, x^\star)$, so by Lemma 3.5, when δ is sufficiently small we have $r_\delta = \frac{\delta}{\|\tilde{w}^\star\|_1} = \frac{\delta}{\|x^\star\|_1 + \|s^\star\|_1}$. □

Finally, we are ready to prove Lemma 3.2.

Proof of Lemma 3.2. Lemma 3.7 establishes that $\hat{D}_\delta \le D_\delta \le 2\hat{D}_\delta$ for sufficiently small δ, so we can deduce $\lim_{\delta\searrow 0} \frac{\hat{D}_\delta}{r_\delta} \le \lim_{\delta\searrow 0} \frac{D_\delta}{r_\delta} \le 2 \cdot \lim_{\delta\searrow 0} \frac{\hat{D}_\delta}{r_\delta}$. Note that $\hat{\Phi} = \lim_{\delta\searrow 0} \frac{D_\delta}{r_\delta}$ as defined in (3.7). Substituting the values of $\hat{D}_\delta$ and r_δ in Lemma 3.7 into (3.1) yields $\Phi = \lim_{\delta\searrow 0} \frac{\hat{D}_\delta}{r_\delta}$. Therefore, $\Phi \le \hat{\Phi} \le 2\Phi$. □

4. Finite-Time Optimal Basis Identification and Fast Local Convergence In this section, we investigate the two-stage performance of rPDHG. It is frequently observed in practice that the behavior of rPDHG transitions to faster local linear convergence in a neighborhood of the optimal solution, in which the support sets of all iterates keep consistent. Lu and Yang [40] study this phenomenon for vanilla PDHG. In the first stage, PDHG converges in a sublinear rate until identifying the active set of the converging solution. In the second stage, PDHG turns to faster local linear convergence. Recently, Lu and Yang [42] propose the restarted Halpern PDHG (rHPDHG), a variant of rPDHG, and prove its two-stage performance behavior. However, the iteration bounds of the two stages proven above (for both PDHG and rHPDHG) are not accessible because they both depend on the Hoffman constant of a linear system that is determined by the converging solution. The Hoffman constant is hard to analyze, challenging to compute, and may be too conservative. And computing the trajectory of the algorithm may also be difficult.

Based on the new iteration bounds obtained in Section 3 for LPs with unique optima, this section will show an accessible refined convergence guarantee of rPDHG that avoids Hoffman constants entirely. Although using the additional assumption of unique optimum, the new iteration bounds for the two stages both have closed-form expressions and are thus straightforward to analyze and compute:

- In Stage I, rPDHG identifies the optimal basis within at most $O(\kappa\Phi \cdot \ln(\kappa\Phi))$ iterations.
- In Stage II, having identified the optimal basis Θ, the behavior of rPDHG transitions to faster local linear convergence that is no longer related to Φ. In this stage, components of index in Θ are sufficiently bounded away from 0, while all other components equal 0.

The following theorem summarizes the iteration bounds of the two stages:

Theorem 4.1. *Suppose Assumption 2.1 and $Ac = 0$. Let Algorithm 1 (rPDHG) run with $\tau = \frac{1}{2\kappa}$, $\sigma = \frac{1}{2\lambda_{\max}\lambda_{\min}}$ and $\beta := 1/e$ to solve the LP. Let T_1 be the total number of OnePDHG iterations required to obtain N_1 such that for all $N \ge N_1$ the positive components of $x^{N,0}$ exactly correspond to the optimal basis. Then,*

$$T_1 \le T_{basis}\ , \quad \textit{in which} \quad T_{basis} := O\left(\kappa\Phi \cdot \ln\left(\kappa\Phi\right)\right). \tag{4.1}$$

Furthermore, let T_2 be the total number of OnePDHG iterations required to obtain the first N_2 for which $w^{N_2,0}$ is ε-optimal. Then,

$$T_2 \le T_{basis} + T_{local}, \quad \textit{in which} \quad T_{local} := O\left(\|B^{-1}\|\|A\| \cdot \max\left\{0,\ \ln\left(\frac{\min_{1\le i\le n}\left\{x_i^\star + s_i^\star\right\}}{\varepsilon}\right)\right\}\right). \tag{4.2}$$

Theorem 4.1 demonstrates that it takes at most T_{basis} iterations for rPDHG to identify the optimal basis (independent of ε), after which it requires at most additional T_{local} iterations to achieve ε-optimality. Both T_{basis} and T_{local} are accessible, because they do not contain any Hoffman constant and are easy to compute and analyze if the optimal solution is known. Furthermore, T_{local} is independent of Φ, which is frequently considerably larger. This provides a partial explanation for why rPDHG often becomes significantly faster in Stage II compared to Stage I.

The first-stage iteration bound T_{basis} exhibits a linear relationship with $\frac{\|w^\star\|_1}{\xi}$ (in the expression of Φ), where we use ξ to denote the smallest nonzero of $x^\star$ and $s^\star$, written as follows:

$$\xi := \min_{1\le i\le n}\left\{x_i^\star + s_i^\star\right\}\ . \tag{4.3}$$

The first-stage iteration bounds of the vanilla PDHG (without restarts) and the restarted Halpern PDHG proven by Lu and Yang [40, 42] also depend on $\frac{\|w^\star\|_1}{\xi}$ (in a different norm), but they are further multiplied by a Hoffman constant of a linear system. A similar dependence on $\frac{\|w^\star\|}{\xi}$ and certain norms of $B^{-1}A$ is also observed in the complexity analysis of finite-termination results for IPMs, such as Anstreicher et al. [3], Potra [50]. In those cases of IPMs, this dependence appears only within logarithmic terms, but it comes with higher per-iteration cost.

Once the optimal basis is identified, the method's behavior automatically transitions into the second stage without changing the algorithm or additional "crossover" operations. Notably, the coefficient $\|B^{-1}\|\|A\|$ of the linear convergence is independent of $w^\star$ and is solely determined by the optimal basis and the constraint matrix. In other words, it can also be upper bounded by a constant that is only determined by the constraint matrix. Similar results also hold for IPMs, proven by Vavasis and Ye [53]. One can observe that a small ξ may slightly decrease T_{local} but simultaneously significantly increase T_{basis} because T_{basis} is linear in $\frac{\|w^\star\|_1}{\xi}$.

The significance of $\kappa\Phi$ and $\|B^{-1}\|\|A\|$ in the two-stage performance will be empirically confirmed in Section 6. The remainder of this section will prove Theorem 4.1. The key of the proof is to utilize the adaptive restart condition (β-restart condition in Line 8 of Algorithm 1) and show that the number of inner-loop iterations of Stage II becomes different than Stage I. To this end, Section 4.1 recalls the adaptive restart condition and Section 4.2 proves Theorem 4.1.

4.1. Adaptive restart condition The β-restart condition is built on a metric of stationarity, called the "normalized duality gap," proposed by Applegate et al. [6].

Definition 4.1 (Normalized duality gap, Applegate et al. [6]). For any $z = (x, y) \in \mathbb{R}^{m+n}$ and a certain $r > 0$, the normalized duality gap of the saddlepoint problem (2.2) is then defined as

$$\rho(r; z) := \frac{1}{r} \sup_{\hat{z} \in B(r;z)} \left[L(x, \hat{y}) - L(\hat{x}, y) \right] \tag{4.4}$$

in which $B(r; z) := \left\{ \hat{z} := (\hat{x}, \hat{y}) : \hat{x} \in \mathbb{R}^n_+ \text{ and } \|\hat{z} - z\|_M \leq r \right\}$.
The normalized duality gap $\rho(r; z)$ is also a valid measure of the optimality and feasibility errors of z, and it can be efficiently computed or approximated in strongly polynomial time as shown by Applegate et al. [6]. Furthermore, let the k-th iterate of PDHG be denoted as z^k, and let the average of the first k iterates be $\bar{z}^k := \frac{1}{k} \sum_{i=1}^k z^i$. The normalized duality gap $\rho(\|z^0 - \bar{z}^k\|_M; \bar{z}^k)$ of the average iterate $\bar{z}^k$ converges to 0 sublinearly. See Applegate et al. [6], Xiong and Freund [56] for more details.

Then the β-restart condition is hold when $n = 0$ and $k = 1$, or

$$\rho(\|\bar{z}^{n,k} - z^{n,0}\|_M; \bar{z}^{n,k}) \leq \beta \cdot \rho(\|z^{n,0} - z^{n-1,0}\|_M; z^{n,0}) , \tag{4.5}$$

for a chosen value of $\beta \in (0, 1)$. This criterion is nearly identical to the condition used by Applegate et al. [6]. It essentially stipulates that the normalized duality gap diminishes by a factor of β between outer loop iterates $z^{n+1,0}$ and $z^{n,0}$.

For LP problems, it can be shown that there exists a constant $\mathcal{L} \geq 0$ such that the following condition holds for all $n \geq 1$:

$$\mathrm{Dist}_M(z^{n,0}, \mathcal{Z}^\star) \leq \rho(\|z^{n,0} - z^{n-1,0}\|_M; z^{n,0}) \cdot \mathcal{L} . \tag{4.6}$$

This condition indicates the M-distance to the optimal solutions is upper bounded by the normalized duality gap multiplied by the fixed constant $\mathcal{L}$. Applegate et al. [6] refer to this as the "sharpness" property of the normalized duality gap. Under this condition, each inner loop requires at most $\left\lceil \frac{8\mathcal{L}}{\beta} \right\rceil$ iterations to achieve sufficient decrease in the normalized duality gap of Applegate et al. [6, Theorem 2] (see Appendix A for a formal statement and proof). Therefore, n outer loops contain at most $O\left(\frac{n\mathcal{L}}{\beta}\right)$ OnePDHG iterations, while reducing the normalized duality gap to a β^n fraction of its initial value. This leads to a linear convergence rate dependent on $\mathcal{L}$; see Applegate et al. [6, Theorem 2] or a more general version in Xiong and Freund [58, Theorem 3.5]. Indeed, Lemma 3.13 of Xiong and Freund [58] demonstrates that (4.6) always holds with $\mathcal{L} = O(\kappa\hat{\Phi})$, which then leads to the global convergence rate in Lemma 3.1.

4.2. Technical lemmas of Theorem 4.1 The linear convergence of rPDHG relies on the condition (4.6). If (4.6) holds for all $n \geq 1$ with a constant $\mathcal{L} \geq 0$, then the global linear convergence is established by showing the number of iterations required for each inner loop does not exceed $O(\frac{\mathcal{L}}{\beta})$. We will now demonstrate the existence of a close neighborhood of the optimal solution $z^\star$ within which (4.6) holds with a potentially much smaller $\mathcal{L}$ than the global $\mathcal{L}$, so rPDHG requires a significantly smaller number of iterations for the inner loops and therefore have a much faster local linear convergence rate in this neighborhood. Our analysis will proceed in two steps with two lemmas, followed by their proofs.

In the first step, we demonstrate that the condition (4.6) indeed holds with an alternative $\mathcal{L}$ for the iterates $\bar{z} = (\bar{x}, \bar{y})$ if (i) each component of $\bar{x}_\Theta$ is sufficiently bounded away from 0 and (ii) each component of $\bar{x}_{\bar{\Theta}}$ equals zero. For simplicity, we still assume that the optimal basis is $\{1, 2, \ldots, m\}$ for simplicity. We will use the following step-size dependent constant:

$$c_{\tau,\sigma} := \max\left\{\frac{1}{\sqrt{\sigma}\lambda_{\min}}, \frac{1}{\sqrt{\tau}}\right\} . \tag{4.7}$$

When $\tau = \frac{1}{2\kappa}$ and $\sigma = \frac{1}{2\lambda_{\max}\lambda_{\min}}$ (the step-sizes used by Theorem 4.1), we have $c_{\tau,\sigma} = \sqrt{2\kappa}$.

Lemma 4.1. *Under Assumption 2.1, for any $\bar{z} = (\bar{x}, \bar{y})$ and $r > 0$ such that*

$$(i)\ \ \bar{x}_i \geq r\sqrt{\tau} \text{ for } i \in [m], \quad \text{and} \quad (ii)\ \bar{x}_{m+j} = 0 \text{ for } j \in [n-m] , \tag{4.8}$$

it holds that

$$\|\bar{z} - z^\star\|_M \leq \sqrt{2}c_{\tau,\sigma}\|B^{-1}\|\sqrt{\frac{1}{\sigma} + \frac{\|A\|^2}{\tau}} \cdot \rho(r; \bar{z}) . \tag{4.9}$$

When the step-sizes are carefully chosen, (4.9) in Lemma 4.1 can be further simplified:

Remark 4.1. With the choice of step-size $\tau = \frac{1}{2\kappa}$ and $\sigma = \frac{1}{2\lambda_{\max}\lambda_{\min}}$, (4.9) becomes

$$\|\bar{z} - z^\star\|_M \leq 4\|B^{-1}\|\|A\| \cdot \rho(r; \bar{z}) . \tag{4.10}$$

In the second step, we show that the two conditions in Lemma 4.1 are automatically satisfied by all average iterations $\bar{z}^{N,k} = (\bar{x}^{N,k}, \bar{y}^{N,k}) = \frac{1}{k}\sum_{i=1}^k z^{N,i}$ once a previous outer loop iteration $z^{n,0}$ (with $n \leq N$) is sufficiently close to the optimal solution $z^\star$ under the M-norm distance.

Lemma 4.2. *Under Assumption 2.1, suppose that Algorithm 1 (rPDHG) has any β-restart condition, and let the step-sizes satisfy* (2.7) *strictly. If there exists n such that*

$$\left\|z^{n,0} - z^\star\right\|_M \leq \bar{\varepsilon} := \frac{\sqrt{1 - \sqrt{\tau\sigma}\|A\|}}{3} \cdot \min\left\{\frac{1}{\sqrt{\tau}}, \sqrt{\tau}\right\} \cdot \left(\min_{1 \leq i \leq n}\ \left\{x_i^\star + s_i^\star\right\}\right) , \tag{4.11}$$

then for any $N \geq n$ and $k \geq 1$, the following conditions hold:

$$(i)\ \ \bar{x}_i^{N,k} \geq \sqrt{\tau}\left\|\bar{z}^{N,k} - z^{N,0}\right\|_M \ \text{ for } i \in [m] , \quad \text{and} \quad (ii)\ \bar{x}_{m+j}^{N,k} = 0 \text{ for } j \in [n-m] . \tag{4.12}$$

Moreover, the positive components of $\bar{x}^{N,k}$ correspond exactly to the optimal basis.

Lemmas 4.1 and 4.2 provide the foundation for a refined complexity analysis of rPDHG. In Stage I, rPDHG converges to a neighborhood of the optimal solution such that condition (4.11) of Lemma 4.2 is satisfied. The number of iterations in this stage is determined by the linear convergence rate established in Lemma 3.1. In Stage II, rPDHG converges to the optimal solution with accelerated local linear convergence, owing to the potentially smaller $\mathcal{L}$ provided by Lemma 4.1. The proof of Theorem 4.1 then uses Theorem 3.5 of Xiong and Freund [58] on both stages with condition (4.6) with different $\mathcal{L}$ values. The complete proof is deferred to Appendix B.

Proofs of Lemmas 4.1 and 4.2. First of all, we defined the $\tilde{M}$-norm

$$\|(x,y)\|_{\tilde{M}} := \sqrt{\frac{1}{\tau}\|x\|^2 + \frac{1}{\sigma}\|y\|^2} \quad \text{where} \quad \tilde{M} := \begin{pmatrix} \frac{1}{\tau} I_n & \\ & \frac{1}{\sigma} I_m \end{pmatrix} . \tag{4.13}$$

When τ and σ are sufficiently small, the M-norm and $\tilde{M}$-norm are equivalent up to well-specified constants related to τ and σ (Proposition 2.8 of Xiong and Freund [56]). For any point $z := (x,y) \in \mathbb{R}^{n+m}$, and $w := (x, c - A^\top y)$, it holds that

$$\sqrt{1 - \sqrt{\tau\sigma}\lambda_{\max}} \cdot \|z\|_{\tilde{M}} \le \|z\|_M \le \sqrt{2} \cdot \|z\|_{\tilde{M}} \le \sqrt{2} c_{\tau,\sigma} \cdot \|w\| \ . \tag{4.14}$$

For example, if $\tau = \frac{1}{2\kappa}$ and $\sigma = \frac{1}{2\lambda_{\max}\lambda_{\min}}$, then (4.14) becomes $\frac{\sqrt{2}}{2}\|z\|_{\tilde{M}} \le \|z\|_M \le \sqrt{2}\|z\|_{\tilde{M}} \le 2\sqrt{\kappa}\|w\|$. This result will be extensively used later.

Proof of Lemma 4.1. Let $\bar{s} = c - A^\top \bar{y}$ and $\bar{w} = (\bar{x}, \bar{s})$. From the definition of $\rho(r;\cdot)$ we have:

$$L(\bar{x}, y) - L(x, \bar{y}) \le r\rho(r;\bar{z}) \text{ for any } z = (x,y) \in B(r;\bar{z}) \tag{4.15}$$

where recall that $B(r;\bar{z})$ is defined in Definition 4.1.

Firstly, we prove that

$$\|\bar{x} - x^\star\| \le \frac{\|B^{-1}\|}{\sqrt{\sigma}} \cdot \rho(r;\bar{z}) \ . \tag{4.16}$$

As the optimal basis is unique, $x^\star$ is represented by its basic and nonbasic parts: $x^\star_{[m]} = B^{-1}b$ and $x^\star_{[n]\setminus[m]} = 0$. Consequently, due to (4.8),

$$\left\|\bar{x}_{[m]} - x^\star_{[m]}\right\| = \left\|\bar{x}_{[m]} - B^{-1}b\right\| \le \|B^{-1}\| \cdot \left\|B\bar{x}_{[m]} - b\right\| \text{ and } \left\|\bar{x}_{[n]\setminus[m]} - x^\star_{[n]\setminus[m]}\right\| = 0 \ . \tag{4.17}$$

Let $u = b - B\bar{x}_{[m]}$ and define $y := \bar{y} + \sqrt{\sigma} r \cdot u/\|u\|$. Let $z := (\bar{x}, y)$, and then $z \in B(r;\bar{z})$. Thus, from (4.15) we obtain

$$r\rho(r;\bar{z}) \ge L(\bar{x}, y) - L(\bar{x}, \bar{y}) = (b - A\bar{x})^\top (y - \bar{y}) = (b - B\bar{x}_{[m]})^\top (y - \bar{y}) = \sqrt{\sigma} r \|u\|, \tag{4.18}$$

implying $\|u\| = \|b - B\bar{x}_{[m]}\| \le \frac{\rho(r;\bar{z})}{\sqrt{\sigma}}$. Substituting this result back into (4.17) yields (4.16).

Secondly, we prove that

$$\|\bar{y} - y^\star\| \le \frac{\|B^{-1}\|}{\sqrt{\tau}} \cdot \rho(r;\bar{z}) \ . \tag{4.19}$$

Given that the optimal basis is $[m]$, we have $B^\top y^\star = c_{[m]}$, and $y^\star = (B^\top)^{-1} c_{[m]}$. Consequently,

$$\left\|\bar{y} - y^\star\right\| = \left\|\bar{y} - (B^\top)^{-1} c_{[m]}\right\| \le \left\|(B^\top)^{-1}\right\| \cdot \left\|B^\top \bar{y} - c_{[m]}\right\| = \left\|B^{-1}\right\| \cdot \left\|B^\top \bar{y} - c_{[m]}\right\| \ . \tag{4.20}$$

Let $v = c_{[m]} - B^\top \bar{y}$ and define x as follows: $x_{[m]} := \bar{x}_{[m]} - \sqrt{\tau} r \cdot \frac{v}{\|v\|}$ and $x_{[n]\setminus[m]} := 0$. Note that due to the condition $\bar{x}_i \ge r\sqrt{\tau}$ for all $i \in [m]$ in (4.8), x remains in $\mathbb{R}^n_+$. Now, let $z := (x, \bar{y})$, and then $z \in B(r;\bar{z})$. Thus, from (4.15), we derive:

$$r\rho(r;\bar{z}) \ge L(\bar{x}, \bar{y}) - L(x, \bar{y}) = (c - A^\top \bar{y})^\top (\bar{x} - x) = (c_{[m]} - B^\top \bar{y})^\top (\bar{x}_{[m]} - x_{[m]}) = \sqrt{\tau} r \|v\| \ , \tag{4.21}$$

implying $\|v\| = \|B^\top \bar{y} - c_{[m]}\| \le \frac{\rho(r;\bar{z})}{\sqrt{\tau}}$. Substituting this result back into (4.20) yields (4.19).

Finally, we can assert that

$$\|\bar{w} - w^\star\|^2 = \|\bar{x} - x^\star\|^2 + \|\bar{s} - s^\star\|^2 \le \|\bar{x} - x^\star\|^2 + \|A\|\|\bar{y} - y^\star\|^2 \overset{(4.16)(4.19)}{\le} \left(\frac{\|B^{-1}\|^2}{\sigma} + \frac{\|B^{-1}\|^2\|A\|^2}{\tau}\right)\rho(r;\bar{z})^2 \tag{4.22}$$

Due to (4.14), $\|\bar{z} - z^\star\|_M \le \sqrt{2} c_{\tau,\sigma} \cdot \|\bar{w} - w^\star\|$. Applying it on the left-hand side of (4.22) proves (4.9). Thus, the proof is complete. □

Before proving Lemma 4.2, we recall the nonexpansive property of rPDHG proven by Applegate et al. [6]. We use the more convenient format presented by Xiong and Freund [58].

LEMMA 4.3 **(Lemma 2.2 of Xiong and Freund [58]).** *For for any $n, k \geq 0$, it holds that $\|\bar{z}^{n,k} - z^\star\|_M \leq \left\|z^{n,0} - z^\star\right\|_M$. For any n_1, n_2 so that $n_2 \geq n_1 \geq 0$, it holds that $\|z^{n_2,0} - z^\star\|_M \leq \|z^{n_1,0} - z^\star\|_M$.*

Proof of Lemma 4.2. The proof contains two steps. In the first step, we prove that when

$$\left\|z^{n,0} - z^\star\right\|_M \leq \frac{\sqrt{1-\sqrt{\tau\sigma}\|A\|}}{3\sqrt{\tau}} \cdot \xi\ , \tag{4.23}$$

then for any $N \geq n$ and $k \geq 1$, item (i) of (4.12) holds.

We begin by establishing a lower bound for $\min_{1\leq i\leq m} \bar{x}_i^{N,k}$, the left-hand side of item (i):

$$\min_{1\leq i\leq m} \bar{x}_i^{N,k} \geq \left(\min_{1\leq i\leq m} x_i^\star\right) - \left\|\bar{x}_{[m]}^{N,k} - x_{[m]}^\star\right\| \geq \xi - \left\|\bar{x}_{[m]}^{N,k} - x_{[m]}^\star\right\| \ . \tag{4.24}$$

The second term on the right-hand side of (4.24) can be bounded as follows:

$$\left\|\bar{x}_{[m]}^{N,k} - x_{[m]}^\star\right\| \leq \left\|\bar{x}^{N,k} - x^\star\right\| \overset{(4.13)}{\leq} \sqrt{\tau}\left\|\bar{z}^{N,k} - z^\star\right\|_{\tilde{M}} \overset{(4.14)}{\leq} \frac{\sqrt{\tau}\left\|\bar{z}^{N,k} - z^\star\right\|_M}{\sqrt{1-\sqrt{\tau\sigma}\|A\|}} \overset{\text{Lemma 4.3}}{\leq} \frac{\sqrt{\tau}\left\|z^{n,0} - z^\star\right\|_M}{\sqrt{1-\sqrt{\tau\sigma}\|A\|}} \tag{4.25}$$

for any $n \leq N$. In addition, due to the nonexpansive property, the right-hand side of item (i) is upper bounded by:

$$\left\|\bar{z}^{N,k} - z^{N,0}\right\|_M \leq \left\|\bar{z}^{N,k} - z^\star\right\|_M + \left\|z^\star - z^{N,0}\right\|_M \leq 2\left\|z^\star - z^{n,0}\right\|_M \leq \frac{2\left\|z^{n,0} - z^\star\right\|_M}{\sqrt{1-\sqrt{\tau\sigma}\|A\|}} \tag{4.26}$$

where the last inequality holds because $\sqrt{1-\sqrt{\tau\sigma}\|A\|} \leq 1$. Finally, we have

$$\begin{aligned}
&\left(\min_{1\leq i\leq m} \bar{x}_i^{N,k}\right) - \sqrt{\tau}\left\|\bar{z}^{N,k} - z^{N,0}\right\|_M \overset{(4.24),(4.25)}{\geq} \xi - \frac{\sqrt{\tau}\left\|z^{n,0} - z^\star\right\|_M}{\sqrt{1-\sqrt{\tau\sigma}\|A\|}} - \sqrt{\tau}\left\|\bar{z}^{N,k} - z^{N,0}\right\|_M \\
&\overset{(4.26)}{\geq} \xi - \frac{\sqrt{\tau}\left\|z^{n,0} - z^\star\right\|_M}{\sqrt{1-\sqrt{\tau\sigma}\|A\|}} - 2\sqrt{\tau}\left\|z^\star - z^{n,0}\right\|_M \geq \xi - \frac{3\sqrt{\tau}\left\|z^{n,0} - z^\star\right\|_M}{\sqrt{1-\sqrt{\tau\sigma}\|A\|}}\ ,
\end{aligned} \tag{4.27}$$

in which the last term is nonnegative when (4.23) holds. This show that when $\left\|z^{n,0} - z^\star\right\|_M$ is small enough and satisfies (4.23), the left-hand side of (4.27) is nonnegative, and item (i) of (4.12) holds. This completes the first step of the proof.

In the second step, we prove that when

$$\left\|z^{n,0} - z^\star\right\|_M \leq \frac{\sqrt{\tau}\sqrt{1-\sqrt{\tau\sigma}\|A\|}}{2} \cdot \xi, \tag{4.28}$$

then for any $N \geq n$ and $k \geq 1$ item (ii) of (4.12) holds.

Let α denote $\left(\min_{1\leq j\leq n-m, k\geq 0} s_{m+j}^{N,k}\right)$, then according to (2.6) (in Line 5 of Algorithm 1),

$$x_{[n]\setminus[m]}^{N,k} = \left(x_{[n]\setminus[m]}^{N,k-1} - \tau s^{N,k-1}\right)^+ \leq \left(x_{[n]\setminus[m]}^{N,k-1} - \tau\alpha\right)^+$$

and applying this inequality recursively as k decreases to 0 yields

$$x_{[n]\setminus[m]}^{N,k} \leq \left(x_{[n]\setminus[m]}^{N,k-1} - \tau\alpha\right)^+ \leq \left(x_{[n]\setminus[m]}^{N,k-2} - 2\tau\alpha\right)^+ \leq \cdots \leq \left(x_{[n]\setminus[m]}^{N,0} - k\tau\alpha\right)^+\ . \tag{4.29}$$

Therefore, if $\alpha \geq 0$ and

$$\left(\max_{1\leq j\leq n-m} x_{m+j}^{N,0}\right) \leq \tau \cdot \left(\min_{1\leq j\leq n-m, k\geq 0} s_{m+j}^{N,k}\right) , \tag{4.30}$$

then $x_{[n]\setminus[m]}^{N,k} = 0$ for all $k \geq 1$ and thus $\bar{x}_{[n]\setminus[m]}^{N,k} = \frac{1}{k}\sum_{i=1}^{k} x_{[n]\setminus[m]}^{N,i} = 0$ for all $k \geq 1$. Intuitively, since Assumption 2.1 holds, and $x^{N,0}$ and $s^{N,k}$ converge to the optimal solution, $x_{[n]\setminus[m]}^{N,0}$ should converge to 0 and $s_{[n]\setminus[m]}^{N,k}$ should stay away from 0. In the rest of the proof we will establish a lower bound of $\left(\min_{1\leq j\leq n-m,k\geq 0} s_{m+j}^{N,k}\right)$ and an upper bound of $\left(\max_{1\leq j\leq n-m} x_{m+j}^{N,0}\right)$.

Similar to the first step, for the dual iteration $y^{N,k}$ (and $s^{N,k} = c - A^\top y^{N,k}$), we have

$$\begin{aligned}\left(\min_{1\leq j\leq n-m} s_{m+j}^{N,k}\right) &\geq \left(\min_{1\leq j\leq n-m} s_{m+j}^{\star}\right) - \left\|s_{[n]\setminus[m]}^{N,k} - s_{[n]\setminus[m]}^{\star}\right\| \\ &= \xi - \left\|A_{[n]\setminus[m]}^\top (y^{N,k} - y^\star)\right\| \geq \xi - \left\|A_{[n]\setminus[m]}\right\| \cdot \left\|y^{N,k} - y^\star\right\| .\end{aligned} \tag{4.31}$$

As for $\left\|y^{N,k} - y^\star\right\|$ in the last term of the above inequality, we derive:

$$\left\|y^{N,k} - y^\star\right\| \overset{(4.13)}{\leq} \sqrt{\sigma}\left\|z^{N,k} - z^\star\right\|_{\tilde{M}} \overset{(4.14)}{\leq} \frac{\sqrt{\sigma}\left\|z^{N,k} - z^\star\right\|_M}{\sqrt{1-\sqrt{\tau\sigma}\|A\|}} \overset{\text{Lemma 4.3}}{\leq} \frac{\sqrt{\sigma}\left\|z^{n,0} - z^\star\right\|_M}{\sqrt{1-\sqrt{\tau\sigma}\|A\|}} . \tag{4.32}$$

Combining (4.31) and (4.32) yields a valid lower bound of $\left(\min_{1\leq j\leq n-m} s_{m+j}^{N,k}\right)$ for all $k \geq 0$:

$$\left(\min_{1\leq j\leq n-m} s_{m+j}^{N,k}\right) \geq \xi - \sqrt{\sigma}\|A_{[n]\setminus[m]}\| \cdot \frac{\left\|z^{n,0} - z^\star\right\|_M}{\sqrt{1-\sqrt{\tau\sigma}\|A\|}} \geq \xi - \frac{1}{\sqrt{\tau}} \cdot \frac{\left\|z^{n,0} - z^\star\right\|_M}{\sqrt{1-\sqrt{\tau\sigma}\|A\|}} , \tag{4.33}$$

where the second inequality holds because $\sqrt{\tau\sigma}\|A_{[n]\setminus[m]}\| \leq \sqrt{\tau\sigma}\|A\| \leq 1$ (due to the step-size requirement (2.7)). The above (4.33) presents a lower bound of $\left(\min_{1\leq j\leq n-m,k\geq 0} s_{m+j}^{N,k}\right)$.

On the other hand, we also have the following upper bound of $\left(\max_{1\leq j\leq n-m} x_{m+j}^{N,0}\right)$ for $N \geq n$:

$$\left(\max_{1\leq j\leq n-m} x_{m+j}^{N,0}\right) \leq \left\|x_{[n]\setminus[m]}^{N,0}\right\| = \left\|x_{[n]\setminus[m]}^{N,0} - x_{[n]\setminus[m]}^{\star}\right\| \leq \|x^{N,0} - x^\star\| \overset{(4.25)}{\leq} \frac{\sqrt{\tau}\left\|z^{n,0} - z^\star\right\|_M}{\sqrt{1-\sqrt{\tau\sigma}\|A\|}} . \tag{4.34}$$

Finally, applying the upper and lower bounds (4.33) and (4.34), we find that (4.30) holds when

$$\frac{\sqrt{\tau}\left\|z^{n,0} - z^\star\right\|_M}{\sqrt{1-\sqrt{\tau\sigma}\|A\|}} \leq \tau\xi - \sqrt{\tau} \cdot \frac{\left\|z^{n,0} - z^\star\right\|_M}{\sqrt{1-\sqrt{\tau\sigma}\|A\|}} ,$$

which is equivalent to (4.28). This completes the second step of the proof.

Finally, when both (4.23) and (4.28) hold, which is satisfied by (4.11), then for any $N \geq n$ and $k \geq 1$, both item (i) and (ii) of (4.12) hold. Furthermore, (4.25) and (4.24) ensure $\bar{x}_{\Theta}^{N,k} > 0$, while the item (ii) ensures $\bar{x}_{\bar{\Theta}}^{N,k} = 0$. Therefore, the positive components of $\bar{x}^{N,k}$ correspond exactly to the optimal basis. This completes the proof. □

5. Relationship of Φ with Stability under Data Perturbations, Proximity to Multiple Optima, and LP Sharpness The previous two sections showed new accessible iteration bounds for rPDHG. Because Φ has a closed-form expression, in this section we further prove that there is an exact relationship between Φ and three equivalent types of condition measures: (i) stability under data perturbations, (ii) proximity to multiple optima, and (iii) the LP sharpness of the instance. Furthermore, this section provides new computational guarantees using these measures, providing new insights into the performance of rPDHG.

We start by defining two quantities of primal and dual stabilities. For the original problem (1.1), let the perturbed problem be as follows:

$$\min\ \tilde{c}^\top x \quad \text{s.t. } Ax = \tilde{b}\ ,\ x \geq 0 \tag{5.1}$$

where $\tilde{c}$ and $\tilde{b}$ might be the perturbed versions of c and b respectively. When Assumption 2.1 holds for (1.1), we define ζ_p and ζ_d as follows:

$$\zeta_p := \inf\left\{\|\Delta c\| :\ \Theta \text{ is not the unique optimal basis for (5.1) with } \tilde{c} = c + \Delta c \text{ and } \tilde{b} = b\right\}\ , \tag{5.2}$$

$$\zeta_p := \inf\left\{\|\Delta b\|_{(AA^\top)^{-1}} :\ \Theta \text{ is not the unique optimal basis for (5.1) with } \tilde{c} = c \text{ and } \tilde{b} = b + \Delta b\right\}\ . \tag{5.3}$$

The ζ_p and ζ_d denote the size of the smallest perturbation on the cost vector c and the right-hand side vector b, respectively, such that the optimal basis becomes different. In other words, the larger they are, the more stable the optimal basis is under data perturbations on b and c. Here ζ_d uses the $(AA^\top)^{-1}$-norm instead of the Euclidean norm because later we will show that ζ_d can be defined in the symmetric way to ζ_p on the symmetric form (2.4).

More importantly, Φ has a close relationship with ζ_p and ζ_d, leading to a new computational guarantee using ζ_p and ζ_d. Below is the first main theorem of this section.

Theorem 5.1. *Suppose Assumption 2.1 holds. The following relationship holds for Φ, ζ_p, and ζ_d:*

$$\Phi = \frac{\|x^\star\|_1 + \|s^\star\|_1}{\min\left\{\zeta_p,\ \zeta_d\right\}}\ . \tag{5.4}$$

Therefore, in the identical setting of Theorem 3.1, the total number of OnePDHG iterations required to compute an ε-optimal solution is at most

$$O\left(\kappa \cdot \frac{\|x^\star\|_1 + \|s^\star\|_1}{\min\left\{\zeta_p,\ \zeta_d\right\}} \cdot \ln\left(\frac{\kappa\frac{\|x^\star\|_1+\|s^\star\|_1}{\min\left\{\zeta_p,\ \zeta_d\right\}} \cdot \|w^\star\|}{\varepsilon}\right)\right).$$

This theorem implies that the less stable the optimal basis is under data perturbations, the larger the value of Φ, and the more iterations rPDHG might require to compute an ε-optimal solution. Actually in Section 6 we will confirm the tightness of these bounds via experiments on LP instances. Small values of $\min\left\{\zeta_p,\ \zeta_d\right\}$ may very significantly affect the performance of rPDHG, because they stay in the denominator in the expression of Φ in (5.4). It should be noted that ζ_p and ζ_d are not intrinsic properties of the constraint matrix as they are also dependent on c and b, so ζ_p and ζ_d do not affect the Stage II iteration bound T_{local} in Theorem 4.1.

In the rest of this section, Section 5.1 introduces three measures on the primal and dual LPs: stability under data perturbation, proximity to multiple optima and the LP sharpness of the instance. Then we present the second main theorem, the equivalence between these three measures. Moreover, ζ_p and ζ_d are actually equivalent to them. These equivalence relationships provide new insights into the performance of rPDHG. Finally, Section 5.2 presents the proof of Theorem 5.1.

5.1. Stability under data perturbations, proximity to multiple optima, LP sharpness Both the primal and dual problems in the symmetric form (2.4) are instances of the following generic form of LP:

$$\min\ g^\top u \qquad \text{s.t. } u \in \mathcal{F}_{generic} := V_{generic} \cap \mathbb{R}^n_+ \tag{5.5}$$

where the feasible set $\mathcal{F}_{generic}$ is the intersection of the nonnegative orthant $\mathbb{R}^n_+$ and an affine subspace $V_{generic}$. The objective function $g^\top u$ is a linear function. We denote the optimal solution of (5.5) by $\mathcal{U}^\star$, in which $u^\star$ is an optimal solution. We let $\mathrm{OPT}(\breve{g})$ denote the set of optimal solutions of the generic LP (5.5) with the objective vector equal to $\breve{g}$. For example, $\mathrm{OPT}(g) = \mathcal{U}^\star$. The primal problem (1.1) is an instantiation of (5.5) with $\mathcal{F}_{generic} = \mathcal{F}_p$, $V_{generic} = V_p$, and $g = c$. Similarly, the dual problem (2.3) is another instantiation of (5.5) with $\mathcal{F}_{generic} = \mathcal{F}_d$, $V_{generic} = V_d$, and $g = q$. Now we define the three measures.

Definitions of the three measures. The first measure is the stability under data perturbation ζ, the size of the smallest data perturbation on g such that a new optimal solution occurs.

Definition 5.1 (Stability under data perturbations). The stability under data perturbations is defined as

$$\zeta := \inf_{\Delta g} \left\{ \|\Delta g\| : \mathrm{OPT}(g+\Delta g) \neq \emptyset \text{ and } \mathrm{OPT}(g+\Delta g) \not\subseteq \mathrm{OPT}(g) \right\} . \tag{5.6}$$

The second measure is the proximity to multiple optima, defined as the size of the smallest perturbation that leads to multiple optimal solutions.

Definition 5.2 (Proximity to multiple optima). When $\mathrm{OPT}(g)$ is a singleton, the proximity to multiple optima is defined as

$$\eta = \min_{\Delta g} \left\{ \|\Delta g\| : |\,\mathrm{OPT}(g+\Delta g)| > 1 \right\} . \tag{5.7}$$

Since having multiple optima is equivalent to having degenerate dual optimal solutions, η can also be interpreted as the proximity to degenerate dual optima.

The third measure is LP sharpness (see Xiong and Freund [56]), which measures how quickly the objective function grows away from the optimal solution set $\mathcal{U}^\star$ (i.e., $\mathrm{OPT}(g)$) among all feasible points.

Definition 5.3 (LP sharpness). The LP sharpness of (5.5) is defined as

$$\mu := \inf_{u \in \mathcal{F}_{generic} \setminus \mathcal{U}^\star} \frac{\mathrm{Dist}(u, V_{generic} \cap \{u \in \mathbb{R}^n : g^\top u = g^\top u^\star\})}{\mathrm{Dist}(u,\ \mathcal{U}^\star)} . \tag{5.8}$$

Sharpness is a useful analytical tool (for example, see Lu and Yang [37], Yang and Lin [59]), and Applegate et al. [6] employ the sharpness of the normalized duality gap for the saddlepoint problem to prove the linear convergence of rPDHG on LP. Sharpness for LP, denoted by LP sharpness, is a more natural and intuitive measure of the original LP instance.

Equivalence between the three measures: ζ, η and μ. We now present the second main theorem of this section, the equivalence between the three measures when $\mathcal{U}^\star$ is a singleton.

Theorem 5.2. *When $\mathcal{U}^\star$ is a singleton, the following equivalence relationship holds for the stability under data perturbations, proximity to multiple optima, and LP sharpness:*

$$\frac{1}{\left\|P_{\vec{V}_{generic}}(g)\right\|} \cdot \zeta = \frac{1}{\left\|P_{\vec{V}_{generic}}(g)\right\|} \cdot \eta = \mu . \tag{5.9}$$

Here ζ is normalized by the norm of $P_{\vec{V}_{generic}}(g)$, indicating that ζ and η are equivalent to μ in a relative sense. This normalization arises because μ is a purely geometric concept that remains invariant under positive scaling of g. We use the norm of $P_{\vec{V}_{generic}}(g)$ rather than g because the complementary part $P_{\vec{V}^\perp_{generic}}(g)$ does not have any impact on the optimal solution and the smallest perturbation Δg must lie in $\vec{V}_{generic}$.

We use μ_p and μ_d to denote the μ for the primal problem (1.1) and the dual problem (2.3), for which $\left\|P_{\vec{V}_p}(c)\right\|$ and $\|q\|$ correspond to $\left\|P_{\vec{V}_{generic}}(g)\right\|$ (Fact 2.1).

We now prove Theorem 5.2. The equivalence between μ and ζ is already proven by Xiong and Freund [56]:

Lemma 5.1 **(Theorem 5.1 of Xiong and Freund [56])**. *LP sharpness is equivalent to stability under data perturbations through the relation:* $\mu = \zeta \cdot \frac{1}{\left\|P_{\vec{V}_{generic}}(g)\right\|}$.

With Lemma 5.1 we can prove Theorem 5.2.

Proof of Theorem 5.2. The second equality of (5.9) directly comes from Lemma 5.1. As for the first equality of (5.9), in the case that $\mathrm{OPT}(g)$ (i.e., $\mathcal{U}^\star = \{u^\star\}$) is a singleton, ζ is the smallest magnitude of the perturbation that leads to multiple optimal solutions at the threshold at which a new solution is added to $\mathrm{OPT}(g+\Delta g)$ while $u^\star$ remains optimal. Therefore, ζ is equal to η. This finishes the proof. □

We have now shown the equivalence between the above three condition measures on the generic LP (5.5). Furthermore, the ζ_p and ζ_d we defined in (5.2) and (5.3) are actually equal to ζ for (1.1) and (2.3).

Lemma 5.2. *Suppose Assumption 2.1 holds. The ζ for* (1.1) *and* (2.3) *is equal to ζ_p and ζ_d, respectively.*

Proof. For ζ of (1.1), $\mathrm{OPT}(c+\Delta c)\nsubseteq\mathrm{OPT}(c)$ if and only if Θ is no longer an optimal basis or the optimal basis is no longer unique for the perturbed problem. This proves ζ_p is equal to ζ of (1.1).

Similarly, ζ of (2.3) is defined as the Euclidean norm of the smallest perturbation Δq on the objective function so that $\bar{\Theta}$ is no longer the unique optimal basis for (2.3). Since all feasible solutions are in the affine subspace $c+\mathrm{Im}(A^\top)$, the smallest perturbation Δq must lie in $\mathrm{Im}(A^\top)$ and there exists $\Delta b_0\in\mathbb{R}^m$ such that $\Delta q = A^\top(AA^\top)^{-1}\Delta b_0$. When applying the right-hand side perturbation Δb_0 on (1.1), the corresponding dual problem (2.3) has the objective vector $A^\top(AA^\top)^{-1}(b+\Delta b_0)$ that is exactly equal to $q+\Delta q$. In this case $\|\Delta q\| = \|A^\top(AA^\top)^{-1}\Delta b\| = \|\Delta b\|_{(AA^\top)^{-1}}$. Therefore, the smallest Euclidean norm of the objective vector perturbation Δq on (2.3) such that $\bar{\Theta}$ is no longer the unique optimal basis is equal to the smallest $(AA^\top)^{-1}$-norm of right-hand side perturbation Δb on (1.1) such that Θ is no longer the unique optimal basis. This proves ζ_d is equal to ζ of (2.3). □

The above equivalence relationships of ζ_p, ζ_d with the three condition measures also provide new insights into the performance of rPDHG. Here we detail them as follows.

It is often observed that rPDHG has good performance in some LP instances with multiple optimal solutions, but a minor data perturbation results in a substantial degradation in rPDHG's performance on the perturbed problem (see Section 6 for examples). Now we have some insight and a partial explanation for this phenomenon. According to Theorem 5.2 and Lemma 5.2, ζ_p and ζ_d are equal to the proximity to multiple optima for (1.1) and (2.3). The perturbed problem still stays in close proximity to the original unperturbed problem, so its $\min\{\zeta_p,\zeta_d\}$ is at most as large as the magnitude of the perturbation. Furthermore, $\min\{\zeta_p,\zeta_d\}$ lies in the denominator of the iteration bound of Theorems 5.1 (and Stage I iteration bound of Theorem 4.1), so a small perturbation may significantly increase the iteration bound. This explanation is complementary with Lu and Yang [40] which study the vanilla PDHG and explain the phenomenon by the size of the region for local fast convergence. Our result shows that perturbations significantly affect the iteration bound for Stage I of rPDHG as well. This effect of perturbations on rPDHG will be confirmed by computational experiments in Section 6.

Moreover, Theorem 5.1 together with Lemmas 5.1 and 5.2 also provide new computational guarantees that use κ, the size of the optimal solutions, and the LP sharpness μ_p and μ_d. They are simpler and more intuitive guarantees than those in Xiong and Freund [56], because the latter also involves the limiting error ratio. This simplification is due to the unique optimum assumption. Notably, despite this additional assumption, our iteration bounds are strictly better than the iteration bounds of Xiong and Freund [57, Corollary 4.3], which exhibit quadratic dependence on the reciprocals of μ_p and μ_d.

5.2. Proof of Theorem 5.1 The key to proving Theorem 5.1 is the following lemma:

LEMMA 5.3. *Suppose that Assumption 2.1 holds. The ζ_p and ζ_d have the following expression:*

$$\zeta_p = \min_{1\le j\le n-m}\frac{s^\star_{m+j}}{\sqrt{\left\|(B^{-1}N)_{\cdot,j}\right\|^2+1}}\ \text{ and }\ \zeta_d = \min_{1\le i\le m}\frac{x^\star_i}{\sqrt{\left\|(B^{-1}N)_{i,\cdot}\right\|^2+1}}\ . \tag{5.10}$$

Before proving it, we recall how to compute the LP sharpness μ of the generic LP (5.5) by computing the smallest sharpness along all of the edges emanating from the optimal solutions.

LEMMA 5.4 **(A restatement of Theorem 5.5 of Xiong and Freund [56])**. *Suppose that Assumption 2.1 holds. Let $\mathcal{U}^\star=\{u^\star\}$ and the directions of the edges emanating from $\mathcal{U}^\star$ be $v^1,v^2,\dots,v^{n-m}$. Then for any given $\bar{\varepsilon}>0$, the LP sharpness μ is characterized as follows:*

$$\mu = \min_{1\le j\le n-m}\frac{\mathrm{Dist}(u^\star+\bar{\varepsilon}\cdot v^j, V_{generic}\cap\{u\in\mathbb{R}^n : g^\top u = g^\top u^\star\})}{\left\|\left(u^\star+\bar{\varepsilon}\cdot v^j\right)-u^\star\right\|} \tag{5.11}$$

Now we are ready to prove Lemma 5.3.

Proof of Lemma 5.3. Let μ_p and μ_d denote the LP sharpness for (1.1) and (2.3) respectively. We first compute μ_p and then $\zeta_p = \left\|P_{\vec{V}_p}(c)\right\|\mu_p$ (Lemma 5.1). For (1.1), $u^\star = x^\star$, $V_{generic} = V_p$ and $\{u \in \mathbb{R}^n : g^\top u = g^\top u^\star\} = \left\{x : \mathrm{E}^{\mathrm{p}}_{\mathrm{obj}}(x) = 0\right\}$ in the generic LP (5.5), with the directions of connected edges given by $\{u^j : 1 \le j \le n-m\}$ as defined in (3.11). Furthermore, the following equalities hold:

$$\begin{aligned}\mathrm{Dist}\left(x^\star + \bar{\varepsilon}\cdot u^j, V_p \cap \left\{x : \mathrm{E}^{\mathrm{p}}_{\mathrm{obj}}(x) = 0\right\}\right) &= \frac{\mathrm{E}^{\mathrm{p}}_{\mathrm{obj}}(x^\star + \bar{\varepsilon}\cdot u^j)}{\left\|P_{\vec{V}_p}(c)\right\|} \\ = \frac{(s^\star)^\top(x^\star + \bar{\varepsilon}\cdot u^j) - (s^\star)^\top x^\star}{\left\|P_{\vec{V}_p}(c)\right\|} = \bar{\varepsilon}\cdot\frac{(s^\star)^\top u^j}{\left\|P_{\vec{V}_p}(c)\right\|} &= \bar{\varepsilon}\cdot\frac{s^\star_{m+j}u^j_{m+j}}{\left\|P_{\vec{V}_p}(c)\right\|} = \bar{\varepsilon}\cdot\frac{s^\star_{m+j}}{\left\|P_{\vec{V}_p}(c)\right\|}\end{aligned} \tag{5.12}$$

for all $j \in [n-m]$. Here the second and third equalities use the result that $\mathrm{E}^{\mathrm{p}}_{\mathrm{obj}}(x) = \mathrm{Gap}(x, s^\star) = (s^\star)^\top x - (s^\star)^\top x^\star$ for any $x \in V_p$. The fourth equality holds because $s^\star_{[m]} = 0$ and $x^j_{[n]\setminus([m]\cup\{m+j\})} = 0$. The final equality uses $u^j_{m+j} = 1$. In addition, for all $j \in [n-m]$ we have

$$\|(x^\star + \bar{\varepsilon}\cdot u^j) - x^\star\| = \bar{\varepsilon}\cdot\|u^j\| = \bar{\varepsilon}\cdot\sqrt{\left\|(B^{-1}N)_{\cdot,j}\right\|^2 + 1}\,. \tag{5.13}$$

Substituting (5.12) and (5.13) into the definition (5.11) of μ_p yields the expression:

$$\mu_p = \frac{1}{\left\|P_{\vec{V}_p}(c)\right\|}\cdot\min_{1\le j\le n-m}\frac{s^\star_{m+j}}{\sqrt{\left\|(B^{-1}N)_{\cdot,j}\right\|^2 + 1}}\,.$$

Finally, substituting it back to $\zeta_p = \left\|P_{\vec{V}_p}(c)\right\|\mu_p$ proves the first half of (5.10).

Next we compute μ_d and then $\zeta_d = \left\|P_{\vec{V}_d}(q)\right\|\mu_d$ (Lemma 5.1). For (2.3), we repeat the above process on (3.24), and then we can symmetrically obtain $\mu_d = \frac{1}{\left\|P_{\vec{V}_d}(q)\right\|}\cdot\min_{1\le i\le m}\frac{x^\star_i}{\sqrt{\left\|(Q_{\bar{\Theta}}^{-1}Q_\Theta)_{\cdot,i}\right\|^2+1}}$. By Lemma 3.6, $Q_{\bar{\Theta}}^{-1}Q_\Theta = -(B^{-1}N)^\top$, and thus $\left\|(Q_{\bar{\Theta}}^{-1}Q_\Theta)_{\cdot,i}\right\| = \|(B^{-1}N)_{i,\cdot}\|$. Combined with $\zeta_d = \left\|P_{\vec{V}_d}(q)\right\|\mu_d$, this completes the proof. □

With Lemma 5.3, we can now prove Theorem 5.1.

Proof of Theorem 5.1. Directly substituting the expressions of ζ_p and ζ_d into the expression (3.1) of Φ completes the proof. □

6. Experimental Confirmation This section shows how our new theories align with the practical performance through computational evaluations. Section 6.1 confirms the reciprocal relationship between rPDHG complexity and the magnitude of perturbations, providing experimental evidence for our findings in Section 5. Section 6.2 confirms the significance of $\kappa\Phi$ and $\|B^{-1}\|\|A\|$ in the two-stage performance of rPDHG, supporting our results in Section 4.

We implement rPDHG (Algorithm 1) on standard-form LPs, adhering precisely to the setting of Theorem 3.1, with one exception: the normalized duality gap uses the $\tilde{M}$-norm that is defined in (4.13), instead of the M-norm. This alternative is proven by Applegate et al. [6], Xiong and Freund [58] equivalent to the original normalized duality gap but significantly more computationally efficient. It is also widely adopted in practice, such as Applegate et al. [4], Lu and Yang [38], Lu et al. [43].

6.1. Effects of Data Perturbations To evaluate rPDHG's sensitivity to perturbations in the objective vector c, we construct a family of standard-form LP instances (1.1) with data $\left(A^1, b^1, c^1\right)$, where

$$A^1 = [1,1,1]\,,\ b^1 = 2,\ \text{ and } c^1 = c^1_\gamma := [2,-1,-1] + \left[0, -\frac{\gamma}{2}, \frac{\gamma}{2}\right]$$

for parameter $\gamma \geq 0$. This LP family, denoted by LP^1_γ, is designed to illustrate the effect of perturbations $[0, -\frac{\gamma}{2}, \frac{\gamma}{2}]$ on the objective vector c^1_0. When $\gamma = 0$, LP^1_γ has multiple optimal solutions along the line segment connecting $(0, 2, 0)$ and $(0, 0, 2)$. For $\gamma > 0$, the problem has a unique optimal solution at $(0, 2, 0)$. Since ζ_p is equivalent to the proximity to multiple optima, we have $\zeta_p \leq O(\gamma)$ when $\gamma > 0$. As for κ, it is always equal to 1 for LP^1_γ instances.

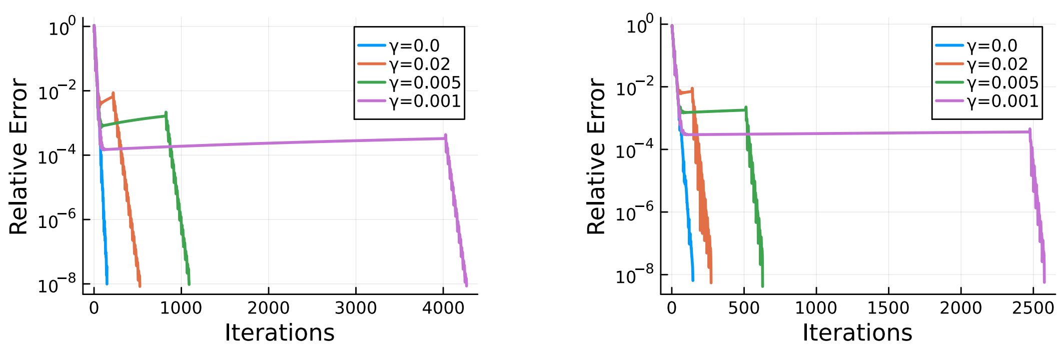
For the family of problems LP^1_γ, the values of Φ for different γ values are as follows:

γ	1e0	1e-1	1e-2	1e-3	1e-4
Φ of LP^1_γ	6.4e0	4.5e1	4.3e2	4.2e3	4.2e4

These values of Φ are clearly in a reciprocal relationship with the value of γ. This observation aligns with (5.4) of Theorem 5.1. Figure 1a shows the convergence performance of rPDHG on LP^1_γ instances for $\gamma \in \{0, 0.02, 0.005, 0.001\}$. The horizontal axis reports the number of iterations, while the vertical axis reports 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 iterates (x, y). We use $\mathcal{E}_r(x, y)$ because it is easy to compute, and applicable when the problem has multiple optima. It is also a widely used standard tolerance (also used in Applegate et al. [4], Lu et al. [43]). The results clearly demonstrate that as $\gamma \searrow 0$, the number of iterations (of Stage I in particular) increases significantly, exhibiting an approximately reciprocal relationship with γ. Notably, the results also indicate that neither γ nor Φ influences the local convergence rate in Stage II. This finding is consistent with Theorem 4.1, which asserts that the local convergence rate is solely determined by $\|B^{-1}\| \|A\|$.

FIGURE 1. Convergence performance of rPDHG on two families of LP instances.

(a) LP^1_γ (b) LP^2_γ



We further construct another family of standard-form LP instances (1.1) with data (A^2, b^2, c^2), where

$$A^2 = \begin{bmatrix} 1 & 1 & -1 \\ 1 & 0 & 1 \end{bmatrix}, \; c^2 = [-0.5, 1, 0.5], \;\text{ and } b^2 = b^2_\gamma := [1, 1] + [\gamma, 2\gamma]$$

for parameter $\gamma \geq 0$. This LP family, denoted by LP^2_γ, is designed to illustrate the effect of perturbations $[\gamma, 2\gamma]$ on the right-hand side vector b^2_0. The value of κ is always equal to 1.22 for all LP^2_γ instances. When $\gamma = 0$, LP^2_γ has a unique optimal primal solution $[1, 0, 0]$. This solution is degenerate, implying multiple dual optimal solutions for LP^2_0. When $\gamma > 0$, the problem only has a unique dual optimal solution. Since ζ_d is also equivalent to the proximity to multiple dual optimal solutions, we have $\zeta_d \leq O(\gamma)$ when $\gamma > 0$.

For the family of problems LP^2_γ, the values of Φ for different γ values are as follows:

γ	1e0	1e-1	1e-2	1e-3	1e-4
Φ of LP^2_γ	6.7e0	3.4e1	3.4e2	3.4e3	3.4e4

These values of Φ still exhibit a clear reciprocal relationship with γ. Figure 1b shows the convergence performance of rPDHG for LP^1_γ instances for $\gamma \in \{0, 0.02, 0.005, 0.001\}$. Although the perturbations are now on the right-hand side vector b, our observations are nearly symmetric to those for LP^1_γ, which match the predictions of Theorems 5.1 and 4.1 again.

6.2. Two-stage Performance of rPDHG This subsection presents experimental confirmations of the implications of κ and Φ in practical computations. We test on LP instances generated according to Todd's random LP model:

Definition 6.1 (Random linear program). Let u be a random variable from the normal distribution $\mathcal{N}(0,1)$ with mean 0 and variance 1. Let the components of the matrix A be independent and identically distributed (i.i.d.) copies of u. The primal and dual solutions $\hat{x}$ and $\hat{s}$ are generated as follows:

$$\hat{x}_\Theta \in \mathbb{R}^m_+, \quad \hat{s}_{\bar{\Theta}} \in \mathbb{R}^{n-m}_+, \quad \hat{x}_{\bar{\Theta}} = 0, \quad \hat{s}_\Theta = 0, \tag{6.1}$$

in which $\Theta = \{1,2,\ldots,m\}$ and the components of $\hat{x}_\Theta$ and $\hat{s}_{\bar{\Theta}}$ are i.i.d. copies of $|u|$. The right-hand side b is generated by $b = A\hat{x}$, and the cost vector c is generated by $c = \hat{s}$. (Optionally, the cost vector $\bar{c}$ with the smallest norm is generated by $\bar{c} := \hat{s} + A^\top \hat{y}$, where $\hat{y} = \arg\min_{y\in\mathbb{R}^m} \|\hat{s} + A^\top \hat{y}\|$.)

The random LP in Definition 6.1 is Model 1 of Todd [51]. Variants of this model have been analyzed by Anstreicher et al. [2, 3], Ye [61] to elucidate the average performance of interior-point methods. One can observe that $\hat{x}$ and $\hat{s}$ are the optimal primal-dual solution because they are feasible and satisfy the complementary slackness condition $\hat{x}^\top\hat{s} = 0$. Components of $\hat{x}_\Theta$ and $\hat{s}_{\bar{\Theta}}$ are all nonzero almost surely, and the LP instance has a unique optimum with optimal basis Θ almost surely. Since the optimal basis and the optimal solution are known prior to solving the problem, $\kappa\Phi$ and $\|B^{-1}\|\|A\|$ used in the iteration bound results can be easily computed. Furthermore, replacing the cost vector c with the smallest-norm cost vector $\bar{c}$ does not influence the optimality and degeneracy of $\hat{x}$ and $\hat{s}$, but $\bar{c}$ lies in Null(A), i.e., $A\bar{c} = 0$. To keep consistent with the theoretical results, we use $\bar{c}$ in the experiments.

We run rPDHG on 100 randomly generated LP problems according to Definition 6.1 with $m = 50$ and $n = 100$. We use LP$_i$ to denote the i-th instance, where $i \in \{1,2,\ldots,100\}$. Let $(x^{n,k}, y^{n,k})$ be the iterations of the rPDHG, and we deem an instance solved when rPDHG computes a solution $(x^{n,k}, y^{n,k})$ whose Euclidean distance to the optimal solution $(x^\star, y^\star)$ is smaller than 10^{-4}. After solving the problem, we define the number of Stage I (optimal basis identification) iterations to be the iteration counter after which the support set of $x^{n,k}$ remains unchanged. And we define the number of Stage II iterations (local convergence) to be the rest of the iterations.

According to Theorem 3.1, the overall number of iterations (in Stage I and Stage II) should be upper bounded by $O\big(\kappa\Phi \cdot \ln(\kappa\Phi \cdot \frac{\|w^\star\|}{\varepsilon})\big)$. Furthermore, Theorem 4.1 predicts that the number of iterations in Stage I and Stage II should be upper bounded by $O\big(\kappa\Phi \cdot \ln(\kappa\Phi)\big)$ and $O\big(\|B^{-1}\|\|A\| \cdot \ln(\frac{\xi}{\varepsilon})\big)$ respectively.

Figure 2a presents a scatter plot of the overall number of OnePDHG iterations and $\kappa\Phi\ln(\kappa\Phi)$, in which the red line represents the linear prediction model

$$\log_{10}(\text{predicted iteration number}) = \log_{10}\big(\kappa\Phi\ln(\kappa\Phi)\big) - 2.42\ . \tag{6.2}$$

The empirical R^2 value of this model is defined to be:

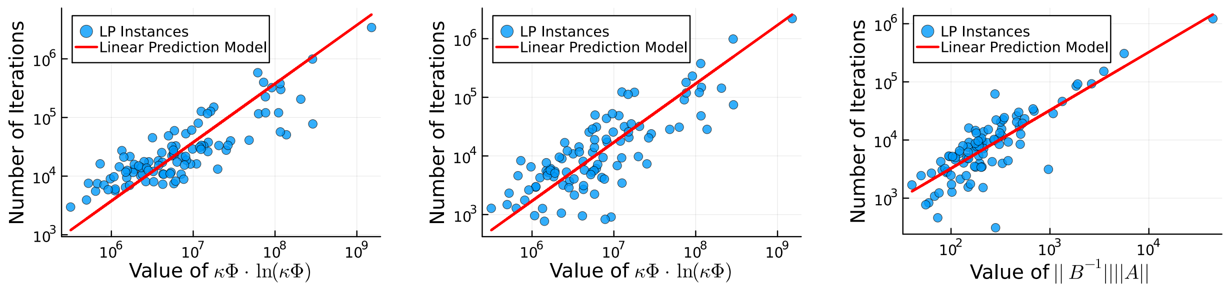
$$1 - \frac{\sum_{i=1}^{100}\big(\log_{10}(\text{actual iteration number of LP}_i) - \log_{10}(\text{predicted iteration number of LP}_i)\big)^2}{\sum_{i=1}^{100}\big(\log_{10}(\text{actual iteration number of LP}_i) - \text{mean of } \log_{10}(\text{actual iteration number})\big)^2}\ ,$$

and is equal to 0.5367. This R^2 value indicates that in these 100 instances, more than half of the variation in $\log_{10}$(actual iteration number) is accounted for by the model in (6.2). Although a few instances fall considerably below the line, notably, no instance significantly exceeds it. On the one hand, this observation suggests that $\kappa\Phi\ln(\kappa\Phi)$ serves as a reliable indicator of the performance of rPDHG. On the other hand, because Φ is equivalent to $\hat{\Phi}$ (Lemma 3.2), it offers direct experimental evidence supporting the role of the level set geometry condition number $\hat{\Phi}$, as proposed by Xiong and Freund [58],

Similarly, we also validate Theorem 4.1 through Figures 2b and 2c. These figures illustrates the relationship between the number of iterations in Stages I and II and the corresponding $\kappa\Phi\ln(\kappa\Phi)$ and $\|B^{-1}\|\|A\|$, respectively. One can also see a clear linear dependence in both stages. The empirical R^2 values of the two linear prediction models in Figures 2b and 2c are equal to 0.6146 and 0.6955, respectively. They are higher than the R^2 value 0.5367 in Figure 2a, indicating a tighter linear relationship. This confirms that the iteration bounds presented in Theorem 4.1 are refined, compared to Theorem 3.1.

FIGURE 2. Number of ONEPDHG iterations in Stage I (optimal basis identification) and Stage II (local convergence) and the corresponding values of $\kappa\Phi\ln(\kappa\Phi)$ and $\|B^{-1}\|\|A\|$.

(a) Stage I + Stage II (b) Stage I (c) Stage II



7. Optimized Reweighting and Step-Size Ratio Having shown the new accessible iteration bounds and confirmed their tightness via experiments, in this section, we show how to use the expression of the new iteration bounds to derive an optimized reweighting on the right-hand side vector b and the cost vector c. The reweighting is also equivalent to an optimized ratio between the primal-dual step-sizes (τ, σ).

Strategic configuration of the step-size ratio, also denoted as the "primal weight" has been observed to yield significant improvements in rPDHG's convergence rate (see, e.g., Applegate et al. [4], Lu and Yang [38], Xiong and Freund [57]). Applegate et al. [4] observe the practical benefits of adaptively adjusting the primal weight to balance the norm of accumulated updates in the primal and dual iterates. Xiong and Freund [56] verify the value of tuning step-sizes by providing a formula for an "optimized" step-size ratio. However, the formula incorporates the LP sharpness terms μ_p and μ_d, which are hard to approximate or compute *a priori*. In this section, we examine the step-size ratio from the perspective of reweighting the primal and dual variables. We present a simple expression of the "optimized" reweights, through which we can validate the heuristic of balancing primal weights.

For any reweights $\omega_1, \omega_2 > 0$, the reweighed problem is as follows:

$$\min_{x\in\mathbb{R}^n} \quad (\omega_1 \cdot c)^\top x \qquad \text{s.t.} \quad Ax = \omega_2 \cdot b \ , \ x \geq 0 \tag{7.1}$$

Compared to the original problem, the primal solutions are (re-)scaled by a factor of ω_2, while the dual solutions are (re-)scaled by a factor of ω_1. We use Φ_{ω_1,ω_2} to denote the Φ value of (7.1), e.g. $\Phi_{1,1}$ is the value of Φ for (1.1). Although κ remains invariant under reweighting, Φ_{ω_1,ω_2} may become different than $\Phi_{1,1}$, leading to different convergence rates of rPDHG. Applegate et al. [6] demonstrate that applying PDHG to the reweighted problem (7.1) with primal-dual step-sizes $(\tilde{\tau}, \tilde{\sigma})$ is equivalent to applying PDHG to the original problem (1.1) with primal-dual step-sizes (τ, σ) when the step-sizes are related by $(\tau, \sigma) = \left(\frac{\omega_1}{\omega_2}\tilde{\tau}, \frac{\omega_2}{\omega_1}\tilde{\sigma}\right)$.

For simplicity of notations, we let $\tilde{x}^\star$ and $\tilde{s}^\star$ represent the optimal solutions of the reweighted problem. Then $\tilde{x}^\star = \omega_2 \cdot x^\star$ and $\tilde{s}^\star = \omega_1 \cdot s^\star$. Thus, rPDHG can directly run on (7.1) and $(x^\star, s^\star)$ is computed by recovering from the convergent solution of rPDHG. We now present the iteration bound of an optimized weighting of this strategy as follows:

THEOREM 7.1. *Let the settings of the LP instance and Algorithm 1 be identical to those in Theorem 3.1. Consider rPDHG applied to* (7.1) *where* $\frac{\omega_1}{\omega_2} = \frac{\|x^\star\|_1}{\|s^\star\|_1}$, *and denote its iterations by* $(\tilde{x}^{n,k}, \tilde{s}^{n,k})$. *Then the total number of* ONEPDHG *iterations required to obtain the first outer iteration* N *such that* $w^{N,0} = \left(\frac{1}{\omega_2}\tilde{x}^{N,0}, \frac{1}{\omega_1}\tilde{s}^{N,0}\right)$ *is* ε*-optimal for the original problem is bounded above by*

$$O\left(\kappa \cdot \max\left\{\frac{\|x^\star\|_1}{\zeta_p}, \frac{\|s^\star\|_1}{\zeta_d}\right\} \cdot \ln\left(\frac{\kappa\frac{\|x^\star\|_1+\|s^\star\|_1}{\min\{\zeta_p,\ \zeta_d\}} \cdot \|w^\star\|}{\varepsilon}\right)\right). \tag{7.2}$$

Furthermore, once $\frac{\omega_1}{\omega_2} = \frac{\|x^\star\|_1}{\|s^\star\|_1}$, then (ω_1, ω_2) is actually an "optimal" reweighting, as it approximately minimizes Φ_{ω_1,ω_2}.

LEMMA 7.1. *When $\hat\omega_1, \hat\omega_2 > 0$ and $\frac{\hat\omega_1}{\hat\omega_2} = \frac{\|x^\star\|_1}{\|s^\star\|_1}$, we have*

$$\Phi_{\hat\omega_1,\hat\omega_2} = 2\cdot\max\left\{\frac{\|x^\star\|_1}{\zeta_p}, \frac{\|s^\star\|_1}{\zeta_d}\right\} \le 2\left(\min_{\omega_1,\omega_2>0}\Phi_{\omega_1,\omega_2}\right) . \tag{7.3}$$

Proof. The constraint matrix and the optimal basis remain unchanged, but the optimal primal solutions $x^\star$ and $s^\star$ are scaled by ω_2 and ω_1, respectively. Applying the definition of Φ to the reweighted problem (7.1) and using Lemma 5.3 yield the formula of Φ_{ω_1,ω_2}:

$$\Phi_{\omega_1,\omega_2} = \left(\omega_2\|x^\star\|_1 + \omega_1\|s^\star\|_1\right)\cdot\max\left\{\frac{1}{\omega_1\zeta_d}, \frac{1}{\omega_2\zeta_d}\right\} . \tag{7.4}$$

Using the condition $\frac{\hat\omega_1}{\hat\omega_2} = \frac{\|x^\star\|_1}{\|s^\star\|_1}$, $\Phi_{\hat\omega_1,\hat\omega_2} = 2\cdot\max\left\{\frac{\|x^\star\|_1}{\zeta_p}, \frac{\|s^\star\|_1}{\zeta_d}\right\}$, which is exactly the first equality of (7.3). On the other hand, (7.4) implies that $\Phi_{\omega_1,\omega_2} \ge \max\left\{\frac{\|x^\star\|_1}{\zeta_p}, \frac{\|s^\star\|_1}{\zeta_d}\right\}$ for all ω_1, ω_2. This proves the inequality of (7.3). □

With this expression of $\Phi_{\hat\omega_1,\hat\omega_2}$ presented above, the proof of Theorem 7.1 is straightforward by applying Theorem 3.1. We delay the proof to Appendix C.

Since $x^\star$ and $s^\star$ are unknown when running rPDHG on real-world instances, the reweighting presented in Theorem 7.1 is still not practical. But this finding demonstrates the practical effectiveness of adaptively adjusting the primal weight to balance primal and dual norms, as observed in Applegate et al. [4], Lu and Yang [38], Xiong and Freund [57]. Furthermore, in comparison to another impractical formula proposed by Xiong and Freund [56], this optimized reweighting approach is more amenable to practical approximation, as it only requires the ℓ_1-norms of the optimal solutions, which can be progressively estimated using the ℓ_1-norms of the iterates.

Comparing Theorem 7.1 and Theorem 5.1, the only difference is the coefficient of the linear convergence iteration bound. The coefficient in Theorem 7.1 is $\kappa\cdot\max\left\{\frac{\|x^\star\|_1}{\zeta_p}, \frac{\|s^\star\|_1}{\zeta_d}\right\}$, which is smaller (and can be much smaller) than $\kappa\cdot\frac{\|x^\star\|_1+\|s^\star\|_1}{\min\{\zeta_p,\zeta_d\}}$ of rPDHG for the original problem (Theorem 5.1). This distinction implies that when $\frac{\|x^\star\|_1}{\zeta_d}$ and $\frac{\|s^\star\|_1}{\zeta_p}$ significantly exceed $\frac{\|x^\star\|_1}{\zeta_p}$ and $\frac{\|s^\star\|_1}{\zeta_d}$, adjusting the reweights (or the equivalent step-size ratio) may have a huge benefit for rPDHG. This observation is also consistent with the experimental validation presented in Figure 3 of Xiong and Freund [56].

It is worth noting that our approach is complementary to the step-size ratio proposed by Xiong and Freund [56]. In fact, the approach of Xiong and Freund [56] can also approximately minimize Φ_{ω_1,ω_2}. See a detailed discussion in Appendix C.

Appendix A: Proofs of Section 3

A.1. Technical lemmas about the normalized duality gap and β-restart condition Note that Section 4.1 has reviewed the basic information of the β-restart condition and the normalized duality gap, which were omitted in Section 3 but will be heavily used throughout the proofs. Here we formally present the sublinear convergence result of the normalized duality gap summarized by Applegate et al. [6], using an equivalent result presented by Xiong and Freund [56]:

LEMMA A.1 **(Corollary 2.4 of Xiong and Freund [56])**. *Suppose that σ, τ satisfy* (2.7). *Then for any $z^0 := (x^0, y^0)$ with $x^0 \in \mathbb{R}^n_+$, the following inequality holds for all $k \ge 1$:*

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

Then an upper bound on the number of iterations for inner loops in rPDHG is as follows.

LEMMA A.2. *Suppose that* (4.6) *holds for* $z^{n,0}$ *and* $z^{n-1,0}$ *with* $\mathcal{L} > 0$. *Whenever* $k \geq \frac{8\mathcal{L}}{\beta}$, *sufficient decrease has been made on the normalized duality gap, i.e., the restart condition* (4.5) *is satisfied.*

Proof. For $k \geq 1$, Lemma A.1 implies:

$$\rho(\|\bar{z}^{n,k} - z^{n,0}\|_M; \bar{z}^{n,k}) \leq \frac{8\,\mathrm{Dist}_M(z^{n,0}, \mathcal{Z}^\star)}{k} . \tag{A.2}$$

If $\rho(\|z^{n,0} - z^{n-1,0}\|_M; z^{n,0}) = 0$, then $z^{n,0}$ is a saddle point of (2.2) and thus $z^{n,0} \in \mathcal{Z}^\star$. In this case, $z^{n,k} = z^{n,0}$ for all $k \geq 1$, and any $k \geq 1$ satisfies (4.5).

If $\rho(\|z^{n,0} - z^{n-1,0}\|_M; z^{n,0}) \neq 0$, dividing both sides of (A.2) by $\rho(\|z^{n,0} - z^{n-1,0}\|_M; z^{n,0})$ yields:

$$\frac{\rho(\|\bar{z}^{n,k} - z^{n,0}\|_M; \bar{z}^{n,k})}{\rho(\|z^{n,0} - z^{n-1,0}\|_M; z^{n,0})} \leq \frac{8}{k} \cdot \frac{\mathrm{Dist}_M(z^{n,0}, \mathcal{Z}^\star)}{\rho(\|z^{n,0} - z^{n-1,0}\|_M; z^{n,0})} \leq \frac{8}{k} \cdot \mathcal{L} \tag{A.3}$$

where the last inequality follows from (4.6). Therefore, when $k \geq \frac{8\mathcal{L}}{\beta}$, the right-hand side is no larger than β, i.e., $\frac{8}{k} \cdot \mathcal{L} \leq \beta$, and the restart condition (4.5) is satisfied. □

This lemma is essentially part of Theorem 2 of Applegate et al. [6].

A.2. Proof of Proposition 3.1

Proof of Proposition 3.1. For the second term of multiplication in the right-hand side of (3.1), we have

$$\begin{aligned}
&\max\left\{\max_{1\leq j\leq n-m} \frac{\sqrt{\left\|(B^{-1}N)_{\cdot,j}\right\|^2+1}}{s^\star_{m+j}}, \max_{1\leq i\leq m} \frac{\sqrt{\left\|(B^{-1}N)_{i,\cdot}\right\|^2+1}}{x^\star_i}\right\} \\
\leq\ & \frac{\max\left\{\max_{1\leq j\leq n-m}\sqrt{\left\|(B^{-1}N)_{\cdot,j}\right\|^2+1}, \max_{1\leq i\leq m}\sqrt{\left\|(B^{-1}N)_{i,\cdot}\right\|^2+1}\right\}}{\min\left\{\min_{1\leq i\leq m} x^\star_i, \min_{1\leq j\leq n-m} s^\star_{m+j}\right\}} \\
=\ & \frac{\max\left\{\sqrt{\|B^{-1}N\|_{1,2}^2+1}, \sqrt{\|B^{-1}N\|_{2,\infty}^2+1}\right\}}{\min_{1\leq i\leq n}\left\{x^\star_i + s^\star_i\right\}}
\end{aligned} \tag{A.4}$$

where the first equality holds because $x^\star$ and $s^\star$ are strictly complementary. Furthermore, because $\|\cdot\|_{1,2}$ and $\|\cdot\|_{2,\infty}$ norms are upper bounded by the $\|\cdot\|_2$ norm, we have the following inequalities:

$$\begin{aligned}
&\max\left\{\sqrt{\|B^{-1}N\|_{1,2}^2+1}, \sqrt{\|B^{-1}N\|_{2,\infty}^2+1}\right\} \leq \sqrt{\|B^{-1}N\|_2^2+1} = \sigma^+_{\max}(B^{-1}NN^\top B^{-\top})+1 \\
&= \sigma^+_{\max}(B^{-1}(NN^\top + BB^\top)B^{-\top}) = \sigma^+_{\max}(B^{-1}AA^\top B^{-\top}) = \|B^{-1}A\|_2^2 .
\end{aligned}$$

Applying these inequalities to the definition of Φ in (3.1) completes the proof. □

A.3. Proof of Lemma 3.1

In this subsection, we prove Lemma 3.1. We begin with the following lemma.

LEMMA A.3. *Suppose that* $Ac = 0$. *Algorithm 1 (rPDHG) is run starting from* $z^{0,0} = (x^{0,0}, y^{0,0}) = (0,0)$, *and the step-sizes* σ *and* τ *satisfy* (2.7). *Then for all* $n \geq 1$, *it holds that*

$$\mathrm{Dist}_M(z^{n,0}, \mathcal{Z}^\star) \leq \sqrt{2}c_{\tau,\sigma} \cdot \mathrm{Dist}(w^{n,0}, \mathcal{W}^\star) \leq (3\sqrt{2}+4)c^2_{\tau,\sigma} \cdot \hat{\Phi} \cdot \rho(\|z^{n,0} - z^{n-1,0}\|_M; z^{n,0}) . \tag{A.5}$$

In other words, condition (4.6) *holds with* $\mathcal{L} = (3\sqrt{2}+4)c^2_{\tau,\sigma}\hat{\Phi}$.

This lemma is Lemma 3.13 of Xiong and Freund [58] by taking limits as δ approaches 0 on both sides.

LEMMA A.4 **(Proposition 3.7 of Xiong and Freund [58])**. *Suppose that* $Ac = 0$ *and* $z^{0,0} = (0,0)$. *Then* $\|w^{0,0} - w^\star\| = \|(0,c) - w^\star\| \leq \|w^\star\|$.

Proof of Lemma 3.1. First of all, Lemma A.3 states:

$$\|z^{n,0}-z^\star\|_M \le \sqrt{2}c_{\tau,\sigma}\cdot\|z^{n,0}-z^\star\| \le (3\sqrt{2}+4)c_{\tau,\sigma}^2\cdot\hat{\Phi}\cdot\rho(\|z^{n,0}-z^{n-1,0}\|_M;z^{n,0}) \tag{A.6}$$

Substituting the step-sizes, we have $c_{\tau,\sigma}=\sqrt{2\kappa}$, $(3\sqrt{2}+4)c_{\tau,\sigma}^2\approx 16.4853\kappa\le 16.5\kappa$, and

$$\|z^{n,0}-z^\star\| \le (3\sqrt{2}+4)\sqrt{\kappa}\hat{\Phi}\cdot\rho(\|z^{n,0}-z^{n-1,0}\|_M;z^{n,0}) \le 8.25\sqrt{\kappa}\hat{\Phi}\cdot\rho(\|z^{n,0}-z^{n-1,0}\|_M;z^{n,0})\ . \tag{A.7}$$

Let $\bar{T}_\varepsilon$ denote the total number of OnePDHG iterations required to obtain the first outer iteration N that satisfies $\rho(\|z^{N,0}-z^{N-1,0}\|_M;z^{N,0})\le\varepsilon$. Theorem 3.5 of Xiong and Freund [58] and (A.6) then guarantee:

$$\bar{T}_{\bar\varepsilon}\ \le\ 23\cdot 16.5\kappa\hat{\Phi}\cdot\ln\left(\frac{23\|z^{0,0}-z^\star\|_M)}{\varepsilon}\right)\le 380\kappa\hat{\Phi}\cdot\ln\left(\frac{23\|z^{0,0}-z^\star\|_M}{\varepsilon}\right)\ . \tag{A.8}$$

Note that (A.7) ensures that when the normalized duality gap is sufficiently small, the distance to the optimal solution is correspondingly small. Therefore,

$$T\le\bar{T}_{\frac{\varepsilon}{8.25\sqrt{\kappa}\hat{\Phi}}}\le 380\kappa\hat{\Phi}\cdot\ln\left(\frac{189.75\sqrt{\kappa}\hat{\Phi}\cdot\|z^{0,0}-z^\star\|_M}{\varepsilon}\right)\le 380\kappa\hat{\Phi}\cdot\ln\left(\frac{379.5\kappa\hat{\Phi}\cdot\|w^{0,0}-w^\star\|}{\varepsilon}\right) \tag{A.9}$$

where the last inequality follows from (4.14) and $c_{\tau,\sigma}=\sqrt{2\kappa}$. Finally, by Lemma A.4, we have $\|w^{0,0}-w^\star\|\le\|w^\star\|$. Consequently, (A.9) leads to (3.6) of Lemma 3.1. □

Furthermore, similar results also exist for the goal of obtaining an iterate $z^{N,0}$ satisfying $\|z^{N,0}-z^\star\|_M$, which will be useful later in Appendix B.

Remark A.1. Under the same conditions as Lemma 3.1, let $\tilde{T}$ denote the total number of OnePDHG iterations required to obtain the first outer iteration N that satisfies both $\rho(\|z^{N-1,0}-z^{N,0}\|_M;z^{N,0})\le\frac{\varepsilon}{16.5\kappa\hat{\Phi}}$ and $\|z^{N,0}-z^\star\|_M\le\varepsilon$. Then, $\tilde{T}\le 380\kappa\hat{\Phi}\cdot\ln\left(\frac{760\kappa^{1.5}\hat{\Phi}\cdot\|w^\star\|}{\varepsilon}\right)$.

The proof of Remark A.1 is almost identical to that of Lemma 3.1, except (A.9) is replaced by the inequality derived from $\tilde{T}\le\bar{T}_{\frac{\varepsilon}{16.5\kappa\hat{\Phi}}}$ (due to (A.6)).

Appendix B: Proof of Theorem 4.1

Proof of Theorem 4.1. We first prove (4.1). According to Lemma 4.2, once N_0 satisfies $\|z^{N_0,0}-z^\star\|_M\le\bar\varepsilon$, which is equivalent to:

$$\|z^{N_0,0}-z^\star\|_M\le\bar\varepsilon=\frac{\sqrt{2}}{6}\cdot\frac{1}{\sqrt{2\kappa}}\cdot\xi=\frac{\xi}{6\sqrt{\kappa}}\ , \tag{B.1}$$

then for all $N>N_0$, we have:

$$(i)\ \ x_i^{N,0}\ge\sqrt{\tau}\|z^{N,0}-z^{N-1,0}\|_M\ \text{ for } i\in[m]\ ,\quad\text{and}\quad (ii)\ x_{m+j}^{N,0}=0 \text{ for } j\in[n-m]\ , \tag{B.2}$$

and the positive components of $x^{N,0}$ correspond exactly to the optimal basis. Therefore, $N_1\le N_0+1$ and T_1 is bounded above by the number of OnePDHG iterations required to obtain $z^{N_0+1,0}$.

According to Remark A.1, the number of OnePDHG iterations needed to obtain such a $z^{N_0,0}$ is upper bounded by the number of iterations $\tilde{T}$ required to obtain $z^{\tilde{N}_0,0}$ such that $\rho(\|z^{\tilde{N}_0-1,0}-z^{\tilde{N}_0,0}\|_M;z^{\tilde{N}_0,0})\le\frac{\bar\varepsilon}{16.5\kappa\hat{\Phi}}$, and

$$\tilde{T}\le 380\kappa\hat{\Phi}\cdot\ln\left(\frac{4560\kappa^2\hat{\Phi}\|w^\star\|}{\xi}\right)\ . \tag{B.3}$$

Furthermore, Lemma A.3 implies that, with step-sizes of Theorem 3.1, for all n:

$$\|z^{n,0}-z^\star\|_M\le 2\sqrt{\kappa}\|w^{n,0}-w^\star\|\le(6\sqrt{2}+8)\kappa\hat{\Phi}\cdot\rho(\|z^{n-1,0}-z^{n,0}\|_M;z^{n,0})\ . \tag{B.4}$$

Lemma A.2 guarantees that the number of additional OnePDHG iterations before obtaining the next outer loop iteration $z^{\tilde{N}_0+1,0}$ is at most $\left\lceil \frac{8\cdot(6\sqrt{2}+8)\kappa\hat{\Phi}}{\beta} \right\rceil$. Overall, since $\tilde{N}_0 \geq N_0$, the number of OnePDHG iterations required before obtaining $z^{\tilde{N}_0+1,0}$ is at most:

$$\tilde{T} + \left\lceil \frac{8\cdot(6\sqrt{2}+8)\kappa\hat{\Phi}}{\beta} \right\rceil \overset{\text{(B.3)}}{\leq} 380\kappa\hat{\Phi}\cdot\ln\left(\frac{4560\kappa^2\hat{\Phi}\|w^\star\|}{\xi}\right) + \left\lceil \frac{8\cdot(6\sqrt{2}+8)\kappa\hat{\Phi}}{\beta} \right\rceil ,$$

which we use T_{basis} to denote. Because Φ is equivalent to $\hat{\Phi}$ as demonstrated by Lemma 3.2 and $\frac{\|w^\star\|}{\xi} \leq \Phi$ from the definition, T_{basis} reduces to $O(\kappa\Phi\ln(\kappa\Phi))$ in (4.1).

Next, we prove (4.2). We first study how large $(N_2 - \tilde{N}_0 - 1)$ could be. We mainly consider the case $N_2 > \tilde{N}_0 + 1$; otherwise $N_2 \leq \tilde{N}_0 + 1$ and $T_2 \leq T_{basis}$. In this case $N_2 > \tilde{N}_0 + 1$, because N_2 is the first iteration such that $\|w^{N_2,0} - w^\star\| \leq \varepsilon$, for the previous iteration we have $\|w^{N_2-1,0} - w^\star\| > \varepsilon$. By (B.4), this implies:

$$\rho\left(\|z^{N_2-1,0} - z^{N_2-2,0}\|_M; z^{N_2-1,0}\right) > \frac{\sqrt{\kappa}\varepsilon}{(3\sqrt{2}+4)\kappa\hat{\Phi}} . \tag{B.5}$$

On the other hand, as shown in the definition of $\tilde{N}_0$,

$$\rho\left(\|z^{\tilde{N}_0,0} - z^{\tilde{N}_0+1,0}\|_M; z^{\tilde{N}_0+1,0}\right) \leq \beta\cdot\frac{\bar{\varepsilon}}{16.5\kappa\hat{\Phi}} = \frac{1}{e}\cdot\frac{1}{16.5\kappa\hat{\Phi}}\cdot\frac{\xi}{6\sqrt{\kappa}} \leq \frac{\xi}{99e\cdot\kappa^{1.5}\hat{\Phi}} \leq \frac{\xi}{269\cdot\kappa^{1.5}\hat{\Phi}} . \tag{B.6}$$

Furthermore, due to the restart condition:

$$\rho\left(\|z^{N_2-1,0} - z^{N_2-2,0}\|_M; z^{N_2-1,0}\right) \leq \beta^{N_2-\tilde{N}_0-2}\cdot\rho(\|z^{N_0,0} - z^{\tilde{N}_0+1,0}\|_M; z^{\tilde{N}_0+1,0}) . \tag{B.7}$$

Substituting (B.5) and (B.6) into (B.7) yields an upper bound of $(N_2 - \tilde{N}_0 - 1)$:

$$\frac{\sqrt{\kappa}\varepsilon}{(3\sqrt{2}+4)\kappa\hat{\Phi}} < \left(\frac{1}{e}\right)^{N_2-\tilde{N}_0-2}\cdot\frac{\xi}{269\cdot\kappa^{1.5}\hat{\Phi}} \Rightarrow N_2 - \tilde{N}_0 - 1 \leq \ln\left(\frac{e\cdot(3\sqrt{2}+4)\kappa\hat{\Phi}}{269\kappa^2\hat{\Phi}}\right) + \ln\left(\frac{\xi}{\varepsilon}\right) \leq \ln\left(\frac{\xi}{\varepsilon}\right) . \tag{B.8}$$

Here the final inequality is due to $\kappa \geq 1$ and $\frac{e\cdot(3\sqrt{2}+4)\kappa\hat{\Phi}}{269\kappa^2\hat{\Phi}} \leq 1$. Since we have assumed $N_2 > \tilde{N}_0 + 1$, the upper bound of $(N_2 - \tilde{N}_0 - 1)$ in (B.8) has to be strictly positive. Once $\ln\left(\frac{\xi}{\varepsilon}\right) \leq 0$, then it is no longer in the case $N_2 > \tilde{N}_0 + 1$ and as previously stated before we already have $N_2 \leq \tilde{N}_0 + 1$ and $T_2 \leq T_{basis}$. In conclusion, we can assert that $N_2 - \tilde{N}_0 - 1 \leq \max\left\{0, \ln\left(\frac{\xi}{\varepsilon}\right)\right\}$.

We now turn our attention to the number of OnePDHG iterations between $z^{N_0+1,0}$ and $z^{N_2,0}$. Remark 4.1 ensures that for all $N \geq \tilde{N}_0 + 1$:

$$\|z^{N,0} - z^\star\|_M \leq 4\|B^{-1}\|\|A\|\cdot\rho\left(\|z^{N,0} - z^{N-1,0}\|_M; z^{N,0}\right) . \tag{B.9}$$

Then due to Lemma A.2, the number of iterations in inner loops is at most $\left\lceil \frac{8\cdot 4\|B^{-1}\|\|A\|}{1/e} \right\rceil$. Therefore, we can bound the overall number of iterations T_2 as follows:

$$T_2 \leq T_{basis} + (N_2 - \tilde{N}_0 - 1)\cdot\left\lceil \frac{8\cdot 4\|B^{-1}\|\|A\|}{1/e} \right\rceil \tag{B.10}$$

Substituting the upper bound $N_2 - \tilde{N}_0 - 1 \leq \max\left\{0, \ln\left(\frac{\xi}{\varepsilon}\right)\right\}$ into (B.10) finishes the proof. □

Appendix C: Proofs of Section 7

Proof of Theorem 7.1. Without loss of generality, we study the case $\varepsilon \le \|w^\star\|$. Otherwise, 0 is sufficiently optimal and the optimization problem is trivial.

Note that $\|\tilde{w}^\star\| \le \max\{\omega_1, \omega_2\} \cdot \|w^\star\|$, and $\|w^{N,0} - w^\star\| \le \varepsilon$ if $\|\tilde{w}^{N,0} - \tilde{w}^\star\| \le \frac{\varepsilon}{\max\{\frac{1}{\omega_1}, \frac{1}{\omega_2}\}} = \varepsilon \cdot \min\{\omega_1, \omega_2\}$. According to Theorem 3.1, finding an $(\varepsilon \cdot \min\{\omega_1, \omega_2\})$-optimal solution in the reweighted problem requires at most

$$O\left(\kappa \cdot \Phi_{\omega_1,\omega_2} \cdot \ln\left(\frac{\kappa \cdot \Phi_{\omega_1,\omega_2} \cdot \frac{\max\{\omega_1,\omega_2\}}{\min\{\omega_1,\omega_2\}} \cdot \|w^\star\|}{\varepsilon}\right)\right) \tag{C.1}$$

iterations. Using $\frac{\omega_1}{\omega_2} = \frac{\|x^\star\|_1}{\|s^\star\|_1}$, we have $\frac{\max\{\omega_1,\omega_2\}}{\min\{\omega_1,\omega_2\}} = \frac{\max\{\|x^\star\|_1, \|s^\star\|_1\}}{\min\{\|x^\star\|_1, \|s^\star\|_1\}} = \max\left\{\frac{\|x^\star\|_1}{\|s^\star\|_1}, \frac{\|s^\star\|_1}{\|x^\star\|_1}\right\}$. Next we study the value of $\Phi_{\omega_1,\omega_2} \cdot \max\left\{\frac{\|x^\star\|_1}{\|s^\star\|_1}, \frac{\|s^\star\|_1}{\|x^\star\|_1}\right\}$, denoted by γ for simplicity.

Note that $\Phi_{\omega_1,\omega_2} = \max\left\{\frac{\|x^\star\|_1}{\zeta_p}, \frac{\|s^\star\|_1}{\zeta_d}\right\}$. Without loss of generality, we assume $\frac{\|x^\star\|_1}{\zeta_p} \ge \frac{\|s^\star\|_1}{\zeta_d}$. If $\frac{\|x^\star\|_1}{\|s^\star\|_1} \le \frac{\|s^\star\|_1}{\|x^\star\|_1}$, then $\gamma = \frac{\|x^\star\|_1}{\zeta_p} \cdot \frac{\|s^\star\|_1}{\|x^\star\|_1} = \frac{\|s^\star\|_1}{\zeta_p} \le \Phi_{1,1}$. If $\frac{\|x^\star\|_1}{\|s^\star\|_1} \ge \frac{\|s^\star\|_1}{\|x^\star\|_1}$, then

$$\gamma \le \gamma \cdot \frac{\|s^\star\|_1}{\zeta_d} = \frac{\|x^\star\|_1}{\zeta_p} \cdot \frac{\|x^\star\|_1}{\|s^\star\|_1} \cdot \frac{\|s^\star\|_1}{\zeta_d} = \frac{\|x^\star\|_1}{\zeta_p} \cdot \frac{\|x^\star\|_1}{\zeta_d} \le \Phi_{1,1}^2 .$$

where the first inequality is due to $\frac{\|s^\star\|_1}{\zeta_d} \ge 1$ according to the formula of ζ_d presented in Lemma 5.3, and the final inequality is due to the definition of $\Phi_{1,1}$, presented by that of Φ in Theorem 5.1. Overall, $\gamma \le \Phi_{1,1}^2$. Finally, because $\kappa \ge 1$ and $\frac{\|w^\star\|}{\varepsilon} \ge 1$, we have

$$\ln\left(\frac{\kappa \cdot \Phi_{\omega_1,\omega_2} \cdot \frac{\max\{\omega_1,\omega_2\}}{\min\{\omega_1,\omega_2\}} \cdot \|w^\star\|}{\varepsilon}\right) \le \ln\left(\frac{\kappa \cdot \Phi_{1,1}^2 \cdot \|w^\star\|}{\varepsilon}\right) \le 2\ln\left(\frac{\kappa \cdot \Phi_{1,1} \cdot \|w^\star\|}{\varepsilon}\right) .$$

Substituting it back to (C.1) completes the proof. □

Comparing Theorem 7.1 with Remark 3.4 of Xiong and Freund [56]. Remark 3.4 of Xiong and Freund [56] proposes the "optimized" step-sizes

$$\tau = \frac{\mu_d \cdot \mathrm{Dist}(0, V_p)}{2\kappa\mu_p \cdot \mathrm{Dist}(0, V_d)} \quad \text{and} \quad \sigma = \frac{\mu_p \cdot \mathrm{Dist}(0, V_d)}{2\lambda_{\max}\lambda_{\min}\mu_d \cdot \mathrm{Dist}(0, V_p)} \tag{C.2}$$

where μ_p and μ_d denote the LP sharpness of the primal and dual problems, as defined in Definition 5.3. As demonstrated by Applegate et al. [6], setting the step-size ratio (C.2) is equivalent to setting primal weights $(\breve{\omega}_1, \breve{\omega}_2)$ such that

$$\frac{\breve{\omega}_1}{\breve{\omega}_2} = \frac{\mu_d \cdot \mathrm{Dist}(0, V_p)}{\mu_p \cdot \mathrm{Dist}(0, V_d)} \tag{C.3}$$

with the standard step-sizes τ, σ of Theorem 3.1. Furthermore, due to Lemmas 5.1 and 5.3, (C.3) is actually equivalent to $\frac{\breve{\omega}_1}{\breve{\omega}_2} = \frac{\zeta_p}{\zeta_d}$. Now, substituting $(\omega_1, \omega_2) = (\breve{\omega}_1, \breve{\omega}_2)$ into (7.4) yields

$$\Phi_{\breve{\omega}_1,\breve{\omega}_2} = \frac{\|x^\star\|_1}{\zeta_p} + \frac{\|s^\star\|_1}{\zeta_d} .$$

This is equivalent to $\Phi_{\hat{\omega}_1,\hat{\omega}_2}$ (when $\frac{\hat{\omega}_1}{\hat{\omega}_2} = \frac{\|x^\star\|_1}{\|s^\star\|_1}$) up to a constant of 2, so Remark 3.4 of Xiong and Freund [56] is also equivalent to approximately optimizing the geometric measure $\hat{\Phi}_{\omega_1,\omega_2}$ via reweighting.

Acknowledgments

The author is grateful to Robert M. Freund for the valuable suggestions while preparing this manuscript. The author also thanks Yinyu Ye, Haihao Lu, Tianyi Lin, Javier Pena and Xin Jiang for the helpful discussions.

References

[1] Alizadeh F, Haeberly JPA, Overton ML (1998) Primal-dual interior-point methods for semidefinite programming: Convergence rates, stability and numerical results. *SIAM Journal on Optimization* 8(3):746–768.

[2] Anstreicher KM, Ji J, Potra FA, Ye Y (1993) Average performance of a self–dual interior point algorithm for linear programming. *Complexity in Numerical Optimization*, 1–15 (World Scientific).

[3] Anstreicher KM, Ji J, Potra FA, Ye Y (1999) Probabilistic analysis of an infeasible-interior-point algorithm for linear programming. *Mathematics of Operations Research* 24(1):176–192.

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

[5] Applegate D, Díaz M, Lu H, Lubin M (2024) Infeasibility detection with primal-dual hybrid gradient for large-scale linear programming. *SIAM Journal on Optimization* 34(1):459–484.

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

[7] Basu K, Ghoting A, Mazumder R, Pan Y (2020) Eclipse: An extreme-scale linear program solver for web-applications. *Proceedings of the 37th International Conference on Machine Learning*, 704–714 (PMLR).

[8] Bertsimas D, Tsitsiklis JN (1997) *Introduction to linear optimization*, volume 6 (Athena scientific Belmont, MA).

[9] Biele A, Gade D (2024) FICO® Xpress solver 9.4. `https://community.fico.com/s/blog-post/a5QQi0000019II5MAM/fico4824`. Accessed: 2024-09-11.

[10] Borgwardt KH (1987) *The simplex method: A probabilistic analysis*, volume 1 (Springer Nature).

[11] Bowman EH (1956) Production scheduling by the transportation method of linear programming. *Operations Research* 4(1):100–103.

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

[13] Charnes A, Cooper WW (1954) The stepping stone method of explaining linear programming calculations in transportation problems. *Management Science* 1(1):49–69.

[14] Chen K, Sun D, Yuan Y, Zhang G, Zhao X (2024) HPR-LP: An implementation of an HPR method for solving linear programming. *arXiv preprint arXiv:2408.12179* .

[15] Cormen TH, Leiserson CE, Rivest RL, Stein C (2022) *Introduction to algorithms* (MIT Press), 4th edition.

[16] Dantzig GB (2002) Linear programming. *Operations Research* 50(1):42–47.

[17] De Rosa A, Khajavirad A (2024) On the power of linear programming for k-means clustering. *arXiv preprint arXiv:2402.01061* .

[18] Deng Q, Feng Q, Gao W, Ge D, Jiang B, Jiang Y, Liu J, Liu T, Xue C, Ye Y, Zhang C (2024) An enhanced alternating direction method of multipliers-based interior point method for linear and conic optimization. *INFORMS Journal on Computing* 0(0).

[19] Drusvyatskiy D, Lewis AS (2011) Generic nondegeneracy in convex optimization. *Proceedings of the American Mathematical Society* 2519–2527.

[20] Esser E, Zhang X, Chan TF (2010) 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.

[21] Fender A (2024) Advances in optimization AI. `https://resources.nvidia.com/en-us-ai-optimization-content/gtc24-s62495`. Accessed: 2024-09-11.

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

[23] Freund RM, Vera JR (1999) Condition-based complexity of convex optimization in conic linear form via the ellipsoid algorithm. *SIAM Journal on Optimization* 10(1):155–176.

[24] Ge D, Hu H, Huangfu Q, Liu J, Liu T, Lu H, Yang J, Ye Y, Zhang C (2024) cuPDLP-C. `https://github.com/COPT-Public/cuPDLP-C`. Accessed: 2024-09-11.

[25] Greene WH (2017) *Econometric analysis* (Pearson), 7th edition.

[26] Güler O, Ye Y (1993) Convergence behavior of interior-point algorithms. *Mathematical Programming* 60(1-3):215–228.

[27] Hanssmann F, Hess SW (1960) A linear programming approach to production and employment scheduling. *Management Science* 1(1):46–51.

[28] Hinder O (2023) Worst-case analysis of restarted primal-dual hybrid gradient on totally unimodular linear programs. *arXiv preprint arXiv:2309.03988* .

[29] Hough M, Vavasis SA (2024) A primal-dual Frank-Wolfe algorithm for linear programming. *arXiv preprint arXiv:2402.18514* .

[30] Huang Y, Zhang W, Li H, Xue W, Ge D, Liu H, Ye Y (2024) Restarted primal-dual hybrid conjugate gradient method for large-scale quadratic programming. *arXiv preprint arXiv:2405.16160* .

[31] Koch T, Berthold T, Pedersen J, Vanaret C (2022) Progress in mathematical programming solvers from 2001 to 2020. *EURO Journal on Computational Optimization* 10:100031.

[32] Li B, Yang L, Chen Y, Wang S, Chen Q, Mao H, Ma Y, Wang A, Ding T, Tang J, Sun R (2024) PDHG-unrolled learning-to-optimize method for large-scale linear programming. *arXiv preprint arXiv:2406.01908* .

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

[34] Liu XW, Dai YH, Huang YK (2022) A primal-dual majorization-minimization method for large-scale linear programs. *arXiv preprint arXiv:2208.03672* .

[35] Lu H, Applegate D (2024) Scaling up linear programming with PDLP. `https://research.google/blog/scaling-up-linear-programming-with-pdlp/`. Accessed: 2024-09-26.

[36] Lu H, Simester D, Zhu Y (2023) Optimizing scalable targeted marketing policies with constraints. *Available at SSRN 4668582* .

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

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

[39] Lu H, Yang J (2023) A practical and optimal first-order method for large-scale convex quadratic programming. *arXiv preprint arXiv:2311.07710* .

[40] Lu H, Yang J (2024) On the geometry and refined rate of primal–dual hybrid gradient for linear programming. *Mathematical Programming* 1–39.

[41] Lu H, Yang J (2024) PDOT: A practical primal-dual algorithm and a GPU-based solver for optimal transport. *arXiv preprint arXiv:2407.19689* .

[42] Lu H, Yang J (2024) Restarted Halpern PDHG for linear programming. *arXiv preprint arXiv:2407.16144* .

[43] Lu H, Yang J, Hu H, Huangfu Q, Liu J, Liu T, Ye Y, Zhang C, Ge D (2023) cuPDLP-C: A strengthened implementation of cuPDLP for linear programming by C language. *arXiv preprint arXiv:2312.14832* .

[44] Mehrotra S, Ye Y (1993) Finding an interior point in the optimal face of linear programs. *Mathematical Programming* 62(1-3):497–515.

[45] Mirrokni V (2023) 2022 & beyond: Algorithmic advances. `https://ai.googleblog.com/2023/02/google-research-2022-beyond-algorithmic.html`. Accessed: 2024-09-11.

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

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

[48] Pena J, Vera JC, Zuluaga LF (2021) New characterizations of Hoffman constants for systems of linear constraints. *Mathematical Programming* 187:79–109.

[49] Pock T, Cremers D, Bischof H, Chambolle A (2009) An algorithm for minimizing the Mumford–Shah functional. *Proceedings of the 2009 International Conference on Computer Vision*, 1133–1140 (IEEE).

[50] Potra FA (1994) A quadratically convergent predictor—corrector method for solving linear programs from infeasible starting points. *Mathematical Programming* 67(1):383–406.

[51] Todd MJ (1991) Probabilistic models for linear programming. *Mathematics of Operations Research* 16(4):671–693.

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

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

[54] Wagner M, Meller J, Elber R (2004) Large-scale linear programming techniques for the design of protein folding potentials. *Mathematical Programming* 101:301–318.

[55] Wang H, Ghosal P, Mazumder R (2023) Linear programming using diagonal linear networks. *arXiv preprint arXiv:2310.02535* .

[56] Xiong Z, Freund RM (2023) Computational guarantees for restarted PDHG for LP based on "limiting error ratios" and LP sharpness. *arXiv preprint arXiv:2312.14774* .

[57] Xiong Z, Freund RM (2023) On the relation between LP sharpness and limiting error ratio and complexity implications for restarted PDHG. *arXiv preprint arXiv:2312.13773* .

[58] Xiong Z, Freund RM (2024) The role of level-set geometry on the performance of PDHG for conic linear optimization. *arXiv preprint arXiv:2406.01942* .

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

[60] Ye Y (1992) On the finite convergence of interior-point algorithms for linear programming. *Mathematical Programming* 57(1):325–335.

[61] Ye Y (1994) Toward probabilistic analysis of interior-point algorithms for linear programming. *Mathematics of Operations Research* 19(1):38–52.

[62] Ye Y (2011) The simplex and policy-iteration methods are strongly polynomial for the Markov decision problem with a fixed discount rate. *Mathematics of Operations Research* 36(4):593–603.