Operations Research Letters 57 (2024) 107199

Contents lists available at ScienceDirect

Operations Research Letters

journal homepage: www.elsevier.com/locate/orl

Worst-case analysis of restarted primal-dual hybrid gradient on totally unimodular linear programs

Oliver Hinder

University of Pittsburgh, Pittsburgh, PA, United States

ARTICLE INFO

Keywords:
Totally unimodular
First-order methods
Linear programming
Primal-dual methods
Worst-case analysis
Restarts
Low memory

ABSTRACT

We analyze restarted PDHG on totally unimodular linear programs. In particular, we show that restarted PDHG finds an ϵ-optimal solution in $O(Hm_1^{2.5}\sqrt{\mathbf{nnz}(A)}\log(Hm_2/\epsilon))$ matrix-vector multiplies where m_1 is the number of constraints, m_2 the number of variables, $\mathbf{nnz}(A)$ is the number of nonzeros in the constraint matrix, H is the largest absolute coefficient in the right hand side or objective vector, and ϵ is the distance to optimality of the outputted solution.

1. Introduction

Consider the following linear program:

$$\underset{x\in\mathbb{R}^{m_2}}{\text{minimize}}\, c^\top x \tag{1a}$$

$$Ax = b \tag{1b}$$

$$x \geq 0 \tag{1c}$$

and its dual maximize$_{y\in\mathbb{R}^{m_1}}$ $b^\top y$ subject to $A^\top y \leq c$, where m_1 and m_2 are positive integers, x and y are the primal and dual variables, and $A \in \mathbb{R}^{m_1\times m_2}$, $c \in \mathbb{R}^{m_2}$, $b \in \mathbb{R}^{m_1}$ are the problem parameters. Traditional methods for solving linear programs such as simplex and interior point methods require linear system factorizations, which have high memory overhead and are difficult to parallelize. Recently, there has been interest in using first-order methods for solving linear programs that use matrix-vector multiplication as their key primitive [16,2,19]. The advantage of matrix-vector multiplication is that it can be efficiently parallelized across multiple cores or machines. Moreover, matrix-vector multiplication has low memory footprint, using minimal additional memory beyond storing the problem data. These properties make these first-order methods suitable for tackling extreme-scale problems.

First-order methods reformulate finding a primal and dual optimal solution to the linear program (1) as solving a minimax problem:

$$\min_{x\geq 0}\max_{y\in\mathbb{R}^{m_1}} \mathcal{L}(x,y) = c^\top x + b^\top y - y^\top Ax \tag{2}$$

and then apply methods designed for solving minimax problems such as primal-dual hybrid gradient (PDHG) [4] or the alternating direction method of multipliers (ADMM) [9]. See Applegate et al. [1] and Applegate et al. [2] for a more exhaustive introduction and list of references.

Empirically, a promising first-order method for solving linear programs is primal-dual hybrid gradient for linear programming (PDLP) [1]. PDLP is based on restarted PDHG, combined with several other heuristics, for example, preconditioning and adaptively choosing the primal and dual step sizes. Restarted PDHG was analyzed by Applegate et al. [2] on linear programs, but their convergence bounds depend on the Hoffman constant of the KKT system. This bound is difficult to interpret and is not easily computable. Recent work by Lu and Yang [17] and Xiong and Freund [25] develop different, more interpretable linear convergence bounds for PDHG. However, it is unclear if these new bounds would be useful for analyzing totally unimodular linear programs.

In this paper, we extend the analysis of Applegate et al. [2] to provide an explicit complexity bound when this method is applied to totally unimodular linear programs [12]. Totally unimodular linear programs are an important subclass of linear programs which, for any integer right hand side and objective coefficients, all extreme points are integer. This subclass is of particular interest to the integer programming community [7]. It also encapsulates the minimum cost flow problem, an important subclass of linear programs, for which almost linear-time algorithms exist [5].

This work analyzes a general-purpose linear programming method on a specialized problem. Many papers perform this style of analysis

E-mail address: ohinder@pitt.edu.

https://doi.org/10.1016/j.orl.2024.107199
Received 2 March 2024; Received in revised form 5 October 2024; Accepted 11 October 2024
Available online 15 October 2024

Table 1
Total number of iterations of restarted PDHG until the distance to optimality contracts by a factor of ten on the totally unimodular linear program $\min_{x_1,x_2\geq 0}(H-1)x_1+x_2$ s.t. $x_1+x_2=H$. The restart length is fixed but instance-wise tuned over the grid $2^1, 2^2, \ldots, 2^{21}$ (with number of total iterations of the best restart length reported). The step size η is set to 0.5 and the primal and dual variables are initialized at the origin.

H	10^2	10^4	10^6
# iterations	1.8×10^2	1.7×10^4	1.7×10^6

for the simplex method. For example, even though the simplex method has worst-case exponential runtime on general linear programs, improved guarantees for the simplex method exist for subclasses such as Markov decision processes [26,22], minimum cost flow [20,8] and totally unimodular linear programs [13,18]. In particular, Kitahara and Mizuno [13] shows that the number of iterations of the simplex method to find an exact optimal solution on a totally unimodular linear program with a nondegenerate primal is $m_2\lceil m_1\|b\|_1\log(m_2\|b\|_1)\rceil$. Better complexities for this problem can be achieved by interior point methods [15], although at the cost of potentially much higher memory usage.

Finally, in independent work, Cole et al. [6] studies the performance of first-order methods for linear programming on totally unimodular linear programs. Their work is strongly related to ours. However, there are a few important differences: (i) our result studies an algorithm with proven practical performance [1], (ii) their bounds depend on $\log(H)$ instead of H but for a fixed H our bound has better dependence on m_1, m_2 and $\mathbf{nnz}(A)$, and (iii) their results also extend to more general A matrices (those with bounded max circuit imbalance measure). Our experiments summarized in Table 1 indicate that restarted PDHG has an iteration bound of at least $\Omega(H)$ on totally unimodular linear programs.

Notation Let $\mathbb{R}$ be the set of real numbers and $\mathbb{N}$ be the set of natural numbers starting from one. Denote $\{1,\ldots,m\}$ by $[m]$. Let $\mathbf{nnz}(A)$ be the number of nonzeros in A. Assume $m_2\geq m_1$ and that $|b_i|\leq H$ for all $i\in[m_1]$ and $|c_j|\leq H$ for all $j\in[m_2]$. Let $\|\cdot\|_2$ be the Euclidean norm for vectors and spectral norm for matrices. Let $\sigma_{\min}(M) := \min_{\|v\|_2=1}\|Mv\|_2$ be the minimum singular value of a matrix M, $Z=\{x\in\mathbb{R}^{m_2} : x\geq 0\}\times\mathbb{R}^{m_1}$, $W_r(z) := \{\hat{z}\in Z : \|z-\hat{z}\|_2\leq r\}$, and $\mathbf{dist}(z,Z) := \min_{\hat{z}\in Z}\|z-\hat{z}\|_2$. Let $X^\star$ be the set of optimal primal solutions to (1) and $Y^\star$ be the set of optimal dual solutions. Define $Z^\star = X^\star\times Y^\star$. Let e_i be a vector containing a one in the ith entry and zero elsewhere. Let $\mathbf{1}$ be a matrix or vector of ones, and $\mathbf{0}$ be a matrix or vector of zeros. Let $(\cdot)^+ = \max\{\cdot,\mathbf{0}\}$ where the max operator is applied element-wise.

Paper outline Section 2 provides background on restarted PDHG, Section 3 provides a new Hoffman bound that we will find useful and Section 4 proves the main result.

2. Background on restarted PDHG

This section introduces concepts from Applegate et al. [2] that will be useful for our analysis. For ease of exposition, we specialize PDHG to linear programming (Algorithm 1). See Chambolle and Pock [4] for the general PDHG formula.

A key concept for restarted PDHG is the *normalized duality gap* [2] defined for $r>0$ as

$$\rho_r(z) := \frac{\max_{\hat{z}\in W_r(z)}\mathcal{L}(x,\hat{y})-\mathcal{L}(\hat{x},y)}{r}$$

where for conciseness, we use the notation $(\hat{x},\hat{y})=\hat{z}$ (this notation is used throughout the paper, i.e., we also have $(x,y)=z$), and for completeness define $\rho_0(z) := \limsup_{r\to 0^+}\rho_r(z)$. The normalized duality gap is preferable over the standard duality gap, $\max_{z\in Z}\mathcal{L}(x,\hat{y})-\mathcal{L}(\hat{x},y)$, because when Z is unbounded the standard duality gap can be infinite even when we are arbitrarily close to an optimal solution in both the primal and dual. Next, we introduce the definition of a primal-dual problem being sharp (Definition 1) and PDHG with adaptive restarts (Algorithm 2) along with three results from Applegate et al. [2] that we will use in this paper.

Algorithm 1: One step of PDHG on (2).

function `OneStepOfPDHG`(z,η)
 $x' \leftarrow \mathbf{proj}_{x\geq 0}\left(x-\eta(c-A^\top y)\right)$;
 $y' \leftarrow y-\eta(b-A(2x'-x))$;
 return (x',y')
end

Definition 1 *(Definition 1 of Applegate et al. [2])*. We say a primal-dual problem (2) is α-sharp on the set $S\subseteq Z$ if for all $r\in(0,\mathbf{diam}(S)]$ and $z\in S$ that $\alpha\,\mathbf{dist}(z,Z^\star)\leq\rho_r(z)$.

Algorithm 2: PDHG with adaptive restarts [2, Algorithm 1].

Input: $z^{0,0},\tau^0,\beta,\eta$
for $n=0,\ldots,\infty$ **do**
 $t\leftarrow 0$
 repeat
  $z^{n,t+1}\leftarrow$ `OneStepOfPDHG`$(z^{n,t},\eta)$;
  $\bar{z}^{n,t+1}\leftarrow\frac{1}{t+1}\sum_{i=1}^{t+1}z^{n,i}$;
  $t\leftarrow t+1$
 until *[$n=0$ and $t\geq\tau^0$] or* $\rho_{\|\bar{z}^{n,t}-z^{n,0}\|_2}(\bar{z}^{n,t})\leq\beta\rho_{\|z^{n,0}-z^{n-1,0}\|_2}(z^{n,0})$;
 $z^{n+1,0}\leftarrow\bar{z}^{n,t}$;
end

Lemma 1. *Algorithm 2 for $z^{0,0}\in Z$ satisfies $z^{n,0}\in W_{\theta\,\mathbf{dist}(z^{0,0},Z^\star)}(z^{0,0})$ for $\theta=2\sqrt{\frac{1+\eta\|A\|_2}{1-\eta\|A\|_2}}$.*

Proof. Proposition 9 of Applegate et al. [2] (which uses the norm $\|z\|_{\eta A} := \|x\|_2^2-2\eta x^\top Ay+\|y\|_2^2$) states that $\|z^{n,0}-z^{0,0}\|_{\eta A}\leq 2\|z^{0,0}-z^\star\|_{\eta A}$ for any starting point $z^{0,0}\in Z$ and optimal solution $z^\star\in Z^\star$. Proposition 7 of Applegate et al. [2] states that $(1-\eta\|A\|_2)\|z\|_2^2\leq\|z\|_{\eta A}\leq(1+\eta\|A\|_2)\|z\|_2^2$ for all $z\in Z$. Combining these two statements yields $(1-\eta\|A\|_2)\|z^{n,0}-z^{0,0}\|_2^2\leq\|z^{n,0}-z^{0,0}\|_{\eta A}^2\leq 2^2\|z^{0,0}-z^\star\|_{\eta A}^2\leq 2^2(1+\eta\|A\|_2)\|z^{0,0}-z^\star\|_2^2$ for all $z^\star\in Z^\star$. Rearranging gives the result. □

Theorem 1 *(Applegate et al. [2]). Consider the sequence $\{z^{n,0}\}_{n=0}^\infty$ and $\{\tau^n\}_{n=1}^\infty$ generated by Algorithm 2 with $\eta\in(0,1/\|A\|_2)$ and $\beta\in(0,1)$. Suppose that there exists a set $S\subseteq Z$ such that $z^{n,0}\in S$ for any $n\geq 0$ and the primal-dual problem (2) is α-sharp on the set S. Then, for each outer iteration $n\in\mathbb{N}$ we have*

- *The restart length, τ^n, is upper bounded by $t^\star$: $\tau^n\leq t^\star := \left\lceil\frac{2C(q+2)}{\alpha\beta}\right\rceil$ with $C := \frac{2}{\eta(1-\eta\|A\|_2)}$ and $q := 4\frac{1+\eta\|A\|_2}{1-\eta\|A\|_2}$.*
- *The distance to the primal-dual optimal solution set decays linearly: $\mathbf{dist}(z^{n,0},Z^\star)\leq\beta^{n\frac{t^\star}{\tau^0}}\mathbf{dist}(z^{0,0},Z^\star)$.*

Proof. See Theorem 2 and Corollary 2 of Applegate et al. [2]. □

Lemma 2. *For all $R\in(0,\infty)$, $z\in Z$, and $r\in(0,R]$ with $\|z\|_2\leq R$ we have*

$$\frac{1}{2}\left\|\begin{pmatrix}\frac{1}{R}(c^\top x-b^\top y)^+\\ Ax-b\\ (A^\top y-c)^+\end{pmatrix}\right\|_2\leq\rho_r(z).$$

Proof. This is a variant of Lemma 4 of Applegate et al. [2]. To prove the result it suffices to combine Equation (22) and (25) of [2]. □

3. A Hoffman bound that explicitly takes into account nonnegativity and inequality constraints

Hoffman bounds guarantee how much the distance to feasibility decreases as the constraint violation decreases. Typical Hoffman bounds consider a linear inequality system of the form: $Kz \leq k$ where K is a matrix and k is a vector. For example, [10, Theorem 4.2.] states that for any matrix K, vector k and vector z we have: $\mathbf{dist}(z, \{w : Kw \leq k\}) \leq \alpha \|(Kz - k)^+\|_2$ where $\alpha > 0$ is the minimum singular value across all nonsingular submatrices of K.

In Applegate et al. [2], the authors employ Hoffman bounds to show that

$$\left\| \begin{pmatrix} (c^\top x - b^\top y)^+ \\ Ax - b \\ (A^\top y - c)^+ \end{pmatrix} \right\|_2$$

is bounded below by a constant times the distance to optimality. Using Lemma 2, this establishes that ρ_r is sharp.

Corollary 1 is a Hoffman bound that explicitly handles both inequality and nonnegativity constraints. Similar types of Hoffman bounds exist in the literature (e.g., [21]) but we were unable to find a bound that could be readily adapted to our purpose. The proof of Corollary 1 is a blackbox reduction to standard Hoffman bounds [10, Theorem 4.2.]. In contrast, Applegate et al. [2] treats the nonnegativity constraints as generic inequality constraints. However, in this paper we take advantage of the fact that the nonnegativity constraints on x are never violated (due to PDHG performing a projection). Explicitly handling these nonnegativity constraints improves the quality of the Hoffman constant. In particular, the nonsingular submatrices considered in the calculation are smaller (because they do not contain the identity block that [2, Equation (20)] introduces). If we did employ the strategy of Applegate et al. [2] the remainder of the paper would remain essentially unchanged but the iteration bound in Theorem 2 would contain $m_2 + m_1$ instead of just m_1 because the nonsingular submatrix G could be much bigger. This alternative worst-case bound is inferior for $m_2 \gg m_1$.

To prove Corollary 1 we will find the following Lemma useful.

Lemma 3. *Define* $M_\lambda := \begin{pmatrix} M_{11} & M_{12} \\ \mathbf{0} & \lambda M_{22} \end{pmatrix}$ *where* M_{22} *is a square matrix, and assume* M_λ *is nonsingular for some* $\lambda \in (0, \infty)$. *Then* M_{11} *and* M_{22} *are nonsingular and*

$$\lim_{\lambda \to \infty} M_\lambda^{-1} = \begin{pmatrix} M_{11}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}.$$

Proof. Since M_λ is nonsingular and M_{22} is square we deduce that M_{11} is also square. Also, we have for all $u \neq \mathbf{0}$ that $\mathbf{0} \neq M_\lambda \begin{pmatrix} u \\ \mathbf{0} \end{pmatrix} = M_{11} u$ which implies that M_{11} is nonsingular. Similarly, for all $v \neq \mathbf{0}$ we have $\mathbf{0} \neq M_\lambda^\top \begin{pmatrix} \mathbf{0} \\ v \end{pmatrix} = \lambda M_{22}^\top v$ which implies that M_{22} is nonsingular. Next, the Schur complement of M_λ with respect to the λM_{22} block is $S := M_{11} - \mathbf{0} \frac{1}{\lambda} M_{11}^{-1} M_{12} = M_{11}$. Using the Schur complement [11] we have

$$M_\lambda^{-1} = \begin{pmatrix} M_{11}^{-1} & \frac{1}{\lambda} M_{11}^{-1} M_{12} M_{22}^{-1} \\ \mathbf{0} & \frac{1}{\lambda} M_{22}^{-1} \end{pmatrix}.$$

Taking $\lambda \to \infty$ yields the desired result. □

Corollary 1. *For any nonnegative integers* m, n *and* p *consider matrices* $D \in \mathbb{R}^{m \times n}$, $F \in \mathbb{R}^{p \times n}$ *and vectors* $d \in \mathbb{R}^m$, $f \in \mathbb{R}^p$. *Let* $U_S := \{u \in \mathbb{R}^n : u_i \geq 0, \forall i \in S\}$ *for some* $S \subseteq [n]$. *Define the polytope,* $P := \{u \in U_S : Du \leq d, Fu = f\}$. *Let* $\mathcal{G}$ *be the set of all nonsingular submatrices of the matrix* $\begin{pmatrix} D \\ F \end{pmatrix}$ *and let* $\alpha = \frac{1}{\max_{G \in \mathcal{G}} \|G^{-1}\|_2}$. *Then, for all* $u \in U_S$ *we have*

$$\alpha \, \mathbf{dist}(u, P) \leq \left\| \begin{pmatrix} (Du - d)^+ \\ Fu - f \end{pmatrix} \right\|_2.$$

Proof. Consider the system

$$\left[K_\lambda := \begin{pmatrix} D \\ F \\ -F \\ -\lambda E \end{pmatrix} \right] u \leq \begin{pmatrix} d \\ f \\ -f \\ \mathbf{0} \end{pmatrix} =: k$$

for $\lambda \in (0, \infty)$, where $-\lambda E u \leq \mathbf{0}$ corresponds to the constraint $u \in U_S$ with each row of E containing exactly one entry (i.e., a one). Let Q be a nonsingular submatrix of K_λ decomposed into submatrices of D, F, $-F$ and E which we call Q_D, Q_F, Q_{-F}, Q_E respectively such that

$$Q = \begin{pmatrix} Q_D \\ Q_F \\ Q_{-F} \\ -\lambda Q_E \end{pmatrix}.$$

As each row of E contains exactly one nonzero entry, each row of Q_E must contain at most one nonzero entry. Since Q is nonsingular it follows that $Q_E^\top v \neq \mathbf{0}$ for all $v \neq \mathbf{0}$. By choosing $v = e_i$ for each row i, we deduce each row of Q_E must contain exactly one nonzero entry. Therefore there exist matrices M_{11}, M_{12} and M_{22} such that M_{11} and M_{12} are submatrices of $\begin{pmatrix} D \\ F \\ -F \end{pmatrix}$, M_{22} is a nonsingular square submatrix of Q_E, and

$$\Pi Q = \begin{pmatrix} M_{11} & M_{12} \\ \mathbf{0} & \lambda M_{22} \end{pmatrix}$$

where Π is some permutation matrix; let $\mathcal{M}$ represent the set of all such matrices. Note that since Q is square and M_{22} is square, we must have M_{11} is square.

Applying [10, Theorem 4.2.] gives $\alpha_\lambda \, \mathbf{dist}(u, P) \leq \|(Ku - k)^+\|_2$ for all $u \in U_S$ with

$$\alpha_\lambda := \max_{M \in \mathcal{M}} \left\| \begin{pmatrix} M_{11} & M_{12} \\ \mathbf{0} & \lambda M_{22} \end{pmatrix}^{-1} \right\|_2.$$

It follows by Lemma 3 that

$$\begin{aligned} \lim_{\lambda \to \infty} \alpha_\lambda &= \lim_{\lambda \to \infty} \max_{M \in \mathcal{M}} \left\| \begin{pmatrix} M_{11} & M_{12} \\ \mathbf{0} & \lambda M_{22} \end{pmatrix}^{-1} \right\|_2 \\ &= \max_{M \in \mathcal{M}} \lim_{\lambda \to \infty} \left\| \begin{pmatrix} M_{11} & M_{12} \\ \mathbf{0} & \lambda M_{22} \end{pmatrix}^{-1} \right\|_2 \\ &= \max_{M \in \mathcal{M}} \|M_{11}^{-1}\|_2. \end{aligned}$$

Recall that M_{22} is square, and that Q and therefore ΠQ is nonsingular. Therefore by Lemma 3 we deduce M_{11} is nonsingular and consequently either the row $-F_{i,\cdot}$ or $F_{i,\cdot}$ appears in M_{11}. Moreover, negating rows of M_{11} does not effect its maximum singular value. Hence, there exists $G \in \mathcal{G}$ such that $\|G^{-1}\|_2 = \|M_{11}^{-1}\|_2$. □

4. Analysis of restarted PDHG on totally unimodular linear programs

This section analyzes the worst-case convergence rate of restarted PDHG on totally unimodular linear programs. For completeness, we first define what it means for a linear program to be totally unimodular.

Definition 2 *([12]).* A matrix A is totally unimodular if every square submatrix is unimodular (i.e., has determinant 0, 1, or −1). The linear program (2) is totally unimodular if A is totally unimodular and the entries of c and b are all integers.

The following standard result will be useful. We include the proof for completeness.

Lemma 4. *If $A \in \mathbb{R}^{m_1 \times m_2}$ is a totally unimodular matrix then (i) $[A\ e_i]$ for any $i \in [m_1]$, (ii) $A^\top$ and (iii) $\begin{pmatrix} A & 0 \\ 0 & A^\top \end{pmatrix}$ are totally unimodular matrices. Moreover, (iv) if A is a nonsingular then $(A^{-1})_{ij} \in \{-1, 0, 1\}$ for all $i \in [m_1]$, $j \in [m_2]$ and $\|A^{-1}\|_2 \le m_2 = m_1$.*

Proof. The proof of (ii) follows from the fact that the determinant is invariant to transpose. For (i) and (iii), we use the well-known formula for the determinant of block matrices (e.g., [23]),

$$\det \begin{pmatrix} P & Q \\ 0 & S \end{pmatrix} = \det(P)\det(S).$$

Thus (i) follows by setting $P = \begin{pmatrix} 1 \end{pmatrix}$, Q to be the ith row of A and S the remaining portion (also using that the absolute value of the determinant is preserved by row and column permutations). Next, (iii) follows by setting $Q = \mathbf{0}$, $P = A$ and $S = A^\top$.

To see (iv), note by Cramer's rule and Lemma 4.i, the inverse of a square nonsingular totally unimodular matrix has all entries equal to either −1, 0 or 1, and therefore $\|A^{-1}\|_2 \le \|A^{-1}\|_F \le m_2 = m_1$. □

The key insight of this paper is Lemma 5, which allows us to reason about matrices that are nonsingular and after removing one row totally unimodular. The proof uses Lemma 4 to decompose the matrix into the sum of a totally unimodular matrix and a rank one component. The Sherman-Morrison formula is then applied to analyze the inverse and its spectral norm.

Lemma 5. *Suppose the matrix $\begin{pmatrix} v^\top \\ V \end{pmatrix}$ is nonsingular where V is a totally unimodular matrix with n rows, and v is a vector of length $n+1$ with rational entries. Let M be a positive integer such that for each $i \in [n+1]$, there exists an integer k_i such that $v_i = k_i / M$. Then*

$$\left\| \begin{pmatrix} v^\top \\ V \end{pmatrix}^{-1} \right\|_2 \le n + 1 + M\left((n+1)^{1.5}\|v\|_2 + n + 1\right).$$

Proof. First observe that $\begin{pmatrix} v^\top \\ V \end{pmatrix} = \begin{pmatrix} e_j^\top \\ V \end{pmatrix} + e_1 (v - e_j)^\top$ where j is chosen such that $W := \begin{pmatrix} e_j^\top \\ V \end{pmatrix}$ is nonsingular. Such a j exists because the rows of V are linearly independent and V has n rows. Therefore, there must exist some e_j that is not in the span of the rows of V, making the dimension of the subspace spanned by the rows of W equal to $n+1$, which implies it is nonsingular.

Note that by Lemma 4, W is totally unimodular. By the Sherman-Morrison formula [24]:

$$\begin{pmatrix} v^\top \\ V \end{pmatrix}^{-1} = W^{-1} - \frac{W^{-1} e_1 (v - e_j)^\top W^{-1}}{1 + (v - e_j)^\top W^{-1} e_1}. \tag{3}$$

As $W^{-1}e_1$ is an integer vector, and $v_i = k_i / M$ where k_i is an integer and M is a positive integer then there exists some integer z such that $1 + (v - e_j)^\top W^{-1} e_1 = z / M$. Since $1 + (v - e_j)^\top W^{-1} e_1 \ne 0$ it follows that

$$\left| 1 + (v - e_j)^\top W^{-1} e_1 \right| \ge \frac{1}{M}. \tag{4}$$

Using Lemma 4.iv we have $\|W^{-1}\|_2 \le n + 1$,

$$\begin{aligned} \|W^{-1} e_1 v W^{-1}\|_2 &\le \|W^{-1} e_1\|_2 \|v W^{-1}\|_2 \\ &\le (n+1)^{0.5} \|W^{-1}\|_2 \|v\|_2 \\ &\le (n+1)^{0.5} (n+1) \|v\|_2, \end{aligned}$$

and $\|W^{-1} e_1 e_j^\top W^{-1}\|_2 \le \|W^{-1} e_1\|_2 \|W^{-1} e_j\|_2 \le n+1$. Therefore, by (3) and (4) we get

$$\left\| \begin{pmatrix} v^\top \\ V \end{pmatrix}^{-1} \right\|_2 \le n + 1 + M((n+1)^{1.5}\|v\|_2 + n + 1). \quad \square$$

Lemma 6 characterizes the sharpness constant of the normalized duality gap. The proof uses Lemma 2 and Lemma 5.

Lemma 6. *Let $R := \lceil 8 m_1^{1.5} H \rceil$. If (1) is a totally unimodular linear program with an optimal solution, then there exists $\alpha > 0$ such that $\alpha = \Omega\left(\frac{1}{m_1^{2.5} H}\right)$ and $\alpha\, \mathbf{dist}(z, Z^\star) \le \rho_r(z)$ for all $z \in W_R(\mathbf{0})$ and $r \in (0, R]$. Moreover, there exists $z^\star \in Z^\star$ such that $\|z^\star\|_2 \le R/4$.*

Proof. First, we get a bound on the norm of an optimal solution. As there is an optimal solution to the linear program there exists an optimal basic feasible solution [3, Chapter 2]. Let B be an optimal basis with corresponding optimal solutions $x^\star$ and $y^\star$. It follows that $\|x^\star\|_2 = \|x_B^\star\|_2 = \|A_B^{-1} b\|_2 \le \|A_B^{-1}\|_2 \|b\|_2 \le m_1 \|b\|_2 \le m_1^{1.5} H$ and $\|y^\star\|_2 = \|(A_B^{-1})^\top c_B\|_2 \le \|(A_B^{-1})^\top\|_2 \|c_B\|_2 \le m_1 \|c\|_2 \le m_1^{1.5} H$ where $\|A_B^{-1}\|_2 \le m_1$ by Lemma 4.iv. With $z^\star = (x^\star, y^\star)$ we conclude $\|z^\star\|_2 \le \|x^\star\|_2 + \|y^\star\|_2 \le 2 m_1^{1.5} H \le R/4$.

Consider a square nonsingular submatrix $\begin{pmatrix} v^\top \\ V \end{pmatrix}$ of the matrix

$$\begin{pmatrix} \frac{1}{R} c^\top & -\frac{1}{R} b^\top \\ A & 0 \\ 0 & A^\top \end{pmatrix}$$

where v is a subvector of $\begin{pmatrix} \frac{1}{R} c^\top & -\frac{1}{R} b^\top \end{pmatrix}$ and V is a submatrix of $\begin{pmatrix} A & 0 \\ 0 & A^\top \end{pmatrix}$. Note that V is totally unimodular by Lemma 4.

We now prove V contains at most $2m_1$ rows. Note A contains at most m_1 rows by definition and the submatrix of V corresponding to $A^\top$ contains at most m_1 columns. Consequently, the submatrix of V corresponding to $A^\top$ contains at most m_1 rows; otherwise, it would be row rank-deficient rendering $\begin{pmatrix} v^\top \\ V \end{pmatrix}$ row rank-deficient and contradicting our assumption that $\begin{pmatrix} v^\top \\ V \end{pmatrix}$ is nonsingular.

By Lemma 5 with $M = R = O(H m_1^{1.5})$, $n = 2m_1$ and using that $\|v\|_2 \le \frac{1}{R}(\|c\|_2 + \|b\|_2) \le 2 H m_1^{0.5} / R$ we get

$$\begin{aligned} \left\| \begin{pmatrix} v^\top \\ V \end{pmatrix}^{-1} \right\|_2 &\le n + 1 + M((n+1)^{1.5}\|v\|_2 + n + 1) \\ &= O\left(m_1^{2.5} H\right). \end{aligned}$$

This implies by Corollary 1 that

$$\left\| \begin{pmatrix} \frac{1}{R}(c^\top x - b^\top y)^+ \\ Ax - b \\ (A^\top y - c)^+ \end{pmatrix} \right\|_2 \ge \Omega\left(\frac{1}{m_1^{2.5} H}\right) \mathbf{dist}(z, Z^\star).$$

Combining this inequality with Lemma 2 shows $\alpha\, \mathbf{dist}(z, Z^\star) \le \rho_r(z)$. □

We are now ready to prove the main result, Theorem 2. Note that $\|A\|_2$ can be readily estimated by power iteration with high probability in $\tilde{O}(1)$ matrix-vector multiplications [14]. Therefore, it is possible to select a step size that meets the requirements of the theorem.

Theorem 2. *If* (1) *is a totally unimodular linear program with an optimal solution, then Algorithm* 2 *starting from* $z^{0,0} = \mathbf{0}$ *with* $\frac{1}{4\|A\|_2} \le \eta \le \frac{1}{2\|A\|_2}$ *and* $\tau^0 = 1$ *requires at most*

$$O\left(H m_1^{2.5} \sqrt{\mathbf{nnz}(A)} \log\left(\frac{m_2 H}{\epsilon}\right)\right)$$

matrix-vector multiplications to obtain a point satisfying $\mathbf{dist}(Z^\star, z^{n,0}) \le \epsilon$.

Proof. By Lemma 1, we have $z^{n,0} \in W_{\theta\,\mathbf{dist}(z^{0,0},Z^\star)}(\mathbf{0})$ for $\theta = 2\sqrt{\frac{1+\eta\|A\|_2}{1-\eta\|A\|_2}} \le 2\sqrt{2} \le 4$. By Lemma 6, $\rho_r(z)$ is $\Omega\left(m_1^{-2.5} H^{-1}\right)$-sharp for all $\|z\|_2 \le 4\,\mathbf{dist}(\mathbf{0}, Z^\star)$ and $r \le 4\,\mathbf{dist}(\mathbf{0}, Z^\star)$. Combining this with Theorem 1 (using $\tau^0 = 1$) gives $t^\star = \left\lceil \frac{2C(q+2)}{\alpha\beta} \right\rceil = O(m_1^{2.5} H \|A\|_2)$ and $\mathbf{dist}(z^{n,0}, Z^\star) \le \beta^n t^\star\, \mathbf{dist}(z^{0,0}, Z^\star)$. Hence, for $n \ge \log_{1/\beta}(t^\star/\epsilon)$ we have $\mathbf{dist}(z^{n,0}, Z^\star) \le \epsilon$ and the total number of iterations is $O\left(H m_1^{2.5} \|A\|_2 \log\left(\frac{m_1 H \|A\|_2}{\epsilon}\right)\right)$. From this bound, we obtain the result since $\|A\|_2 \le \|A\|_F \le \sqrt{\mathbf{nnz}(A)}$ and $\log\left(\frac{m_1 H \|A\|_2}{\epsilon}\right) \le \log\left(\frac{m_2^2 H}{\epsilon}\right) \le 2\log\left(\frac{m_2 H}{\epsilon}\right)$ because $\|A\|_2 \le \sqrt{\mathbf{nnz}(A)} \le \sqrt{m_1 m_2} \le m_2$ where the last inequality uses the assumption that $m_1 \le m_2$. $\square$

CRediT authorship contribution statement

Oliver Hinder: Writing – review & editing, Writing – original draft, Formal analysis, Conceptualization.

Acknowledgements

The author thanks Yinyu Ye and Javier Peña for useful feedback. The author was supported by a Google Research Scholar award, AFOSR grant #FA9550-23-1-0242 and by the NSF-BSF program, under NSF grant #2239527.

References

[1] D. Applegate, M. Díaz, O. Hinder, H. Lu, M. Lubin, B. O'Donoghue, W. Schudy, Practical large-scale linear programming using primal-dual hybrid gradient, Adv. Neural Inf. Process. Syst. 34 (2021) 20243–20257.
[2] D. Applegate, O. Hinder, H. Lu, M. Lubin, Faster first-order primal-dual methods for linear programming using restarts and sharpness, Math. Program. (2022) 1–52.
[3] D. Bertsimas, J.N. Tsitsiklis, Introduction to Linear Optimization, vol. 6, Athena Scientific, Belmont, MA, 1997.
[4] A. Chambolle, T. Pock, A first-order primal-dual algorithm for convex problems with applications to imaging, J. Math. Imaging Vis. 40 (2011) 120–145.
[5] L. Chen, R. Kyng, Y.P. Liu, R. Peng, M.P. Gutenberg, S. Sachdeva, Maximum flow and minimum-cost flow in almost-linear time, in: 2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS), IEEE, 2022, pp. 612–623.
[6] R. Cole, C. Hertrich, Y. Tao, L.A. Végh, A first order method for linear programming parameterized by circuit imbalance, arXiv:2311.01959, 2023.
[7] M. Conforti, G. Cornuéjols, G. Zambelli, et al., Integer Programming, vol. 271, Springer, Switzerland, 2014.
[8] G.B. Dantzig, Application of the simplex method to a transportation problem. Activity analysis and production and allocation, 1951.
[9] J. Douglas, H.H. Rachford, On the numerical solution of heat conduction problems in two and three space variables, Trans. Am. Math. Soc. 82 (1956) 421–439.
[10] O. Güler, A.J. Hoffman, U.G. Rothblum, Approximations to solutions to systems of linear inequalities, SIAM J. Matrix Anal. Appl. 16 (1995) 688–696.
[11] E.V. Haynsworth, On the Schur complement, Technical Report, Basel Univ. (Switzerland) Mathematics Inst., 1968.
[12] A.J. Hoffmann, J.B. Kruskal, Integral boundary points of convex polyhedra, Linear Inequal. Relat. Syst. (1956) 223–246.
[13] T. Kitahara, S. Mizuno, A bound for the number of different basic solutions generated by the simplex method, Math. Program. 137 (2013) 579–586.
[14] J. Kuczyński, H. Woźniakowski, Estimating the largest eigenvalue by the power and Lanczos algorithms with a random start, SIAM J. Matrix Anal. Appl. 13 (1992) 1094–1122.
[15] Y.T. Lee, A. Sidford, Path finding methods for linear programming: solving linear programs in $\tilde{O}(\sqrt{\text{rank}})$ iterations and faster algorithms for maximum flow, in: 2014 IEEE 55th Annual Symposium on Foundations of Computer Science, IEEE, 2014, pp. 424–433.
[16] T. Lin, S. Ma, Y. Ye, S. Zhang, An ADMM-based interior-point method for large-scale linear programming, Optim. Methods Softw. 36 (2021) 389–424.
[17] H. Lu, J. Yang, On the geometry and refined rate of primal-dual hybrid gradient for linear programming, arXiv:2307.03664, 2023.
[18] S. Mizuno, The simplex method using Tardos' basic algorithm is strongly polynomial for totally unimodular LP under nondegeneracy assumption, Optim. Methods Softw. 31 (2016) 1298–1304.
[19] B. O'Donoghue, E. Chu, N. Parikh, S. Boyd, Conic optimization via operator splitting and homogeneous self-dual embedding, J. Optim. Theory Appl. 169 (2016) 1042–1068.
[20] J.B. Orlin, A polynomial time primal network simplex algorithm for minimum cost flows, Math. Program. 78 (1997) 109–129.
[21] J. Pena, J.C. Vera, L.F. Zuluaga, New characterizations of Hoffman constants for systems of linear constraints, Math. Program. 187 (2021) 79–109.
[22] I. Post, Y. Ye, The simplex method is strongly polynomial for deterministic Markov decision processes, Math. Oper. Res. 40 (2015) 859–868.
[23] M. Taboga, Determinant of a block matrix, Lectures on matrix algebra, https://www.statlect.com/matrix-algebra/determinant-of-block-matrix, 2021.
[24] J. Sherman, W.J. Morrison, Adjustment of an inverse matrix corresponding to a change in one element of a given matrix, Ann. Math. Stat. 21 (1950) 124–127.
[25] Z. Xiong, R.M. Freund, Computational guarantees for restarted PDHG for LP based on limiting error ratios and LP sharpness, arXiv:2312.14774, 2023.
[26] Y. Ye, The simplex and policy-iteration methods are strongly polynomial for the Markov decision problem with a fixed discount rate, Math. Oper. Res. 36 (2011) 593–603.