Identifying an optimal basis in linear programming

Stephen A. Vavasis*   Yinyu Ye †

October 18, 1995

Abstract

We propose a sufficient condition that allows an optimal basis to be identified from a central path point in a linear programming problem. This condition can be applied when there is a "gap" in the sorted list of slack values. Unlike previously known conditions, this condition is valid for real-number data and does not involve the number of bits in the data.

*Department of Computer Science, Upson Hall, Cornell University, Ithaca, NY 14853. Email: vavasis@cs.cornell.edu. This work is supported in part by the National Science Foundation, the Air Force Office of Scientific Research, and the Office of Naval Research, through NSF grant DMS-8920550. Also supported in part by an NSF Presidential Young Investigator award with matching funds received from AT&T and Xerox Corp. Part of this work was done while the author was visiting Sandia National Laboratories, supported by the U.S. Department of Energy under contract DE-AC04-76DP00789.

†Department of Management Sciences, The University of Iowa, Iowa City, IA 52242. E-mail: bbuyin@vaxa.weeg.uiowa.edu. This author is supported in part by NSF Grant DDM-9207347. Part of this work was done while the author was on a sabbatical leave from the University of Iowa and visiting the Cornell Theory Center, Cornell University, Ithaca, NY 14853, supported in part by the Cornell Center for Applied Mathematics and by the Advanced Computing Research Institute, a unit of the Cornell Theory Center, which receives major funding from the National Science Foundation and IBM Corporation, with additional support from New York State and members of its Corporate Research Institute.

1 The central path

The problem we consider is linear programming written in dual format:

$$\begin{array}{ll} \text{minimize} & \mathbf{b}^T\mathbf{y} \\ \text{subject to} & A^T\mathbf{y} - \mathbf{s} = \mathbf{c}, \\ & \mathbf{s} \geq \mathbf{0}. \end{array} \tag{1}$$

Here, A is an $m \times n$ matrix assumed to have rank m, $\mathbf{b} \in \mathbb{R}^m$ and $\mathbf{c} \in \mathbb{R}^n$ are given vectors, and $\mathbf{y} \in \mathbb{R}^m$ is the unknown vector. The vector $\mathbf{s}$ denotes the *slack variables*.

Let us assume that the set defined by $A^T\mathbf{y} \geq \mathbf{c}$ is bounded and has an interior feasible point. Then the *central path* point $(\mathbf{s}(\mu), \mathbf{y}(\mu))$, for parameter $\mu > 0$ is defined [1, 2, 3, 8, 9, 12] to be the unique minimizer of

$$\begin{array}{ll} \text{minimize} & \mathbf{b}^T\mathbf{y} - \mu \sum_{i=1}^n \ln s_i \\ \text{subject to} & A^T\mathbf{y} - \mathbf{s} = \mathbf{c}, \\ & \mathbf{s} > \mathbf{0}. \end{array} \tag{2}$$

The solution to (2) is also the unique solution to the nonlinear equations

$$\begin{aligned} \mu A S^{-1}\mathbf{e} &= \mathbf{b}, \\ A^T\mathbf{y} - \mathbf{s} &= \mathbf{c} \end{aligned}$$

that satisfies $\mathbf{s} > \mathbf{0}$. Here and for the rest of the paper, S denotes $\mathrm{diag}(\mathbf{s})$ and $\mathbf{e}$ denotes the vector of all 1's.

As $\mu \to 0^+$, the central path point tends to an optimal solution of the original linear programming problem. In fact, it is known [8] that as $\mu \to 0^+$, the point $(\mathbf{s}(\mu), \mathbf{y}(\mu))$ tends to the analytic center $(\mathbf{s}^*, \mathbf{y}^*)$ of the optimal face of (1). Let I be the set of indices i such that $s_i^* = 0$. In the case that $(\mathbf{s}^*, \mathbf{y}^*)$ is not the only optimum, it is known that if $(\mathbf{s}', \mathbf{y}')$ is any other optimum, then $s_i' = 0$ for all $i \in I$. It is also known that $\mu S^{-1}\mathbf{e}$ converges to a vector $\mathbf{x}^*$ that is an optimal solution for the primal, i.e., $\mathbf{x}^*$ solves

$$\begin{array}{ll} \text{maximize} & \mathbf{c}^T\mathbf{x} \\ \text{subject to} & A\mathbf{x} = \mathbf{b}, \\ & \mathbf{x} \geq \mathbf{0}. \end{array} \tag{3}$$

Finally, it is known that $(\mathbf{x}^*, \mathbf{s}^*)$ are strictly complementary, i.e., the set of indices such that $x_i^* = 0$ is precisely $\tilde{I}$, the complement of I [4].

A path-following interior point method [11] generates iterates that approximately satisfy the central path equations. The question arises: under what circumstances have we identified the optimal solution, given a central path point? The traditional answer, which applies only to integer data, is that the optimal solution may be identified when $\mu < 2^{-O(L)}$ where 'L' denotes the number of bits necessary to write $A, \mathbf{b}, \mathbf{c}$ [6, 7]. In Section 3 we propose some conditions depending only on A for real number data under which the optimum solution may be identified. These conditions may be applied in the case that $\mathbf{s}$ contains some "small" entries and some "large" entries separated by a gap. The idea of separating the slacks of an interior point method into "small" and "large" entries has been used before [5, 15] but primarily for speeding up the computation of the interior point step rather than for obtaining new optimality conditions.

2 The parameter $\bar{\chi}_A$

Let A be an $m \times n$ matrix of rank m, and let $\mathcal{D}$ be the set of all positive definite $n \times n$ diagonal matrices. Define

$$\bar{\chi}_A = \sup\{\|A^T(ADA^T)^{-1}AD\| : D \in \mathcal{D}\}. \tag{4}$$

The norm used in this equation, and for the rest of the paper, is the 2-norm. We start with the following result, discovered independently by Stewart and Todd.

Theorem 1 ([13, 14]) *For any $m \times n$ matrix A of rank m, $\bar{\chi}_A$ is finite.*

This leads to the following two lemmas, which also appear in some of our other work. The first part of Lemma 1 is also due to [10, 13].

Lemma 1 *With A, $\mathcal{D}$ and $\bar{\chi}_A$ as above, let $\mathbf{z}$ lie in the nullspace of A, let $\mathbf{w}$ lie in the image of A^T, and let $D \in \mathcal{D}$ be chosen arbitrarily. Then*

$$\|\mathbf{w}\| \leq \bar{\chi}_A \cdot \|\mathbf{w} + D\mathbf{z}\|$$

and

$$\|\mathbf{z}\| \leq (\bar{\chi}_A + 1) \cdot \|\mathbf{z} + D\mathbf{w}\|.$$

PROOF. Since $\mathbf{w}$ is in the image of A^T, write $\mathbf{w} = A^T\mathbf{x}$ for some $\mathbf{x}$. For the first statement, observe that $AD^{-1}(\mathbf{w} + D\mathbf{z}) = AD^{-1}\mathbf{w} + A\mathbf{z} = AD^{-1}\mathbf{w} = AD^{-1}A^T\mathbf{x}$. Thus,

$$\mathbf{x} = (AD^{-1}A^T)^{-1}AD^{-1}(\mathbf{w} + D\mathbf{z})$$

and

$$\mathbf{w} = A^T\mathbf{x} = A^T(AD^{-1}A^T)^{-1}AD^{-1}(\mathbf{w} + D\mathbf{z})$$

By (4) we conclude that

$$\|\mathbf{w}\| \le \bar{\chi}_A \cdot \|\mathbf{w} + D\mathbf{z}\|.$$

For the second part, observe that $A(\mathbf{z} + D\mathbf{w}) = AD\mathbf{w} = ADA^T\mathbf{x}$, hence

$$\mathbf{x} = (ADA^T)^{-1}A(\mathbf{z} + D\mathbf{w}).$$

Thus,

$$D\mathbf{w} = DA^T\mathbf{x} = DA^T(ADA^T)^{-1}A(\mathbf{z} + D\mathbf{w}).$$

Notice that the matrix $DA^T(ADA^T)^{-1}A$ is the transpose of the matrix whose supremum we compute in (4). Since matrix 2-norms are preserved under transpose, we conclude that

$$\|D\mathbf{w}\| \le \bar{\chi}_A \cdot \|\mathbf{z} + D\mathbf{w}\|.$$

Thus, since $\mathbf{z} = \mathbf{z} + D\mathbf{w} - D\mathbf{w}$,

$$\|\mathbf{z}\| \le \|\mathbf{z} + D\mathbf{w}\| + \|D\mathbf{w}\| \le (\bar{\chi}_A + 1) \cdot \|\mathbf{z} + D\mathbf{w}\|.$$

∎

Lemma 2 *Let $A, \mathbf{w}, \mathbf{z}$ be as in the previous lemma. Let $K \subset \{1, \ldots, n\}$ index a set of m columns of A that are a basis of $\mathbb{R}^m$, let $\mathbf{w}_K$ denote the m-subvector of $\mathbf{w}$ indexed by K, and let $\mathbf{z}_{\tilde{K}}$ denote the $(n-m)$-subvector of $\mathbf{z}$ indexed by the complement of K. Then*

$$\|\mathbf{w}\| \le \bar{\chi}_A \|\mathbf{w}_K\|$$

and

$$\|\mathbf{z}\| \le (\bar{\chi}_A + 1)\|\mathbf{z}_{\tilde{K}}\|.$$

Proof. For the first part, since K is a basis, there is a null-vector $\mathbf{z}'$ of A that agrees componentwise with $\mathbf{w}$ in positions indexed by $\tilde{K}$ (and the entries of $\mathbf{z}'$ in K are then uniquely determined by the equation $A\mathbf{z}' = \mathbf{0}$). Let D be the diagonal matrix with 1's in positions indexed by $\tilde{K}$ and $\epsilon > 0$ (very small) in positions indexed by K. By the preceding lemma,

$$\|\mathbf{w}\| \le \bar{\chi}_A \|\mathbf{w} - D\mathbf{z}'\|.$$

However, by construction of D and $\mathbf{z}'$, vector $\mathbf{w} - D\mathbf{z}'$ is arbitrarily close to $\mathbf{w}_K$.

For the second statement, there is an image vector $\mathbf{w}'$ that agrees with $\mathbf{z}$ componentwise in positions indexed by K. Then repeat the preceding argument, and apply the second part of the Lemma 1. ∎

3 The main results

Let $(\mathbf{s}, \mathbf{y})$ be a point on the central path for parameter μ. (The results in this section can be generalized to handle approximate centering as well.) Reorder the equations so that

$$s_1 \le s_2 \le \cdots \le s_n.$$

Suppose that there exists an index j such that

$$s_j < s_{j+1}/g.$$

Here, g is a "gap size" that we define to be

$$g = (\bar{\chi}_A + 1)\sqrt{n}.$$

It is possible that there is more than one such j with this property; in this case, the results proved in this section hold for any such choice of j.

Let $J = \{1, \ldots, j\}$. We note that such a "gap" in the slacks must always occur for μ sufficiently small. This is because the s_n, which is equal to $\|A^T\mathbf{y} - \mathbf{c}\|_1$, cannot become arbitrarily small (unless the original problem (1) is completely degenerate and has a solution $A^T\mathbf{y} = \mathbf{c}$; this can be detected in a preliminary phase). On the other hand, some simple norm estimates show that s_1 must tend to zero proportionally to μ.

Let $(\mathbf{s}^*, \mathbf{y}^*)$ and $\mathbf{x}^*$ be the complementary optimal solutions defined in Section 1. As in that section, let I denote the set of indices i such that $s_i^* = 0$. We have the following result.

Theorem 2 *With I, J defined as above, suppose $\{\mathbf{a}_j : j \in J\}$ are linearly independent vectors in $\mathbb{R}^m$. Then $J \subset I$.*

PROOF. Let

$$\mathbf{w} = A^T(\mathbf{y} - \mathbf{y}^*) = \mathbf{s} - \mathbf{s}^*$$

and

$$\mathbf{z} = \mu S^{-1}\mathbf{e} - \mathbf{x}^*.$$

Notice that $\mathbf{w}$ is in the image of A^T, and $\mathbf{z}$ is the nullspace of A. This latter statement follows from the central path condition that $\mu AS^{-1}\mathbf{e} = \mathbf{b}$ and $A\mathbf{x}^* = \mathbf{b}$.

Let $\theta = s_j$, where j is the distinguished index identified above. We can make the following table of bounds for entries of $\mathbf{w}$ and $\mathbf{z}$; these bounds are easily derived from hypotheses above (letting $\tilde{I}$ denote the complement of I and $\tilde{J}$ the complement of J):

If i is in:	then w_i satisfies:	and z_i satisfies:
$I \cap J$	$0 < w_i \leq \theta$	(no bound)
$\tilde{I} \cap J$	$w_i \leq \theta$	$z_i \geq \mu/\theta$
$I \cap \tilde{J}$	$w_i > g\theta$	$z_i < \mu/(g\theta)$
$\tilde{I} \cap \tilde{J}$	(no bound)	$0 < z_i < \mu/(g\theta)$

Now, suppose that $I \not\subset J$, i.e., suppose there is an index i in $\tilde{I} \cap J$. This means $\|\mathbf{z}\| \geq \mu/\theta$. Consider a vector $\hat{\mathbf{w}}$ defined as follows. For $i \in I \cap J$, define $\hat{w}_i = 0$ if $z_i < 0$, and $\hat{w}_i = -1$ otherwise. For $i \in \tilde{I} \cap J$, define $\hat{w}_i = -1$. Now, since the columns of A indexed by J are linearly independent, we can extend $\hat{\mathbf{w}}$ in positions indexed in $\tilde{J}$ to a vector lying in the row span of A^T. (First, extend J to a basis K by filling in zeros, and then the extension of $\hat{\mathbf{w}}$ to $\tilde{K}$ is uniquely defined.) By Lemma 2, we know that $\|\hat{\mathbf{w}}\| \leq \bar{\chi}_A \|\hat{\mathbf{w}}_J\| \leq \bar{\chi}_A \sqrt{n}$.

Now consider $\mathbf{w} + \theta' \hat{\mathbf{w}}$, where θ' is slightly larger than θ (but still small enough so that $w_i > g\theta'$ for $i \in I \cap \tilde{J}$). Observe that for positions indexed by J, this vector has the opposite sign pattern as $\mathbf{z}$ by construction of $\hat{\mathbf{w}}$. In positions $I \cap \tilde{J}$, $\mathbf{w} + \theta'\hat{\mathbf{w}}$ is positive (because $w_i > \bar{\chi}_A \sqrt{n}\theta'$ but $|\theta' \hat{w}_i| \leq \bar{\chi}_A \sqrt{n}\theta'$). Finally, observe that $\mathbf{w} + \theta'\hat{\mathbf{w}}$ is in the column space of A^T. Let $\mathbf{w}'$ denote $\mathbf{w} + \theta'\hat{\mathbf{w}}$.

Now we pick a positive definite diagonal matrix D as follows. For indices in J, we pick d_{ii} so that $d_{ii} w_i' = -z_i$. For indices in $\tilde{I} \cap \tilde{J}$ we pick d_{ii} arbitrarily small. For indices in $I \cap \tilde{J}$ there are two subcases. If $z_i > 0$ (and hence has absolute value less than $\mu/(g\theta)$) we pick d_{ii} arbitrarily small. If $z_i < 0$ we pick d_{ii} so that $d_{ii}w'_i = -z_i$.

Now we observe that $\mathbf{z} + D\mathbf{w}'$ has small entries in every position; by construction, $\mathbf{z} + D\mathbf{w}'$ is zero in positions indexed by J, and is bounded by $\mu/(g\theta)$ in position indexed by $\tilde{J}$. Thus,

$$\|\mathbf{z} + D\mathbf{w}'\| < \sqrt{n}\mu/(g\theta).$$

But this contradicts the second part of Lemma 1 because, as mentioned above, $\|\mathbf{z}\| \geq \mu/\theta$ and $g = (\bar{\chi}_A + 1)\sqrt{n}$. ∎

The next theorem is a kind of "dual" case; the proof is also dual.

Theorem 3 *With J defined as above, suppose $\{\mathbf{a}_j : j \in J\}$ are vectors that span $\mathbb{R}^m$. Then $I \subset J$.*

Proof. Use the same notation and the same table as above. Suppose there is an index i in $I \cap \tilde{J}$. This means $\|\mathbf{w}\| > g\theta$. Consider a vector $\hat{\mathbf{z}}$ defined as follows. For $i \in I \cap \tilde{J}$, we define $\hat{z}_i = -1$. For $i \in \tilde{I} \cap \tilde{J}$ we define $\hat{z}_i = -1$ if $w_i > 0$ and $\hat{z}_i = 0$ otherwise. Now $\hat{\mathbf{z}}$ is defined in $\tilde{J}$. We can now define $\hat{\mathbf{z}}$ in positions indexed by J as well in such a way that $A\hat{\mathbf{z}} = \mathbf{0}$. This is done by letting K be a basis subset of J. Since $\hat{\mathbf{z}}$ is defined on $\tilde{J}$, we extend it to $\tilde{K}$ by filling in zeros. Then there is a unique way to fill in entries in K to obtain the equation $A\hat{\mathbf{z}} = \mathbf{0}$. By Lemma 2, $\|\hat{\mathbf{z}}\| \leq (\bar{\chi}_A + 1)\sqrt{n}$.

Now consider $\mathbf{z} + \mu\hat{\mathbf{z}}/(g\theta')$, which we will denote as $\mathbf{z}'$. Here, θ' is slightly larger than θ. Observe that $\mathbf{z}'$ is in the nullspace of A. Furthermore, in every position indexed by $\tilde{J}$, $\mathbf{z}'$ has the opposite sign as $\mathbf{w}$. In every position indexed by $\tilde{I} \cap J$, z'_i is positive, and w_i is either small (bounded by θ) or negative. Then, as in the previous proof, we can pick D so that $-D\mathbf{z}'$ agrees with $\mathbf{w}$ in every position indexed by $\tilde{J}$, and furthermore, so that

$$\|\mathbf{w} + D\mathbf{z}'\| \leq \theta\sqrt{n}.$$

But this contradicts the first part of Lemma 1, since $\|\mathbf{w}\| > g\theta$. ∎

Theorem 4 *With I, J as above, suppose J indexes a basic set (i.e., a linearly independent and spanning set) of columns of A. Then $I = J$, and (1) has a unique solution.*

PROOF. The previous two theorems imply that $I = J$. Since the constraints indexed by J are a basis, the system of equations $A_I^T \mathbf{y} = \mathbf{c}_I$ has a unique solution $\mathbf{y}^*$. This solution must therefore be the unique optimum. ∎

4 Remarks

It is tempting to turn the theorems in the last section into an algorithm. In particular, one could iterate the interior point method until a gap of size g occurs. Then, in the case that A_J is spanning, we can delete all constraints not indexed by J and continue. In the case that A_J is linearly independent, we can solve for constraints in J and again reduce the size of the problem.

The difficulty is with the case that A_J is neither spanning nor linearly independent. We have constructed examples of this type in which J seems to carry no information at all about I.

The solution to this difficulty is to continue iterating until another gap appears in the sorted slacks. This is essentially the idea of our much longer paper [16]. A key observation in that paper is that if the structure of the slacks is not changing, then the central path is close to linear and may be followed with a line search. Thus, by using a new "layered step" line-search algorithm, we bound the running time between appearance of successive gaps in terms of A only.

References

[1] D. A. Bayer and J. C. Lagarias. The nonlinear geometry of linear programming. I. Affine and projective scaling trajectories. *Transactions of the AMS*, 314:499–526, 1989.

[2] D. A. Bayer and J. C. Lagarias. The nonlinear geometry of linear programming. II. Legendre transform coordinates and central trajectories. *Transactions of the AMS*, 314:527–581, 1989.

[3] D. A. Bayer and J. C. Lagarias. The nonlinear geometry of linear programming. III. Projective Legendre transform coordinates and Hilbert geometry. *Transactions of the AMS*, 320:193–225, 1990.

[4] O. Güler and Y. Ye. Convergence behavior of interior point algorithms. *Mathematical Programming*, 60:215–228, 1993.

[5] J. A. Kaliski and Y. Ye. A short-cut potential reduction algorithm for linear programming. *Management Science*, 39:757–773, 1993.

[6] N. Karmarkar. A new polynomial-time algorithm for linear programming. *Combinatorica*, 4:373–395, 1984.

[7] L. G. Khachiyan. A polynomial algorithm in linear programming. *Dokl. Akad. Nauk SSSR*, 244:1093–1086, 1979. Translated in *Soviet Math. Dokl.* 20:191–194.

[8] L. McLinden. An analogue of Moreau's proximation theorem, with applications to the nonlinear complementarity problem. *Pacific Journal of Mathematics*, 88:101–161, 1980.

[9] N. Megiddo. Pathways to the optimal set in linear programming. In N. Megiddo, editor, *Progress in Mathematical Programming: Interior Point and Related Method*, pages 131–158. Springer, New York, 1989.

[10] D. P. O'Leary. On bounds for scaled projections and pseudoinverses. *Linear Algebra and its Applications*, 132:115–117, 1990.

[11] J. Renegar. A polynomial-time algorithm based on Newton's method for linear programming. *Mathematical Programming*, 40:59–94, 1988.

[12] G. Sonnevend. An analytical center for polyhedrons and new classes of global algorithms for linear (smooth, convex) programming. In *Lecture Notes in Control and Information Sciences 84*, pages 866–876. Springer–Verlag, New York, 1985.

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

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

[15] K. Tone. An active–set strategy in an interior point method for linear programming. *Mathematical Programming*, 59:345–360, 1993.

[16] S. A. Vavasis and Y. Ye. An accelerated interior point method whose running time depends only on A. Technical Report 93-1391, Department of Computer Science, Cornell University, Ithaca, NY, 1993.