Approximate Farkas Lemmas and Stopping Rules for Iterative Infeasible-Point Algorithms for Linear Programming

Michael J. Todd* Yinyu Ye †

November 1994, revised January 1996

*School of Operations Research and Industrial Engineering, Cornell University, Ithaca, New York 14853. E-mail: miketodd@cs.cornell.edu. Research supported in part by NSF, AFOSR and ONR through NSF Grant DMS-8920550. Some of this work was done while the author was on a sabbatical leave from Cornell University visiting the Department of Mathematics at the University of Washington. Partial support from the department and from NSF Grants CCR-9103804 and DMS-9303772 is gratefully acknowledged.

†Department of Management Sciences, The University of Iowa, Iowa City, IA 52242. E-mail: yyye@dollar.biz.uiowa.edu. This author is supported in part by NSF Grant DDM-9207347. Some 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.

Abstract

In exact arithmetic, the simplex method applied to a particular linear programming problem instance either shows that it is infeasible, shows that its dual is infeasible, or generates optimal solutions to both problems. Interior-point methods do not provide such clear-cut information. We provide general tools (extensions of the Farkas Lemma) for concluding that a problem or its dual is likely (in a certain well-defined sense) to be infeasible, and apply them to develop stopping rules for a homogeneous self-dual algorithm and for a generic infeasible-interior-point method for linear programming. These rules allow precise conclusions to be drawn about the linear programming problem and its dual: either near-optimal solutions are produced, or we obtain "certificates" that all optimal solutions, or all feasible solutions to the primal or dual, must have large norm. Our rules thus allow more definitive interpretation of the output of such an algorithm than previous termination criteria. We give bounds on the number of iterations required before these rules apply. Our tools may also be useful for other iterative methods for linear programming.

Key words: linear programming, Farkas lemma, infeasible-interior-point methods, stopping rules

Running Header: Approximate Farkas and Stopping Rules

1 Introduction

This paper is concerned with what can be concluded about a linear programming problem and its dual when an infinite iterative method for its solution is terminated after a finite number of iterations. Our interest is mainly in recent interior-point methods, originating from the projective algorithm of Karmarkar [8]. Our concern is with practical methods that can be applied to problems with data given by real numbers; thus we do not consider methods whose initialization requires artificial variables or artificial constraints with "big M" coefficients related to the length of the input.

If a variant of the simplex method (a finite algorithm) is run to completion, it provides relatively complete information about the linear programming problem and its dual, at least if we assume that exact arithmetic is used. If it does not generate optimal solutions to both problems, it shows why: it either demonstrates that the original problem is infeasible (and provides a short certificate of its infeasibility via the well-known Farkas lemma) or proves similarly that the dual problem is infeasible (and thus the primal problem either infeasible or unbounded). For a discussion of this property of the simplex method and of the Farkas lemma, see for instance Schrijver [18].

When initial strictly feasible points are given (for the primal or for the primal and dual) then a (primal or primal-dual) interior-point algorithm will generate a sequence of strictly feasible points. If the method is terminated, feasible points are at hand, and usually some measure of their proximity to optimality is also available. Moreover, bounds are known for several methods on the number of iterations required to attain a given level of accuracy in the objective value. For recent overviews of such methods, see Gonzaga [3], den Hertog [6], and Todd [20]. This is a reasonably satisfactory situation.

On the other hand, very often such initial feasible points are not provided or easily found, and then in practice a so-called infeasible-interior-point method is used. Such an algorithm generates a sequence of iterates that satisfy all nonnegativity constraints strictly, but may fail to satisfy equality constraints (general inequality constraints are converted to equations using nonnegative slack variables). When an algorithm of this type is terminated finitely, we usually have iterates that are in some sense approximately feasible and approximately optimal. But the original problem or its dual might be infeasible, and we would like the algorithm to provide some indication of this fact. Early infeasible-interior-point methods merely assumed that optimal solutions existed. More recently, indications of the infeasibility of the problem or its dual, often in the form that, if optimal solutions did exist, they would have to be of very large norm, have been derived for certain algorithms. But the information provided is not totally satisfactory. There is little or no indication of which of the primal and dual problems is infeasible. This paper is an attempt to obtain more useful information from the output of such an iterative method.

We are usually content with approximate answers when we look for optimal solutions. Thus the conclusion that we have nearly feasible solutions to the primal and dual problems with small total complementarity can be viewed as the availability of nearly optimal solutions to slightly perturbed problems; this is similar to backward error analysis in numerical analysis. Indeed, practical simplex implementations only provide this sort of information, since tolerances are used to compensate for rounding errors. We should ask for similar information when we stop with an indication of infeasibility; that is, we would like a proof that a slightly perturbed problem (or its dual) is infeasible. Since a proof of infeasibility for a system of linear inequalities is given by a solution to an alternative system as in the classical Farkas lemma, we are interested in approximate solutions to alternative systems and their interpretations for the original system. In Section 2, we state the linear programming problems we are concerned with and give two approximate Farkas lemmas that can be used to give indications of infeasibility or of feasibility. These results follow easily from more general results of Freund [2] and Renegar [17], but the form in which they are presented and the use we make of them seem to be new.

In the next two sections we consider a homogeneous self-dual algorithm and a generic infeasible-interior-point method for linear programming problems with no known strictly feasible starting points. Algorithms of this type include those proposed by Kojima, Megiddo, and Mizuno [10], Lustig and Lustig, Marsten, and Shanno [11, 12], Mizuno [13], Potra [15, 16], Ye, Todd, and Mizuno [22], and Zhang [23]. We use the results of Section 2 to provide stopping rules which allow us to make definitive conclusions about the problem and its dual, and we investigate the number of iterations required for some specific algorithms before one of a set of stopping rules applies. If we seek either approximate optimality or an indication of infeasibility of the form that all feasible solutions of the problem or its dual (if any) must have large norm, then our results give better bounds for the homogeneous self-dual algorithm than for other infeasible-interior-point methods; of course, this could be due to our analysis rather than differences in the methods.

The final section contains some concluding remarks.

2 Approximate Farkas Lemmas

We consider a linear programming problem given in the standard form

$$(P) \qquad \begin{array}{rrcl} \min_x & c^T x & & \\ & Ax & = & b \\ & x & \geq & 0, \end{array}$$

with its dual written with explicit slacks as

$$(D) \qquad \begin{array}{rrcccl} \max_{y,s} & b^T y & & & & \\ & A^T y & + & s & = & c \\ & & & s & \geq & 0. \end{array}$$

These problems are defined by the matrix $A \in \mathbb{R}^{m \times n}$ and the vectors $b \in \mathbb{R}^m$ and $c \in \mathbb{R}^n$. We collect these into the *data vector* $d := (A, b, c)$.

We use arbitrary norms (both denoted by $\|\cdot\|$) on $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$, and the corresponding dual norms (denoted by $\|\cdot\|^*$) on c and s in $\mathbb{R}^n$ and b and Ax in $\mathbb{R}^m$. We also use the corresponding operator norm $\|A\| := \max\{\|Ax\|^* : \|x\| \leq 1\}$ on A. While $\|A\|$ may be hard to compute, we can easily evaluate the norms of rank-one perturbations. Indeed, if $u \in \mathbb{R}^m$ and $v \in \mathbb{R}^n$, then $\|uv^T\| = \|u\|^* \|v\|^*$. We define $\|d\| := \max\{\|A\|, \|b\|^*, \|c\|^*\}$. For definiteness, we sometimes restrict ourselves to the case where $\|x\| := \|x\|_\infty$ (so that $\|c\|^* = \|c\|_1$ and similarly for s) and $\|y\| := \|y\|_1$ (so that $\|b\|^* = \|b\|_\infty$). Then $\|A\| = \max_i \sum_j |a_{ij}|$. We call this choice for norms the *standard case.*

We say d is *primal feasible (infeasible)* if the primal problem (P) defined by the data in d is feasible (infeasible), and similarly for dual feasibility or infeasibility.

In this section, we show that if (P) or (D) is "approximately infeasible" in the sense that we have an approximate solution to an alternative system as in the Farkas lemma, then any feasible solution must have large norm. Conversely, if (P) or (D) is "approximately feasible" in the sense that we have an approximate solution to their equations and inequalities, then any vector proving their infeasibility must be large. As a corollary of these lemmas, we obtain versions of results in the very general theory of Renegar [17] on perturbation for linear programming. Our results can be proved from Renegar's general theory, but we give a simple proof based on the gauge program duality of Freund [2] (other related results can be seen in Terlaky [19] and the references therein). Both of our results include the Farkas lemma.

Below, we use the usual convention that the minimum over an empty set is $+\infty$; we also find it useful to set $0 \cdot +\infty$ and $+\infty \cdot 0$ conventionally to 1.

Lemma 2.1 *(Approximate Farkas Lemma I) Let*

$$\begin{aligned}
\alpha_x &:= \min\{\|x\| : Ax = b,\ x \geq 0\}, \\
\alpha_y &:= \min\{\|y\| : A^T y + s = c,\ s \geq 0\}, \text{ and} \\
\alpha_s &:= \min\{\|s\|^* : A^T y + s = c,\ s \geq 0\}.
\end{aligned}$$

Let

$$\begin{aligned}
\beta_u &:= \min\{\|u\|^* : A^T y \leq u,\ b^T y = 1\}, \\
\beta_v &:= \min\{\|v\|^* : Ax = v,\ c^T x = -1,\ x \geq 0\}, \text{ and} \\
\beta_w &:= \min\{\|w\| : Ax = 0,\ c^T x = -1,\ x \geq -w\}.
\end{aligned}$$

Then $\alpha_x \beta_u = \alpha_y \beta_v = \alpha_s \beta_w = 1$.

Proof. We derive these relationships from the general gauge duality results of Freund [2]. If $\|\cdot\|$ and $\|\cdot\|^*$ are dual norms, Theorem 2 of [2] shows that

$$\alpha := \min\{\|z\| : Mz \geq d\}$$

and

$$\beta := \min\{\|t\|^* : M^T r = t,\ d^T r = 1,\ r \geq 0\}$$

satisfy $\alpha\beta = 1$. From this it is easy to derive the first two equations of the lemma (we use standard tricks, for example writing $Ax = b$ as $Ax \geq b,\ -Ax \geq -b$). For the last, it is perhaps most straightforward to write α_s as the optimal value of

$$\min\{\|s\|^* : Gs = Gc,\ s \geq 0\},$$

where the rows of G form a basis for the null space of A. Then $\alpha_s \beta'_w = 1$, where β'_w is the optimal value of

$$\min\{\|w\| : G^T r + q = w,\ c^T G^T r = 1,\ q \geq 0\}.$$

But then if we write $x := q - w = -G^T r$, we find exactly the problem defining β_w, since x lies in the range of G^T iff it lies in the null space of A. $\Box$

Remark. We derived our lemma from the gauge duality results of Freund [2]. In addition, Lemma 3.13 in Renegar [17] easily shows that $\alpha_x \beta_u \leq 1$, in fact in a much more general setting. This is the hard part of the result; to show the reverse inequality (which can be viewed as a weak duality result), note that if x and (y, u) are feasible in the problems defining α_x and β_u respectively, then

$$\|u\|^* \|x\| \geq u^T x \geq y^T A x = b^T y = 1.$$

Similar short proofs hold for $\alpha_y \beta_v \geq 1$ and $\alpha_s \beta_w \geq 1$. There are also close connections between the lemma above and results bounding the distance from a given point to a polyhedron, first due to Hoffman [7]. A recent paper giving explicit bounds closely related to our results is Güler, Hoffman, and Rothblum [4]; see also Dax [1].

This lemma has implications related to Renegar's results on perturbation of linear programming problems. Indeed, let us suppose first that β_u is small. Then, if (y, u) solves the corresponding problem, we see that

$$(A + \Delta A)^T y \leq 0,\ b^T y = 1,$$

where $\Delta A := -bu^T$, with $\|\Delta A\| = \|b\|^* \|u\|^* = \beta_u \|b\|^*$. Hence d is close to (indeed, within $\beta_u \|b\|^* \leq \beta_u \|d\|$ of) the primal infeasible data vector $(A + \Delta A, b, c)$. In the notation of [17], $\mathrm{dist}(d, \mathrm{Pri}\emptyset)$ is at most $\beta_u \|b\|^*$; here $\mathrm{Pri}\emptyset$ denotes the set of primal infeasible data vectors. Using the lemma, we deduce that, if $\mathrm{dist}(d, \mathrm{Pri}\emptyset)$ is positive, then there is a feasible solution to (P) with norm at most $\|b\|^* / \mathrm{dist}(d, \mathrm{Pri}\emptyset)$; this is part (1) of Theorem 1.1 of [17].

Next, suppose β_v is small, and let (x, v) solve the corresponding problem. Then we find that

$$(A + \Delta A)x = 0,\ c^T x = -1,\ x \geq 0,$$

where now $\Delta A := vc^T$, with $\|\Delta A\| = \|v\|^*\|c\|^* = \beta_v\|c\|^*$. Hence d is now within $\beta_v\|c\|^*$ of a dual infeasible data vector, so $\mathrm{dist}(d, \mathrm{Dual}\emptyset) \le \beta_v\|c\|^*$, where $\mathrm{Dual}\emptyset$ denotes the set of dual infeasible data vectors. The lemma then implies that, if $\mathrm{dist}(d, \mathrm{Dual}\emptyset)$ is positive, then there is a feasible solution to (D) with the norm of y at most $\|c\|^*/\mathrm{dist}(d, \mathrm{Dual}\emptyset)$; this is again an instance of the result quoted above.

Finally, suppose β_w is small, with (x, w) a solution to the corresponding problem. Assume that $\beta_w\|c\|^* < 1/2$, so that we may set $x' := (x + w)/(1 - c^T w) \ge 0$. We then find that

$$Ax' = v, \;\; c^T x' = -1, \;\; x' \ge 0,$$

where now $v := Aw/(1 - c^T w)$, with $\|v\|^* \le \|A\|\|w\|/(1 - \|c\|^*\|w\|) \le 2\beta_w\|A\|$. We next proceed as in the previous paragraph to find a nearby dual infeasible data vector; we obtain $\mathrm{dist}(d, \mathrm{Dual}\emptyset) \le 2\beta_w\|A\|\|c\|^*$. This implies that, as long as $\mathrm{dist}(d, \mathrm{Dual}\emptyset)$ is positive, there is a feasible solution to (D) with the norm of s at most $2\|A\|\|c\|^*/\mathrm{dist}(d, \mathrm{Dual}\emptyset)$. This last result is less satisfactory than the previous two because of the presence of both $\|A\|$ and $\|c\|^*$ in the bound. This seems to be due to the way in which the vectors s and w enter into their respective problems. If we scale the problems initially so that $\|A\| = 1$ (if A is zero, the problems are trivial), $\|b\|^* = 1$ or $b = 0$, and $\|c\|^* = 1$ or $c = 0$, then more expected "relative perturbation" results are obtained.

Now we provide a complementary result.

Lemma 2.2 *(Approximate Farkas Lemma II) Let*

$$\begin{aligned} \gamma_x &:= \min\{\|x\| : Ax = 0, \; c^T x = -1, \; x \ge 0\}, \\ \gamma_y &:= \min\{\|y\| : A^T y + s = 0, \; b^T y = 1, \; s \ge 0\}, \; \textit{and} \\ \gamma_s &:= \min\{\|s\|^* : A^T y + s = 0, \; b^T y = 1, \; s \ge 0\}. \end{aligned}$$

Let

$$\begin{aligned} \delta_u &:= \min\{\|u\|^* : A^T y + s = c, \; s \ge -u\}, \\ \delta_v &:= \min\{\|v\|^* : Ax - b = v, \; x \ge 0\}, \; \textit{and} \\ \delta_w &:= \min\{\|w\| : Ax = b, \; x \ge -w\}. \end{aligned}$$

Then $\gamma_x\delta_u = \gamma_y\delta_v = \gamma_s\delta_w = 1$.

Proof. These results follow from Lemma 2.1. To show $\gamma_x \delta_u = 1$, we rewrite the equality constraints of the problem defining γ_x as

$$\begin{pmatrix} A \\ c^T \end{pmatrix} x = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$$

and apply the conclusion $\alpha_x \beta_u = 1$ of Lemma 2.1 to this system.

For the other parts, we write the constraints of the problems defining γ_y and γ_s as

$$\begin{pmatrix} A^T \\ b^T \\ -b^T \end{pmatrix} y + \begin{pmatrix} s \\ \sigma \\ \tau \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}, \quad \begin{pmatrix} s \\ \sigma \\ \tau \end{pmatrix} \geq 0,$$

and write the objective functions as

$$\|y\| \quad \text{and} \quad \|(s, \sigma, \tau)\|^* := \max\{\|s\|^*, |\sigma|, |\tau|\}$$

(noting that $\sigma = \tau = 0$ for all feasible solutions). Now apply the $\alpha_y \beta_v = 1$ and $\alpha_s \beta_w = 1$ parts of Lemma 2.1 to these systems. For the second, note that the norm dual to that above is

$$\|(w, \theta, \phi)\| = \|w\| + |\theta| + |\phi|.$$

Thus the appropriate value of β_v and β_w for these systems are

$$\begin{aligned} &\min\{\|v\|^* : Ax + \lambda b - \mu b = v, \lambda - \mu = -1,\ (x, \lambda, \mu) \geq 0\} \\ &= \min\{\|v\|^* : Ax - b = v,\ x \geq 0\} \end{aligned}$$

and

$$\begin{aligned} &\min\{\|w\| + |\theta| + |\phi| : Ax + \lambda b - \mu b = 0, \lambda - \mu = -1, x \geq -w, \lambda \geq -\theta, \mu \geq -\phi\} \\ &= \min\{\|w\| : Ax = b,\ x \geq -w\} \end{aligned}$$

(by choosing $\lambda = 0$, $\mu = 1$, $\theta = 0$, and $\phi = 0$), which equal δ_v and δ_w, respectively. $\square$

This result shows that, if (D) is "almost feasible" (δ_u is small), then if it fails to be exactly feasible, any proof of its infeasibility must have large norm (γ_x is large); moreover, the dual infeasible data vector d is close to the dual feasible data vector $(A, b, c + u)$. Similarly, if (P) is"almost feasible" (δ_v or

δ_w small), then, if it is not feasible, any certificate of its infeasibility must be large in norm (γ_y or γ_s large); moreover, d is close to the primal feasible data vector $(A, b+v, c)$ or $(A, b+Aw, c)$. This kind of argument has been used in the ellipsoid algorithm to relax the right-hand sides of a system of linear inequalities with integer data a small amount so that either both systems are feasible or both are infeasible; see Schrijver [18], p. 169.

3 Homogeneous Self-dual Algorithms

Our goal for an iterative algorithm to solve (P) and (D) is that it produce the following:

- either approximately feasible solutions x and (y, s) to (P) and (D) with $x^T s$ small; in this case we can conclude that we have almost optimal solutions to a linear programming pair that is close to the given one, and (using the second approximate Farkas lemma) that if (P) or (D) is infeasible, any proof of that fact via the Farkas lemma must have large norm;

- or a vector x or a vector y that approximately demonstrates the infeasibility of (D) or (P) respectively; in this case we know that a problem near to (P) or (D) is infeasible, and (using the first approximate Farkas lemma) that if the corresponding original problem is feasible, any feasible solution must have large norm.

Here we address the extent to which we can achieve this goal for an algorithm applied to the following homogeneous and self-dual (artificial) linear program (HLP) related (P) and (D) (this formulation was developed in [22]):

$$
\begin{array}{llrrrrll}
(HLP) & \min & & & & ((x^0)^T s^0 + 1)\theta & & \\
& \text{s.t.} & & Ax & -b\tau & +\bar{b}\theta & = 0, & (3.1)\\
& & -A^T y & & +c\tau & -\bar{c}\theta & \geq 0, & (3.2)\\
& & b^T y & -c^T x & & +\bar{z}\theta & \geq 0, & (3.3)\\
& & -\bar{b}^T y & +\bar{c}^T x & -\bar{z}\tau & & = -(x^0)^T s^0 - 1, & (3.4)\\
& & y \text{ free}, & x \geq 0, & \tau \geq 0, & \theta \text{ free}, & &
\end{array}
$$

where

$$
\bar{b} = b - Ax^0, \quad \bar{c} = c - A^T y^0 - s^0, \quad \bar{z} = c^T x^0 + 1 - b^T y^0. \tag{3.5}
$$

Here $\bar{b}$, $\bar{c}$, and $\bar{z}$ represent the "infeasibility" of the initial primal point, dual point, and primal-dual "gap", respectively. Let $s \in I\!R^n$ and $\kappa \in I\!R$ denote the surplus in equations (3.2) and (3.3). By combining these equations ($-(y^0)^T$ times the first, $-(x^0)^T$ times the second, and (-1) times the third and fourth), the last equality can be written as

$$(s^0)^T x + (x^0)^T s + \tau + \kappa - ((x^0)^T s^0 + 1)\theta = (x^0)^T s^0 + 1, \tag{3.6}$$

which serves as a normalizing constraint for (HLP). Let $\mathcal{F}_h$ denote the set of feasible solutions to (3.1)–(3.4), along with the corresponding s and κ, such that all nonnegative variables are strictly positive. Note that, by construction, $(y^0, x^0, \tau^0 := 1, \theta^0 := 1, s^0, \kappa^0 := 1) \in \mathcal{F}_h$, and we therefore choose this as our starting point. If $(y, x, \tau, \theta, s, \kappa) \in \mathcal{F}_h$, then (y, x, τ, θ) is easily seen to be feasible in the dual of (HLP), with slacks s and κ (since (HLP) is self-dual), and so the duality gap for this pair of solutions equals the total complementarity, or (if we divide by two)

$$((x^0)^T s^0 + 1)\theta = x^T s + \tau\kappa =: (n+1)\mu. \tag{3.7}$$

For simplicity, for the rest of this section we assume $x^0 = s^0 = e$ and $y^0 = 0$, so that $\theta = \mu$ from the equation above. (This assumption is equivalent to starting at a central point, where the products of corresponding components of x^0 and s^0 are all the same. We can first make y^0 zero by translating c to $c - A^T y^0$. Then, by scaling the columns of A and hence the entries of the initial points, we can make all components of x^0 and s^0 the same. Finally, we can scale b and c to make these components all one.)

Since (HLP) is easily seen to be self-dual and has a strictly feasible solution, its optimal value is zero. Thus the constraint (3.4) (or (3.6)) only serves as a convenient normalization; the problem (HLP) is essentially homogeneous. To stress their use of this formulation, we call algorithms for (P) and (D) that apply feasible-interior-point methods to (HLP) homogeneous self-dual algorithms.

We seek an (approximately) optimal solution to (HLP). Note that if $(y^*, x^*, \tau^*, 0, s^*, \kappa^*)$ is an (exactly) optimal solution with τ^* and κ^* not both zero, it is also an optimal solution to the dual of (HLP), and thus by complementary slackness exactly one of τ^* and κ^* is zero. If $\tau^* > 0$ and $\kappa^* = 0$, then it is easy to see that x^*/τ^* and $(y^*/\tau^*, s^*/\tau^*)$ are optimal solutions to (P) and (D). If $\tau^* = 0$ and $\kappa^* > 0$, then either $b^T y^* > 0$ or $c^T x^* < 0$, and this shows infeasibility of (P) or (D) by the Farkas lemma. We want to investigate what we can conclude from a near-optimal solution.

We assume a homogeneous self-dual algorithm generates a sequence $\{(y^k, x^k, \tau^k, \theta^k, s^k, \kappa^k)\}$ of iterates satisfying

$$\begin{aligned} (y^k, x^k, \tau^k, \theta^k, s^k, \kappa^k) &\in \mathcal{F}_h \quad \text{and} & (3.8) \\ \tau^k \kappa^k &\geq (1-\beta)\mu^k = (1-\beta)\theta^k & (3.9) \end{aligned}$$

for some fixed $0 < \beta < 1$. The first condition applies to any feasible-interior-point method applied to (HLP). The second requirement is a minimal centrality condition, and holds for instance for all path-following algorithms and some potential-reduction methods (see, e.g., [14]).

Given such an iterate $(y, x, \tau, \theta, s, \kappa)$, the corresponding approximately feasible primal and dual solutions are

$$\hat{x} := x/\tau; \quad (\hat{y}, \hat{s}) := (y, s)/\tau.$$

Note that we omit the superscript k for simplicity here and below.

We can now state some stopping criteria. Observe that, here and in the next section, we only state stopping criteria based on the iterates produced. In practice, it might be very useful to halt the algorithm based on the directions produced during an iteration. For example, suppose the direction for y is d_y. Then if $b^T d_y$ is safely positive while the largest component of $A^T d_y$ is nonpositive or very small, then by scaling d_y we show that β_u is zero or very small, an indication that (P) is likely to be infeasible. Similar conclusions are possible from the primal direction d_x. We do not state these stopping criteria precisely because we do not know how to estimate the number of iterations necessary until they apply; we can only do this for criteria based on the iterates themselves. However, if a direction would have caused an algorithm to stop, it is probable that the resulting iterate will do so too (because a large step can be taken).

In what follows, all ϵ's are (small) positive tolerances and all ρ's are (large) positive numbers.

- *Stopping Rule* 1*a*. Stop if

$$\|Ax - b\tau\|^* \leq \epsilon_p \tau, \ \|A^T y + s - c\tau\|^* \leq \epsilon_d \tau, \ \text{and} \ c^T x - b^T y \leq \epsilon'_o \tau.$$

In this case, the corresponding primal and dual solutions as above are feasible in perturbed primal and dual problems $(\tilde{P})$ and $(\tilde{D})$, and, with an error of at most ϵ'_o, the reversed weak duality inequality holds for the corresponding objective functions, *using the original coefficients*. (As mentioned at the beginning of this section, this implies that if the primal or dual problem is infeasible, any certificate of its infeasibility using the Farkas lemma would have to have large norm.)

However, we might wish the derived almost-feasible solutions to be ϵ''_o-optimal in these perturbed problems, which suggests

- *Stopping Rule* $1a'$. Stop if

$$\|Ax - b\tau\|^* \leq \epsilon_p \tau, \ \|A^T y + s - c\tau\|^* \leq \epsilon_d \tau, \ \text{and } x^T s \leq \epsilon''_o \tau^2.$$

It is not clear which of these two rules is more appropriate for a general termination criterion. It is easy to see that stopping rule $1a$ holds if

$$\frac{\theta}{\tau} \leq \epsilon_{pd} := \min\{\frac{\epsilon_p}{\|\bar{b}\|^*}, \frac{\epsilon_d}{\|\bar{c}\|^*}\} \quad \text{and} \quad \frac{\theta}{\tau} \leq \frac{\epsilon'_o}{\max(0, \bar{z})}. \tag{3.10}$$

Since $x^T s \leq (n+1)\mu = (n+1)\theta$, stopping rule $1a'$ applies if

$$\frac{\theta}{\tau} \leq \epsilon_{pd} \quad \text{and} \quad \frac{\theta}{\tau^2} \leq \frac{\epsilon''_o}{n+1}. \tag{3.11}$$

The next rule applies when τ is small; we can then conclude that there are no small optimal solutions to (P) and (D):

- *Stopping Rule* $2a$. For some $\rho^* > 0$, stop if

$$\tau \leq \frac{1-\beta}{1+\rho^*}.$$

We have the following result.

Proposition 3.1 *If stopping rule* $2a$ *applies, then there is no optimal solution pair* x^* *and* (y^*, s^*) *for (P) and (D) with* $\|x^*\|_1 + \|s^*\|_1 < \rho^*$.

Proof. We follow the proof technique of Güler and Ye [5]. Assume that such a pair exists. Then

$$\bar{y} := \alpha y^*, \quad \bar{x} := \alpha x^*, \quad \bar{\tau} := \alpha, \quad \bar{\theta} := 0, \quad \bar{s} := \alpha s^*, \quad \bar{\kappa} := 0,$$

where

$$\alpha = \frac{n+1}{e^T x^* + e^T s^* + 1} > 0,$$

is an optimal solution to (HLP). Now we use

$$(x - \bar{x})^T (s - \bar{s}) + (\tau - \bar{\tau})(\kappa - \bar{\kappa}) = 0,$$

which follows by subtracting the constraints of (HLP) for $(\bar{y}, \cdots, \bar{\kappa})$ from those for $(y, \cdots, \kappa)$ and using the skew-symmetry of the constraint matrix. This can be rewritten as

$$x^T \bar{s} + s^T \bar{x} + \tau \bar{\kappa} + \kappa \bar{\tau} = (n+1)\mu.$$

which implies (using (3.9)) that

$$\tau \geq \frac{\tau\kappa}{(n+1)\mu}\bar{\tau} \geq \frac{1-\beta}{n+1}\bar{\tau} = \frac{1-\beta}{e^T x^* + e^T s^* + 1} > \frac{1-\beta}{1+\rho^*},$$

contradicting stopping rule $2a$. □

There is also a complementary result, although since we do not use it in what follows, we omit the similar proof.

Proposition 3.2 *Suppose there is a certificate of primal infeasibility, i.e.,* (y_*, s_*) *with* $A^T y_* + s_* = 0$, $b^T y_* = 1$, $s_* \geq 0$, *or a certificate of dual infeasibility, i.e.,* x_* *with* $A x_* = 0$, $c^T x_* = -1$, $x_* \geq 0$, *with*

$$\|x_*\|_1 < \rho_* \text{ or } \|s_*\|_1 < \rho_*.$$

Then, for all iterates,

$$\kappa > \frac{1-\beta}{1+\rho_*}.$$

□

The contrapositive of this result shows that, if we ever observe a small value of κ during the course of the algorithm, then there are no small certificates of infeasibility for (P) or (D) and hence, from Lemma 2.2, there are almost feasible solutions to both problems.

Suppose we use a feasible-interior-point algorithm to solve (HLP) that yields iterates satisfying (3.8)–(3.9) and guarantees $\theta \leq \epsilon$ in $O(\sqrt{n}\ln(1/\epsilon))$ iterations. Then we have

Theorem 3.1 *For such an algorithm, either stopping rule* $1a$ *or* $2a$ *will be activated within*

$$O\left(\sqrt{n}\ln[\max\{\frac{1+\rho^*}{\epsilon_{pd}}, \frac{\max(0,\bar{z})(1+\rho^*)}{\epsilon'_o}\}]\right)$$

iterations. Also, either stopping rule $1a'$ *or* $2a$ *will be activated within*

$$O\left(\sqrt{n}\ln[\max\{\frac{1+\rho^*}{\epsilon_{pd}}, \frac{n(1+\rho^*)^2}{\epsilon''_o}\}]\right)$$

iterations.

Proof. Since $\tau \geq \Omega(1/(1+\rho^*))$ while stopping rule $2a$ fails to hold, the result follows from (3.10) and (3.11). $\square$

Note that the dependence on ρ^* and $1/\epsilon_{pd}$, $1/\epsilon'_o$, and $1/\epsilon''_o$ is only logarithmic in these bounds. Thus choosing, say, $\rho^* = 10^6$ does not drastically change the estimate of the number of iterations necessary.

We remark that stopping rule $2a$ only gives an indication that any *optimal* solutions of (P) and (D) must be large, rather than a suggestion of infeasibility. We would like to add rules that indicate that any *feasible* solution must be large.

- *Stopping Rule* $3a_p$. Stop if

$$b^T y \geq \|\tau c - \theta\bar{c}\|^* \bar{\rho}_p.$$

- *Stopping Rule* $3a_d$. Stop if

$$c^T x \leq -\|\tau b - \theta\bar{b}\|^* \bar{\rho}_d.$$

Proposition 3.3 *If the condition in stopping rule* $3a_p$ *holds, then any feasible solution to (P) has norm at least* $\bar{\rho}_p$*; if that in stopping rule* $3a_d$ *holds, then any feasible solution to (D) has* $\|y\|$ *at least* $\bar{\rho}_d$*.*

Proof. For the first part, set $\bar{y} := y/b^T y$. Since $A^T y + s = \tau c - \theta\bar{c}$, we see that $\bar{y}$ is feasible in the problem defining β_u in Lemma 2.1 with $u := (\tau c - \theta\bar{c})/b^T y$, which implies $\beta_u \leq 1/\bar{\rho}_p$ and hence $\alpha_x \geq \bar{\rho}_p$.

For the second part, note that $c^T x$ is negative and so $\bar{x} := x/(-c^T x)$ is feasible in the problem defining β_v in Lemma 2.1 with $v = (\tau b - \theta\bar{b})/(-c^T x)$, which implies $\beta_v \leq 1/\bar{\rho}_d$ and hence $\alpha_y \geq \bar{\rho}_d$. □

We want to bound the number of iterations until we get near-optimal solutions or an indication of likely infeasibility of the primal or dual problem. Let

$$\bar{\rho} := \max\{\bar{\rho}_p\|c\|^*, \bar{\rho}_p\|\bar{c}\|^*, \bar{\rho}_d\|b\|^*, \bar{\rho}_d\|\bar{b}\|^*, \bar{z}\} \tag{3.12}$$

and

$$\bar{\epsilon} := \min\{\frac{2}{3}, \epsilon_{pd}, \frac{\epsilon'_o}{\max(0, \bar{z})}\}. \tag{3.13}$$

Lemma 3.1 *If*

$$\theta \leq \frac{(1-\beta)\bar{\epsilon}^2}{4\bar{\rho}}, \tag{3.14}$$

then one of stopping rules $1a$, $3a_p$, *and* $3a_d$ *holds.*

Proof. Suppose stopping rule $1a$ does not apply. Then we know that $\theta/\tau > \bar{\epsilon}$. Thus, using (3.9), we have $\kappa \geq (1-\beta)(\theta/\tau) > (1-\beta)\bar{\epsilon}$. Now use the bound on θ to get

$$\begin{aligned}
\frac{\kappa}{\theta} &> \frac{4\bar{\rho}}{\bar{\epsilon}} \geq [1 + 2(1 + \frac{1}{\bar{\epsilon}})]\bar{\rho} \\
&\geq \bar{z} + 2(1 + \frac{\tau}{\theta})\bar{\rho} \\
&\geq \bar{z} + \frac{1}{\theta}[(\tau\|c\|^* + \theta\|\bar{c}\|^*)\bar{\rho}_p + (\tau\|b\|^* + \theta\|\bar{b}\|^*)\bar{\rho}_d].
\end{aligned}$$

But note that $b^T y - c^T x = \kappa - \bar{z}\theta$, so the inequality above shows that one of stopping rules $3a_p$ and $3a_d$ holds. □

Suppose we use a feasible-interior-point algorithm to solve (HLP) that yields iterates satisfying (3.8)–(3.9) and guarantees $\theta \leq \epsilon$ in $O(\sqrt{n}\ln(1/\epsilon))$ iterations. Then an immediate consequence of the lemma above is

Theorem 3.2 *For such an algorithm, one of stopping rules* $1a$, $3a_p$, *and* $3a_d$ *will be activated within* $O(\sqrt{n}\ln(\bar{\rho}/\bar{\epsilon}))$ *iterations.*

□

Note that again the dependence on $\bar{\rho}$ and $1/\bar{\epsilon}$ is only logarithmic.

Unfortunately, we see no way to bound the number of iterations until one of stopping rules $1a'$, $3a_p$, and $3a_d$ holds, if we apply the algorithm directly. The reason is that, according to (3.11), stopping rule $1a'$ might not hold even though θ/τ is very small, as long as θ/τ^2 is still too large. However, this seeming drawback is not critical. As we mentioned above, it is not clear which of rules $1a$ and $1a'$ is more appropriate in a particular situation. Secondly, even if we want to satisfy the condition in rule $1a'$, we can terminate the algorithm applied to (HLP) when θ satisfies (3.14) (in a number of iterations as given above), setting ϵ'_o very large. In fact, let us modify the step size in the last iteration so that θ satisfies (3.14) with equality. Then, if one of rules $3a_p$ and $3a_d$ applies, we stop. Otherwise, $\theta/\tau \leq \bar{\epsilon}$ and rule $1a$ applies, so we have nearly feasible primal and dual solutions. We can then replace b and c by their perturbed values so that $(x/\tau, s/\tau)$ is feasible, and continue with a standard primal-dual *feasible*-interior-point method. (This technique is similar to that in Section 5 of [10].) Since we have

$$\frac{x^T s}{\tau^2} \leq \frac{(n+1)\theta}{\tau^2} \leq \frac{(n+1)\bar{\epsilon}}{\tau} \leq \frac{(n+1)\bar{\epsilon}^2}{\theta} = \frac{4(n+1)\bar{\rho}}{1-\beta},$$

we need at most $O(\sqrt{n}\ln(n\bar{\rho}/\epsilon''_o))$ iterations to satisfy the criterion in stopping rule $1a'$. Thus, for this hybrid method, one of stopping rules $1a'$, $3a_p$, and $3a_d$ will be activated within $O(\sqrt{n}\ln(\bar{\rho}/\hat{\epsilon}))$ iterations in total, where

$$\hat{\epsilon} := \min\{\frac{2}{3}, \epsilon_{pd}, \frac{\epsilon''_o}{n}\}.$$

Again, the dependence on $\bar{\rho}$ and $1/\hat{\epsilon}$ is logarithmic.

4 Infeasible-interior-point Algorithms

Despite their attractive theoretical properties, homogeneous self-dual methods have not been used much in practice (but see [21]). One possible reason is that the formulation (HLP) ties together the primal and dual iterates (through the linking variables τ and θ), thus restricting flexibility in taking separate step sizes for the primal and dual. In this section, we consider the more widely-used infeasible-interior-point method originally proposed by Lustig [11] and shown to be related to taking Newton steps to reach a point on the central path by Lustig, Marsten, and Shanno [12]. Kojima, Megiddo, and Mizuno [10] proved global convergence for a method of this type, and polynomial-time variants were analyzed by Zhang [23], Mizuno [13], and Potra [15]; see also [9, 16, 24].

We provide a number of stopping rules for a very general generic infeasible-interior-point method, and discuss the conclusions that can be reached using the results of Section 2 when these rules apply. Then, for more specific methods, we give bounds on the number of iterations necessary until one of these rules holds.

We suppose the algorithm generates a sequence $\{x^k\}$ of primal iterates and $\{y^k, s^k\}$ of dual iterates satisfying

$$\begin{array}{rlrl} x^0 &= \rho_0 e, & s^0 &= \rho_0 e; \\ x^k &> 0, & s^k &> 0; \\ Ax^k - b &= \theta_p^k (Ax^0 - b), & A^T y^k + s^k - c &= \theta_d^k (A^T y^0 + s^0 - c); \\ (x^k)^T s^k &\geq \gamma_p \theta_p^k (x^0)^T s^0, & (x^k)^T s^k &\geq \gamma_d \theta_d^k (x^0)^T s^0. \end{array} \tag{4.1}$$

for some $\theta_p^k, \theta_d^k \in [0,1]$ and some positive constants ρ_0, γ_p, and γ_d. For example, the algorithms in [13] and [23] satisfy these conditions with $\theta_p^k = \theta_d^k$ for all k (because they take equal step sizes in the primal and dual problems) and $\gamma_p = \gamma_d = 1$. The algorithm of [15] furthermore has equality in the last two inequalities in (4.1) for all k. The algorithms strive to drive to zero the total complementarity $(x^k)^T s^k$, but this may not be possible if (P) or (D) is infeasible.

These conditions are not too restrictive. The requirement on the initial iterates is equivalent to starting at a central point, where the products of corresponding components of x^0 and s^0 are all the same. In this case, as in the last section, by scaling the columns of A and hence the entries of the initial points, they can be brought into the form above. The second condition is natural for an interior-point method, and the third follows if the iterates are generated by a line search using directions derived from some Newton system to achieve feasibility as well as centering and/or reduced complementarity. Finally, the last conditions seem to be necessary to derive convergence results; the total complementarity should not approach zero faster than primal or dual infeasibility.

Below we state some stopping rules and their consequences. As in Section 3, we omit the iteration superscript k in stating these rules.

- *Stopping Rule* 1*b*. Stop if

$$\|Ax - b\|^* \le \epsilon_p, \ \|A^T y + s - c\|^* \le \epsilon_d, \text{ and } x^T s \le \epsilon_o.$$

In this case, clearly the current iterates are ϵ_o-optimal in perturbed problems $(\tilde{P})$ and $(\tilde{D})$, where b is perturbed by a vector of norm at most ϵ_p and c by one of norm at most ϵ_d.

The second rule and the result and proof below are extensions of those of Mizuno [13]. They are also closely related to lemmas proved by Kojima [9].

- *Stopping Rule* 2*b*. For some $\rho^* \ge \rho_0$, stop if

$$\theta_p e^T s + \theta_d e^T x \ge \frac{\rho_0 + \max\{\gamma_p^{-1}, \gamma_d^{-1}\}\rho^*}{\rho_0^2} x^T s.$$

Proposition 4.1 *If stopping rule* 2*b applies, then there is no optimal solution pair* x^* *and* (y^*, s^*) *for (P) and (D) with* $\|x^*\| < \rho^*$ *and* $\|s^*\|^* < \rho^*$.

Proof. Assume that such a pair exists. Using (4.1), we find that

$$A(\theta_p x^0 + (1-\theta_p)x^*) = Ax \text{ and } A^T(\theta_d y^0 + (1-\theta_d)y^*) + (\theta_d s^0 + (1-\theta_d)s^*) = A^T y + s,$$

so that

$$[(\theta_p x^0 + (1-\theta_p)x^*) - x]^T[(\theta_d s^0 + (1-\theta_d)s^*) - s] = 0, \text{ or}$$

$$(\theta_p x^0 + (1-\theta_p)x^*)^T s + (\theta_d s^0 + (1-\theta_d)s^*)^T x = (\theta_p x^0 + (1-\theta_p)x^*)^T(\theta_d s^0 + (1-\theta_d)s^*) + x^T s.$$

Hence, using $x^0 = s^0 = \rho_0 e$ repeatedly, we obtain

$$\begin{array}{rcl}
\rho_0(\theta_p e^T s + \theta_d e^T x) & \le & (\theta_p x^0 + (1-\theta_p)x^*)^T s + (\theta_d s^0 + (1-\theta_d)s^*)^T x \\
& = & (\theta_p x^0 + (1-\theta_p)x^*)^T(\theta_d s^0 + (1-\theta_d)s^*) + x^T s \\
& < & n \max\{\theta_p, \theta_d\}\rho_0 \rho^* + x^T s.
\end{array}$$

But, using (4.1) again, $\theta_p \le \gamma_p^{-1} x^T s/(\rho_0^2 n)$ and similarly $\theta_d \le \gamma_d^{-1} x^T s/(\rho_0^2 n)$, so this last inequality gives a contradiction to the stopping rule. $\Box$

This stopping rule and result do not lead to any indication of infeasibility, only that any *optimal* solutions will have to be large. We would like rules indicating likely infeasibility of either the primal or dual problems.

For these, we let

$$\hat{x} := A^T(AA^T)^{-1}b, \quad \hat{y} := (AA^T)^{-1}Ac, \quad \text{and } \hat{c} := c - A^T\hat{y}.$$

- *Stopping Rule* $3b_p$.
  Let $\tilde{y} := y - [(1-\theta_d)\hat{y} + \theta_d y^0]$. Then, for some $\bar{\rho}_p > 0$, stop if
  $$b^T\tilde{y} \geq \|(1-\theta_d)\hat{c} + \theta_d s^0\|^* \bar{\rho}_p.$$

- *Stopping Rule* $3b_d$.
  Let $\tilde{x} := x - [(1-\theta_p)\hat{x} + \theta_p x^0]$. Then, for some $\bar{\rho}_d > 0$, stop if
  $$c^T\tilde{x} \leq -\|(1-\theta_p)\hat{x} + \theta_p x^0\|\bar{\rho}_d.$$

Proposition 4.2 *If the condition in stopping rule* $3b_p$ *holds, then any feasible solution to (P) has norm at least* $\bar{\rho}_p$*; if that in stopping rule* $3b_d$ *holds, then any feasible solution to (D) has* $\|s\|^*$ *at least* $\bar{\rho}_d$*.*

Proof. The proof is like that of Proposition 3.3. In the first case,

$$A^Ty + s - c = (1-\theta_d)(A^T\hat{y} + \hat{c} - c) + \theta_d(A^Ty^0 + s^0 - c)$$

by (4.1), so that

$$A^T\tilde{y} + s = (1-\theta_d)\hat{c} + \theta_d s^0.$$

Thus $\tilde{y}/b^T\tilde{y}$ satisfies the constraints of the problem defining β_u in Lemma 2.1 with $u := [(1-\theta_d)\hat{c} + \theta_d s^0]/b^T\tilde{y}$, which implies that $\beta_u \leq 1/\bar{\rho}_p$ and hence $\alpha_x \geq \bar{\rho}_p$.

In the second case, a similar argument using (4.1) shows that $A\tilde{x} = 0$, so that $\tilde{x}/(-c^T\tilde{x})$ is feasible in the constraints of the problem defining β_w in Lemma 2.1 with $w := [(1-\theta_p)\hat{x} + \theta_p x^0]/(-c^T\tilde{x})$. Hence $\beta_w \leq 1/\bar{\rho}_d$ and thus $\alpha_s \geq \bar{\rho}_d$. □

For our next result, we assume that for each k, θ_p^k and θ_d^k are equal (usually corresponding to choosing equal step sizes in the primal and dual), and use θ^k or just θ to denote their common value. We also suppose γ_p and γ_d are equal and write simply γ. We show that, if ρ^* is chosen sufficiently large, then whenever θ is not too small and the "optimality unlikely" stopping rule $2b$ applies, then so does one of the "feasibility unlikely" rules $3b_p$ and $3b_d$.

Lemma 4.1 *Suppose* $\theta \geq \underline{\theta} > 0$ *and*

$$\rho^* \geq \rho_0 + \frac{2}{n\underline{\theta}} \max\{\max(\|\hat{c}\|^*/\rho_0, \|e\|^*)\bar{\rho}_p, \max(\|\hat{x}\|/\rho_0, \|e\|)\bar{\rho}_d\}.$$

Then if stopping rule $2b$ *applies, so does either* $3b_p$ *or* $3b_d$*.*

Proof. We use repeatedly the fact that $Ax = (1-\theta)b + \theta A x^0$ and $A^T y + s = (1-\theta)c + \theta(A^T y^0 + s^0)$, from (4.1). Then

$$\begin{aligned}
x^T s &= (A^T y + s)^T x - (Ax)^T y \\
&= ((1-\theta)c + \theta(A^T y^0 + s^0))^T x - ((1-\theta)b + \theta A x^0)^T y \\
&= (1-\theta)(c^T x - b^T y) + \theta (y^0)^T A x + \theta (s^0)^T x - \theta (x^0)^T A^T y \\
&= (1-\theta)(c^T x - b^T y) + \theta (y^0)^T ((1-\theta)b + \theta A x^0) + \theta (s^0)^T x \\
&\quad - \theta (x^0)^T ((1-\theta)c + \theta(A^T y^0 + s^0) - s) \\
&= (1-\theta)(c^T x - b^T y) + \theta\rho_0(e^T x + e^T s) - \theta^2\rho_0^2 n + (1-\theta)(\theta b^T y^0 - \theta c^T x^0) \\
&= (1-\theta)(c^T x - b^T y) + \theta\rho_0(e^T x + e^T s) - \theta^2\rho_0^2 n \\
&\quad + (1-\theta)(\theta b^T y^0 + (1-\theta) b^T \hat{y} - (1-\theta) c^T \hat{x} - \theta c^T x^0) \\
&= (1-\theta)(c^T \tilde{x} - b^T \tilde{y}) + \theta\rho_0(e^T x + e^T s) - \theta^2\rho_0^2 n.
\end{aligned}$$

Hence we have

$$\begin{aligned}
(1-\theta)(c^T \tilde{x} - b^T \tilde{y}) &= x^T s + \theta^2\rho_0^2 n - \theta\rho_0(e^T x + e^T s) \\
&\leq (1 + \gamma^{-1} - (1 + \gamma^{-1}\frac{\rho^*}{\rho_0}))x^T s \\
&= -\gamma^{-1}(\frac{\rho^*}{\rho_0} - 1)x^T s \\
&\leq -(\rho^* - \rho_0)\theta\rho_0 n,
\end{aligned}$$

where we used the facts that $\theta^2\rho_0^2 n \leq \theta\rho_0^2 n = \theta(x^0)^T s^0 \leq \gamma^{-1} x^T s$ and that stopping rule $2b$ applies. Since the right-hand side is nonpositive, we may remove the factor $(1-\theta)$ on the left-hand side, and conclude that either

$$b^T \tilde{y} \geq (\rho^* - \rho_0)\theta\rho_0 n/2$$

or

$$c^T \tilde{x} \leq -(\rho^* - \rho_0)\theta\rho_0 n/2.$$

Using the formula for ρ^* it is then easy to see that either stopping rule $3b_p$ or $3b_d$ applies. $\square$

The remainder of this section obtains bounds on the number of iterations required by Potra's algorithm [15] or Mizuno's method [13] until one of the stopping criteria is activated. The analysis is technical and assumes familiarity with the proofs in [15, 13]. The reader may prefer to skip the details and just read the theorems below and the explanatory paragraphs immediately following them.

We first consider Potra's algorithm in [15]. While the steps of this algorithm are somewhat complicated (each step is composed of three substeps) it is simple in that the primal and dual infeasibilities and the total complementarity are all reduced at the same rate. Thus we write θ for the common value of θ_p and θ_d and choose $\gamma_p = \gamma_d = 1$. We also suppose for the rest of this section that the standard norms are used, so that $\|u\|$ denotes the ℓ_∞-norm and $\|u\|^*$ denotes the ℓ_1-norm for $u \in I\!R^n$, while $\|v\|$ denotes the ℓ_1-norm and $\|v\|^*$ denotes the ℓ_∞-norm for $v \in I\!R^m$. Let

$$\chi^* := \frac{\rho^*}{\rho_0} \text{ and } \epsilon := \min\{\frac{\epsilon_p}{\|Ax^0 - b\|^*}, \frac{\epsilon_d}{\|A^T y^0 + s^0 - c\|^*}, \frac{\epsilon_o}{(x^0)^T s^0}\}. \tag{4.2}$$

Then the algorithm will terminate according to stopping rule $1b$ iff $\theta \leq \epsilon$. Moreover, if stopping rule $2b$ has not been activated, we can assume that

$$\theta e^T s + \theta e^T x \leq \frac{\rho_0 + \rho^*}{\rho_0^2} x^T s, \tag{4.3}$$

and hence, using $x^T s = \theta (x^0)^T s^0 = \theta \rho_0^2 n$, that

$$\rho_0 (e^T s + e^T x) \leq (1 + \chi^*) \rho_0^2 n.$$

We can now use this inequality in place of the inequality in Lemma 3.1 of [15]. Suppose we choose our starting points by

$$y^0 = 0 \text{ or } y^0 = \hat{y},\ x^0 = s^0 = \rho_0 e,\ \rho_0 \geq \max\{\|\hat{x}\|_\infty, \|c\|_\infty, \|\hat{c}\|_\infty\}. \tag{4.4}$$

Then we obtain, following the proof of Lemma 3.4 in [15], that the steps are bounded as stated there but with

$$\eta := \nu := 2(1-\beta)^{-1/2}(1 + \chi^*) n.$$

Then Lemma 3.5 of [15] remains true with these new values of η and ν. We find that (using the notation there) $\sqrt{|\tau|}$ and $|\gamma|$ are $O(n^{1/2}\chi^*)$ and hence that γ_4, γ_5, γ_6 and γ_8 are all $O(n^2(\chi^*)^2)$, and finally that $1/\hat{\gamma} = O(n\chi^*)$. Thus θ is reduced at each iteration by a factor of the form $1 - 1/O(n\chi^*)$, so that $O(n\chi^* \ln(1/\epsilon))$ iterations are sufficient. To summarize, using our new estimates in the analysis of Potra we can establish the following:

Theorem 4.1 *Let Potra's algorithm be applied to (P) and (D) starting with the initial solutions given by (4.4), and let χ^* and ϵ be defined by (4.2). Then either stopping rule 1b or stopping rule 2b will be activated within $O(n\chi^* \ln(1/\epsilon))$ iterations. In particular, if we choose $\rho_0 = \Omega(\rho^*)$, then $\chi^* = O(1)$ and $O(n \ln(1/\epsilon))$ iterations suffice.*

□

It is important to realize that, if ρ^* is large and we choose $\rho_0 = \Omega(\rho^*)$, then ϵ is much smaller than ϵ_p, ϵ_d, and ϵ_o. However, the added complexity is only a *logarithmic* term in ρ^*. To be precise, if we assume that $\|A\|$, $\|b\|_\infty$, and $\|c\|_1$ are $O(1)$, then the final bound can be replaced by $O(n(\max\{\ln(1/\epsilon_p), \ln(1/\epsilon_d), \ln(1/\epsilon_o)\} + \ln \rho^*))$. This dependence is like that in Section 3.

Now suppose we do not want to terminate until we have either near-optimal solutions or an indication that (P) or (D) is infeasible. We use Lemma 4.1. Suppose that the problems are scaled so that the right-hand side of the inequality in (4.4) is $O(1)$. We also suppose that

$$\bar{\rho}_p = \bar{\rho}_d / n =: \bar{\rho}$$

(this is reasonable since $\bar{\rho}_p$ is related to the ℓ_∞-norm of x and $\bar{\rho}_d$ to the ℓ_1-norm of s). Then the condition in Lemma 4.1 becomes $\rho^* \geq \rho_0 + 2(n\epsilon)^{-1} n\bar{\rho}$, where we have taken $\bar{\theta} = \epsilon$, or

$$\chi^* \geq 1 + \frac{2}{\epsilon} \cdot \frac{\bar{\rho}}{\rho_0}.$$

Now we cannot achieve $\chi^* = O(1)$, since if we try to take $\rho_0 = \Omega(\bar{\rho})$, or even larger, ϵ must be correspondingly smaller to obtain a comparable accuracy in feasibility and optimality. The best we can do is to take $\rho_0 = \Theta(1)$, and then $\chi^* = O(\bar{\rho}/\epsilon)$. Thus, using Lemma 4.1 and the theorem above, we get

Theorem 4.2 *Suppose Potra's algorithm is applied to (P) and (D) starting with the initial solutions given by (4.4), with $\rho_0 = \Theta(1)$, and let ϵ be defined by (4.2). Then either stopping rule $1b$ or one of stopping rules $3b_p$ or $3b_d$ will be activated within $O(n\bar{\rho}\epsilon^{-1}\ln(1/\epsilon))$ iterations.*

□

Note here the very unwelcome appearance of the terms $\bar{\rho}$ and ϵ^{-1}, not logarithmically but as factors.

To conclude this section, we examine Mizuno's algorithms in [13]. He also chooses equal step sizes so that $\theta_p = \theta_d$, but only requires the total complementarity to decrease at most as fast as the infeasibilities, rather than at the same rate. Thus (4.1) holds with $\gamma_p = \gamma_d = 1$. Suppose the starting points are again chosen by (4.4). Then $\|x^1 - u^1\|_\infty$ and $\|z^1 - w^1\|_\infty$, in the notation of [13], can be bounded by $2\rho_0$ (our ρ_0 corresponds to $\gamma_0\rho$ in [13], while our ρ^* corresponds to ρ there). Using this in the analysis above Lemma 3.3 in [13], and assuming (4.3) which implies

$$e^T s + e^T x \le (\theta\rho_0)^{-1}(1+\chi^*)x^T s,$$

we obtain a bound of $O(\sqrt{n}\chi^*)\sqrt{x^T s}$ on the norms of the scaled steps ($\|D^{-1}\Delta x\|_2$ and $\|D\Delta z\|_2$ in [13]). This yields complexities of $O(n^2(\chi^*)^2\ln(1/\epsilon))$ and $O(n(\chi^*)^2\ln(1/\epsilon))$ steps for the first and second algorithms in [13]:

Theorem 4.3 *Suppose Mizuno's first (second) algorithm is applied to (P) and (D) starting with the initial solutions given by (4.4), and let χ^* and ϵ be defined by (4.2). Then either stopping rule $1b$ or stopping rule $2b$ will be activated within $O(n^2(\chi^*)^2\ln(1/\epsilon))$ ($O(n(\chi^*)^2\ln(1/\epsilon))$, respectively) iterations. In particular, if we choose $\rho_0 = \Omega(\rho^*)$, then $\chi^* = O(1)$ and $O(n^2\ln(1/\epsilon))$ ($O(n\ln(1/\epsilon))$, respectively) iterations suffice.*

□

If we want to stop with near-optimal solutions or an indication of infeasibility, we have some difficulties using Lemma 4.1, since θ might be very small while stopping rule $1b$ still does not apply, because $x^T s$ is not sufficiently small. Suppose we modify this rule, and consider

- *Stopping Rule* $1b'$. Stop if

$$\|Ax - b\|^* \le \epsilon_p \text{ and } \|A^T y + s - c\|^* \le \epsilon_d.$$

In this case, we have near-feasible solutions to (P) and (D). We can perturb b and c slightly and continue with a *feasible*-interior-point method from the current iterates until a near-optimal pair of solutions is obtained (compare with the end of the last section and also Section 5 of [10]). Now, if this rule has not been activated, we can assume that $\theta \geq \epsilon$, and thus take $\bar{\theta} = \epsilon$ in Lemma 4.1. Again we suppose that the problems are scaled so that the right-hand side of the inequality in (4.4) is $O(1)$, and that

$$\bar{\rho}_p = \bar{\rho}_d / n =: \bar{\rho}.$$

Then proceeding as above Theorem 4.2 we obtain

Theorem 4.4 *Suppose Mizuno's first (second) algorithm is applied to (P) and (D) starting with the initial solutions given by (4.4), with* $\rho_0 = \Theta(1)$*, and let* ϵ *be defined by (4.2). Then either stopping rule* $1b'$ *or one of stopping rules* $3b_p$ *or* $3b_d$ *will be activated within* $O(n^2\bar{\rho}^2\epsilon^{-2}\ln(1/\epsilon))$ $(O(n\bar{\rho}^2\epsilon^{-2}\ln(1/\epsilon))$*, respectively) iterations.*

$\square$

Once again we note the (even worse) polynomial, rather than logarithmic, appearance of $\bar{\rho}$ and ϵ^{-1} in the complexity bounds.

5 Concluding Remarks

In Section 2, we gave two general lemmas that make precise how certain near-feasible solutions to systems of inequalities indicate the likely infeasibility of alternative systems, extending the Farkas lemma. We used these lemmas at the beginning of Section 3 to suggest what an iterative method for linear programming problems should produce as its output.

In Sections 3 and 4, we described a number of stopping rules for interior-point algorithms. We showed that certain combinations of these stopping rules can lead to one of two conclusions. The first is that either a near-optimal pair of solutions to a linear programming problem and its dual is at hand, or alternatively that all optimal solutions have large norm. The second replaces the latter alternative with a certificate that all feasible solutions to the primal, or all feasible solutions to the dual, have large norm. Previous studies on interior-point methods have generally only been able to give conclusions of the first type. We have given bounds on the number of iterations required before one of these conclusions can be made. For infeasible-interior-point algorithms, this bound is polynomial, rather than logarithmic, in the size of the large norm in the case that achieving near-optimality or indicating infeasibility is the goal. In contrast, applying a feasible-interior-point algorithm to the homogeneous self-dual formulation provides such conclusions in a number of iterations that is logarithmic in this size.

References

[1] A. Dax. On minimum norm solutions. *Journal of Optimization Theory and Applications*, 76:183–193, 1993.

[2] R. M. Freund. Dual gauge programs, with applications to quadratic programming and the minimum-norm problem. *Mathematical Programming*, 38:47–67, 1987.

[3] C. C. Gonzaga. Path following methods for linear programming. *SIAM Review*, 34(2):167–227, 1992.

[4] O. Güler, A. J. Hoffman, and U. G. Rothblum. Approximations to solutions to systems of linear inequalities. *SIAM Journal on Matrix Analysis and Applications*, 16:688–696, 1995.

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

[6] D. den Hertog. *Interior Point Approach to Linear, Quadratic and Convex Programming: Algorithms and Complexity.* Kluwer Academic Publishers, Dordrecht, The Netherlands, 1994.

[7] A. J. Hoffman. On approximate solutions of systems of linear inequalities. *Journal of Research of the National Bureau of Standards*, 49:263–265, 1952.

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

[9] M. Kojima. Basic lemmas in polynomial-time infeasible-interior-point algorithms for linear programs. Research Reports on Information Sciences, Ser. B : Operations Research B–268, Dept. of Information Sciences, Tokyo Institute of Technology, Oh–Okayama, Meguro–ku, Tokyo 152, Japan, June 1993.

[10] M. Kojima, N. Megiddo, and S. Mizuno. A primal-dual infeasible-interior-point algorithm for linear programming. *Mathematical Programming*, 61:263–280, 1993.

[11] I. J. Lustig. Feasibility issues in a primal-dual interior point method for linear programming. *Mathematical Programming*, 49:145–162, 1990/91.

[12] I. J. Lustig, R. E. Marsten, and D. F. Shanno. Computational experience with a primal-dual interior point method for linear programming. *Linear Algebra and Its Applications*, 152:191–222, 1991.

[13] S. Mizuno. Polynomiality of infeasible-interior-point algorithms for linear programming. *Mathematical Programming*, 67:109–119, 1994.

[14] S. Mizuno, M. J. Todd, and Y. Ye. On adaptive step primal-dual interior-point algorithms for linear programming. *Mathematics of Operations Research*, 18:964–981, 1993.

[15] F. A. Potra. An infeasible interior-point predictor-corrector algorithm for linear programming. *SIAM Journal on Optimization*, 6(1), 1996.

[16] F. A. Potra. A quadratically convergent predictor-corrector method for solving linear programs from infeasible starting points. *Mathematical Programming*, 67:383–406, 1994.

[17] J. Renegar. Some perturbation theory for linear programming. *Mathematical Programming*, 65:73–82, 1994.

[18] A. Schrijver. *Theory of Linear and Integer Programming*, J. Wiley, Chichester, 1986.

[19] T. Terlaky. On l_p programming. *European Journal of Operations Research*, 22:70–100, 1985.

[20] M. J. Todd. Potential-reduction methods in mathematical programming. Technical Report 1050, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY 14853–3801, USA, January 1995.

[21] X. Xu, P.-F. Hung, and Y. Ye. A simplified homogeneous and self-dual linear programming algorithm and its implementation. Working Paper, College of Business Administration, The University of Iowa, Iowa City, IA, 1993, to appear in *Annals of Operations Research.*

[22] Y. Ye, M. J. Todd, and S. Mizuno, An $O(\sqrt{n}L)$-iteration homogeneous and self-dual linear programming algorithm. *Mathematics of Operations Research*, 19:52–67, 1994.

[23] Y. Zhang. On the convergence of a class of infeasible interior-point methods for the horizontal linear complementarity problem. *SIAM Journal on Optimization*, 4:208–227, 1994.

[24] Y. Zhang and D. Zhang. On polynomiality of the Mehrotra-type predictor-corrector interior-point algorithm. *Mathematical Programming*, 68:303–318, 1995.