Detecting Infeasibility in Infeasible-Interior-Point Methods for Optimization

M. J. Todd*

January 16, 2003

Abstract

We study interior-point methods for optimization problems in the case of infeasibility or unboundedness. While many such methods are designed to search for optimal solutions even when they do not exist, we show that they can be viewed as implicitly searching for well-defined optimal solutions to related problems whose optimal solutions give certificates of infeasibility for the original problem or its dual. Our main development is in the context of linear programming, but we also discuss extensions to more general convex programming problems.

*School of Operations Research and Industrial Engineering, Cornell University, Ithaca, New York 14853, USA (miketodd@cs.cornell.edu). This author was supported in part by NSF through grant DMS-0209457 and ONR through grant N00014-02-1-0057.

1 Introduction

The modern study of optimization began with G. B. Dantzig's formulation of the linear programming problem and his development of the simplex method in 1947. Over the more than five decades since then, the sizes of instances that could be handled grew from a few tens (in numbers of variables and of constraints) into the hundreds of thousands and even millions. During the same interval, many extensions were made, both to integer and combinatorial optimization and to nonlinear programming. Despite a variety of proposed alternatives, the simplex method remained the workhorse algorithm for linear programming, even after its non-polynomial nature in the worst case was revealed. In 1979, L. G. Khachiyan showed how the ellipsoid method of D. B. Yudin and A. S. Nemirovskii could be applied to yield a polynomial-time algorithm for linear programming, but it was not a practical method for large-scale problems. These developments are well described in Dantzig's and Schrijver's books [4, 25] and the edited collection [18] on optimization.

In 1985, Karmarkar [9] proposed a new polynomial-time method for linear programming which did lead to practically useful algorithms, and this led to a veritable industry of developing so-called interior-point methods for linear programming problems and certain extensions. One highlight was the introduction of the concept of self-concordant barrier functions and the resulting development of polynomial-time interior-point methods for a large class of convex nonlinear programming problems by Nesterov and Nemirovskii [19]. Efficient codes for linear programming were developed, but at the same time considerable improvements to the simplex method were made, so that now both approaches are viable for very large-scale instances arising in practice: see Bixby [3]. These advances are described for example in the books of Renegar and S. Wright [24, 33] and the survey articles of M. Wright, Todd, and Forsgren et al. [32, 26, 27, 5].

Despite their very nice theoretical properties, interior-point methods do not deal very gracefully with infeasible or unbounded instances. The simplex method (a finite, combinatorial algorithm) first determines whether a linear programming instance is feasible: if not, it produces a so-called certificate of infeasibility (see Section 2.4). Then it determines whether the instance is unbounded (in which case it generates a certificate of infeasibility for the dual problem, see Section 2), and if not, produces optimal solutions for the original problem (called the primal) and its dual. By contrast, most interior-point methods (infinite iterative algorithms) assume that the instance has an optimal solution: if not, they usually give iterates that diverge to infinity, from which certificates of infeasibility can often be obtained, but without much motivation or theory. Our goal is to have a interior-point method that, in the case that optimal solutions exist, will converge to such solutions; but if not, it should produce in the limit a certificate of infeasibility for the primal or dual problem. Moreover, the algorithm should achieve this goal without knowing the status of the original problem, and in just one "pass."

The aim of this paper is to show that infeasible-interior-point methods, while apparently striving only for optimal solutions, can be viewed in the infeasible or unbounded case as implicitly searching for certificates of infeasibility. Indeed, under suitable conditions, the "real" iterates produced by such an algorithm correspond to "shadow" iterates that are generated by another interior-point method applied to a related linear programming problem whose optimal solution gives the desired certificate of infeasibility. Hence in some sense these algorithms do achieve our goal. Our main development is in the context of linear programming, but we also discuss extensions to more general convex programming problems.

Section 2 discusses linear programming problems. We define the dual problem, give optimality conditions, describe a generic primal-dual feasible-interior-point method, and discuss certificates of infeasibility. In Section 3, we describe a very attractive theoretical approach (Ye, Todd, and Mizuno [35]) to handling infeasibility in interior-point methods. The original problem and its dual are embedded in a larger self-dual problem which always has a feasible solution. Moreover, suitable optimal solutions of the larger problem can be processed to yield either optimal solutions to the original problem and its dual or a certificate of infeasibility to one of these. This approach seems to satisfy all our goals, but it does have some practical disadvantages, which we discuss.

The heart of the paper is Section 4, where we treat so-called infeasible-interior-point methods. Our main results are Theorems 4.1–4.4, which relate an interior-point iteration in the "real" universe to one applied to a corresponding iterate in a "shadow" universe, where the goal is to obtain a certificate of infeasibility. Thus we see that, in the case of primal or dual infeasibility, the methods can be viewed not as pursuing a chimera (optimal solutions to the primal and dual problems, which do not exist), but as implicitly following a well-defined path to optimal solutions to related problems that yield infeasibility certificates. This helps to explain the observed practical success of such methods in detecting infeasibility.

In Section 5 we discuss convergence issues. While Section 4 provides a conceptual framework for understanding the behavior of infeasible-interior-point methods in case of infeasibility, we do not have rules for choosing the parameters involved in the algorithm (in particular, step sizes) in such a way as to guarantee good progress in both the original problem and its dual *and* a suitable related problem and its dual as appropriate. We obtain results on the iterates produced by such algorithms and a convergence result (Theorem 5.1) for the method of Kojima, Megiddo, and Mizuno [10], showing that it does produce approximate certificates of infeasibility under suitable conditions.

Section 6 studies a number of interior-point methods for more general convex conic programming problems, showing (Theorem 6.1) that the results of Section 4 remain true in these settings also. We make some concluding remarks in Section 7.

2 Linear Programming

For most of the paper, we confine ourselves to linear programming. Thus we consider the standard-form primal problem

$$(P) \quad \text{minimize} \quad c^T x, \qquad Ax = b, \quad x \geq 0,$$

of minimizing a linear function of the nonnegative variables x subject to linear equality constraints (any linear programming problem can be rewritten in this form). Closely related, and defined from the same data, is the dual problem

$$(D) \quad \text{maximize} \quad b^T y, \qquad A^T y + s = c, \quad s \geq 0.$$

Here A, an $m \times n$ matrix, $b \in \mathbf{R}^m$, and $c \in \mathbf{R}^n$ form the data; $x \in \mathbf{R}^n$ and $(y, s) \in \mathbf{R}^m \times \mathbf{R}^n$ are the variables of the problems. For simplicity, and without real loss of generality, we henceforth assume that A has full row rank.

2.1 Optimality conditions

If x is feasible in (P) and (y, s) in (D), then we obtain the weak duality inequality

$$c^T x - b^T y = (A^T y + s)^T x - (Ax)^T y = s^T x \geq 0, \tag{2.1}$$

so that the objective value corresponding to a feasible primal solution is at least as large as that corresponding to a feasible dual solution. It follows that, if we have feasible solutions with equal objective values, or equivalently with $s^T x = 0$, then these solutions are optimal in their respective problems. Since $s \geq 0$ and $x \geq 0$, $s^T x = 0$ in fact implies the seemingly stronger conditions that $s_j x_j = 0$ for all $j = 1, \ldots, n$, called complementary slackness. We therefore have the following optimality conditions:

$$(OC) \quad \begin{array}{rcccl} & A^T y & + & s & = c, \qquad s \geq 0, \\ Ax & & & & = b, \qquad x \geq 0, \\ & SXe & & & = 0, \end{array} \tag{2.2}$$

where S (resp., X) denotes the diagonal matrix of order n containing the components of s (resp., x) down its diagonal, and $e \in \mathbf{R}^n$ denotes the vector of ones. These conditions are in fact necessary as well as sufficient for optimality (strong duality: see [25]).

2.2 The central path

The optimality conditions above consist of $m + 2n$ mildly nonlinear equations in $m + 2n$ variables, along with extra inequalities. Hence Newton's method seems ideal to approximate a solution, but since this necessarily has zero components, the nonnegativities cause problems. Newton's method is better suited to the following perturbed system, called the *central path equations*:

$$(CPE_\nu) \quad \begin{array}{rcccl} & A^T y & + & s & = c, \qquad (s > 0) \\ Ax & & & & = b, \qquad (x > 0) \\ & SXe & & & = \nu e, \end{array} \tag{2.3}$$

for $\nu > 0$, because if it does have a positive solution, then we can keep the iterates positive by using line searches, i.e., by employing a damped Newton method. This is the basis of primal-dual path-following methods: a few (often just one) iterations of a damped Newton method are applied to (CPE_ν) for a given $\nu > 0$, and then ν is decreased and the process continued. See, e.g., Wright [33]. We will give more details of such a method in the next subsection.

For future reference, we record the changes necessary if (P) also includes free variables. Suppose the original problem and its dual are

$$(\hat{P}) \quad \text{minimize} \quad \begin{array}{ccccl} c^T x & + & d^T z, & & \\ Ax & + & Bz & = & b, \quad x \geq 0, \end{array}$$

and

$$(\hat{D}) \quad \text{maximize} \quad b^T y, \quad \begin{array}{rcrcll} A^T y & + & s & = & c, & s \geq 0, \\ B^T y & & & = & d. & \end{array}$$

Here B is an $m \times p$ matrix and $z \in \mathbf{R}^p$ a free primal variable. Assume that B has full column rank and $[A, B]$ full row rank, again without real loss of generality. The original problems are retrieved if B is empty.

The optimality conditions are then

$$(\widehat{OC}) \quad \begin{array}{ccccccccl} & & & A^T y & + & s & = & c, & s \geq 0, \\ & & & B^T y & & & = & d, & \\ Ax & + & Bz & & & & = & b, & x \geq 0, \\ & & & SXe & & & = & 0, & \end{array} \tag{2.4}$$

and the central path equations

$$(\widehat{CPE}_\nu) \quad \begin{array}{ccccccccl} & & & A^T y & + & s & = & c, & (s > 0) \\ & & & B^T y & & & = & d, & \\ Ax & + & Bz & & & & = & b, & (x > 0) \\ & & & SXe & & & = & \nu e. & \end{array} \tag{2.5}$$

If (2.5) has a solution, then $(\hat{P})$ and $(\hat{D})$ must have strictly feasible solutions, where the variables that are required to be nonnegative (x and s) are in fact positive. Further, the converse is true (see [33]):

Theorem 2.1 *Suppose $(\hat{P})$ and $(\hat{D})$ have strictly feasible solutions. Then, for every positive ν, there is a unique solution $(x(\nu), z(\nu), y(\nu), s(\nu))$ to (2.5). These solutions, for all $\nu > 0$, form a smooth path, and as ν approaches 0, $(x(\nu), z(\nu))$ and $(y(\nu), s(\nu))$ converge to optimal solutions to $(\hat{P})$ and $(\hat{D})$ respectively. Moreover, for every $\nu > 0$, $(x(\nu), z(\nu))$ is the unique solution to the primal barrier problem*

$$\min \quad c^T x + d^T z - \nu \sum_j \ln x_j, \qquad Ax + Bz = b, \quad x > 0,$$

and $(y(\nu), s(\nu))$ the unique solution to the dual barrier problem

$$\max \quad b^T y + \nu \sum_j \ln s_j, \qquad A^T y + s = c, \quad B^T y = d, \quad s > 0.$$

□

We call $\{(x(\nu), z(\nu)) : \nu > 0\}$ the *primal central path*, $\{(y(\nu), s(\nu)) : \nu > 0\}$ the *dual central path*, and $\{(x(\nu), z(\nu), y(\nu), s(\nu)) : \nu > 0\}$ the *primal-dual central path.*

2.3 A generic primal-dual feasible-interior-point method

Here we describe a simple interior-point method, leaving out the details of initialization and termination. We suppose we are solving $(\hat{P})$ and $(\hat{D})$, and that B has full column and $[A, B]$ full row rank. Let the current strictly feasible iterates be (x, z) for $(\hat{P})$ and (y, s) for $(\hat{D})$, and let μ denote $s^T x/n$. The next iterate is obtained by approximating the point on the central path corresponding to $\nu := \sigma\mu$ for some $\sigma \in [0,1]$ by taking a damped Newton step. Thus the search direction is found by linearizing the central path equations at the current point, so that $(\Delta x, \Delta z, \Delta y, \Delta s)$ satisfies the Newton system

$$(NS) \qquad \begin{array}{rcccccl} & & & A^T\Delta y & + & \Delta s & = & c - A^T y - s = 0, \\ & & & B^T\Delta y & & & = & d - B^T y = 0, \\ A\Delta x & + & B\Delta z & & & & = & b - Ax - Bz = 0, \\ S\Delta x & & & & + & X\Delta s & = & \nu e - SXe. \end{array} \tag{2.6}$$

Since X and S are positive definite diagonal matrices, our assumptions on A and B imply that this system has a unique solution. We then update our current iterate to

$$x_+ := x + \alpha_P \Delta x, \quad z_+ := z + \alpha_P \Delta z, \quad y_+ := y + \alpha_D \Delta y, \quad s_+ := s + \alpha_D \Delta s,$$

where $\alpha_P > 0$ and $\alpha_D > 0$ are chosen so that x_+ and s_+ are also positive. This concludes the iteration.

We wish to give as much flexibility to our algorithm as possible, so we will not describe rules for choosing the parameter σ and the step sizes α_P and α_D in detail. However, let us mention that, if the initial iterate is suitably close to the central path, then we can choose $\sigma := 1 - 0.1/\sqrt{n}$ and $\alpha_P = \alpha_D = 1$ and the next iterate will be strictly feasible and also suitably close to the central path. Thus these parameters can be chosen at every iteration, and this leads to a polynomial (but very slow) method; practical methods choose much smaller values for σ on most iterations. Finally, if $\alpha_P = \alpha_D$, then the duality gap $s_+^T x_+$ at the next iterate is smaller than the current one by the factor $1 - \alpha_P(1-\sigma)$, so we would like to choose σ small and the α's large. The choice of these parameters is discussed in [33].

2.4 Certificates of infeasibility

In the previous subsection, we assumed that feasible, and even strictly feasible, solutions existed, and were available to the algorithm. However, it is possible that no such feasible solutions exist (often because the problem was badly formulated), and we would like to know that this is the case. Here we revert to the original problems (P) and (D), or equivalently we assume that the matrix B is null.

It is clear that, if we have $(\bar{y}, \bar{s})$ with $A^T\bar{y} + \bar{s} = 0$, $\bar{s} \geq 0$, and $b^T\bar{y} > 0$, then (P) can have no feasible solution x, for if so we would have

$$0 \geq -\bar{s}^T x = (A^T\bar{y})^T x = (Ax)^T \bar{y} = b^T\bar{y} > 0,$$

a contradiction. The well-known Farkas Lemma [25] asserts that this condition is necessary as well as sufficient:

Theorem 2.2 *The problem (P) is infeasible iff there exists $(\bar{y}, \bar{s})$ with*

$$A^T\bar{y} + \bar{s} = 0, \quad \bar{s} \geq 0, \quad \textit{and} \quad b^T\bar{y} > 0. \tag{2.7}$$

□

We call such a $(\bar{y}, \bar{s})$ a *certificate of infeasibility for (P)*.

There is a similar result for dual infeasibility:

Theorem 2.3 *The problem* (D) *is infeasible iff there exists* $\tilde{x}$ *with*

$$A\tilde{x} = 0, \quad \tilde{x} \geq 0, \quad \text{and} \quad c^T\tilde{x} < 0. \tag{2.8}$$

□

We call such an $\tilde{x}$ a *certificate of infeasibility for* (D). It can be shown that, if (P) is feasible, the infeasibility of (D) is equivalent to (P) being unbounded, i.e., having feasible solutions of arbitrarily low objective function value: indeed, arbitrary positive multiples of a solution $\tilde{x}$ to (2.8) can be added to any feasible solution to (P). Similarly, if (D) is feasible, the infeasibility of (P) is equivalent to (D) being unbounded, i.e., having feasible solutions of arbitrarily high objective function value.

Below we are interested in cases where the inequalities of (2.7) or (2.8) hold strictly: in this case we shall say that (P) or (D) is *strictly infeasible.* It is not hard to show, using linear programming duality, that (P) is strictly infeasible iff it is infeasible and, for every $\tilde{b}$, the set $\{x : Ax = \tilde{b}, x \geq 0\}$ is either empty or bounded, and similarly for (D). Note that, if (P) is strictly infeasible, then (D) is strictly feasible (and unbounded), because we can add any large multiple of a strictly feasible solution to (2.7) to the point $(0, c)$; similarly, if (D) is strictly infeasible, then (P) is strictly feasible (and unbounded), because we can we can add any large multiple of a strictly feasible solution to (2.8) to a point x with $Ax = b$. Finally, we remark that, if (P) is infeasible but not strictly infeasible, then an arbitrarily small perturbation to A renders (P) strictly infeasible, and similarly for (D).

3 The Self-Dual Homogeneous Approach

As we mentioned in the introduction, our goal is a practical interior-point method which, when (P) and (D) are feasible, gives iterates approaching optimality for both problems; and when either is infeasible, yields a suitable certificate of infeasibility in the limit. Here we show how this can be done via a homogenization technique due to Ye, Todd, and Mizuno [35], based on work of Goldman and Tucker [6].

First consider the Goldman-Tucker system

$$\begin{array}{rcrcrcrcl} s & = & & - & A^T y & + & c\tau & \geq & 0, \\ & & Ax & & & - & b\tau & = & 0, \\ \kappa & = & -c^T x & + & b^T y & & & \geq & 0, \\ & & x \geq 0, & & y \text{ free} & & \tau \geq 0. & & \end{array} \tag{3.9}$$

This system is "self-dual" in that the coefficient matrix is skew-symmetric, and the inequality constraints correspond to nonnegative variables while the equality constraints correspond to unrestricted variables. The system is homogeneous, but we are interested in nontrivial solutions. Note that any solution (because of the skew-symmetry) has $s^T x + \kappa\tau = 0$, and the nonnegativity then implies that $s^T x = 0$ and $\kappa\tau = 0$. If τ is positive (and hence κ zero), then scaling (x, y, s) by τ gives feasible solutions to (P) and (D) satisfying $c^T x = b^T y$, and because of weak duality, these solutions are necessarily optimal. On the other hand, if κ is positive (and hence τ zero), then either $b^T y$ is positive, which with $A^T y + s = 0$, $s \geq 0$ implies that (P) is infeasible, or $c^T x$ is negative, which with $Ax = 0$, $x \geq 0$ implies that (P) is infeasible (or both). Thus this self-dual system attacks both the optimality and the infeasibility problem together. However, it is not clear how to apply an interior-point method directly to this system.

Hence consider the linear programming problem

$$
\begin{array}{lrrrrrrrrrrrl}
(HLP) & \min & & & & & & & & \bar{h}\theta & & \\
& s = & & - & A^T y & + & c\tau & - & & \bar{c}\theta & \geq & 0, \\
& & Ax & & & - & b\tau & + & & \bar{b}\theta & = & 0, \\
& \kappa = & -c^T x & + & b^T y & & & + & & \bar{g}\theta & \geq & 0, \\
& & \bar{c}^T x & - & \bar{b}^T y & - & \bar{g}\tau & & & & = & -\bar{h}, \\
& & x \geq 0, & & y \text{ free}, & & \tau \geq 0, & & \theta \text{ free}, & & &
\end{array}
$$

where

$$\bar{b} := b\tau^0 - Ax^0,\ \bar{c} := c\tau^0 - A^T y^0 - s^0,\ \bar{g} := c^T x^0 - b^T y^0 + \kappa^0,\ \bar{h} := (s^0)^T x^0 + \kappa^0 \tau^0,$$

for some initial $x^0 > 0$, y^0, $s^0 > 0$, $\tau^0 > 0$, and $\kappa^0 > 0$. Here we have added an extra artificial column to the Goldman-Tucker inequality system so that $(x^0, y^0, s^0, \tau^0, \theta^0, s^0, \kappa^0)$ is strictly feasible. To keep the skew symmetry, we also need to add an extra row. Finally, the objective function is to minimize the artificial variable θ, so as to obtain a feasible solution to (3.9).

Because of the skew symmetry, (HLP) is self-dual, i.e., equivalent to its dual, and this implies that its optimal value is attained and is zero. We can therefore apply a feasible-interior-point method to (HLP) to obtain in the limit a solution to (3.9). Further, it can be shown (see Güler and Ye [7]) that many path-following methods will converge to a strictly complementary solution, where either τ or κ is positive, and thus we can extract either optimal solutions to (P) and (D) or a certificate of infeasibility, as desired.

This technique seems to address all our concerns, since it unequivocally determines the status of the primal-dual pair of linear programming problems. However, it does have some disadvantages. First, it appears that (HLP) is of considerably higher dimension than (P), and thus that the linear system that must be solved at every iteration to obtain the search direction is of twice the dimension as that for (P). However, as long as we initialize the algorithm with corresponding solutions for (HLP) and its (equivalent) dual, we can use the self-duality to show that in fact the linear system that needs to be solved has only a few extra rows and columns compared to that for (P). Second, (HLP) links together the original primal and dual problems through the variables θ, τ, and κ, so equal step sizes must be taken in the primal and dual problems. This is definitely a drawback, since in many applications, one of the feasible regions is "fat," so that a step size of one can be taken without losing feasibility, while the other is "thin" and necessitates quite small steps. There are methods allowing different step sizes [30, 34], but they are more complicated. Thirdly, only in the limit is feasibility attained, while the method of the next section allows early termination with often feasible, but not optimal, solutions.

4 Infeasible-Interior-Point Methods

For the reasons just given, many codes take a simpler and more direct approach to the unavailability of initial strictly feasible solutions to (P) and (D). Lustig et al. [12, 13] proceed almost as in Section 2.3, taking a Newton step towards the (feasible) central path, but now from a point that may not be feasible for the primal or the dual. We call a triple (x, y, s) with x and s positive, but where x and/or (y, s) may not satisfy the linear equality constraints of (P) and (D), an *infeasible interior point.*

We describe this algorithm (the infeasible-interior-point (IIP) method) precisely in the next subsection. Because its aim is to find a point on the central path, it is far from clear how this method will behave when applied to a pair of problems where either the primal or the dual is infeasible. We would like it to produce a certificate of infeasibility, but there seems little reason why it should. However, in practice, the method is amazingly successful in producing certificates of infeasibility by just scaling the iterates generated, and we wish to understand why this is. In the following subsection, we suppose that (P) is strictly infeasible, and we show that the IIP method is in fact implicitly searching for a certificate of primal infeasibility by taking damped Newton steps. Then we outline the analysis for dual strictly infeasible problems, omitting details.

4.1 The primal-dual infeasible-interior-point method

The algorithm described here is almost identical to the generic feasible algorithm outlined in Section 2.3. The only changes are to account for the fact that the iterates are typically infeasible interior points. For future reference, we again assume we wish to solve the more general problems $(\hat{P})$ and $(\hat{D})$, for which an infeasible interior point is a quadruple (x, z, y, s) with x and s positive.

We start at such a point (x_0, z_0, y_0, s_0). (We use subscripts for both iteration indices and components, but the latter only rarely: no confusion should arise.) At some iteration, we have a (possibly) infeasible interior point $(x, z, y, s) := (x_k, z_k, y_k, s_k)$ and, as in the feasible algorithm, we attempt to find the point on the central path corresponding to $\nu := \sigma\mu$, where $\sigma \in [0, 1]$ and $\mu := s^T x/n$, by taking a damped Newton step. The search direction is determined from

$$(NS - IIP) \qquad \begin{array}{ccccccccc} & & & & A^T\Delta y & + & \Delta s & = & c - A^T y - s, \\ & & & & B^T\Delta y & & & = & d - B^T y, \\ A\Delta x & + & B\Delta z & & & & & = & b - Ax - Bz, \\ S\Delta x & & & & & + & X\Delta s & = & \nu e - SXe, \end{array} \tag{4.10}$$

whose only difference from the system (NS) is that the first three right-hand sides may be nonzero. (However, this does cause a considerable difference in the theoretical analysis, which is greatly simplified by the orthogonality of Δs and Δx in the feasible case.) Again, this system has a unique solution under our assumptions. We then update our current iterate to

$$x_+ := x + \alpha_P \Delta x, \quad z_+ := z + \alpha_P \Delta z, \quad y_+ := y + \alpha_D \Delta y, \quad s_+ := s + \alpha_D \Delta s,$$

where $\alpha_P > 0$ and $\alpha_D > 0$ are chosen so that x_+ and s_+ are also positive. This concludes the iteration. Note that, if it is possible to choose α_P equal to one, then (x_+, z_+) (and all subsequent primal iterates) will be feasible in $(\hat{P})$, and if α_D equals one, (y_+, s_+) (and all subsequent dual iterates) will be feasible in $(\hat{D})$.

As in the feasible case, there are many strategies for choosing the parameter σ and the step sizes α_P and α_D. Lustig et al. [12, 13] choose σ close to zero and α_P and α_D as a large multiple (say .9995) of the largest step to keep x and s positive respectively, except that steps larger than 1 are not chosen. Kojima, Megiddo, and Mizuno [10] choose a fixed $\sigma \in (0,1)$ and α_P and α_D to stay within a certain neighborhood of the central path, to keep the complementarity $s^T x$ bounded below by multiples of the primal and dual infeasibilities, and to decrease the complementarity by a suitable ratio. (More details are given in Section 5.2 below.) They are thus able to prove finite convergence, either to a point that is nearly feasible with small complementarity (and hence feasible and nearly optimal in nearby problems), or to a large enough iterate that one can deduce that there are no strictly feasible solutions to $(\hat{P})$ and $(\hat{D})$ in a large region.

Zhang [36], Mizuno [14], and Potra [23] provide extensions of Kojima et al.'s results, giving polynomial bounds to generate near-optimal solutions or guarantees that there are no optimal solutions in a large region.

These results are quite satisfactory when $(\hat{P})$ and $(\hat{D})$ are strictly feasible, but they are not as pleasant when one of these is infeasible – we would prefer to generate certificates of infeasibility, as in the method of the previous section. In the rest of this section, we show that, in the strictly infeasible case, there are "shadow iterates" that seem to approximately indicate infeasibility. Thus in the primal infeasible case, instead of thinking of $(NS-IIP)$ as giving Newton steps towards a nonexistent primal-dual central path, we can think of it as providing a step in the shadow iterates that is a damped Newton step towards a well-defined central path for another optimization problem, which yields a primal certificate of infeasibility. This interpretation explains in some sense the practical success of infeasible-interior-point methods in detecting infeasibility.

4.2 The primal strictly infeasible case

Let us suppose that (P) is strictly infeasible, so that there is a solution to

$$A^T \bar{y} + \bar{s} = 0, \quad \bar{s} > 0, \quad b^T \bar{y} = 1. \tag{4.11}$$

As we showed in Section 2.4, this implies that the dual problem (D) is strictly feasible, and indeed its feasible region is unbounded. When applied to such a primal-dual pair of problems, the IIP method usually generates a sequence of iterates where (y, s) becomes feasible after a certain iteration, and $b^T y$ tends to ∞. It is easy to see that, as the iterations progress, Ax always remains a convex combination of its original value Ax_0 and its "goal" b, but since the problem is infeasible, the weight on the first vector must remain positive. Let us therefore make the following

Assumption 4.1 *The current iterate (x, y, s) has (y, s) strictly feasible in (D) and $\beta := b^T y > 0$. In addition,*

$$Ax = \phi A x_0 + (1-\phi) b, \quad x > 0, \quad \phi > 0.$$

If $\beta = b^T y$ is large, then $(y, s)/\beta$ will be an approximate solution to the Farkas system above. This will be part of our "shadow iterate," but since our IIP method is primal-dual, we also want a primal and dual for our shadow iterate. We therefore turn the Farkas system into an optimization problem, using the initial solution (x_0, y_0, s_0). Let us therefore consider

$$\begin{array}{llrcrcl} (\bar{D}) & \max & (Ax_0)^T\bar{y} & & & & \\ & & A^T\bar{y} & + & \bar{s} & = & 0, \\ & & b^T\bar{y} & & & = & 1, \\ & & & & \bar{s} \geq 0. & & \end{array}$$

We call this $(\bar{D})$ since it is a homogeneous form of (D) with a normalizing constraint and a new objective function, and regard it as a dual problem of the form $(\hat{D})$. From our assumption that (P) is strictly infeasible, $(\bar{D})$ is strictly feasible. Its dual is

$$\begin{array}{llrcrcl} (\bar{P}) & \min & & & \bar{\zeta} & & \\ & & A\bar{x} & + & b\bar{\zeta} & = & Ax_0, \\ & & \bar{x} \geq 0. & & & & \end{array}$$

We will always use bars to indicate the variables of $(\bar{D})$ and $(\bar{P})$. Note that, from our assumption on the current iterate, $(x/\phi, -(1-\phi)/\phi)$ is a strictly feasible solution to $(\bar{P})$. Hence we make the

Definition 4.1 *The shadow iterate corresponding to* (x, y, s) *is given by*

$$(\bar{x}, \bar{\zeta}) := (\frac{x}{\phi}, -\frac{1-\phi}{\phi}), \quad (\bar{y}, \bar{s}) := (\frac{y}{\beta}, \frac{s}{\beta}).$$

(We note that the primal iterate x is infeasible, while the dual iterate (y, s) is feasible; these conditions are reversed in the shadow universe, where $(\bar{x}, \bar{\zeta})$ is feasible and $(\bar{y}, \bar{s})$ is typically infeasible in the first equation, while satisfying the second.)

Since ϕ and β are linear functions of x and (y, s) respectively, the transformations from the original iterates to the shadow iterates is a *projective* one. Projective transformations were used in Karmarkar's original interior-point algorithm [9], but have not been used much since, although they are implicit in the homogeneous approach and are used in Mizuno and Todd's analysis [15] of such methods.

We now wish to compare the results of applying one iteration of the IIP method from (x, y, s) for (P) and (D), and from $(\bar{x}, \bar{\zeta}, \bar{y}, \bar{s})$ for $(\bar{P})$ and $(\bar{D})$.

The idea is shown in the figure below. While the step from (x, y, s) to (x_+, y_+, s_+) is in some sense "following a nonexistent central path," the shadow iterates follow the central path for the strictly feasible pair $(\bar{P})$ and $(\bar{D})$. Indeed, the figure can be viewed as a "commutative diagram." Our main theorem below shows that the point $(\bar{x}_+, \bar{\zeta}_+, \bar{y}_+, \bar{s}_+)$ can be obtained either as the shadow iterate corresponding to the result of a damped Newton step for (P) and (D) from (x, y, s), or as the result of a damped Newton step for $(\bar{P})$ and $(\bar{D})$ from the shadow iterate corresponding to (x, y, s).

For a chosen value for $\sigma \in [0, 1]$, let $(\Delta x, \Delta y, \Delta s)$ be the search direction of the first of these, and let α_P and α_D be the chosen positive step sizes, with (x_+, y_+, s_+) being the next iterate. Then according to the algorithm in Section 4.1, we have

$$\begin{array}{rcrcrcl} & & A^T\Delta y & + & \Delta s & = & 0, \\ A\Delta x & & & & & = & b - Ax, \\ S\Delta x & & & + & X\Delta s & = & \sigma\mu e - SXe \end{array} \tag{4.12}$$

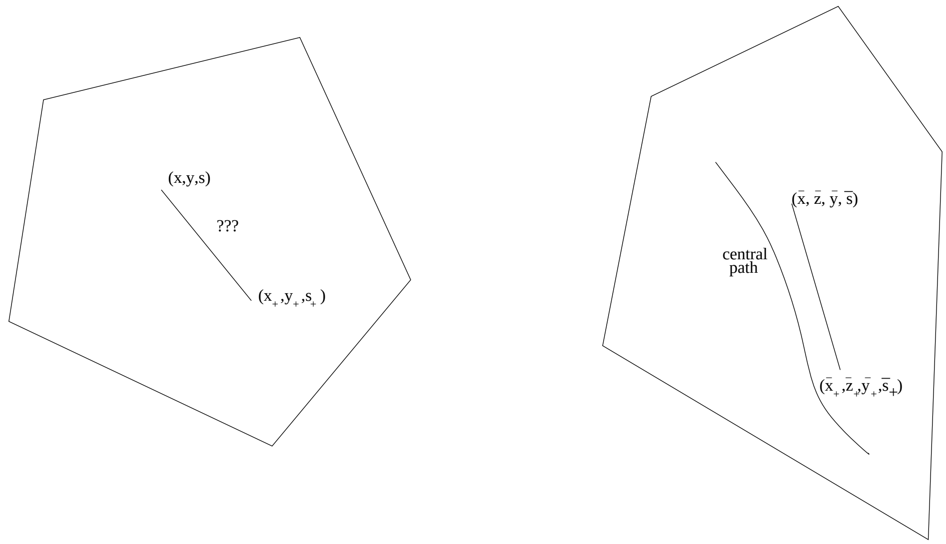

Figure 1. Comparing the real and shadow iterations: a "commutative diagram."

(note that B is empty and the dual iterate is feasible), where $\mu := s^T x/n$, and

$$x_+ := x + \alpha_P \Delta x, \quad y_+ := y + \alpha_D \Delta y, \quad s_+ := s + \alpha_D \Delta s.$$

The corresponding iteration for $(\bar{P})$ and $(\bar{D})$ also comes from Section 4.1, where now B is the single column b, but we postpone stating it until we have generated trial search directions from those above. Before doing so, we note the easily derived and well-known fact that $\Delta y = (AXS^{-1}A^T)^{-1}b - \sigma\mu(AXS^{-1}A^T)^{-1}AS^{-1}e$. Thus

$$\Delta\beta := b^T \Delta y = b^T (AXS^{-1}A^T)^{-1}b - \sigma\mu b^T (AXS^{-1}A^T)^{-1}AS^{-1}e,$$

and it follows (since infeasibility implies that b is nonzero) that $\Delta\beta$ is positive for small enough σ, depending on x and s. Henceforth, we make the

Assumption 4.2 *$\Delta\beta$ is positive.*

From Assumption 4.1, the definition of x_+, and (4.12), we find that

$$Ax_+ = \phi(Ax_0) + (1-\phi)b + \alpha_P(b - \phi(Ax_0) - (1-\phi)b) = \phi_+(Ax_0) + (1-\phi_+)b,$$

where $\phi_+ := (1-\alpha_P)\phi > 0$ (since (P) is infeasible). Also, $\beta_+ := b^T y_+ = \beta + \alpha_D \Delta\beta > 0$ from our assumptions. Hence our new shadow iterates are

$$(\bar{x}_+, \bar{\zeta}_+) := (\frac{x_+}{\phi_+}, -\frac{1-\phi_+}{\phi_+}), \quad (\bar{y}_+, \bar{s}_+) := (\frac{y_+}{\beta_+}, \frac{s_+}{\beta_+}),$$

with ϕ_+ and β_+ as above. We then find

$$\begin{aligned} \bar{x}_+ &= \frac{x + \alpha_P \Delta x}{(1-\alpha_P)\phi} \\ &= \frac{x}{\phi} + \left(\frac{\alpha_P}{1-\alpha_P} . \frac{\Delta\beta}{\beta}\right)\left(\frac{\beta}{\phi\Delta\beta}(\Delta x + x)\right) \\ &= \bar{x} + \bar{\alpha}_P \Delta\bar{x}, \end{aligned}$$

where

$$\bar{\alpha}_P := \frac{\alpha_P}{1-\alpha_P}.\frac{\Delta\beta}{\beta}, \quad \Delta\bar{x} := \frac{\beta}{\phi\Delta\beta}(\Delta x + x), \tag{4.13}$$

and

$$\begin{aligned}
\bar{\zeta}_+ &= -\frac{1-(1-\alpha_P)\phi}{(1-\alpha_P)\phi} \\
&= -\frac{1-\phi}{\phi} + \bar{\alpha}_P\left(-\frac{\beta}{\phi\Delta\beta}\right) \\
&= \bar{\zeta} + \bar{\alpha}_P\Delta\bar{\zeta},
\end{aligned}$$

where

$$\Delta\bar{\zeta} := -\frac{\beta}{\phi\Delta\beta}. \tag{4.14}$$

Note that the choice of $\bar{\alpha}_P$ and hence the scale of $\Delta\bar{x}$ and $\Delta\bar{\zeta}$ is somewhat arbitrary: the particular choice made will be justified in the following theorem. Similarly the choice of $\bar{\alpha}_D$ is somewhat arbitrary below.

We also have

$$\begin{aligned}
\bar{y}_+ &= \frac{y+\alpha_D\Delta y}{\beta+\alpha_D\Delta\beta} \\
&= \frac{y}{\beta} + \left(\frac{\alpha_D\Delta\beta}{\beta+\alpha_D\Delta\beta}\right)\left(\frac{\Delta y}{\Delta\beta} - \frac{y}{\beta}\right) \\
&= \bar{y} + \bar{\alpha}_D\Delta\bar{y},
\end{aligned}$$

where

$$\bar{\alpha}_D := \frac{\alpha_D\Delta\beta}{\beta+\alpha_D\Delta\beta}, \quad \Delta\bar{y} := \frac{\Delta y}{\Delta\beta} - \bar{y}, \tag{4.15}$$

and similarly

$$\bar{s}_+ = \bar{s} + \bar{\alpha}_D\Delta\bar{s},$$

where

$$\Delta\bar{s} := \frac{\Delta s}{\Delta\beta} - \bar{s}. \tag{4.16}$$

Theorem 4.1 *The directions $(\Delta\bar{x}, \Delta\bar{\zeta}, \Delta\bar{y}, \Delta\bar{s})$ defined in (4.13)–(4.16) solve the Newton system for $(\bar{P})$ and $(\bar{D})$ given below:*

$$\begin{array}{ccccccc}
 & & A^T\Delta\bar{y} & + & \Delta\bar{s} & = & -A^T\bar{y} - \bar{s}, \\
 & & b^T\Delta\bar{y} & & & = & 0, \\
A\Delta\bar{x} & + & b\Delta\bar{\zeta} & & & = & 0, \\
\bar{S}\Delta\bar{x} & & & + & \bar{X}\Delta\bar{s} & = & \bar{\sigma}\bar{\mu}e - \bar{S}\bar{X}e,
\end{array} \tag{4.17}$$

for the value

$$\bar{\sigma} := \frac{\beta}{\Delta\beta}\sigma. \tag{4.18}$$

Here $\bar{\mu} := \bar{s}^T\bar{x}/n$.

Proof: We establish the equations of (4.17) in order. First,

$$A^T\Delta\bar{y} + \Delta\bar{s} = A^T\left(\frac{\Delta y}{\Delta\beta} - \bar{y}\right) + \left(\frac{\Delta s}{\Delta\beta} - \bar{s}\right) = -A^T\bar{y} - \bar{s},$$

using the first equation of (4.12). Next,

$$b^T\Delta\bar{y} = b^T\left(\frac{\Delta y}{\Delta\beta} - \bar{y}\right) = 1 - b^T\bar{y} = 1 - b^Ty/\beta = 0$$

from the definition of $\Delta\beta$. For the third equation,

$$A\Delta\bar{x} + b\Delta\bar{\zeta} = \left(\frac{\beta}{\phi\Delta\beta}\right)(A(\Delta x + x) - b) = 0,$$

using the second equation of (4.12). Finally, we find

$$\begin{aligned}
\bar{S}\Delta\bar{x} + \bar{X}\Delta\bar{s} &= \frac{1}{\beta}.\frac{\beta}{\phi\Delta\beta}.(S\Delta x + Sx) + \frac{1}{\phi}\left(\frac{1}{\Delta\beta}X\Delta s - \frac{1}{\beta}Xs\right) \\
&= \left(\frac{\beta}{\Delta\beta}\right)\frac{1}{\beta\phi}(S\Delta x + X\Delta s + SXe) - \frac{1}{\beta\phi}SXe \\
&= \left(\frac{\beta}{\Delta\beta}\right)\left(\frac{1}{\beta\phi}\sigma\mu e\right) - \frac{1}{\beta\phi}SXe \\
&= \left(\frac{\beta}{\Delta\beta}\sigma\right)\bar{\mu}e - \bar{S}\bar{X}e,
\end{aligned}$$

using the last equation of (4.12). □

This theorem substantiates our main claim that, although the IIP method in the strictly infeasible case may be aiming towards a central path that doesn't exist, it is in fact implicitly trying to generate certificates of infeasibility. Indeed, the shadow iterates are being generated by damped Newton steps for the problems $(\bar{P})$ and $(\bar{D})$, for which the central path exists.

Since $(\bar{P})$ and $(\bar{D})$ are better behaved than (P) and (D), and therefore the behavior of the IIP method better understood, it is important to note that this correspondence can be reversed, to give the iteration for (P) and (D) from that for $(\bar{P})$ and $(\bar{D})$. So assume we are given $(\bar{x}, \bar{\zeta}, \bar{y}, \bar{s})$ with $A\bar{x} + b\bar{\zeta} = Ax_0$, $\bar{x} > 0$, $\bar{\zeta} \leq 0$ and $A^T\bar{y} + \bar{s} = c/\beta$, $b^T\bar{y} = 1$, $\bar{s} > 0$ for some positive β. Then we can define $\phi := 1/(1 - \bar{\zeta}) \in (0, 1]$ so that $\bar{\zeta} = -(1 - \phi)/\phi$, and make the

Definition 4.2 *The "real" iterate corresponding to* $(\bar{x}, \bar{\zeta}, \bar{y}, \bar{s})$ *is given by*

$$x := \phi\bar{x}, \quad (y, s) := \beta(\bar{y}, \bar{s}).$$

Thus $Ax = \phi(Ax_0) + (-\phi\bar{\zeta})b = \phi(Ax_0) + (1 - \phi)b$, $x > 0$ and $A^Ty + s = c$, $s > 0$.

Suppose $(\Delta\bar{x}, \Delta\bar{\zeta}, \Delta\bar{y}, \Delta\bar{s})$ is the solution to (4.17), and also make the

Assumption 4.3 $\Delta\bar{\zeta}$ *is negative.*

This also automatically holds if $\bar{\sigma}$ is sufficiently small, and is in a sense more reasonable than Assumption 4.2 since we are now presumably close (if β is large) to a well-defined central path, and from the form of $(\bar{P})$, the assumption just amounts to monotonicity of the objective in the primal shadow problem (see Mizuno et al. [16]).

We now define our new shadow iterate $(\bar{x}_+, \bar{\zeta}_+, \bar{y}_+, \bar{s}_+)$ by taking steps in this direction, $\bar{\alpha}_P > 0$ for $(\bar{x}, \bar{\zeta})$ and $\bar{\alpha}_D > 0$ for $(\bar{y}, \bar{s})$. (We can assume that $\bar{\alpha}_D$ is less than one, since otherwise $(\bar{y}_+, \bar{s}_+)$ is a certificate of primal infeasibility for (P) and we stop.) We set $\phi_+ := 1/(1-\bar{\zeta}_+) = 1/(1-\bar{\zeta}-\bar{\alpha}_P\Delta\bar{\zeta})$ (positive by Assumption 4.3) and $\beta_+ = \beta/(1-\bar{\alpha}_D) > 0$ so that $A^T\bar{y}_+ + \bar{s}_+ = c/\beta_+$. Then we define

$$\begin{aligned} x_+ &:= \phi_+\bar{x}_+ = \frac{\bar{x}+\bar{\alpha}_P\Delta\bar{x}}{1-\bar{\zeta}-\bar{\alpha}_P\Delta\bar{\zeta}} \\ &= x + \left(\frac{-\bar{\alpha}_P\Delta\bar{\zeta}}{1-\bar{\zeta}-\bar{\alpha}_P\Delta\bar{\zeta}}\right)\left(\frac{\Delta\bar{x}}{-\Delta\bar{\zeta}} - x\right) \\ &=: x + \alpha_P\Delta x \end{aligned}$$

(i.e., α_P and Δx are defined by the expressions in parentheses in the penultimate line);

$$\begin{aligned} y_+ &:= \beta_+\bar{y}_+ = \frac{\beta\bar{y}+\beta\bar{\alpha}_D\Delta\bar{y}}{1-\bar{\alpha}_D} \\ &= y + \left(\frac{-\bar{\alpha}_D\phi\Delta\bar{\zeta}}{1-\bar{\alpha}_D}\right)\left(\frac{\beta}{-\phi\Delta\bar{\zeta}}(\Delta\bar{y}+\bar{y})\right) \\ &=: y + \alpha_D\Delta y; \end{aligned}$$

and similarly

$$\begin{aligned} s_+ &:= \beta_+\bar{s}_+ \\ &= s + \left(\frac{-\bar{\alpha}_D\phi\Delta\bar{\zeta}}{1-\bar{\alpha}_D}\right)\left(\frac{\beta}{-\phi\Delta\bar{\zeta}}(\Delta\bar{s}+\bar{s})\right) \\ &=: s + \alpha_D\Delta s. \end{aligned}$$

It is straightforward to check

Theorem 4.2 *The directions $(\Delta x, \Delta y, \Delta s)$ defined above solve the Newton system (4.12) for (P) and (D) for the value $\sigma := \bar{\sigma}/(-\phi\Delta\bar{\zeta})$.*

□

We note that $b^T\Delta y = \beta/(-\phi\Delta\bar{\zeta})$, which is positive under Assumption 4.3. This and (4.14) show that Assumptions 4.2 and 4.3 are equivalent.

The relationship between α_P and $\bar{\alpha}_P$, α_D and $\bar{\alpha}_D$, and σ and $\bar{\sigma}$ will be discussed further in the next section. For example, if we suspect that (P) is infeasible, we may want to choose α_P and α_D so that $\bar{\alpha}_P$ and $\bar{\alpha}_D$ are close to 1, so that we are taking near-Newton steps in terms of the shadow iterates.

4.3 The dual strictly infeasible case

Now we sketch the analysis for the dual strictly infeasible case, omitting details. We suppose there is a solution to

$$A\tilde{x} = 0, \quad \tilde{x} > 0, \quad c^T\tilde{x} = -1.$$

In this case, the IIP algorithm usually generates a sequence of iterates where x becomes feasible after a certain iteration, and $c^T x$ tends to $-\infty$. $A^T y + s$ always remains a convex combination of its original value $A^T y_0 + s_0$ and its goal c. Thus we make the following

Assumption 4.4 *The current iterate (x, y, s) has x feasible in (P) and $\gamma := -c^T x > 0$. In addition,*

$$A^T y + s = \psi(A^T y_0 + s_0) + (1-\psi)c, \quad s > 0, \quad \psi > 0.$$

If $c^T x$ is large and negative, then x/γ will be an approximate solution to the Farkas system above. We formulate the optimization problem

$$\begin{array}{lll} (\tilde{P}) & \min & (A^T y_0 + s_0)^T \tilde{x} \\ & & A\tilde{x} = 0, \\ & & -c^T \tilde{x} = 1, \\ & & \tilde{x} \geq 0. \end{array}$$

($(\tilde{P})$ is a modified homogeneous form of (P).) This is strictly feasible. Its dual is

$$\begin{array}{lll} (\tilde{D}) & \max & \tilde{\kappa} \\ & & A^T \tilde{y} - c\tilde{\kappa} + \tilde{s} = A^T y_0 + s_0, \\ & & \tilde{s} \geq 0. \end{array}$$

We will use tildes to indicate the variables of $(\tilde{P})$ and $(\tilde{D})$. Note that, from our assumption on the current iterate, $(y/\psi, (1-\psi)/\psi, s/\psi)$ is a strictly feasible solution to $(\tilde{D})$. Hence we make the

Definition 4.3 *The shadow iterate corresponding to (x, y, s) is given by*

$$\tilde{x} := x/\gamma, \ \text{where } \gamma := -c^T x, \quad (\tilde{y}, \tilde{\kappa}, \tilde{s}) := (\frac{y}{\psi}, \frac{1-\psi}{\psi}, \frac{s}{\psi}).$$

We now wish to compare the results of applying one iteration of the IIP method from (x, y, s) for (P) and (D), and from $(\tilde{x}, \tilde{y}, \tilde{\kappa}, \tilde{s})$ for $(\tilde{P})$ and $(\tilde{D})$.

Let $(\Delta x, \Delta y, \Delta s)$ be the search direction of the first of these, and let α_P and α_D be the chosen positive step sizes, with (x_+, y_+, s_+) being the next iterate. Then according to the algorithm, we have

$$\begin{array}{rcl} A^T \Delta y + \Delta s & = & c - A^T y - s, \\ A\Delta x & = & 0, \\ S\Delta x + X\Delta s & = & \sigma\mu e - SXe. \end{array} \tag{4.19}$$

and

$$x_+ := x + \alpha_P \Delta x, \quad y_+ := y + \alpha_D \Delta y, \quad s_+ := s + \alpha_D \Delta s,$$

where $\mu := s^T x/n$. The corresponding iteration for $(\tilde{P})$ and $(\tilde{D})$ also comes from Section 4.1, where now A is augmented by the row $-c^T$, but we postpone stating it until we have generated trial search directions from those above. Before doing so, we note that

$$\Delta x = -(I - XS^{-1}A^T(AXS^{-1}A^T)^{-1}A)XS^{-1}c + \sigma\mu(I - XS^{-1}A^T(AXS^{-1}A^T)^{-1}A)S^{-1}e,$$

and so

$$\begin{array}{rcl}\Delta\gamma & := & -c^T\Delta x \\ & = & [c^TXS^{-1}c - c^TXS^{-1}A^T(AXS^{-1}A^T)^{-1}AXS^{-1}c] \\ & & -\sigma\mu(c^TS^{-1}e - c^TXS^{-1}A^T(AXS^{-1}A^T)^{-1}AS^{-1}e),\end{array}$$

and it follows (since dual infeasibility implies that c is not in the range of A^T) that $\Delta\gamma$ is positive for small enough σ. Henceforth, we make the

Assumption 4.5 *$\Delta\gamma$ is positive.*

We find that

$$A^Ty_+ + s_+ = \psi(A^Ty_0+s_0)+(1-\psi)c+\alpha_D(c-\psi(A^Ty_0+s_0)-(1-\psi)c) = \psi_+(A^Ty_0+s_0)+(1-\psi_+)c,$$

where $\psi_+ := (1-\alpha_D)\psi > 0$ (since (D) is infeasible). Also, $\gamma_+ := -c^Tx_+ = \gamma + \alpha_P\Delta\gamma > 0$ from our assumptions. Hence our new shadow iterates are

$$\tilde{x}_+ := \frac{x_+}{\gamma_+}, \quad (\tilde{y}_+, \tilde{\kappa}_+, \tilde{s}_+) := (\frac{y_+}{\psi_+}, \frac{1-\psi_+}{\psi_+}, \frac{s_+}{\psi_+}).$$

We then obtain

$$\begin{array}{rcl}\tilde{x}_+ & = & \dfrac{x+\alpha_P\Delta x}{\gamma+\alpha_P\Delta\gamma} \\ & = & \dfrac{x}{\gamma} + \left(\dfrac{\alpha_P\Delta\gamma}{\gamma+\alpha_P\Delta\gamma}\right)\left(\dfrac{\Delta x}{\Delta\gamma} - \dfrac{x}{\gamma}\right) \\ & = & \tilde{x} + \tilde{\alpha}_P\Delta\tilde{x},\end{array}$$

where

$$\tilde{\alpha}_P := \frac{\alpha_P\Delta\gamma}{\gamma+\alpha_P\Delta\gamma}, \quad \Delta\tilde{x} := \frac{\Delta x}{\Delta\gamma} - \tilde{x}. \tag{4.20}$$

We also have

$$\begin{array}{rcl}\tilde{y}_+ & = & \dfrac{y+\alpha_D\Delta y}{(1-\alpha_D)\psi} \\ & = & \dfrac{y}{\psi} + \left(\dfrac{\alpha_D}{1-\alpha_D}.\dfrac{\Delta\gamma}{\gamma}\right)\left(\dfrac{\gamma}{\psi\Delta\gamma}(\Delta y + y)\right) \\ & = & \tilde{y} + \tilde{\alpha}_D\Delta\tilde{y},\end{array}$$

where

$$\tilde{\alpha}_D := \frac{\alpha_D}{1-\alpha_D}.\frac{\Delta\gamma}{\gamma}, \quad \Delta\tilde{y} := \frac{\gamma}{\psi\Delta\gamma}(\Delta y + y). \tag{4.21}$$

Similarly, $\tilde{s}_+ = \tilde{s} + \tilde{\alpha}_D\Delta\tilde{s}$, where

$$\Delta\tilde{s} := \frac{\gamma}{\psi\Delta\gamma}(\Delta s + s), \tag{4.22}$$

and $\tilde{\kappa}_+ = \tilde{\kappa} + \tilde{\alpha}_D\Delta\tilde{\kappa}$, where

$$\Delta\tilde{\kappa} := \frac{\gamma}{\psi\Delta\gamma}. \tag{4.23}$$

Theorem 4.3 *The directions* $(\Delta\tilde{x}, \Delta\tilde{y}, \Delta\tilde{\kappa}, \Delta\tilde{s})$ *defined in (4.20) – (4.23) above solve the Newton system for* $(\tilde{P})$ *and* $(\tilde{D})$ *given below:*

$$\begin{array}{rcccccl} & A^T\Delta\tilde{y} & - & c\Delta\tilde{\kappa} & + & \Delta\tilde{s} & = & 0, \\ A\Delta\tilde{x} & & & & & & = & -A\tilde{x}, \\ -c^T\Delta\tilde{x} & & & & & & = & 0, \\ \tilde{S}\Delta\tilde{x} & & & & + & \tilde{X}\Delta\tilde{s} & = & \tilde{\sigma}\tilde{\mu}e - \tilde{X}\tilde{S}e, \end{array} \tag{4.24}$$

for the value

$$\tilde{\sigma} := \frac{\gamma}{\Delta\gamma}\sigma. \tag{4.25}$$

Here $\tilde{\mu} := \tilde{s}^T\tilde{x}/n$.

□

This argument can also be reversed. Given $(\tilde{x}, \tilde{y}, \tilde{\kappa}, \tilde{s})$, where we assume that $A\tilde{x} = b/\gamma$, $-c^T\tilde{x} = 1$, $\tilde{x} > 0$ for some positive γ and $A^T\tilde{y} - c\tilde{\kappa} + \tilde{s} = A^T y_0 + s_0$, $\tilde{s} > 0$, and $\tilde{\kappa} \geq 0$, we define $\psi := 1/(1+\tilde{\kappa}) \in (0, 1]$ so that $\tilde{\kappa} = (1-\psi)/\psi$, and hence the "real" iterate given by $x := \gamma\tilde{x}$, $(y, s) := \psi(\tilde{y}, \tilde{s})$. We compute the search direction from (4.24) and take steps of size $\tilde{\alpha}_P$ (assumed less than one, otherwise we have a certificate of dual infeasibility) and $\tilde{\alpha}_D$ to obtain new shadow iterates. The appropriate requirement is

Assumption 4.6 $\Delta\tilde{\kappa}$ *is positive,*

which turns out to be equivalent to our previous assumption that $\Delta\gamma > 0$. Then the new real iterates corresponding to the new shadow iterates are obtained from the old real iterates by using the step sizes and directions given below:

$$\alpha_P := \frac{\tilde{\alpha}_P\psi\Delta\tilde{\kappa}}{1-\tilde{\alpha}_P}, \qquad \Delta x := \frac{\gamma}{\psi\Delta\tilde{\kappa}}(\Delta\tilde{x} + \tilde{x}),$$

$$\alpha_D := \frac{\tilde{\alpha}_D\Delta\tilde{\kappa}}{1+\tilde{\kappa}+\tilde{\alpha}_D\Delta\tilde{\kappa}}, \qquad \Delta y := \frac{\Delta\tilde{y}}{\Delta\tilde{\kappa}} - y, \qquad \Delta s := \frac{\Delta\tilde{s}}{\Delta\tilde{\kappa}} - s.$$

Again it is easy to check

Theorem 4.4 *The directions* $(\Delta x, \Delta y, \Delta s)$ *defined above solve the Newton system (4.19) for* (P) *and* (D) *for the value* $\sigma := \bar{\sigma}/(\psi\Delta\tilde{\kappa})$.

□

5 Convergence and Implications

Here we give further properties of the iterates in the infeasible case, discuss the convergence of IIP methods in case of strict infeasibility, and consider the implications of our equivalence between real and shadow iterations for designing an efficient IIP method. In Section 5.1 we discuss the boundedness of the iterates in the infeasible case, while in Section 5.2 we consider the Kojima-Megiddo-Mizuno algorithm and convergence issues. Finally, Section 5.3 addresses the implications of our equivalence results for IIP methods.

5.1 Boundedness and unboundedness

Here we will assume that (P) is strictly infeasible, so that there is a solution to (4.11), which we repeat here:

$$A^T\bar{y} + \bar{s} = 0, \quad \bar{s} > 0, \quad b^T\bar{y} = 1.$$

(Similar results can be obtained in the dual strictly infeasible case.)

Note that any primal-dual IIP method has iterates (x_k, y_k, s_k) that satisfy

$$Ax_k = b_k := \phi_k(Ax_0) + (1-\phi_k)b, \quad 0 \le \phi_k \le 1, \tag{5.26}$$

and

$$A^T y_k + s_k = c_k := \psi_k(A^T y_0 + s_0) + (1-\psi_k)c, \quad 0 \le \psi_k \le 1, \tag{5.27}$$

for all k.

Proposition 5.1 *In the primal strictly infeasible case, we have*

$$\phi_k \ge (1 + \bar{s}^T x_0)^{-1}, \qquad \bar{s}^T x_k \le \bar{s}^T x_0 \tag{5.28}$$

for all $k \ge 0$. Hence all x_k's lie in a bounded set. Further, for any $\tilde{b}$ with $\|\tilde{b} - b\| < 1/\|\bar{y}\|$, the system $Ax = \tilde{b}$, $x \ge 0$ is infeasible.

Proof: For the first part, premultiply (5.26) by $-\bar{y}^T$ to get

$$\begin{aligned} \bar{s}^T x_k \;=\; -\bar{y}^T A x_k &= \phi_k(-\bar{y}^T A x_0) + (1-\phi_k)(-b^T\bar{y}) \\ &= \phi_k \bar{s}^T x_0 - 1 + \phi_k \\ &= \phi_k(1 + \bar{s}^T x_0) - 1. \end{aligned}$$

Since $\bar{s}^T x_k > 0$, we obtain the lower bound on ϕ_k. From $\phi_k \le 1$, the upper bound on $\bar{s}^T x_k$ holds. For the second part, note that $\tilde{b}^T\bar{y} = b^T\bar{y} + (\tilde{b} - b)^T\bar{y} \ge 1 - \|\tilde{b} - b\|\|\bar{y}\| > 0$, so that $(\bar{y}, \bar{s})$ certifies the infeasibility of $Ax = \tilde{b}$, $x \ge 0$. □

Proposition 5.2 *Suppose that in addition the sequence $\{(x_k, y_k, s_k)\}$ satisfies $s_k^T x_k \le s_0^T x_0$ and $\|s_k\| \to \infty$. Then $b^T y_k \to \infty$.*

Proof: Indeed, we have

$$\begin{aligned} s_0^T x_0 \;\ge\; s_k^T x_k \;=\; (c_k - A^T y_k)^T x_k & \\ = c_k^T x_k - y_k^T[\phi_k(Ax_0) + (1-\phi_k)b] & \\ = c_k^T x_k - \phi_k(A^T y_k)^T x_0 - (1-\phi_k) b^T y_k & \\ = c_k^T x_k - \phi_k(c_k - s_k)^T x_0 - (1-\phi_k) b^T y_k & \\ = [c_k^T x_k - \phi_k c_k^T x_0] + \phi_k s_k^T x_0 - (1-\phi_k) b^T y_k. & \end{aligned} \tag{5.29}$$

Now, by Proposition 5.1, the quantity in brackets remains bounded, while $\phi_k \ge (1 + \bar{s}^T x_0)^{-1} > 0$ and $s_k^T x_0 \to \infty$. Thus we must have $b^T y_k \to \infty$. □

5.2 The Kojima-Megiddo-Mizuno algorithm and convergence

Kojima, Megiddo, and Mizuno [10] (henceforth KMM) devised a particular IIP method that correctly detected infeasibility, but without generating a certificate of infeasibility in the usual sense. Here we show that their algorithm does indeed generate certificates of infeasibility in the limit (in the strictly infeasible case). We also see how their method relates to the assumptions and shadow iterates we studied in Section 4.2.

KMM's algorithm uses special rules for choosing σ, α_P, and α_D at each iteration, and employs a special neighborhood: for (P) and (D), this is defined to be

$$\begin{array}{lcl} \mathcal{N} & := & \mathcal{N}(\gamma_0, \gamma_P, \gamma_D, \epsilon_P, \epsilon_D) := \mathcal{N}_0 \cap \mathcal{N}_P \cap \mathcal{N}_D, \quad \text{where} \\ \mathcal{N}_0 & := & \{(x,y,s) \in \mathbf{R}^n_{++} \times \mathbf{R}^m \times \mathbf{R}^n_{++} : s_j x_j \geq \gamma_0 s^T x/n, \text{ for all } j\}, \\ \mathcal{N}_P & := & \{(x,y,s) \in \mathbf{R}^n_{++} \times \mathbf{R}^m \times \mathbf{R}^n_{++} : \|Ax - b\| \leq \max(\epsilon_P, s^T x/\gamma_P)\}, \\ \mathcal{N}_D & := & \{(x,y,s) \in \mathbf{R}^n_{++} \times \mathbf{R}^m \times \mathbf{R}^n_{++} : \|A^T y + s - c\| \leq \max(\epsilon_D, s^T x/\gamma_D)\}. \end{array} \tag{5.30}$$

Here ϵ_P and ϵ_D are small positive constants, and $\gamma_0 < 1$, γ_P, and γ_D are positive constants chosen so that $(x_0, y_0, s_0) \in \mathcal{N}$. KMM maintain all iterates in $\mathcal{N}$.

They choose parameters $0 < \sigma_1 < \sigma_2 < \sigma_3 < 1$. At every iteration, σ is chosen to be σ_1 to generate search directions $(\Delta x, \Delta y, \Delta s)$ from the current iterate $(x, y, s) \in \mathcal{N}$. (In fact, it suffices for their arguments to choose σ from the interval $[\sigma_1', \sigma_1'']$, possibly with different choices at each iteration, where $0 < \sigma_1' < \sigma_1'' < \sigma_2 < \sigma_3 < 1$.) Next, a step size $\bar{\alpha}$ is chosen as the largest $\tilde{\alpha} \leq 1$ so that

$$\begin{array}{c} (x + \alpha \Delta x, y + \alpha \Delta y, s + \alpha \Delta s) \in \mathcal{N} \qquad \text{and} \\ (s + \alpha \Delta s)^T (x + \alpha \Delta x) \leq [1 - \alpha(1 - \sigma_2)] s^T x \end{array}$$

for all $\alpha \in [0, \tilde{\alpha}]$. Finally, $\alpha_P \leq 1$ and $\alpha_D \leq 1$ are chosen so that

$$\begin{array}{c} (x + \alpha_P \Delta x, y + \alpha_D \Delta y, s + \alpha_D \Delta s) \in \mathcal{N} \qquad \text{and} \\ (s + \alpha_D \Delta s)^T (x + \alpha_P \Delta x) \leq [1 - \bar{\alpha}(1 - \sigma_3)] s^T x \end{array} \tag{5.31}$$

Note that a possible choice is $\alpha_P = \alpha_D = \bar{\alpha}$. However, the relaxation provided by choosing $\sigma_3 > \sigma_2$ allows other options; in particular, it might be possible to choose one of α_P and α_D as 1 (thus attaining primal or dual feasibility) while the other is necessarily small (because the dual or primal problem is infeasible).

The algorithm is terminated whenever an iterate (x, y, s) is generated satisfying

$$s^T x \leq \epsilon_0, \quad \|Ax - b\| \leq \epsilon_P, \quad \text{and } \|A^T y + s - c\| \leq \epsilon_D \tag{5.32}$$

(an approximately optimal point; more precisely, x and (y, s) are ϵ_0-optimal in the nearby problems where b is replaced by Ax and c by $A^T y + s$), or

$$\|(x, s)\|_1 > \omega^*, \tag{5.33}$$

for suitable positive (small) ϵ_0, ϵ_P, ϵ_D and (large) ω^*. KMM argue (Section 4 of [10]) that, in the latter case, there is no feasible solution in a large region of $\mathbf{R}^n_{++} \times \mathbf{R}^m \times \mathbf{R}^n_{++}$. A slight modification of their algorithm (Section 5 of [10]) yields stronger conclusions, but neither version appears to generate a certificate of infeasibility.

KMM prove (Section 4 of [10]) that for given positive ϵ_0, ϵ_P, ϵ_D and ω^*, their algorithm terminates finitely. We now show how their method can provide approximate certificates of infeasibility. Suppose that (P) is strictly infeasible, and that ϵ_P is chosen sufficiently small that there is no nonnegative solution to $\|Ax - b\| \leq \epsilon_P$ (see Proposition 5.1).

Theorem 5.1 *Suppose the KMM algorithm is applied to a primal strictly infeasible instance, with ϵ_P chosen as above and the large norm termination criterion (5.33) disabled. Then $\|s_k\| \to \infty$, $\beta_k := b^T y_k \to \infty$ and there is a subsequence along which $(y_k/\beta_k, s_k/\beta_k) \to (\bar{y}, \bar{s})$, with the latter a certificate of primal infeasibility.*

Proof: By our choice of ϵ_P, the algorithm cannot terminate due to (5.32), and we have disabled the other termination criterion, so that the method generates an infinite sequence of iterates (x_k, y_k, s_k). KMM show that, if $\|(x_k, s_k)\|_1 \leq \omega$, for any positive ω, then there is some $\underline{\alpha} > 0$, depending on ω, such that $\bar{\alpha}_k \geq \underline{\alpha}$, and hence, by (5.31), the total complementarity $s^T x$ decreases at least by the factor $[1 - \underline{\alpha}(1 - \sigma_3)] < 1$ at this iteration. On every iteration, the total complementarity does not increase. Hence, if there is an infinite number of iterations with $\|(x_k, s_k)\|_1 \leq \omega$, $s_k^T x_k$ converges to zero, and since all iterates lie in $\mathcal{N}$, $\|Ax_k - b\|$ also tends to zero. But this contradicts strict primal infeasibility, so that there cannot be such an infinite subsequence. This holds for any positive ω, and thus $\|(x_k, s_k)\| \to \infty$. By Proposition 5.1, $\{x_k\}$ remains bounded, so $\|s_k\| \to \infty$. By the rules of the KMM algorithm, $s_k^T x_k \leq s_0^T x_0$ for all k. Hence by Proposition 5.2, $\beta_k \to \infty$. From (5.29) we see that $(s_k/\beta_k)^T x_0$ and thus $\bar{s}_k := s_k/\beta_k$ remain bounded, so that there is a infinite subsequence K with $\lim_K \bar{s}_k := \lim_{k \in K, k \to \infty} \bar{s}_k = \bar{s}$ for some $\bar{s} \geq 0$. Further, $\bar{y}_k := y_k/\beta_k$ satisfies $A^T \bar{y}_k = c_k/\beta_k - \bar{s}_k$, which converges to $-\bar{s}$ along K, since c_k remains bounded. Hence $\bar{y}_k$ converges to $\bar{y} := -(AA^T)^{-1} A\bar{s}$ along this subsequence. We therefore have

$$A^T \bar{y} + \bar{s} = \lim_K (A^T y_k + s_k)/\beta_k = \lim_K c_k/\beta_k = 0, \quad \bar{s} \geq 0, \quad b^T \bar{y} = \lim_K b^T y_k/\beta_k = 1,$$

as desired. □

While an exact certificate of infeasibility is obtained only in the limit (except under the happy circumstance that $A^T \bar{y}_k \leq 0$ for some k), $(\bar{y}_k, \bar{s}_k)$ is an approximate such certificate for large k, and we can conclude that there is no feasible x in a large region, and that a nearby problem with slightly perturbed A matrix is primal infeasible; see Todd and Ye [28].

The results above shed light on our assumptions in Section 4.2. Indeed, we showed that $b^T y_k \to \infty$, which justifies our supposition that $\beta > 0$ in Assumption 4.1. As we noted in Section 4.2, Assumption 4.2 (or equivalently 4.3) holds if σ (or $\bar{\sigma}$) is sufficiently small (depending on the current iterate), although this may contradict the KMM choice of σ. In practice, even with empirical rules for choosing the parameters, the assumptions that $\beta > 0$ and $\Delta\beta > 0$ seem to hold after the first few iterations. The main assumption left is that (y, s) is feasible, and we have not been able to establish rules for choosing α_P and α_D that will assure this (it is necessary to have $\alpha_D = 1$ at some iteration, unless (y_0, s_0) is itself feasible). As we noted, this assumption does seem to hold in practice. Moreover, if $A^T y_k + s_k$ converges to c but never equals it, then eventually $\|A^T y_k + s_k - c\| \leq \epsilon_D$, and then KMM's modified algorithm (Section 5 of [10]) replaces c by $c_k = A^T y_k + s_k$, so that the dual iterates are from now on feasible in the perturbed problem.

Finally, let us relate the neighborhood conditions for an iterate in the "real" universe to those for the corresponding shadow iterate. Let us suppose that the current iterate (x, y, s) satisfies Assumption 4.1, and let $(\bar{x}, \bar{\zeta}, \bar{y}, \bar{s})$ be the corresponding shadow iterate. We define the neighborhood $\bar{\mathcal{N}}$ in the shadow universe using parameters $\bar{\gamma}_0$, $\bar{\gamma}_P$, $\bar{\gamma}_D$,

$\bar{\epsilon}_P$, and $\bar{\epsilon}_D$ in the obvious way, with the centering condition involving only the $\bar{x}$- and $\bar{s}$-variables, since $\bar{\zeta}$ is free.

Proposition 5.3 *Suppose $\epsilon_P \leq s^T x/\gamma_P$ and $\bar{\epsilon}_D \leq \bar{s}^T \bar{x}/\bar{\gamma}_D$. Then, if $\bar{\gamma}_0 = \gamma_0$ and $\bar{\gamma}_D = (\|Ax_0 - b\|/\|c\|)\gamma_P$, $(x, y, s) \in \mathcal{N}$ if and only if $(\bar{x}, \bar{\zeta}, \bar{y}, \bar{s}) \in \bar{\mathcal{N}}$.*

(Note that our requirements on the γ's are natural; γ_0 and $\bar{\gamma}_0$ are dimension-free, while we expect γ_P to be inversely proportional to a typical norm for $Ax - b$, such as $\|Ax_0 - b\|$, and $\bar{\gamma}_D$ to be inversely proportional to a typical norm for $A^T\bar{y} + \bar{s} - 0$, such as $\|c\|$.)
Proof: Since $\bar{x} = x/\phi$ and $\bar{s} = s/\beta$, we have $\bar{s}^T\bar{x} = s^T x/(\beta\phi)$ and $\bar{\mu} = \mu/(\beta\phi)$ where $\mu := s^T x/n$ and $\bar{\mu} := \bar{s}^T\bar{x}/n$. Thus, for each j,

$$\bar{s}_j\bar{x}_j \geq \bar{\gamma}_0\bar{\mu} \text{ iff } s_j x_j/(\beta\phi) \geq \bar{\gamma}_0\mu/(\beta\phi) \text{ iff } s_j x_j \geq \gamma_0\mu.$$

Next, $\|A^T y + s - c\| = 0 \leq \max(\epsilon_D, s^T x/\gamma_D)$ and $\|A\bar{x} + \bar{\zeta} - Ax_0\| = 0 \leq \max(\bar{\epsilon}_P, \bar{s}^T\bar{x}/\bar{\gamma}_P)$. Finally, $Ax - b = \phi(Ax_0 - b)$, so

$$\phi\|Ax_0 - b\| = \|Ax - b\| \leq \max(\epsilon_P, s^T x/\gamma_P) = s^T x/\gamma_P$$

if and only if

$$\phi \leq s^T x/(\gamma_P\|Ax_0 - b\|);$$

whereas $b^T\bar{y} - 1 = 0$ and $A^T\bar{y} + \bar{s} - 0 = c/\beta$, so

$$\|c\|/\beta = \|A^T\bar{y} + \bar{s} - 0\| \leq \max(\bar{\epsilon}_D, \bar{s}^T\bar{x}/\bar{\gamma}_D) = \bar{s}^T\bar{x}/\bar{\gamma}_D = s^T x/(\beta\phi\bar{\gamma}_D)$$

if and only if

$$\phi \leq s^T x/(\bar{\gamma}_D\|c\|).$$

By our conditions on γ_P and $\bar{\gamma}_D$, these conditions are equivalent. □

Let us summarize what we have shown (and not shown) about the convergence of IIP methods. (Of course, analogous results for the dual strictly infeasible case can easily be established.) Theorem 5.1 shows that the original KMM algorithm will provide certificates of infeasibility in the limit for strictly infeasible instances. However, our development of Sections 4.2 and 4.3 suggests a more ambitious goal. We would like a strategy for choosing the centering parameter σ and the step sizes α_P and α_D at each iteration so that:

(a) In case (P) and (D) are feasible, the iterates converge to optimal solutions to these problems;

(b) In case (P) is strictly infeasible, the iterates become dual feasible, $b^T y$ becomes positive, and thenceforth the shadow iterates converge to optimal solutions of $(\bar{P})$ and $(\bar{D})$, unless a certificate of primal infeasibility is generated;

(c) In case (D) is strictly infeasible, the iterates become primal feasible, $c^T x$ becomes negative, and thenceforth the shadow iterates converge to optimal solutions of $(\tilde{P})$ and $(\tilde{D})$, unless a certificate of dual infeasibility is generated.

Of course, the algorithm should proceed without knowing which case obtains. We would further like some sort of polynomial bound on the number of iterations required in each case.

Unfortunately, we are a long way from achieving this goal. We do not know how to achieve dual (primal) feasibility in case (b) (case (c)). And we do not know how to choose the parameter σ and corresponding $\bar{\sigma}$ and the step sizes α_P and α_D and corresponding $\bar{\alpha}_P$ and $\bar{\alpha}_D$ to achieve simultaneously good progress in (P) and (D) and $(\bar{P})$ and $(\bar{D})$ (which we would like since we do not know which of cases (a) and (b) holds). The next subsection gives some practical guidance for choosing these parameters, without any guarantees of convergence.

5.3 Implications for the design of IIP methods

Suppose we are at a particular iteration of an IIP method and we suspect that (P) is strictly infeasible. (Similar considerations of course apply in the dual case.) For example, we might check that ϕ (the proportion of primal infeasibility remaining at the current iterate x) is at least .01 whereas the dual iterate (y, s) is (approximately) feasible, and $\beta := b^T y$ and $\Delta\beta := b^T \Delta y$ are positive. It then might make practical sense to choose the parameters to determine the next iterates with some attention to the (presumably better behaved) shadow iterates. Let us recall the relationship between the parameters in the real and shadow universes:

$$\bar{\sigma} := \frac{\beta}{\Delta\beta}\sigma, \quad \bar{\alpha}_P := \frac{\alpha_P}{1-\alpha_P}.\frac{\Delta\beta}{\beta}, \quad \bar{\alpha}_D := \frac{\alpha_D \Delta\beta}{\beta + \alpha_D \Delta\beta}.$$

In the case that we expect, $\Delta\beta$ will be considerably larger than β, so that $\bar{\sigma}$ will be much smaller than σ. Practical rules for choosing σ might lead to a value close to 1, since poor progress is being made in achieving feasibility in (P); but $\bar{\sigma}$ may still be quite small, indicating good progress toward optimality in $(\bar{P})$ and $(\bar{D})$. Indeed, it seems reasonable to choose σ quite large, so that $\bar{\sigma}$ is not too small — recall that merely achieving feasibility in $(\bar{D})$ yields a certificate of primal infeasibility; an optimal solution is not required. Of course, $\Delta\beta$ itself depends on σ by the relation above Assumption 4.2, but choosing a larger σ is likely to increase $\bar{\sigma}$.

Having thus chosen σ, we need to choose the step size parameters α_P and α_D. Because primal feasibility cannot be achieved, $\alpha_P < 1$, but again, the resulting $\bar{\alpha}_P$ may be much larger, indeed even bigger than 1. In such a case it seems reasonable to make α_P even smaller, so that the corresponding $\bar{\alpha}_P = 1$, using the formula above. A reverse situation occurs for α_D and $\bar{\alpha}_D$. If we limit α_D to 1, the corresponding $\bar{\alpha}_D$ may be quite small, whereas we would like to have $\bar{\alpha}_D = 1$ to obtain a certificate of infeasibility. Such a value corresponds to $\alpha_D = \infty$, so that it seems reasonable to take α_D as a large fraction of the distance to the boundary, even if this exceeds 1. If it is possible to choose $\alpha_D = \infty$, then $(\Delta y, \Delta s)$ is itself a certificate of primal infeasibility.

Modifications of this kind in the software package SDPT3 (see [31]) seem quite useful to detect infeasibility; in particular, allowing α_D (α_P) to be very large when primal (dual) infeasibility is suspected usually gives a very good certificate of primal (dual) infeasibility at the next iteration.

6 Extensions to Conic Programming

All of our discussion so far has concentrated on the linear programming case. In this section we show that the results of Section 4 extend to many IIP methods for more general conic programming problems of the form

$$(\check{P}) \quad \text{minimize} \quad \langle c, x \rangle, \qquad Ax = b, \quad x \in K.$$

Here K is a closed, convex, pointed (i.e., containing no line), and solid (with nonempty interior) cone in a finite-dimensional real vector space E with dual E^*, and $c \in E^*$: $\langle s, x \rangle$ denotes the result of $s \in E^*$ acting on $x \in E$. A is a surjective linear transformation from E to the dual Y^* of another finite-dimensional real vector space Y, and $b \in Y^*$. In particular, this includes the case of semidefinite programming (SDP), where $E = E^*$ is the space of symmetric matrices of order n with the inner product $\langle s, x \rangle := \text{Trace}(s^T x)$ and K is the positive semidefinite cone. It also contains the case of second-order cone programming (SOCP), where $E = E^* = \mathbf{R}^n$ with the usual inner product and $K = K_1 \times \ldots \times K_q$, with $K_i := \{x_i \in \mathbf{R}^{n_i} : x_i = (x_{i1}; \bar{x}_i), x_{i1} \geq \|\bar{x}_i\|\}$ and $\sum_i n_i = n$. (Here we have used Matlab notation: $(u; v)$ is the column vector obtained by concatenating the column vectors u and v. Hence the first component of x_i is required to be at least the Euclidean norm of the vector of the remaining components.)

Both of these classes of optimization problems have nice theory and wide-ranging applications: see, e.g., Ben-Tal and Nemirovski [2] or Todd [27].

The problem dual to $(\check{P})$ is

$$(\check{D}) \quad \text{maximize} \quad \langle b, y \rangle, \qquad A^* y + s = c, \quad s \in K^*,$$

where $A^* : Y \to E^*$ is the adjoint transformation to A and $K^* := \{s \in E^* : \langle s, x \rangle \geq 0 \text{ for all } x \in K\}$ is the cone dual to K. In the two cases above, K is self-dual, so that $K^* = K$ (we have identified E and E^*).

Given a possibly infeasible interior point $(x, y, s) \in \text{int}\, K \times Y \times \text{int}\, K^*$, a primal-dual IIP method (see, e.g., [19, 20, 21, 27, 2]) takes steps in the directions $(\Delta x, \Delta y, \Delta s)$ obtained from a linear system of the form

$$\begin{array}{rcrcrcl} & & A^*\Delta y & + & \Delta s & = & c - A^*y - s, \\ A\Delta x & & & & & = & b - Ax, \\ \mathcal{E}\Delta x & & & + & \mathcal{F}\Delta s & = & \sigma g - h, \end{array} \tag{6.34}$$

for certain operators $\mathcal{E} : E \to V$ and $\mathcal{F} : E^* \to V$ (V is another real vector space of the same dimension as E) and certain $g, h \in V$, depending on the current iterates x and s, and for a certain parameter $\sigma \in [0, 1]$; compare with (4.10).

We are again interested in the case that $(\check{P})$ or $(\check{D})$ is infeasible, and again we concentrate on the primal case, the dual being similar. We note that a sufficient condition for primal infeasibility is the existence of $(\bar{y}, \bar{s}) \in Y \times E^*$ with

$$A^*\bar{y} + \bar{s} = 0, \quad \bar{s} \in K^*, \quad \langle b, \bar{y} \rangle = 1, \tag{6.35}$$

but in the general nonpolyhedral case this is no longer necessary. We will say that $(\check{P})$ is strictly infeasible if there is such a certificate with $\bar{s} \in \operatorname{int} K^*$ (again, this implies that $(\check{D})$ is strictly feasible). Henceforth we suppose that $(\check{P})$ is strictly infeasible and that a sequence of iterations from the initial infeasible interior point (x_0, y_0, s_0) has led to the current iterate (x, y, s) where the analogue of Assumption 4.1 holds (the only change is that $\beta := \langle b, y\rangle$ is positive). We consider the Farkas-like problem

$$
\begin{array}{lllllll}
(\bar{D}) & \max & \langle Ax_0, \bar{y}\rangle & & & & \\
& & A^*\bar{y} & + & \bar{s} & = & 0, \\
& & \langle b, \bar{y}\rangle & & & = & 1, \\
& & & & \bar{s} \in K^*, & &
\end{array}
$$

with dual

$$
\begin{array}{llllll}
(\bar{P}) & \min & & & \bar{\zeta} & \\
& & A\bar{x} & + & b\bar{\zeta} & = \; Ax_0, \\
& & \bar{x} \in K. & & &
\end{array}
$$

We define the shadow iterate $(\bar{x}, \bar{\zeta}, \bar{y}, \bar{s})$ of (x, y, s) exactly as in Definition 4.1. We will show that, assuming $\mathcal{E}$, $\mathcal{F}$, g, and h depend on x and s suitably, once again an iteration from (x, y, s) corresponds appropriately to a shadow iteration from $(\bar{x}, \bar{\zeta}, \bar{y}, \bar{s})$. We define $\bar{\mathcal{E}}$, $\bar{\mathcal{F}}$, $\bar{g}$, and $\bar{h}$ from the shadow iterate as their unbarred versions were defined from (x, y, s).

Since we are assuming (y, s) feasible, our directions $(\Delta x, \Delta y, \Delta s)$ solve (6.34) with the first right-hand side replaced by zero. We again assume that $\Delta\beta := \langle b, \Delta y\rangle$ is positive. Having chosen positive step sizes α_P and α_D to obtain the new iterate (x_+, y_+, s_+), we define $\bar{\alpha}_P$, $\Delta\bar{x}$, $\Delta\bar{\zeta}$, $\bar{\alpha}_D$, $\Delta\bar{y}$, and $\Delta\bar{s}$ exactly as in Section 4.2.

Theorem 6.1 *Let us suppose that $\mathcal{E}$, $\mathcal{F}$, g, and h and $\mathcal{E}$, $\mathcal{F}$, g, and h are related in one of the following ways:*

(a) $\bar{\mathcal{E}} = \mathcal{E}/\beta$, $\bar{\mathcal{F}} = \mathcal{F}/\phi$, $\bar{g} = g/(\beta\phi)$, and $\bar{h} = h/(\beta\phi)$;

(b) $\bar{\mathcal{E}} = \mathcal{E}$, $\bar{\mathcal{F}} = (\beta/\phi)\mathcal{F}$, $\bar{g} = g/\phi$, and $\bar{h} = h/\phi$; or

(c) $\bar{\mathcal{E}} = (\phi/\beta)\mathcal{E}$, $\bar{\mathcal{F}} = \mathcal{F}$, $\bar{g} = g/\beta$, and $\bar{g} = g/\beta$.

Suppose also that $\mathcal{E}x = \mathcal{F}s = h$. Then the directions $(\Delta\bar{x}, \Delta\bar{\zeta}, \Delta\bar{y}, \Delta\bar{s})$ solve the Newton system for $(\bar{P})$ and $(\bar{D})$ given below:

$$
\begin{array}{rcccccl}
& & A^*\Delta\bar{y} & + & \Delta\bar{s} & = & -A^*\bar{y} - \bar{s}, \\
& & \langle b, \Delta\bar{y}\rangle & & & = & 0, \\
A\Delta\bar{x} & + & b\Delta\bar{\zeta} & & & = & 0, \\
\bar{\mathcal{E}}\Delta\bar{x} & & & + & \bar{\mathcal{F}}\Delta\bar{s} & = & \bar{\sigma}\bar{g} - \bar{h},
\end{array}
\qquad (6.36)
$$

for the value $\bar{\sigma} := \frac{\beta}{\Delta\beta}\sigma$.

Proof: The derivation is exactly as in the proof of Theorem 4.1 except for that of the last equation. In case (a) we obtain

$$
\bar{\mathcal{E}}\Delta\bar{x} + \bar{\mathcal{F}}\Delta\bar{s} \;=\; \frac{1}{\beta}\mathcal{E}\left(\frac{\beta}{\phi\Delta\beta}(\Delta x + x)\right) + \frac{1}{\phi}\mathcal{F}\left(\frac{\Delta s}{\Delta\beta} - \frac{s}{\beta}\right)
$$

$$
\begin{aligned}
&= \frac{1}{\phi\Delta\beta}(\mathcal{E}\Delta x + \mathcal{F}\Delta s) + \frac{1}{\phi\Delta\beta}\mathcal{E}x - \frac{1}{\beta\phi}\mathcal{F}s \\
&= \frac{1}{\phi\Delta\beta}(\sigma g - h) + \frac{1}{\phi\Delta\beta}h - \frac{1}{\beta\phi}h \\
&= \left(\frac{\beta}{\Delta\beta}\sigma\right)\left(\frac{g}{\beta\phi}\right) - \left(\frac{h}{\beta\phi}\right) = \bar{\sigma}\bar{g} - \bar{h},
\end{aligned}
$$

as desired. Cases (b) and (c) are exactly the same after dividing the last equation by β (case (b)) or ϕ (case (c)). □

Note that case (a) covers any situation where $\mathcal{E}$ scales with s, $\mathcal{F}$ with x, and g and h with both x and s. (As long as we also have $\mathcal{E}x = \mathcal{F}s = h$.) This includes our previous linear programming analysis, where $\mathcal{E} = S$, $\mathcal{F} = X$, $\gamma = \mu e$, and $h = SXe$. It also includes the Alizadeh-Haeberly-Overton [1] direction for SDP, where $\mathcal{E}$ is the operator $v \to (sv + vs)/2$, $\mathcal{F}$ the operator $v \to (vx + xv)/2$, g $\langle s, x\rangle/n$ times the identity, and $h = (sx + xs)/2$. (We write direction instead of method here and below, to stress that we are concerned here with the Newton system, which defines the direction; many different methods can use this direction, depending on their choices of the centering parameter and the step sizes.)

As a second example, the HRVW/KSH/M direction for SDP (see [8, 11, 17]) has $\mathcal{E}$ the identity, $\mathcal{F}$ the operator $v \to (xvs^{-1} + s^{-1}vx)/2$, g $\langle s, x\rangle/n$ times s^{-1}, and $h = x$. It is easily seen that these choices satisfy the conditions of case (b), as well as the extra condition. Another instance of case (b) is the Nesterov-Todd (NT) direction for SDP — see [20, 21]. Here $\mathcal{F}$ is the operator $v \to wvw$, where $w := x^{1/2}[x^{1/2}sx^{1/2}]^{-1/2}x^{1/2}$ is the unique positive definite matrix with $wsw = x$, and $\mathcal{E}$, g, and h are as above. Then, if $\bar{w}\bar{s}\bar{w} = \bar{x}$, it is easy to see that $\bar{w} = (\beta/\phi)^{1/2}w$, so again the conditions are simple to check.

The dual HRVW/KSH/M direction for SDP (see [11, 17]) is an instance of case (c). Here $\mathcal{E}$ takes v to $(svx^{-1} + x^{-1}vs)/2$, $\mathcal{F}$ is the identity, g $\langle s, x\rangle/n$ times x^{-1}, and $h = s$.

We presented the NT direction above in the form that is most useful for computing the directions, and only for SDP. But it is applicable in more general self-scaled conic programming (including SOCP), using a self-scaled barrier function, and can be given in a form as above satisfying the conditions of case (b), or another form that conforms to case (c).

Lastly, the presentations of the HRVW/KSH/M and NT directions for SOCP in Tsuchiya [29] use different forms of the Newton system: and it is easy to see that these fit into case (a) of the theorem.

Let us finally note that our results on boundedness and unboundedness in Section 5.1 also hold for general conic programming problems. The key simple fact is that, if $s \in \operatorname{int} K^*$, then $\{x \in K : \langle s, x\rangle \le \delta\}$ is bounded for any positive δ. Hence analogues of Propositions 5.1 and 5.2 hold.

7 Concluding Remarks

We have shown that there is a surprising connection between the iterates of an IIP method, applied to a dual pair of problems (P) and (D) in the case that one of them is strictly infeasible, and those of another IIP method applied to a related pair of strictly feasible problems whose optimal solutions give a certificate of infeasibility for (P) or (D). This connection involves a projective transformation from the original setting of the problems to a "shadow" universe, where the corresponding iterates lie. It holds not only for linear programming, but also for a range of methods for certain more general conic programming problems, including semidefinite and second-order cone programming problems. We hope that an intriguing glimpse of this connection has been provided, but it is clear that much work remains to be done to understand the convergence of IIP methods.

Acknowledgement

The author would like to thank Bharath Rangarajan for helpful discussions; in particular, he improved my earlier analysis of the Kojima-Megiddo-Mizuno algorithm, in which I only proved that the limit superior of the dual objective functions $b^T y_k$ was infinite, not that the whole sequence diverged.

References

[1] F. Alizadeh, J.-P. A. Haeberly, and M. L. Overton. Primal-dual interior-point methods for semidefinite programming: convergence rates, stability and numerical results. *SIAM J. Optim.*, 8:746–768, 1998.

[2] A. Ben-Tal and A. S. Nemirovski. *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*. SIAM Publications. SIAM, Philadelphia, USA, 2001.

[3] R. E. Bixby. Solving real-world linear programs: a decade and more of progress. *Oper. Res.*, 50:3–15, 2001.

[4] G. B. Dantzig. *Linear Programming and Extensions*. Princeton University Press. Princeton, 1963.

[5] A. Forsgren, P. E. Gill, and M. H. Wright. Interior methods for nonlinear optimization. *SIAM Rev.*, 44:525–597, 2002.

[6] A. J. Goldman and A. W. Tucker. Theory of linear programming. In H. W. Kuhn and A. W. Tucker, editors, *Linear Equalities and Related Systems*, pages 53–97. Princeton University Press, Princeton, N. J., 1956.

[7] O. Güler and Y. Ye. Convergence behavior of interior point algorithms. *Math. Progr.*, 60:215–228, 1993.

[8] C. Helmberg, F. Rendl, R. Vanderbei, and H. Wolkowicz. An interior-point method for semidefinite programming. *SIAM J. Optim.*, 6:342–361, 1996.

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

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

[11] M. Kojima, S. Shindoh, and S. Hara. Interior-point methods for the monotone semidefinite linear complementarity problem in symmetric matrices. *SIAM J. Optim.*, 7:86–125, 1997.

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

[13] I. J. Lustig, R. E. Marsten, and D. F. Shanno. Computational experience with a primal–dual interior point method for linear programming. *Lin. Alg. Appl.*, 152:191–222, 1991.

[14] S. Mizuno. Polynomiality of infeasible–interior–point algorithms for linear programming. *Math. Progr.*, 67:109–119, 1994.

[15] S. Mizuno and M. J. Todd. On two homogeneous self-dual systems for linear programming and its extensions. *Math. Progr.*, 89:517–534, 2001.

[16] S. Mizuno, M. J. Todd, and L. Tunçel. Monotonicity of primal and dual objective values in primal–dual interior–point algorithms. *SIAM J. Optim.*, 4:613–625, 1994.

[17] R. D. C. Monteiro. Primal-dual path-following algorithms for semidefinite programming. *SIAM J. Optim.*, 7:663–678, 1997.

[18] G. L. Nemhauser, A. H. G. Rinnooy Kan, and M. J. Todd, editors. *Optimization*, volume 1 of *Handbooks in Operations Research and Management Science*, North Holland. Amsterdam, The Netherlands, 1989.

[19] Yu. E. Nesterov and A. S. Nemirovskii. *Interior Point Polynomial Methods in Convex Programming: Theory and Algorithms*. SIAM Publications. SIAM, Philadelphia, USA, 1994.

[20] Yu. E. Nesterov and M. J. Todd. Self-scaled barriers and interior-point methods for convex programming. *Math. Oper. Res.*, 22:1–42, 1997.

[21] Yu. E. Nesterov and M. J. Todd. Primal-dual interior-point methods for self-scaled cones. *SIAM J. Optim.*, 8:324–364, 1998.

[22] Yu. E. Nesterov, M. J. Todd, and Y. Ye. Infeasible-start primal-dual methods and infeasibility detectors for nonlinear programming problems. *Math. Progr.*, 84:227–267, 1999.

[23] F. A. Potra. An infeasible interior–point predictor–corrector algorithm for linear programming. *SIAM J. Optim.*, 6:19–32, 1996.

[24] J. Renegar. *A Mathematical View of Interior-Point Methods in Convex Optimization*. SIAM Publications. SIAM, Philadelphia, USA, 2001.

[25] A. Schrijver. *Theory of Linear and Integer Programming*. John Wiley & Sons, New York, 1986.

[26] M. J. Todd. The many facets of linear programming. *Math. Progr.*, 91:417–436, 2002.

[27] M. J. Todd. Semidefinite optimization. *Acta Numer.*, 10:515–560, 2001.

[28] M. J. Todd and Y. Ye. Approximate Farkas lemmas and stopping rules for iterative infeasible-point algorithms for linear programming. *Math. Progr.*, 81:1–21, 1998.

[29] T. Tsuchiya. A convergence analysis of the scaling-invariant primal-dual path-following algorithms for second-order cone programming. *Optim. Meth. Soft.*, 11/12:141–182, 1999.

[30] R. H. Tütüncü. An infeasible-interior-point potential-reduction algorithm for linear programming. *Math. Progr.*, 86:313–334, 1999.

[31] R. H. Tütüncü, K. C. Toh, and M. J. Todd. Solving semidefinite-quadratic-linear programs using SDPT3. *Math. Progr.*, to appear (2003).

[32] M. H. Wright. Interior methods for constrained optimization. *Acta Numer.*, 1:341–407, 1992.

[33] S. J. Wright. *Primal-Dual Interior-Point Methods*. SIAM, Philadelphia, 1997.

[34] X. Xu, P. F. Hung, and Y. Ye. A simplified homogeneous and self-dual linear programming algorithm and its implementation. *Ann. Oper. Res.*, 62:151–171, 1996.

[35] Y. Ye, M. J. Todd, and S. Mizuno, An $O(\sqrt{n}L)$-iteration homogeneous and self-dual linear programming algorithm. *Math. Oper. Res.*, 19:53–67, 1994.

[36] Y. Zhang. On the convergence of a class of infeasible interior–point methods for the horizontal linear complementarity problem. *SIAM J. Optim.*, 4:208–227, 1994.