SCHOOL OF OPERATIONS RESEARCH
AND INDUSTRIAL ENGINEERING
COLLEGE OF ENGINEERING
CORNELL UNIVERSITY
ITHACA, NEW YORK 14853-3801

TECHNICAL REPORT NO. 978

August 1991
Revised September 1992

INTERIOR-POINT ALGORITHMS
FOR SEMI-INFINITE PROGRAMMING*

by

Michael J. Todd

*Research supported in part by NSF, AFORS and ONR through NSF grant DMS-8920550.

Abstract

In order to study the behavior of interior-point methods on very large-scale linear programming problems, we consider the application of such methods to continuous semi-infinite linear programming problems in both primal and dual form. By considering different discretizations of such problems we are led to a certain invariance property for (finite-dimensional) interior-point methods. We find that while many methods are invariant, several, including all those with the currently best complexity bound, are not. We then devise natural extensions of invariant methods to the semi-infinite case. Our motivation comes from our belief that for a method to work well on large-scale linear programming problems, it should be effective on fine discretizations of a semi-infinite problem and it should have a natural extension to the limiting semi-infinite case.

1. Introduction

This paper is motivated by the recent research of Ferris and Philpott [10,11] and Powell [37] on applying interior-point methods to semi-infinite linear programming problems. The former used the dual affine-scaling method of Adler, Resende, Veiga and Karmarkar [1] while the latter applied a natural generalization of Karmarkar's projective-scaling method [26], both for problems with a finite number of unconstrained variables subject to an infinite number of inequality constraints. In addition, Christiansen and Kortanek [8] applied the primal affine-scaling method of Dikin [9] to a discretization of an infinite problem.

Our aim here is to discuss what interior-point algorithms can naturally be extended to the semi-infinite setting, in the form above and its dual. Since interior-point methods are supposed to be efficient for large-scale problems, it is natural to consider the limit as one dimension -- n , the number of inequality constraints -- becomes infinite. We will not deal here with questions of convergence -- indeed, Powell [37] shows that his generalized Karmarkar algorithm can converge to a non-optimal point. We will show that affine-scaling and projective-scaling methods can be so extended, whereas most path-following methods cannot. Finally, some of the recent (primal, dual, or primal-dual) potential-reduction methods can be applied to these problems. These methods are described, for instance, in Goldfarb and Todd [17], Todd [45], and den Hertog and Roos [24]; the latter gives explicit rules for forming the search directions in dozens of interior-point methods, and we will use it extensively -- see also the earlier study of Gonzaga [20]. It is of some interest that none of the methods requiring $O(\sqrt{n} L)$ iterations for a standard-form problem with n variables and integer data of bit length L can be extended to the semi-infinite setting, while many $O(nL)$ -iteration algorithms can. We are not proposing our algorithms as practical methods of solution for semi-infinite linear programming problems; for one thing, our search directions cannot be computed analytically in most cases, only approximated. Rather, our interest is more theoretical; we seek insights into the behavior of interior-point methods for large-scale finite (n -dimensional) linear programming. Other methods for semi-infinite programming are described in, e.g., Anderson and Nash [3], Glashoff and Gustafson [16], and Hettich and Kortanek [25].

We will deal exclusively with the continuous semi-infinite linear programming problem defined by continuous functions $a: T \rightarrow \mathbb{R}^m$ and $\gamma: T \rightarrow \mathbb{R}$, and $b \in \mathbb{R}^m$. Here T is a compact subset of some $\mathbb{R}^p$ of measure 1. Then the semi-infinite problem in dual form is

$$\begin{aligned} \text{(SID)} \quad & \max_y b^T y \\ & a(t)^T y \leq \gamma(t), \quad \text{for all } t \in T. \end{aligned}$$

We will use Greek lower-case letters for both scalars and scalar-valued functions. Operations and relations will usually be considered pointwise, so that the constraints of (SID) can be written as $a^T y \leq \gamma$; here $a^T y$ denotes the function on T with $(a^T y)(t) = a(t)^T y$, and the inequality holds for all $t \in T$. If ξ and σ are two continuous real-valued functions on T ($\xi, \sigma \in C(T)$), we write

$$\langle \xi, \sigma \rangle := \int_T \xi(t) \sigma(t) dt,$$

while ξ^2 and $\xi\sigma^{-1}$ denote the functions with $(\xi^2)(t) := (\xi(t))^2$ and $(\xi\sigma^{-1})(t) := \xi(t)/\sigma(t)$.

We often require the slack function σ for (SID). Using the conventions above, we can then rewrite (SID) in the form

$$\begin{aligned} \text{(SID)} \quad & \max_{y, \sigma} b^T y \\ & a^T y + \sigma = \gamma, \\ & \sigma \geq 0. \end{aligned}$$

Note that, since γ and $a^T y$ lie in $C(T)$ for any y , we have $\sigma \in C(T)$ also. As discussed in Anderson and Nash [3, Section 4.4], the appropriate space for the problem dual to (SID) is the space $M_r[0,1]$ of regular Borel measures on $[0,1]$. However, to discuss interior-point methods, we will find it convenient to restrict ourselves to measures with continuous density functions $\xi(t)$. We therefore consider the semi-infinite problem in primal form as

$$(SIP) \quad \min \langle \gamma, \xi \rangle$$

$$\int_T a(t)\xi(t)dt = b$$

$$\xi \in C(T), \quad \xi(t) \geq 0;$$

this is a restriction of the problem dual to (SID). We also write $\int_T a(t)\xi(t)dt$ simply as $\int_T a\xi$, etc., or as $\int a\xi$; integrals are always over T unless otherwise indicated. In this space, there may be no optimal solution, as simple examples demonstrate.

We will derive algorithms for (SID) and (SIP) by considering the related finite linear programming problems

$$(D) \quad \begin{aligned} \max \quad & b^T y \\ & A^T y + s = c, \\ & s \geq 0, \end{aligned}$$

and

$$(P) \quad \begin{aligned} \min \quad & c^T x \\ & Ax = b, \\ & x \geq 0, \end{aligned}$$

which can be viewed as discretizations of (SID) and (SIP) respectively. Here A is $m \times n$ and c $n \times 1$. Think of A as $[a(t_1), \dots, a(t_n)]$ and c as $[\gamma(t_1), \dots, \gamma(t_n)]^T$, where t_j represents the j th region in a partition of T into n regions; for example, if $T = [0, 1]$, the j th region can be $[(j-1)/n, j/n]$, represented by $t_j = (2j-1)/2n$. Then the slack s_j in the j th constraint of (D) is $\sigma(t_j)$, while x_j in (P) corresponds to the integral of ξ over the j th region. We usually think of each region as having measure $1/n$. Conversely, (D) and (P) can be represented as semi-infinite problems. Let T be the union of n closed disjoint intervals of length $1/n$ each on the real line, and let $a(t)$ be the j th column of A and $\gamma(t)$ be c_j for t in the j th interval. Then (D) and (P) are equivalent to (SID) and (SIP).

We can now describe in a little more detail the motivation for this study. Interior-point methods seem to be very effective on very large-scale problems -- see for instance Bixby et al. [6] and Lustig et al. [32]. On the other hand, theoretical worst case bounds on the number of iterations required grow as either $\sqrt{n}$ or n , where n is the number of inequality constraints. In practice, the growth seems much smaller -- the most convincing documentation comes from [32], where several subproblems obtained by taking subsets of the columns of a 13 million variable problem (with 8 million distinct variables) are solved, and growth of the number of iterations roughly as $\log n$ is observed. (Perhaps n is the “wrong” parameter in terms of which to estimate the complexity; to find the “right” parameter we remove n from consideration by studying semi-infinite problems.) In attacking the discrepancy either empirically or theoretically, it is clearly useful to have a way to generate “homogeneous” problems of varying n . One method, as described above, is to take “chunks” of a fixed, huge problem. We study another natural avenue; consider discretizations of a given semi-infinite programming problem. Here we might expect (or hope for), after a certain point, no growth in the number of iterations with the fineness of the discretization; we would hope that the finite-dimensional algorithm is mimicking the performance of some limiting algorithm on the semi-infinite problem. In this context, the invariance of the algorithm under the “duplication” operation discussed below and in section 3 seems a natural question. Closely related is the mesh-independence concept of Allgower et al. [2].

This avenue has already been explored by Powell, who shows that these expectations may well be unfounded; indeed, as already noted, a natural generalization of Karmarkar’s algorithm converges to a non-optimal point for a particular semi-infinite programming problem [37], while the number of iterations required for discretizations increases linearly with n [38,39]. We consider Powell’s example further in section 4.

As we mentioned above, this paper contains no convergence results for our algorithms. Nevertheless, our goals are to replace the parameter n in bounds with a measure of the “roughness” (inverse of “smoothness”) of the problem data. First, this roughness might determine the discretization

necessary before the invariance property applies, and hence affects the number of iterations to solve finite-dimensional approximations. Second, the dependence on n arises in complexity analyses in the finite-dimensional case from relations between the 1-, 2-, and ∞ -norm of various vectors. Loosely speaking, an average value (2-norm) of the components must be large if any value (∞ -norm) is. This can be translated directly into bounds for piecewise-constant functions, constant on pieces of measure $1/n$. In the general infinite-dimensional case, $\int_T \xi^2 (= \|\xi\|_2^2)$ must be large if $\max\{|\xi(t)|: t \in T\}$ ($=: \|\xi\|_\infty$) is, if ξ is smooth, since then ξ^2 is large for a subset of significant measure. It is far from clear how precise bounds can be obtained in general, but we feel that this is an important conceptual framework. The follow-up paper of Tuncel and Todd [48] provides such an analysis under certain restrictions.

Of course, other aspects of the problem besides the smoothness of the functions may have a significant aspect on convergence properties of an interior-point algorithm. In the finite case, nondegeneracy plays an important part, and the related notion of strong uniqueness in the semi-infinite case may have a significant effect. Indeed, this was observed in the computational results in [11].

In section 2 we discuss briefly a number of interior-point methods for the finite problems (P) and (D). Section 3 considers the invariance of these methods under the following simple operation. Suppose (D) arises from (SID) by discretizing as above using n values $t_1, \dots, t_n$ in $T = [0, 1]$. If instead we discretized using $2n$ values of t , the resulting problem (using the continuity of $a(\cdot)$ and $\gamma(\cdot)$) would be very close to

$$\begin{aligned}
 (D_d) \quad & \max b^T y \\
 & A^T y \leq c, \\
 & A^T y \leq c,
 \end{aligned}$$

where the subscript “d” refers to duplication of the constraints. The dual of (D_d) is

$$\begin{aligned}
(P_d) \quad & \min c^T x' + c^T x'' \\
& Ax' + Ax'' = b \\
& x', x'' \geq 0,
\end{aligned}$$

resulting from “splitting” the primal variables x . Note that (P_d) is an approximation to a discretization of (SIP) using $2n$ values of t ; if $T = [0,1]$ think of x'_j and x''_j as corresponding to integrals of $\xi(t)$ over the intervals $[(2j-2)/2n, (2j-1)/2n]$ and $[(2j-1)/2n, 2j/2n]$ respectively. From the continuity of $\xi(\cdot)$, we consider as corresponding to the feasible solution x of (P), the feasible solution (x', x'') of (P_d) with $x'_j := x''_j := \frac{1}{2} x_j$. A similar interpretation for the operation taking (P) and (D) to (P_d) and (D_d) can be made if T is higher-dimensional. Finally, if we construct from (P) and (D) instances of (SIP) and (SID) with piecewise-constant functions a and γ , as described above, and then discretize the resulting problems using $2n$ values of t , two in each of the small intervals, then we get exactly (P_d) and (D_d) .

We shall see that primal and dual affine-scaling and projective-scaling algorithms, as well as some large-step path-following and potential-reduction methods, are invariant under this operation. However, no algorithm using the ℓ_2 -norm as a centering condition or as a signal to update lower bounds, and none using a parameter $n + \nu\sqrt{n}$ for constant ν (e.g. many of the potential-reduction methods) are so invariant.

Finally, in section 4 we describe extensions of several interior-point algorithms to (SID) and (SIP), and section 5 contains some concluding remarks.

2. Interior-point algorithms

Here we discuss several algorithms for the finite linear programming problems (P) and (D). These can roughly be classified as falling into one of the following four categories:

(i) affine-scaling methods, first developed by Dikin [9] in 1967, and rediscovered by several authors including Adler, Resende, Veiga and Karmarkar [1], Barnes [5], Cavalier and Soyster [7], Kortanek and Shi [30], and Vanderbei, Meketon and Freedman [50];

(ii) projective-scaling methods, starting with Karmarkar's seminal contribution [26], with extensions by Anstreicher [4], Gay [14], de Ghellinck and Vial [15], Gonzaga [18,20], Jensen and Steger [43], Todd and Burrell [46], Yamashita [51], and Ye and Kojima [53], among others;

(iii) path-following methods, starting with Renegar [40], with contributions by Gonzaga [19], Kojima, Mizuno and Yoshise [27,28], Monteiro and Adler [35,36], Roos and Vial [41], Sonnevend, Stoer and Zhao [42], and Vaidya [49]; and

(iv) affine potential-reduction methods, introduced by Gonzaga [22]; see also Freund [13], Gonzaga [23], Kojima, Mizuno and Yoshise [29], and Ye [52].

For a compendious bibliography of these methods, see Kranich [31].

We discuss some of these methods first for the standard form problem (P). Corresponding methods for (D) can be derived by the following general technique. First, eliminate the free variables y : if F is a matrix whose rows span the null space of A , $A^T y + s = c$ for some y iff $Fs = Fc$. Next, assuming $A\hat{x} = b$ for some (not necessarily nonnegative) $\hat{x}$, the objective function can be written as $c^T \hat{x} - \hat{x}^T s$. Hence (D) is equivalent to

$$\begin{array}{ll}
 \min & \hat{x}^T s \\
 (D') & \\
 & Fs = Fc \\
 & s \geq 0,
 \end{array}$$

which is in standard form. Now we can apply the primal form of the algorithm to (D'), and then reexpress the iterates in terms of the variables y . (For an explicit description of this process, see for example the analysis of the author leading to the dual affine-scaling algorithm that appears in Monma and Morton [34].)

Let Λ be a positive definite diagonal matrix, and consider the changes to (P) and (D) when we make the change of variables $x \rightarrow \Lambda^{-1}x$. Then $c \rightarrow \Lambda c$, $A \rightarrow A\Lambda$, and $s \rightarrow \Lambda s$. Given $\hat{x} \in F_+(P) := \{x: Ax = b, x > 0\}$ and/or $\hat{s} \in F_+(D) := \{s: \exists y, A^T y + s = c, s > 0\}$, we may choose Λ to transform $\hat{x}$ to $e := (1, 1, \dots, 1)^T$ ($\Lambda = \text{diag}(\hat{x})$, primal scaling), to transform $\hat{s}$ to e ($\Lambda = (\text{diag}(\hat{s}))^{-1}$, dual scaling), or to transform both $\hat{x}$ and $\hat{s}$ to $\hat{v}$ with $\hat{v}_j = (\hat{x}_j \hat{s}_j)^{\frac{1}{2}}$ ($\Lambda = (\text{diag}(\hat{x}))^{\frac{1}{2}} (\text{diag}(\hat{s}))^{-\frac{1}{2}}$, symmetric primal-dual scaling; think of $\hat{v}$ as an approximation to e). Almost all interior-point methods amount to first scaling the problem as above, then computing (in the transformed space) the gradient of a suitable potential function, next taking as a search direction (in the transformed space) the projection into the appropriate subspace of the negative of this gradient, and finally scaling this direction back to the original space.

As our first example, suppose we use primal scaling and the objective function as the potential function. The scaled problem is then

$$\begin{aligned} \min \bar{c}^T \bar{x} \\ \bar{A} \bar{x} &= b \\ \bar{x} &\geq 0 \end{aligned}$$

in terms of the scaled variables $\bar{x} = \hat{X}^{-1}x$, where $\bar{c} := \hat{X}c$ and $\bar{A} := A\hat{X}$; here $\hat{X}$ denotes $\text{diag}(\hat{x})$. In the transformed space, the current iterate is e and the gradient of the objective function $\bar{c}$. The negative gradient, projected into the null space of the constraint matrix $\bar{A}$, is then $-P_{\bar{A}} \bar{c} = -P_{A\hat{X}} \hat{X}c$, where P_M denotes the orthogonal projector into the null space of M . Transforming this search direction back to the original space yields $-\hat{X}P_{A\hat{X}} \hat{X}c$. For typographic convenience, we assume the current iterate is x rather than $\hat{x}$. Hence we have the affine-scaling direction

$$P_{\text{AFF}} := -XP_{AX}Xc. \quad (1)$$

We will assume throughout that $\text{rank}(A) = m$, so that we can alternatively write

$$p_{\text{AFF}} = -X(I - XA^T(AX^2A^T)^{-1}AX)Xc. \quad (2)$$

Besides decreasing the cost, interior-point methods seek to stay far away from the boundary of the feasible region. This motivates the other key direction that arises in several algorithms, the centering direction which comes from using the centering function $-\ell_n(x) := -\sum_j \ell_n(x_j)$ as the “potential function.” Note that the transformed function $f(\bar{x}) := -\ell_n(X\bar{x}) = -\ell_n(x) - \ell_n(\bar{x})$ has $-e$ as its gradient at the transformed current iterate $\bar{x} = e$, so we get the direction

$$\begin{aligned} p_{\text{CEN}} &:= XP_{AX}e \\ &= X(I - XA^T(AX^2A^T)^{-1}AX)e. \end{aligned} \quad (3)$$

If we use the technique described above (using the equivalence of (D) and (D')) to develop corresponding directions for the dual problem, we obtain the dual affine-scaling direction

$$d_{\text{AFF}} := (AS^{-2}A^T)^{-1}b \quad (4)$$

where (y, s) is our current iterate, $s > 0$, and $S = \text{diag}(s)$, and the dual centering direction

$$d_{\text{CEN}} := -(AS^{-2}A^T)^{-1}AS^{-1}e. \quad (5)$$

Finally, if we employ primal-dual scaling, so that $\Lambda = X^{\frac{1}{2}}S^{-\frac{1}{2}}$ where x and (y, s) are our current primal and dual iterates, we have

$$\begin{aligned} p'_{\text{AFF}} &:= -\Lambda P_{\Lambda\Lambda}\Lambda c \\ &= -\Lambda P_{\Lambda\Lambda}X^{\frac{1}{2}}S^{\frac{1}{2}}e \\ &= -\Lambda(I - \Lambda A^T(\Lambda A^2A^T)^{-1}\Lambda\Lambda)X^{\frac{1}{2}}S^{\frac{1}{2}}e, \end{aligned} \quad (6)$$

$$\begin{aligned}
p'_{\text{CEN}} &:= \Lambda P_{\Lambda\Lambda} X^{-\frac{1}{2}} S^{-\frac{1}{2}} e, \\
&= \Lambda(I - \Lambda\Lambda^\top (\Lambda\Lambda^2\Lambda^\top)^{-1} \Lambda\Lambda) X^{-\frac{1}{2}} S^{-\frac{1}{2}} e
\end{aligned} \tag{7}$$

$$d'_{\text{AFF}} := (\Lambda\Lambda^2\Lambda^\top)^{-1} b, \text{ and} \tag{8}$$

$$d'_{\text{CEN}} := -(\Lambda\Lambda^2\Lambda^\top)^{-1} \Lambda S^{-1} e. \tag{9}$$

The derivation of these directions is similar. Note that, since (y, s) is dual feasible, Λc and $X^{\frac{1}{2}} S^{\frac{1}{2}} e = \Lambda s$ differ by a vector in the range space of $\Lambda\Lambda^\top$, so their projections into the null space of $\Lambda\Lambda$ coincide.

To our knowledge, all (but one) interior-point algorithms that iterate feasible primal and/or dual solutions use search directions that are combinations of these, to effect a compromise between improving the objective ($p_{\text{AFF}}, d_{\text{AFF}}, p'_{\text{AFF}}$ or d'_{AFF}) and staying away from the boundary of the feasible region (p_{CEN} , etc). However, the way these methods are derived can be very different, so we only provide motivation for the particular combination chosen in a few cases. In others, the fact that the search direction is of this form requires some analysis. As our first example, the primal affine-scaling algorithm uses p_{AFF} , while the dual affine-scaling algorithm uses d_{AFF} .

Before describing several other search directions, we list the merit or potential functions they are based on. Let z_* denote the optimal value of (P) and (D) and z_ℓ and z_u denote lower and upper bounds on it. Let $\mu > 0$ denote a barrier parameter and $\rho \geq n$ a parameter to be used in logarithmic potential functions. First we have the objective functions

$$c^\top x \text{ and } b^\top y. \tag{10}$$

Then, consider the primal or dual potential functions

$$\rho \ln(c^\top x - z_\ell) - \ln(x) \text{ and } \rho \ln(z_u - b^\top y) - \ln(s) \tag{11}$$

which combine a monotonic function of the optimality gap $c^\top x - z_\ell$ or $z_u - b^\top y$ with the

centering function. The classic example is Karmarkar's potential function [26] which used $z_\ell = z_*$ and $\rho = n$. Standard-form variants of the projective-scaling algorithm use such functions with $z_\ell \leq z_*$ or $z_u \geq z_*$ and either $\rho = n$ (Gonzaga [18]) or $\rho = n+1$ (Anstreicher [4], Gay [14], de Ghellinck and Vial [15], Steger [43], Yamashita [51] and Ye and Kojima [53]); the extra “1” comes from the addition of a homogenizing variable. Finally, potential functions (11) are used by affine potential-reduction methods with $\rho \geq n + \sqrt{n}$ (Gonzaga [22,23], Freund [13]). Next we have the primal or dual centering functions (with objective bounds)

$$-n \ln(z_u - c^T x) - \ln(x) \text{ and } -n \ln(b^T y - z_\ell) - \ln(s) \quad (12)$$

of Renegar [34] and Vaidya [49]; and the classical primal or dual barrier functions

$$c^T x - \mu \ln(x) \text{ and } b^T y + \mu \ln(s) \quad (13)$$

used, e.g., by Gonzaga [19] and Roos and Vial [41]. In the primal-dual context we have the duality gap

$$c^T x - b^T y = x^T s \quad (14)$$

and the primal-dual potential function

$$\rho \ln(x^T s) - \ln(x) - \ln(s) \quad (15)$$

of Tanabe [44] and Todd and Ye [47]. For primal-dual path-following methods we have the centering measures

$$\|\frac{Xs}{\mu} - e\|_2 \text{ and } \|\frac{Xs}{\mu} - e\|_\infty \quad (16)$$

used by Kojima, Mizuno and Yoshise ([27] with ℓ_∞ -norm, [28] with ℓ_2 -norm), Monteiro and Adler [35, 36] (ℓ_2 -norm) and Mizuno, Todd and Ye [33] (both norms); these measure the distance to the central trajectory, the set of feasible pairs with $Xs = \mu e$ for some $\mu > 0$.

For primal methods, a norm of the projection of the potential function gradient in the transformed space is often used, hence

$$\| \frac{\rho}{c^T x - z_\ell} X^{-1} p_{\text{AFF}} + X^{-1} p_{\text{CEN}} \|_2 \text{ or } \| \frac{\rho}{c^T x - z_\ell} X^{-1} p_{\text{AFF}} + X^{-1} p_{\text{CEN}} \|_\infty; \quad (17)$$

the first is used in Gonzaga [22, 23], Freund [13], Ye [52], and implicitly in Gonzaga [19] and Roos and Vial [41] for certain values of ρ and z_ℓ , and the second in Gonzaga [23].

Now we describe the search directions of various methods -- derivations of most may be found in den Hertog and Roos [24].

The primal standard-form projective method uses the direction

$$p_{\text{PRO}} := p_{\text{AFF}} + \nu_{\text{PRO}} p_{\text{CEN}}, \quad (18)$$

where

$$\nu_{\text{PRO}} := \frac{c^T x - c^T p_{\text{CEN}} - z_\ell}{(1+b)^T (AX^2 A^T)^{-1} b}, \quad (19)$$

and where the “1+” arises in the variants [4, 14, 15, 43, 53] (who use a reformulation as a homogeneous problem with one extra variable) and not in Gonzaga’s variant [18]. Here z_ℓ can be z_* if it is known, and otherwise is updated as follows. Let ν_+ solve

$$\begin{aligned} \min_{\nu} \quad & \nu \\ \text{Xs}(\nu) := P_{\text{AX}} Xc + (e - P_{\text{AX}} e)\nu & \geq 0 \end{aligned} \quad (20)$$

($= \infty$ if the problem is infeasible), and hence obtain

$$\begin{aligned} z_+ &:= c^T x - e^T (P_{\text{AX}} Xc + \nu_+ (e - P_{\text{AX}} e)) \\ &= c^T x - c^T p_{\text{CEN}} - \nu_+ \|e - P_{\text{AX}} e\|^2. \end{aligned} \quad (21)$$

Then set $z_\ell \leftarrow \max\{z_+, z_\ell\}$, where on the right hand side we use the previous value of z_ℓ . By (20), for any ν , $\text{Xs}(\nu)$ differs from Xc by a vector in the range of XA^T . It follows that $\text{s}(\nu_+)$ is a

feasible dual slack, so that z_+ is a valid lower bound. In fact, equation (20) seems to be the most general way to obtain feasible dual slacks from the computation of p_{AFF} and p_{CEN} , and then (21) suggests that the way to get the highest lower bound is to choose the smallest ν with $s(\nu)$ feasible, motivating the choice of ν_+ . Note that $z_\ell = z_+$ implies $\nu_{\text{PRO}} = \nu_+$ (if the “1+” is omitted), since $b^\top (AX^2 A^\top)^{-1} b = e^\top X A^\top (AX^2 A^\top)^{-1} A X e = e^\top (I - P_{AX}) e = \|e - P_{AX} e\|^2$.

For the dual projective-scaling algorithm we have the direction

$$d_{\text{PRO}} := d_{\text{AFF}} + \hat{\nu}_{\text{PRO}} d_{\text{CEN}}, \quad (22)$$

where now

$$\hat{\nu}_{\text{PRO}} := \frac{z_u - b^\top y + b^\top d_{\text{CEN}}}{(1 +) \|P_{AS^{-1}} e\|^2}, \quad (23)$$

where the “1+” occurs in the variants of Gay [14], Gonzaga ([21], section 3.3), and Yamashita [51] and not in that of Freund [12] and Gonzaga ([21], section 3.2). The upper bound z_u is now updated as follows. Let $\hat{\nu}_+$ solve

$$\min_{\hat{\nu}} Sx(\hat{\nu}) := S^{-1} A^\top d_{\text{AFF}} + (e + S^{-1} A^\top d_{\text{CEN}}) \hat{\nu} \geq 0 \quad (24)$$

($=\infty$ if the problem is infeasible), hence obtain

$$\begin{aligned} \hat{z}_+ &:= b^\top y + e^\top (S^{-1} A^\top d_{\text{AFF}} + \hat{\nu}_+ (e + S^{-1} A^\top d_{\text{CEN}})) \\ &= b^\top y - b^\top d_{\text{CEN}} + \hat{\nu}_+ \|P_{AS^{-1}} e\|^2, \end{aligned} \quad (25)$$

and set $z_u = \min\{\hat{z}_+, z_u\}$. Here $x(\hat{\nu}_+)$ is a feasible primal solution, and if $z_u = \hat{z}_+$, $\nu_{\text{PRO}} = \hat{\nu}_+$ (if the “1+” is omitted).

For the method of centers of Renegar [40] and Vaidya [49], in the dual form, the direction is

$$d_{\text{REN}} := d_{\text{AFF}} + \nu_{\text{REN}} d_{\text{CEN}} \quad (26)$$

where

$$\nu_{\text{REN}} := \frac{(\mathbf{b}^\top \mathbf{y} - z_\ell)^2}{\mathbf{n}} + \mathbf{b}^\top \mathbf{d}_{\text{AFF}}, \quad (27)$$

where now the lower bound z_ℓ is explicitly adjusted via

$$z_\ell \leftarrow \left(1 - \frac{\beta}{\sqrt{\mathbf{n}}}\right) z_\ell + \frac{\beta}{\sqrt{\mathbf{n}}} \mathbf{b}^\top \mathbf{y} \quad (28)$$

for some constant β . There is also a corresponding primal form.

For other path-following methods we have the directions

$$\mathbf{p}_{\text{PAT}} := \mathbf{p}_{\text{AFF}} + \mu \mathbf{p}_{\text{CEN}} \quad (29)$$

and

$$\begin{pmatrix} \mathbf{p}'_{\text{PAT}} \\ \mathbf{d}'_{\text{PAT}} \end{pmatrix} := \begin{pmatrix} \mathbf{p}'_{\text{AFF}} \\ \mathbf{d}'_{\text{AFF}} \end{pmatrix} + \mu \begin{pmatrix} \mathbf{p}'_{\text{CEN}} \\ \mathbf{d}'_{\text{CEN}} \end{pmatrix} \quad (30)$$

where μ is explicitly adjusted according to

$$\mu \leftarrow \gamma \mu \quad (31)$$

([27, 32, 33]) or

$$\mu \leftarrow \left(1 - \frac{\beta}{\sqrt{\mathbf{n}}}\right) \mu \quad (32)$$

[19, 28, 35, 36, 41]. (Here, $\beta > 0$ and $\gamma \in (0,1)$ are constants.) These directions arise as the scaled projected negative gradients of the barrier functions in (13).

Finally, for the potential-reduction methods, we have first the direction

$$\mathbf{p}_{\text{PRP}} := \kappa_{\text{PRP}} \mathbf{p}_{\text{AFF}} + \mathbf{p}_{\text{CEN}} \quad (33)$$

with

$$\kappa_{\text{PRP}} := \frac{\rho}{\mathbf{c}^\top \mathbf{x} - z_\ell} \quad (34)$$

and various means of updating z_ℓ in the primal algorithms of Gonzaga [22,23], Freund [13] and Ye

[52, section 6]. This comes from the gradient of the primal potential function in (11). Closely related is the direction

$$p_{\text{PRY}} := \kappa_{\text{PRY}} p_{\text{AFF}} + p_{\text{CEN}} \quad (35)$$

with

$$\kappa_{\text{PRY}} := \frac{\rho}{x^T s} \quad (36)$$

and a technique to update s , used in the primal algorithm of Ye [52; section 3]. In the primal-dual setting, we have the direction

$$\begin{pmatrix} p'_{\text{PRPD}} \\ d'_{\text{PRPD}} \end{pmatrix} := \kappa_{\text{PRPD}} \begin{pmatrix} p'_{\text{AFF}} \\ d'_{\text{AFF}} \end{pmatrix} + \begin{pmatrix} p'_{\text{CEN}} \\ d'_{\text{CEN}} \end{pmatrix} \quad (37)$$

with

$$\kappa_{\text{PRPD}} := \frac{\rho}{x^T s}, \quad (38)$$

used in the primal-dual potential reduction method of Kojima, Mizuno and Yoshise [29]. The directions (35) and (37) come from the gradient of the primal-dual potential function (15).

3. Invariance

Here we consider how the algorithms outlined in the previous section change if (P) or (D) is replaced by

$$\begin{aligned} (P_d) \quad & \min c^T x' + c^T x'' \\ & Ax' + Ax'' = b \\ & x', x'' \geq 0 \end{aligned}$$

or

$$\begin{aligned} (D_d) \quad & \max b^T y \\ & A^T y + s_1 = c \\ & A^T y + s_2 = c \\ & s_1, s_2 \geq 0, \end{aligned}$$

corresponding to finer discretizations of (SIP) and (SID). We assume that we have a current solution $x \in F_+(P)$ and/or $s \in F_+(D)$, and we consider the corresponding solutions $(x', x'') = (\frac{1}{2}x, \frac{1}{2}x) \in F_+(P_d)$ and $(s_1, s_2) = (s, s) \in F_+(D_d)$. (In fact, the results of this section remain valid with appropriate changes to the proofs if the variables of (P) and the constraints of (D) are replicated k times, where k is any positive integer.)

An algorithm is invariant if, given corresponding starting points, it generates corresponding iterates when applied to (P) and (P_d) or to (D) and (D_d) .

We say a quantity is invariant if it is the same (or duplicated if necessary, for vector and matrix quantities) when calculated at the current solution of (P) and/or (D) using that data or at the current solution of (P_d) and/or (D_d) using the augmented data. Hence, for instance, nx is invariant since $2n(x', x'') = 2n(\frac{1}{2}x, \frac{1}{2}x) = (nx, nx)$. So are y and s .

Proposition 3.1. The following directions are invariant: $n^2 p_{\text{AFF}}, np_{\text{CEN}}, nd_{\text{AFF}}, d_{\text{CEN}}, np'_{\text{AFF}}, p'_{\text{CEN}}, d'_{\text{AFF}}$, and $n^{-1}d'_{\text{CEN}}$.

Proof. In the system (P_d) , the projection matrix P_{AX} becomes

$$\begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix} - \begin{pmatrix} \frac{1}{2} XA^T \\ \frac{1}{2} XA^T \end{pmatrix} (A(\frac{1}{4} X^2)A^T + A(\frac{1}{4} X^2)A^T)^{-1} (\frac{1}{2} AX \frac{1}{2} AX) \\ = \begin{pmatrix} I - \frac{1}{2} XA^T(AX^2A^T)^{-1}AX & -\frac{1}{2} XA^T(AX^2A^T)^{-1}AX \\ -\frac{1}{2} XA^T(AX^2A^T)^{-1}AX & I - \frac{1}{2} XA^T(AX^2A^T)^{-1}AX \end{pmatrix}.$$

Hence, if v is an invariant n -vector, then it becomes $\begin{pmatrix} v \\ v \end{pmatrix}$ in the larger system, and so $P_{\text{AX}}v$ is also invariant.

Thus the invariance of nXc implies that $nP_{\text{AX}}Xc$ is invariant, and hence so is $-[nX][nP_{\text{AX}}Xc] = n^2 p_{\text{AFF}}$. Similarly, $np_{\text{CEN}} = [nX][P_{\text{AX}}e]$ is invariant. (We frequently use square brackets in this way to show invariant factors.)

Next, in the system (D_d) , the matrix $AS^{-2}A^\top$ becomes $(A \ A) \begin{pmatrix} S^{-2} & 0 \\ 0 & S^{-2} \end{pmatrix} \begin{pmatrix} A^\top \\ A^\top \end{pmatrix} = 2AS^{-2}A^\top$, so $n(AS^{-2}A^\top)^{-1}$ is invariant. Hence so are $nd_{\text{AFF}} = [n(AS^{-2}A^\top)^{-1}]b$ and $d_{\text{CEN}} = -[n(AS^{-2}A^\top)^{-1}][n^{-1}AS^{-1}e]$, since $n^{-1}AS^{-1}e$ is easily seen to be invariant.

Finally, for $\Lambda = X^{\frac{1}{2}}S^{-\frac{1}{2}}$, we easily obtain that, if v is invariant, so are $P_{A\Lambda}v$, $n^{\frac{1}{2}}\Lambda v$ and $n^{-\frac{1}{2}}X^{-\frac{1}{2}}S^{-\frac{1}{2}}v$, and that $(\Lambda\Lambda^2A^\top)^{-1}$ is invariant. Hence so are $np'_{\text{AFF}} = -[n^{\frac{1}{2}}\Lambda][P_{A\Lambda}n^{\frac{1}{2}}\Lambda c]$, $p'_{\text{CEN}} = [n^{\frac{1}{2}}\Lambda][P_{A\Lambda}n^{-\frac{1}{2}}X^{-\frac{1}{2}}S^{-\frac{1}{2}}e]$, $d'_{\text{AFF}} = [(\Lambda\Lambda^2A^\top)^{-1}][b]$, and $n^{-1}d'_{\text{CEN}} = -[(\Lambda\Lambda^2A^\top)^{-1}][n^{-1}AS^{-1}e]$. $\square$

Next we turn to potential functions and measures of centrality.

Proposition 3.2. The following are invariant:

- (i) $c^\top x$ and $b^\top y$;
- (ii) if z_ℓ and z_u and ρ/n are invariant, $n^{-1}(\rho \ell n(c^\top x - z_\ell) - \ell n(x) - n \ell n(n))$ and $n^{-1}(\rho \ell n(z_u - b^\top y) - \ell n(s))$;
- (iii) if z_ℓ and z_u are invariant, $n^{-1}(-n \ell n(z_u - c^\top x) - \ell n(x) - n \ell n(n))$ and $n^{-1}(-n \ell n(b^\top y - z_\ell) - \ell n(s))$;
- (iv) if $n\mu$ is invariant, $c^\top x - \mu \ell n(x) - n\mu \ell n(n)$ and $b^\top y + \mu \ell n(s)$;
- (v) $x^\top s$;
- (vi) if ρ/n is invariant, $n^{-1}(\rho \ell n(x^\top s) - \ell n(x) - \ell n(s) - n \ell n(n))$;
- (vii) if $n\mu$ is invariant, $n^{-\frac{1}{2}}\|\frac{Xs}{\mu} - e\|_2$ and $\|\frac{Xs}{\mu} - e\|_\infty$; and
- (viii) if ρ/n and z_ℓ are invariant, $n^{-\frac{1}{2}}\|\frac{\rho}{c^\top x - z_\ell} X^{-1}p_{\text{AFF}} + X^{-1}p_{\text{CEN}}\|_2$ and $\|\frac{\rho}{c^\top x - z_\ell} X^{-1}p_{\text{AFF}} + X^{-1}p_{\text{CEN}}\|_\infty$.

Proof. The following remarks help to show that these are all straightforward. We note that if v is an invariant n -vector, then $\ell n(v) := \sum_j \ell n(v_j)$ is not invariant, but $n^{-1}\ell n(v)$ is. Also, nx is invariant, hence so is $n^{-1}\ell n(nx) = n^{-1}(\ell n(x) + n \ell n(n))$. Next, if v is invariant, then so are

$n^{-\frac{1}{2}}\|v\|_2$ (since $\|(v;v)\|_2 = \sqrt{2} \|v\|_2$) and $\|v\|_\infty$. Assembling these rules gives the desired invariance; for example, we can write the vector in (viii) as

$$\left[\frac{\rho/n}{c^T x - z_\ell} \right] [(nX)^{-1}][n^2 p_{\text{AFF}}] + [(nX)^{-1}][np_{\text{CEN}}],$$

which is invariant. $\square$

We now examine the algorithms of section 2. We will not deal with the issue of step sizes; however, the standard choices (a fixed proportion of the way to the boundary, or to minimize a suitable potential function) lead to invariant algorithms, as long as ρ/n is invariant in methods using a potential function.

First, proposition 3.1 shows that the affine-scaling algorithm is invariant. We write the iteration

$$x^+ = x + \alpha p_{\text{AFF}}$$

in the invariant form

$$[nx^+] = [nx] + [\alpha/n][n^2 p_{\text{AFF}}],$$

from which it is clear that α/n should be invariant. For the dual variant, we have

$$[y^+] = [y] + [\alpha/n][nd_{\text{AFF}}].$$

For algorithms that use a combination p of the directions p_{AFF} and p_{CEN} , or a combination d of d_{AFF} and d_{CEN} , it is clearly enough to have $n^2 p$ or np invariant, or nd or d invariant. Hence we must examine the linear combination used and compare with proposition 3.1.

Consider first the primal projective-scaling method. If z_ℓ is invariant, then so is the numerator of ν_{PRO} in (19). It is easily seen that nAX^2A^T is invariant, so $n\nu_{\text{PRO}}$ is invariant if the “1+” is omitted. Then we find

$$n^2 p_{\text{PRO}} = [n^2 p_{\text{AFF}}] + [n\nu_{\text{PRO}}][np_{\text{CEN}}]$$

is invariant. Now $z_\ell = z_*$ is clearly invariant. Suppose we use the updating method of (20), (21). From the proof of Proposition 3.1, $nP_{AX}Xc$ is invariant, as is $e - P_{AX}e$ and $n^{-\frac{1}{2}}\|e - P_{AX}e\|$. Hence $n\nu_+$ is invariant, as is $[n\nu_+][n^{-\frac{1}{2}}\|e - P_{AX}e\|]^2$, so z_+ and thus z_ℓ are invariant (assuming the initial value of z_ℓ chosen is invariant).

Now turning to the dual projective-scaling method, we find by analogous arguments that nd_{PRO} is invariant if $n\hat{\nu}_{\text{PRO}}$ is (see (22)-(23)), and this follows if the “1+” is omitted, and if $z_u = z_*$ or z_u is updated as in (24)-(25).

Next we consider path-following methods. For the method of centers, (27) and Propositions 3.1 and 3.2 show that $n\nu_{\text{REN}}$ and hence nd_{REN} are invariant as long as z_ℓ is. But for this to hold we cannot use the theoretically attractive update formula (28) but must instead use the bolder choice

$$z_\ell \leftarrow (1 - \hat{\beta})z_\ell + \hat{\beta}b^T y$$

where $\hat{\beta} \in (0,1)$ is invariant. For primal path-following methods we find from (29) that $n^2 p_{\text{PAT}}$ is invariant if $n\mu$ is, and then so is

$$\begin{pmatrix} np'_{\text{PAT}} \\ d'_{\text{PAT}} \end{pmatrix} = \begin{pmatrix} np'_{\text{AFF}} \\ d'_{\text{AFF}} \end{pmatrix} + [n\mu] \begin{pmatrix} p'_{\text{CEN}} \\ n^{-1}d'_{\text{CEN}} \end{pmatrix},$$

so that

$$\begin{pmatrix} nx^+ \\ y^+ \end{pmatrix} = \begin{pmatrix} nx \\ y \end{pmatrix} + \alpha \begin{pmatrix} np'_{\text{PAT}} \\ d'_{\text{PAT}} \end{pmatrix}$$

is invariant if also α is. However, the update rule (32) does not give an invariant $n\mu$, and instead we must use the more ambitious update (31) where $\gamma \in (0,1)$ is invariant. Path-following methods also use a neighborhood of the central path to define their iterates. By Proposition 3.2(vii), centering conditions such as $\|\frac{Xs}{\mu} - e\|_2 \leq \beta$ for invariant $\beta \in (0,1)$ [19,28,35,36,41] are not invariant, while

those using ∞ -norms (or the even weaker condition

$$\|\frac{\mathbf{X}s}{\mu} - \mathbf{e}\|_{\infty}^{-} := \|(\frac{\mathbf{X}s}{\mu} - \mathbf{e})^{-}\|_{\infty} \leq \beta$$

[27,33], where $\mathbf{v}^{-} = (\min\{0, v_j\})$, are invariant. Hence it seems that, of all theoretical path-following methods, only those of Kojima, Mizuno and Yoshise [27] and Algorithm 2 of Mizuno, Todd and Ye [33] (with neighborhood $\mathcal{N}$ equal to $\mathcal{N}_{\infty}(\beta)$ or $\mathcal{N}_{\infty}^{-}(\beta)$ and invariant $\beta \in (0,1)$) are invariant.

Finally, we consider potential-reduction methods. From (33) and (34), we see that npPRP is invariant if ρ/n and $\mathbf{z}_{\ell}$ are. Moreover, npPRY is invariant if ρ/n and $\mathbf{x}^T \mathbf{s}$ are, and in this case so is

$$\begin{pmatrix} \text{pPRPD} \\ n^{-1} d'_{\text{PRPD}} \end{pmatrix}.$$

(Note that, in the context of npPRY , $\mathbf{s}$ is a vector constructed in a particular way at a certain iteration--like $\mathbf{s}(\nu)$ in (20)--and hence we cannot conclude that $\mathbf{x}^T \mathbf{s}$ is invariant from Proposition 3.2(v); the latter only applies to primal-dual algorithms that iterate the pair $(\mathbf{x}, \mathbf{s})$.) Now for ρ/n to be invariant, we must have ρ equal to a constant times n , rather than the values $n + \nu\sqrt{n}$ or $2n + \nu\sqrt{n}$ for constant ν which lead to improved theoretical complexity bounds. For $\mathbf{z}_{\ell}$ or $\mathbf{x}^T \mathbf{s}$ to be invariant, by Proposition 3.2(viii) updating rules cannot be based on the Euclidean norm of the vector in (17) falling below a fixed constant, as in the algorithms of Gonzaga [23, section 5], Ye [52, section 3], or Freund [13].

However, Ye's practical algorithm in [52, section 6] updates $\mathbf{z}_{\ell}$ roughly as in the projective algorithm of (20)-(21) (Ye includes the extra constraint $\mathbf{b}^T \mathbf{y}(\mathbf{z}) - \mathbf{z} \geq 0$, which is unnecessary), and this gives an invariant $\mathbf{z}_{\ell}$; also Ye suggests larger ρ in this section, so ρ/n invariant (e.g., $\rho = 2n$ or $\rho = 5n$) is permitted. Also, Gonzaga's large-step potential-reduction method in [23; section 4] allows updating based on the ∞ -norm of the vector in (17), and suggests a large value of ρ (his q).

Let $\mathbf{h}$ denote the vector in (17):

$$\mathbf{h} = \frac{\rho}{\mathbf{c}^\top \mathbf{x} - z_\ell} \mathbf{X}^{-1} \mathbf{p}_{\text{AFF}} + \mathbf{X}^{-1} \mathbf{p}_{\text{CEN}}. \quad (39)$$

Gonzaga's rule is then (choosing $\epsilon = 1$ in [23]):

$$\begin{aligned} & \text{if } \|\mathbf{h}\|_\infty < 1, \text{ set} \\ & z_\ell \leftarrow \mathbf{c}^\top \mathbf{x} - \frac{(1 + n^{-\frac{1}{2}} \|\mathbf{h}\|_2)}{\rho/n} (\mathbf{c}^\top \mathbf{x} - z_\ell). \end{aligned} \quad (40)$$

By Proposition 3.2(viii), this update rule is invariant if ρ/n is. The motivation for this update is as follows. Note that $\mathbf{h} = \mathbf{X}^{-1} \mathbf{p}_{\text{PRP}}$. Let $\Delta = \mathbf{c}^\top \mathbf{x} - z_\ell$ (before the update), and define

$$\begin{aligned} \mathbf{s} &= \frac{\Delta}{\rho} (\mathbf{X}^{-1} \mathbf{e} - \mathbf{X}^{-1} \mathbf{h}) \\ &= \frac{\Delta}{\rho} (\mathbf{X}^{-1} \mathbf{e} - \mathbf{X}^{-2} \mathbf{p}_{\text{PRP}}) \\ &= -\mathbf{X}^{-2} \mathbf{p}_{\text{AFF}} + \frac{\Delta}{\rho} (\mathbf{X}^{-1} \mathbf{e} - \mathbf{X}^{-2} \mathbf{p}_{\text{CEN}}). \end{aligned}$$

From the first expression and $\|\mathbf{h}\|_\infty < 1$ we see $\mathbf{s} > 0$, while from the last $\mathbf{s} = \mathbf{c} - \mathbf{A}^\top \mathbf{y}$ for $\mathbf{y} = (\mathbf{A} \mathbf{X}^2 \mathbf{A}^\top)^{-1} (\mathbf{A} \mathbf{X}^2 \mathbf{c} - \frac{\Delta}{\rho} \mathbf{A} \mathbf{X} \mathbf{e})$. Hence $(\mathbf{y}, \mathbf{s})$ is feasible in (D) and the resulting duality gap is

$$\begin{aligned} \mathbf{x}^\top \mathbf{s} &= \frac{\Delta}{\rho} \mathbf{e}^\top (\mathbf{e} - \mathbf{h}) = \frac{\Delta}{\rho} (n - \mathbf{e}^\top \mathbf{h}) \\ &\leq \frac{\Delta}{\rho} (n + \sqrt{n} \|\mathbf{h}\|_2) = \frac{(1 + \|\mathbf{h}\|_2 / \sqrt{n})}{\rho/n} \Delta. \end{aligned}$$

This final quantity is the duality gap after the update. Since $\|h\|_\infty < 1$ implies $\|h\|_2 < \sqrt{n}$, the duality gap is reduced by at least the constant factor $\frac{2}{\rho/n}$ if we choose $\rho/n > 2$.

Finally, Kojima-Mizuno-Yoshise's primal-dual potential-reduction algorithm [29] is invariant if ρ is chosen to be a constant multiple of n rather than $n + \sqrt{n}$.

In summary, affine- and projective-scaling algorithms are invariant. Certain path-following methods, using measures of centrality based on ∞ -norms and large reductions in the parameter μ , are invariant, as are some large-step primal or primal-dual potential-reduction methods. In particular, to our knowledge, no $O(\sqrt{n}L)$ -iteration algorithm is invariant, while some $O(nL)$ -iteration methods are.

4. Algorithms for Semi-Infinite Problems

Here we describe extensions of the invariant methods from the previous section to the semi-infinite problems

$$\begin{aligned}
 & \min \langle \gamma, \xi \rangle \\
 \text{(SIP)} \quad & \int a\xi = b \\
 & \xi \in C(T), \quad \xi(t) \geq 0,
 \end{aligned}$$

and

$$\begin{aligned}
 & \max b^T y \\
 \text{(SID)} \quad & a^T y + \sigma = \gamma, \\
 & \sigma \geq 0.
 \end{aligned}$$

Of course, it may be possible to define also limits for algorithms that are not invariant, but these usually degenerate into either extensions of other invariant methods or nonsensical algorithms. For example, if $\rho = n + \nu\sqrt{n}$ for some invariant ν , the limit will be the same as that for the choice $\rho = n$. On the other hand, if $\rho = n + \nu n^2$, say, the limit is the same as that with ρ infinite, which

usually corresponds to the affine-scaling method. Similarly, algorithms using centering conditions or update rules based on the Euclidean norm falling below some constant will have limits where the centering condition or update rule never obtains, and thus will become nonsensical. On the other hand, invariant algorithms usually have natural extensions, obtained by taking suitable limits of their invariant ingredients.

Many of the methods described are only conceptual, in that the directions cannot be computed analytically but only approximated. One might then ask why one would develop algorithms in this general framework instead of applying the standard algorithms to finite-dimensional approximations. One reason is that thereby one assures some consistency in these finite-dimensional approximations, in that they are all approximating some conceptual infinite-dimensional method. Another is that studying these algorithms in the general setting may give insight into the behavior of similar methods applied to very large-scale problems. Finally, in certain cases the directions can be computed analytically; in particular, this is true for dual algorithms (where σ lies in an m -dimensional affine space in $C(T)$) for a special problem considered by Powell [37] where $m = 2$.

First we note that, since x_j in (P) corresponds to the integral of ξ over a subregion of T of measure $\frac{1}{n}$ containing t_j , nx_j approximates $\xi(t_j)$, and we replace the invariant vector nx by the function ξ . The vector y is unchanged, while the invariant vector s becomes the function σ .

We consider first the primal affine-scaling algorithm. We are given $\xi \in C(T)$, feasible to (SIP) with $\xi > 0$ (i.e., $\xi(t) > 0$ all $t \in T$). Now we extend the scaled direction

$$n^2 p_{\text{AFF}} = - (nX)^2 c + (nX)^2 A^T (nAX^2 A^T)^{-1} (nAX^2 c)$$

to

$$\pi_{\text{AFF}} := -\xi^2(\gamma - a^T [\int \xi^2 aa^T]^{-1} [\int a \xi^2 \gamma]). \quad (41)$$

To see how this limiting process works, we show how $nAX^2 A^T$ is replaced by $\int_0^1 a(\tau)a(\tau)^T \xi^2(\tau) d\tau$ in the case $T = [0,1]$. Indeed, if $t_j := (2j-1)/2n$ represents the interval $[(j-1)/n, j/n]$, then the

integral is approximated by $\sum_{j=1}^n a(t_j)a(t_j)^\top \xi^2(t_j) \frac{1}{n} = nAX^2A^\top$, where A has the $a(t_j)$'s as columns and $x_j := \xi(t_j)/n$ approximates the integral of ξ over this interval. A similar argument applies for more general T . The reader can similarly satisfy herself that the other expressions in (39) and in the formulae below naturally arise as suitable limits when A , x and c are viewed as the appropriate discretizations of the data in (SIP) and (SID). Next we would compute

$$\alpha_{\text{MAX}} := \max\{\alpha > 0: \xi + \alpha\pi_{\text{AFF}} \geq 0\}$$

and set

$$\xi^+ := \xi + .9\alpha_{\text{MAX}}\pi_{\text{AFF}}, \quad (42)$$

say. The computation of $\pi_{\text{AFF}}(t)$, in particular, of the integrals, can be performed either numerically, by some quadrature, or in certain cases analytically. In the former case, feasibility will be maintained only approximately and some correction scheme may be required. Even if π_{AFF} can be computed or approximated, obtaining (an approximation to) α_{MAX} may also be a difficult problem, as noted by Ferris and Philpott [11].

For the dual affine-scaling method, we assume we have $y \in \mathbb{R}^m$, $\sigma \in C(T)$, feasible to (SID) with $\sigma > 0$. We extend the scaled direction

$$\text{nd}_{\text{AFF}} = [n^{-1}AS^{-2}A^\top]^{-1}b$$

to

$$\tilde{d}_{\text{AFF}} := \left[\int \sigma^{-2}aa^\top \right]^{-1}b; \quad (43)$$

the corresponding direction for σ is

$$\delta_{\text{AFF}} := -a^\top \tilde{d}_{\text{AFF}}. \quad (44)$$

(The notation here is chosen to suggest that the direction $\tilde{d}_{\text{AFF}}$ is the natural extension of the affine-

scaling direction $\tilde{d}_{\text{AFF}}$ in the finite case; however, the tilde is added because to make the limit exist, d_{AFF} must be scaled by n .) Next we would compute

$$\hat{\alpha}_{\text{MAX}} := \max\{\hat{\alpha} > 0: \sigma + \hat{\alpha} \delta_{\text{AFF}} \geq 0\}$$

and set, for example,

$$\begin{aligned} y^+ &= y + .9\hat{\alpha}_{\text{MAX}}\tilde{d}_{\text{AFF}}, \\ \sigma^+ &= \sigma + .9\hat{\alpha}_{\text{MAX}}\delta_{\text{AFF}}. \end{aligned} \tag{45}$$

The computation of $\tilde{d}_{\text{AFF}}$ and $\hat{\alpha}_{\text{MAX}}$ presents similar problems as in the primal case. This dual algorithm is exactly that suggested by Ferris and Philpott [11] (the analogue of their algorithm 2).

For generalizing other algorithms, we note that

$$np_{\text{CEN}} = [nX]e - [nX]^2 A^T [nAX^2 A^T]^{-1} b$$

extends to

$$\pi_{\text{CEN}} := \xi - \xi^2 a^T [\int \xi^2 aa^T]^{-1} b \tag{46}$$

and

$$d_{\text{CEN}} = -[n^{-1}AS^{-2}A^T]^{-1}[n^{-1}AS^{-1}e]$$

extends to

$$\tilde{d}_{\text{CEN}} := -[\int \sigma^{-2} aa^T]^{-1} [\int \sigma^{-1} a], \tag{47}$$

with corresponding direction

$$\delta_{\text{CEN}} := -a^T \tilde{d}_{\text{CEN}} \tag{48}$$

for σ . Finally, primal-dual algorithms use the extensions

$$\begin{pmatrix} \pi'_{\text{AFF}} \\ \tilde{d}'_{\text{AFF}} \end{pmatrix} \text{ of } \begin{pmatrix} \pi'_{\text{AFF}} \\ d'_{\text{AFF}} \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \pi'_{\text{CEN}} \\ \tilde{d}'_{\text{CEN}} \end{pmatrix} \text{ of } \begin{pmatrix} \pi'_{\text{CEN}} \\ n^{-1} d'_{\text{CEN}} \end{pmatrix},$$

where

$$\pi'_{\text{AFF}} := -\xi + \xi \sigma^{-1} \mathbf{a}^\top [\int \xi \sigma^{-1} \mathbf{a} \mathbf{a}^\top]^{-1} \mathbf{b} \quad (49)$$

$$\tilde{d}'_{\text{AFF}} := [\int \xi \sigma^{-1} \mathbf{a} \mathbf{a}^\top]^{-1} \mathbf{b}, \quad (50)$$

$$\pi'_{\text{CEN}} := \sigma^{-1} - \xi \sigma^{-1} \mathbf{a}^\top [\int \xi \sigma^{-1} \mathbf{a} \mathbf{a}^\top]^{-1} [\int \sigma^{-1} \mathbf{a}], \quad (51)$$

and

$$\tilde{d}'_{\text{CEN}} := -[\int \xi \sigma^{-1} \mathbf{a} \mathbf{a}^\top]^{-1} [\int \sigma^{-1} \mathbf{a}]. \quad (52)$$

Let us turn to projective-scaling algorithms. For the primal, we use

$$\pi_{\text{PRO}} := \pi_{\text{AFF}} + \nu_{\text{PRO}} \pi_{\text{CEN}}, \quad (53)$$

where

$$\nu_{\text{PRO}} := \frac{\langle \gamma, \xi \rangle - \langle \gamma, \pi_{\text{CEN}} \rangle - z_\ell}{\mathbf{b}^\top [\int \xi^2 \mathbf{a} \mathbf{a}^\top]^{-1} \mathbf{b}}. \quad (54)$$

(Compare with (18), (19).) Here z_ℓ can be the optimal value z_* of (SIP) if known, or it can be updated as follows. Let ν_+ solve

$$\begin{aligned} \min_{\nu} \quad & \nu \\ \xi^2 \sigma_\nu &:= -\pi_{\text{AFF}} + (\xi - \pi_{\text{CEN}}) \nu \geq 0, \quad t \in [0,1] \end{aligned} \quad (55)$$

($= \infty$ if the problem is infeasible), and hence obtain (since σ_{ν_+} is a feasible dual slack)

$$z_+ := \langle \gamma, \xi \rangle - \langle \xi, \sigma_{\nu_+} \rangle. \quad (56)$$

Now substitute from (55) and use the easily-checked facts

$$\langle \xi^{-1}, \pi_{\text{AFF}} \rangle = \langle \gamma, \pi_{\text{CEN}} \rangle \quad \text{and} \quad \langle \eta, \eta - \xi^{-1} \pi_{\text{CEN}} \rangle = \langle \eta - \xi^{-1} \pi_{\text{CEN}}, \eta - \xi^{-1} \pi_{\text{CEN}} \rangle$$

where $\eta(t) := 1$ for all t . We find

$$z_+ = \langle \gamma, \xi \rangle - \langle \gamma, \pi_{\text{CEN}} \rangle - \nu_+ \langle \eta - \xi^{-1} \pi_{\text{CEN}}, \eta - \xi^{-1} \pi_{\text{CEN}} \rangle; \quad (57)$$

compare with (21). Finally, set $z_\ell \leftarrow \max\{z_+, z_\ell\}$. Using (57) we can show that, if $z_\ell = z_+$, then $\nu_{\text{PRO}} = \nu_+$, exactly as in the finite-dimensional case. This uses $\mathbf{b}^T [\int \xi^2 \mathbf{a} \mathbf{a}^T]^{-1} \mathbf{b} = \langle \eta - \xi^{-1} \pi_{\text{CEN}}, \eta - \xi^{-1} \pi_{\text{CEN}} \rangle$, which follows directly from the definition (46) of π_{CEN} . Once we have π_{PRO} , we set

$$\xi^+ := \xi + \alpha \pi_{\text{PRO}}, \quad (58)$$

where α can be chosen as in (42) or to approximately minimize the potential function

$$\ell_n(\langle c, \xi \rangle - z_\ell) - \int \ell_n \xi. \quad (59)$$

(Note that the parameter ρ in (11) has disappeared. In fact, (59) corresponds to the invariant potential function given in (ii) of Proposition 3.2 if $\rho = n$ as in Karmarkar [26] and Gonzaga [18].)

It can be shown that π_{PRO} is a descent direction for this function.

For the dual projective-scaling algorithm, suppose we have $\mathbf{y} \in \mathbb{R}^m$, $\sigma \in C(T)$ feasible for (SID) with $\sigma > 0$. Then let

$$\tilde{d}_{\text{PRO}} := \tilde{d}_{\text{AFF}} + \hat{\nu}_{\text{PRO}} \tilde{d}_{\text{CEN}}, \quad (60)$$

where now

$$\hat{\nu}_{\text{PRO}} = \frac{z_u - b^T y + b^T \tilde{d}_{\text{CEN}}}{1 - \left(\int \sigma^{-1} a \right)^T \left(\int \sigma^{-2} a a^T \right)^{-1} \left(\int \sigma^{-1} a \right)}. \quad (61)$$

(Compare with (22), (23). The denominator in (61) corresponds to that in (23) (without the “1+”); the “1” in (61) is the integral of the function identically equal to one over T , which has measure 1.) Here z_u can be the optimal value z_* of (SID) if known, or can be updated as follows. Let $\hat{\nu}_+$ solve

$$\min_{\hat{\nu}} \sigma^2 \xi_{\hat{\nu}} := a^T \tilde{d}_{\text{AFF}} + (\sigma + a^T \tilde{d}_{\text{CEN}}) \hat{\nu} \geq 0 \quad (62)$$

($= \infty$ if the problem is infeasible), hence obtain

$$\hat{z}_+ = b^T y + \langle \xi_{\hat{\nu}_+}, \sigma \rangle, \quad (63)$$

which is again easily shown to equal

$$b^T y - b^T \tilde{d}_{\text{CEN}} + \hat{\nu}_+ \langle \eta + \sigma^{-1} a^T \tilde{d}_{\text{CEN}}, \eta + \sigma^{-1} a^T \tilde{d}_{\text{CEN}} \rangle \quad (64)$$

where again $\eta(t) := 1$ for all t , and finally set $z_u = \min\{\hat{z}_+, z_u\}$. Again, if $z_u = \hat{z}_+$, then $\hat{\nu}_{\text{PRO}} = \hat{\nu}_+$; this uses the fact that the denominator in (61) equals the inner product in (64), from the definition (47) of $\tilde{d}_{\text{CEN}}$. Then we update to

$$y^+ := y + \hat{\alpha} \tilde{d}_{\text{PRO}}, \quad (65)$$

where $\hat{\alpha}$ can be chosen as in (45) or to approximately minimize the potential function

$$\ell_n(z_u - b^T y) - \int \ell_n \sigma; \quad (66)$$

again, d_{PRO} can be shown to be a descent direction for this function, which corresponds to the invariant dual potential function in Proposition 3.2(ii) where $\rho = n$.

We have compared this algorithm to that derived by Powell [37] for the problem

$$\begin{aligned} \max \quad & -y_2 \\ \text{subject to} \quad & \cos(2\pi t)y_1 + \sin(2\pi t)y_2 \leq 1, \quad t \in [0,1]. \end{aligned}$$

(Here π is the trigonometric constant, not the primal direction!) Powell proceeds directly from the potential function (66), using $z_u = z_* = 1$. (The optimal solution is clearly $y^* = (0, -1)^T$.) The direction he obtains coincides with our $\tilde{d}_{\text{PRO}}$ above if we use $z_u = 1$, and hence yields identical iterates, including the sequence that converges to a nonoptimal point starting with $y^0 = (.599, .8)^T$. However, if we start with a strict upper bound sufficiently above z_* then it appears that our algorithm converges to the optimal solution. For example, with the y^0 above and the initial $z_u = 10$, the upper bound is first improved at iteration 3 and the algorithm terminates (duality gap less than 10^{-4}) at iteration 8 with $y^8 = (1.01 \times 10^{-4}, -1 + 5.52 \times 10^{-9})^T$ and $z_u = 1 + 3.12 \times 10^{-5}$. It is interesting to note that for some feasible y and upper bounds z_u the resulting $\hat{\nu}_{\text{PRO}}$ is negative, but this never occurred in our experiments. We also point out that the dual affine-scaling method fails disastrously on this problem. From the same starting point, the iterates converge to $(.6445, .7646)^T$. In all cases, the search directions were computed analytically; the projective-scaling method chose $\hat{\alpha}$ to minimize (66) using 100 iterations of a bisection algorithm, while the affine-scaling method used (45).

Next we turn to path-following methods. The only algorithms in this class based on ℓ_∞ -neighborhoods of the central path are primal-dual methods, so we confine our attention to such methods. For $\beta \in (0,1)$, we let $\mathcal{N}_\infty(\beta)$ be the set of all pairs (ξ, σ) where ξ is feasible in (SIP), (y, σ) is feasible in (SID) for some y , $\xi > 0$, $\sigma > 0$, and

$$(1 - \beta)\langle \xi, \sigma \rangle \eta \leq \xi \sigma \leq (1 + \beta)\langle \xi, \sigma \rangle \eta; \quad (67)$$

similarly $\mathcal{N}_\infty^-(\beta)$ is the set of all such pairs satisfying just the left hand inequality of (67). Given such a pair (ξ, σ) (in $\mathcal{N}_\infty$, equal to $\mathcal{N}_\infty(\beta)$ or $\mathcal{N}_\infty^-(\beta)$), we define $\mu := \theta\langle \xi, \sigma \rangle$ for some fixed

$\theta \in (0,1)$ and set

$$\begin{pmatrix} \pi'_{\text{PAT}} \\ \tilde{d}'_{\text{PAT}} \end{pmatrix} := \begin{pmatrix} \pi'_{\text{AFF}} \\ \tilde{d}'_{\text{AFF}} \end{pmatrix} + \mu \begin{pmatrix} \pi'_{\text{CEN}} \\ \tilde{d}'_{\text{CEN}} \end{pmatrix}. \quad (68)$$

Then let

$$\bar{\alpha} := \max\{\alpha \in [0,1]: (\xi_\alpha, \sigma_\alpha) := (\xi + \alpha\pi'_{\text{PAT}}, \sigma - \alpha a^T \tilde{d}'_{\text{PAT}}) \in \mathcal{N}\}$$

and set $(\xi_+, \sigma_+) = (\xi_{\bar{\alpha}}, \sigma_{\bar{\alpha}})$. It is not hard to show, using the continuity of ξ , π'_{PAT} , σ and $a^T \tilde{d}'_{\text{PAT}}$ and the compactness of T , that $\bar{\alpha}$ is positive. Also

$$\begin{aligned} \langle \xi_\alpha, \sigma_\alpha \rangle &= \langle \xi, \sigma \rangle - \alpha(\langle \xi, a^T \tilde{d}'_{\text{PAT}} \rangle - \langle \pi'_{\text{PAT}}, \sigma \rangle) - \alpha^2 \langle \pi'_{\text{PAT}}, a^T \tilde{d}'_{\text{PAT}} \rangle \\ &= \langle \xi, \sigma \rangle - \alpha(\langle \xi, a^T \tilde{d}'_{\text{PAT}} \rangle - \langle \pi'_{\text{PAT}}, \sigma \rangle) \\ &= \langle \xi, \sigma \rangle - \alpha(\langle \xi, a^T \tilde{d}'_{\text{AFF}} \rangle - \langle \pi'_{\text{AFF}}, \sigma \rangle + \mu(\langle \xi, a^T \tilde{d}'_{\text{CEN}} \rangle - \langle \pi'_{\text{CEN}}, \sigma \rangle)) \\ &= \langle \xi, \sigma \rangle (1 - \alpha + \alpha\theta), \end{aligned}$$

so that the duality gap $\langle \gamma, \xi \rangle - b^T y = \langle \gamma, \xi \rangle - \langle \xi, a^T y \rangle = \langle \xi, \sigma \rangle$ is reduced by the factor $1 - \bar{\alpha} + \bar{\alpha}\theta < 1$ in moving from (ξ, σ) to (ξ_+, σ_+) . This is exactly the generalization of algorithm 2 of Mizuno, Todd and Ye [33] to the semi-infinite case, and is closely related to the method of Kojima, Mizuno and Yoshise [27], who use a different choice of step size.

Finally, we consider potential-reduction methods. For the primal, we suppose we have a feasible $\xi > 0$ and a lower bound z_ℓ . Let $\Delta := \langle \gamma, \xi \rangle - z_\ell$ be the corresponding duality gap. Then the direction is

$$\pi_{\text{PRP}} := \frac{\tau}{\Delta} \pi_{\text{AFF}} + \pi_{\text{CEN}}, \quad (69)$$

where τ (corresponding to ρ/n in the finite dimensional case) is at least two. We now show that a step in this direction decreases the primal potential function

$$\phi(\xi; \mathbf{z}_\ell) := \tau \ell n(\langle \gamma, \xi \rangle - \mathbf{z}_\ell) - \langle \eta, \ell n \xi \rangle.$$

For simplicity, we write π for π_{PRP} .

Proposition 4.1. Suppose $0 < \alpha \leq \frac{1}{2}$ is such that $\alpha |\xi^{-1} \pi| \leq \frac{1}{2} \eta$, so that $\xi + \alpha \pi = \xi(\eta + \alpha \xi^{-1} \pi) \geq \frac{1}{2} \xi$. Then

$$\begin{aligned} \phi(\xi + \alpha \pi; \mathbf{z}_\ell) &\leq \phi(\xi; \mathbf{z}_\ell) - \frac{1}{2} \alpha \langle \xi^{-1} \pi, \xi^{-1} \pi \rangle \\ &=: \phi(\xi; \mathbf{z}_\ell) - \frac{1}{2} \alpha \|\xi^{-1} \pi\|_2^2. \end{aligned}$$

Proof. We have

$$\begin{aligned} \phi(\xi + \alpha \pi; \mathbf{z}_\ell) - \phi(\xi; \mathbf{z}_\ell) &= \tau \ell n \left(\frac{\langle \gamma, \xi + \alpha \pi \rangle - \mathbf{z}_\ell}{\langle \gamma, \xi \rangle - \mathbf{z}_\ell} \right) - \langle \eta, \ell n(\xi^{-1}(\xi + \alpha \pi)) \rangle \\ &\leq \alpha \frac{\tau}{\Delta} \langle \gamma, \pi \rangle - \langle \eta, \alpha \xi^{-1} \pi - (\alpha \xi^{-1} \pi)^2 (2(\eta - |\alpha \xi^{-1} \pi|))^{-1} \rangle \end{aligned}$$

$$\text{(using } \ell n(1 + \theta) \geq \theta - \frac{\theta^2}{2(1 - |\theta|)} \text{ for } |\theta| < 1)$$

$$\leq \alpha \langle \frac{\tau}{\Delta} \xi \gamma - \eta, \xi^{-1} \pi \rangle + \alpha^2 \langle \xi^{-1} \pi, \xi^{-1} \pi \rangle$$

$$\text{(since } \eta - |\alpha \xi^{-1} \pi| \geq \frac{1}{2} \eta \text{)}.$$

But, for any $\mathbf{v}$, $\langle \xi \mathbf{a}^\top \mathbf{v}, \xi^{-1} \pi \rangle = 0$, because $\langle \xi \mathbf{a}^\top \mathbf{v}, \xi^{-1} \pi_{\text{AFF}} \rangle = \langle \xi \mathbf{a}^\top \mathbf{v}, \xi^{-1} \pi_{\text{CEN}} \rangle = 0$, as $\mathbf{v} \perp \mathbf{a}$ easily checked. Hence, using the definition (69) of π , we obtain

$$\langle \frac{\tau}{\Delta} \xi \gamma - \eta, \xi^{-1} \pi \rangle = -\langle \xi^{-1} \pi, \xi^{-1} \pi \rangle,$$

and the result follows since $0 < \alpha \leq \frac{1}{2}$ implies $-\alpha + \alpha^2 \leq -\frac{1}{2} \alpha$. $\square$

The amount the potential function decreases depends on $\|\xi^{-1}\pi\|_2$, whose size depends on the update rule for z_ℓ . Ye [52, section 6] basically updates z_ℓ as in the projective-scaling algorithm above--see (55)-(57). Suppose after the update we have $\|\xi^{-1}\pi\|_\infty := \max\{|\xi^{-1}\pi|\} < 1$. (Here $\max(\lambda)$ denotes $\max\{\lambda(t): t \in T\}$ for $\lambda \in C(T)$.) Then $-\eta < \xi^{-1}\pi < \eta$, so that

$$\sigma := \frac{\Delta}{\tau} (\xi^{-1}\eta - \xi^{-2}\pi) > 0. \quad (71)$$

However, by definition of π ,

$$\begin{aligned} \sigma &= -\xi^{-2}\pi_{\text{AFF}} + \frac{\Delta}{\tau} (\xi^{-1}\eta - \xi^{-2}\pi_{\text{CEN}}) \\ &= \gamma - \mathbf{a}^T \mathbf{y}, \end{aligned}$$

where $\mathbf{y} = [\int \xi^2 \mathbf{a} \mathbf{a}^T]^{-1} (\int \mathbf{a} \xi^2 \gamma - \frac{\Delta}{\tau} \mathbf{b})$. Thus $(\mathbf{y}, \sigma)$ is feasible in (SID), and the corresponding duality gap is $\langle \xi, \sigma \rangle = \frac{\Delta}{\tau} (\langle \eta, \eta \rangle - \langle \eta, \xi^{-1}\pi \rangle) < \frac{\Delta}{\tau} (\langle \eta, \eta \rangle + \langle \eta, \eta \rangle) = (\frac{2}{\tau})\Delta \leq \Delta$. But this contradicts the update for z_ℓ ; $\sigma \geq 0$ shows that $\nu = \frac{\Delta}{\tau}$ is feasible in (55) ($\sigma = \sigma_\nu$ for this ν), and it gives a better duality gap than the current one, so z_ℓ would have been further increased. Hence we must have $\|\xi^{-1}\pi\|_\infty \geq 1$.

Now by choosing $\alpha = 1/2\|\xi^{-1}\pi\|_\infty \leq \frac{1}{2}$, we find a reduction in ϕ of at least

$$\frac{1}{4} \|\hat{\pi}\|_2^2 \|\xi^{-1}\pi\|_\infty \geq \frac{1}{4} \|\hat{\pi}\|_2^2, \quad (72)$$

where $\hat{\pi} := \xi^{-1}\pi/\|\xi^{-1}\pi\|_\infty$ has $\|\hat{\pi}\|_\infty = 1$. If all the data and the current solution ξ are piecewise constant on subsets of T of measure at least $\frac{1}{n}$ (as in the case where (SIP) models (P)), then $\hat{\pi}$ is also piecewise constant, and $\|\hat{\pi}\|_\infty = 1$ implies $\|\hat{\pi}\|_2^2 \geq \frac{1}{n}$. This yields a decrease in ϕ of at least $\frac{1}{4n}$, which corresponds to the constant decrease in the finite dimensional case, since ϕ here is one n th the potential function in that case. In general, we cannot give a lower bound on $\|\hat{\pi}\|_2^2$, but

clearly it is larger if $\hat{\pi}$ is smoother, for then $\hat{\pi}$ is close to 1 in absolute value on a larger subset of T .

Now suppose we update the lower bound by an extension of Gonzaga's procedure [23, section 4], given in (39)-(40) above. In this case, suppose τ is strictly greater than 2. The update rule becomes

$$\begin{aligned} & \text{if } \|\xi^{-1}\pi\|_\infty < 1, \text{ set} \\ & z_\ell \leftarrow \langle \gamma, \xi \rangle - \frac{(1 + \|\xi^{-1}\pi\|_2)}{\tau} (\langle \gamma, \xi \rangle - z_\ell). \end{aligned} \tag{73}$$

To show that this update is valid, note that if $\|\xi^{-1}\pi\|_\infty < 1$, σ in (71) is a feasible dual slack and the corresponding duality gap is $\langle \xi, \sigma \rangle = \frac{\Delta}{\tau} (\langle \eta, \eta \rangle - \langle \eta, \xi^{-1}\pi \rangle) \leq \frac{\Delta}{\tau} (1 + \|\xi^{-1}\pi\|_2)$. After updating z_ℓ and hence Δ , we can recompute π from (69) and apply (73) again. The duality gap is bounded below by $\langle \gamma, \xi \rangle - z_* > 0$, while every update decreases it by the factor $(1 + \|\xi^{-1}\pi\|_2)/\tau \leq 2/\tau < 1$. Hence a finite number of applications of (73) lead to a direction π with $\|\xi^{-1}\pi\|_\infty \geq 1$. Note that the final update of the duality gap yields a value not less than $\langle \xi, \sigma \rangle$ for σ as in (71); since $\sigma = \sigma_\nu$ for $\nu = \Delta/\tau$, this value of ν is feasible in (55) and hence Ye's update is always at least as good as our extended (repeated) Gonzaga's update.

Now we consider the dual case. We have a feasible (y, σ) for (SID) with $\sigma > 0$ and an upper bound z_u . Let $\Delta := z_u - b^T y$ be the duality gap, and choose as search direction

$$\tilde{d}_{\text{PRD}} := \frac{\tau}{\Delta} \tilde{d}_{\text{AFF}} + \tilde{d}_{\text{CEN}}, \tag{74}$$

where τ is again a constant that is at least 2. The corresponding direction for σ is $-a^T \tilde{d}_{\text{PRD}}$. We show that a step in this direction decreases the dual potential function

$$\psi(y, \sigma; z_u) := \tau \ln(z_u - b^T y) - \langle \eta, \ln \sigma \rangle. \tag{75}$$

For simplicity, we again write $\tilde{d}$ for $\tilde{d}_{\text{PRD}}$.

Proposition 4.2. Suppose $0 < \alpha \leq \frac{1}{2}$ is such that $|\alpha \sigma^{-1} \mathbf{a}^\top \tilde{\mathbf{d}}| \leq \frac{1}{2} \eta$, so that $\sigma - \alpha \mathbf{a}^\top \tilde{\mathbf{d}} = \sigma(\eta - \alpha \sigma^{-1} \mathbf{a}^\top \tilde{\mathbf{d}}) \geq \frac{1}{2} \sigma$. Then

$$\begin{aligned} \psi(y + \alpha \tilde{\mathbf{d}}, \sigma - \alpha \mathbf{a}^\top \tilde{\mathbf{d}}; \mathbf{z}_u) &\leq \psi(y, \sigma; \mathbf{z}_u) - \frac{1}{2} \alpha \langle \sigma^{-1} \mathbf{a}^\top \tilde{\mathbf{d}}, \sigma^{-1} \mathbf{a}^\top \tilde{\mathbf{d}} \rangle \\ &= \psi(y, \sigma; \mathbf{z}_u) - \frac{1}{2} \alpha \|\sigma^{-1} \mathbf{a}^\top \tilde{\mathbf{d}}\|_2^2. \end{aligned}$$

Proof. We proceed as in the proof of Proposition 4.1.

$$\begin{aligned} &\psi(y + \alpha \tilde{\mathbf{d}}, \sigma - \alpha \mathbf{a}^\top \tilde{\mathbf{d}}; \mathbf{z}_u) - \psi(y, \sigma; \mathbf{z}_u) \\ &= \tau \ln \left(\frac{\mathbf{z}_u - \mathbf{b}^\top (y + \alpha \tilde{\mathbf{d}})}{\mathbf{z}_u - \mathbf{b}^\top y} \right) - \langle \eta, \ln(\sigma^{-1}(\sigma - \alpha \mathbf{a}^\top \tilde{\mathbf{d}})) \rangle \\ &\leq \alpha \frac{\tau}{\Delta} (-\mathbf{b}^\top \tilde{\mathbf{d}}) - \langle \eta, -\alpha \sigma^{-1} \mathbf{a}^\top \tilde{\mathbf{d}} - (\alpha \sigma^{-1} \mathbf{a}^\top \tilde{\mathbf{d}})^2 (2(\eta - |\alpha \sigma^{-1} \mathbf{a}^\top \tilde{\mathbf{d}}|))^{-1} \rangle \\ &\leq \alpha \left(\frac{\tau}{\Delta} (-\mathbf{b}^\top \tilde{\mathbf{d}}) + \langle \eta, \sigma^{-1} \mathbf{a}^\top \tilde{\mathbf{d}} \rangle \right) + \alpha^2 \langle \sigma^{-1} \mathbf{a}^\top \tilde{\mathbf{d}}, \sigma^{-1} \mathbf{a}^\top \tilde{\mathbf{d}} \rangle \end{aligned}$$

since $|\alpha \sigma^{-1} \mathbf{a}^\top \tilde{\mathbf{d}}| \leq \frac{1}{2} \eta$. But

$$\begin{aligned} \langle \sigma^{-1} \mathbf{a}^\top \tilde{\mathbf{d}}, \sigma^{-1} \mathbf{a}^\top \tilde{\mathbf{d}} \rangle &= \langle \tilde{\mathbf{d}}^\top (\sigma^{-1} \mathbf{a}), \sigma^{-1} \mathbf{a}^\top \tilde{\mathbf{d}} \rangle \\ &= \tilde{\mathbf{d}}^\top \left(\int \sigma^{-2} \mathbf{a} \mathbf{a}^\top \right) \tilde{\mathbf{d}} \\ &= \tilde{\mathbf{d}}^\top \left(\frac{\tau}{\Delta} \mathbf{b} - \int \mathbf{a} \sigma^{-1} \right) \\ &= - \left(\frac{\tau}{\Delta} (-\mathbf{b}^\top \tilde{\mathbf{d}}) + \langle \eta, \sigma^{-1} \mathbf{a}^\top \tilde{\mathbf{d}} \rangle \right) \end{aligned}$$

by the definition of $\tilde{\mathbf{d}}$, and so the decrease in ψ is at least

$$(\alpha - \alpha^2) \langle \sigma^{-1} \mathbf{a}^\top \tilde{\mathbf{d}}, \sigma^{-1} \mathbf{a}^\top \tilde{\mathbf{d}} \rangle$$

which yields the desired result since $0 < \alpha \leq \frac{1}{2}$. $\square$

Again, the decrease in the potential function depends on $\|\sigma^{-1}\mathbf{a}^\top\tilde{\mathbf{d}}\|_2^2$, and this in turn is a function of the update rule for z_u . Analogous to Ye's technique in the primal case, we can use the update as in the dual projective algorithm. Suppose after the update $\|\sigma^{-1}\mathbf{a}^\top\tilde{\mathbf{d}}\|_\infty < 1$. Then

$$\xi := \frac{\Delta}{\tau} (\sigma^{-1}\eta + \sigma^{-2}\mathbf{a}^\top\tilde{\mathbf{d}}) > 0. \quad (76)$$

Then we find that ξ is feasible in (SIP), and the corresponding duality gap is $\langle \xi, \sigma \rangle = \frac{\Delta}{\tau} (\langle \eta, \eta \rangle + \langle \eta, \sigma^{-1}\mathbf{a}^\top\tilde{\mathbf{d}} \rangle) < \frac{\Delta}{\tau} (\langle \eta, \eta \rangle + \langle \eta, \eta \rangle) = (\frac{2}{\tau}) \Delta \leq \Delta$. Again, this contradicts the update for z_u ; $\xi \geq 0$ shows that $\hat{\nu} = \Delta/\tau$ is feasible in (62), and it gives a better duality gap than the current one, so z_u would have been further decreased. This contradiction shows that $\|\sigma^{-1}\mathbf{a}^\top\tilde{\mathbf{d}}\|_\infty \geq 1$.

If we choose $\alpha = 1/2\|\sigma^{-1}\mathbf{a}^\top\tilde{\mathbf{d}}\|_\infty \leq \frac{1}{2}$, we obtain a reduction in ψ of at least

$$\frac{1}{4} \|\hat{\chi}\|_2^2 \|\sigma^{-1}\mathbf{a}^\top\tilde{\mathbf{d}}\|_\infty \geq \frac{1}{4} \|\hat{\chi}\|_2^2, \quad (77)$$

where $\hat{\chi} := \sigma^{-1}\mathbf{a}^\top\tilde{\mathbf{d}}/\|\sigma^{-1}\mathbf{a}^\top\tilde{\mathbf{d}}\|_\infty$ has $\|\hat{\chi}\|_\infty = 1$. Again, if all the data in (SID) and hence the current σ are piecewise-constant on subsets of $\mathcal{T}$ of measure at least $\frac{1}{n}$ (as in the case where (SID) models (D)), then $\hat{\chi}$ is also piecewise constant, and $\|\hat{\chi}\|_\infty = 1$ implies $\|\hat{\chi}\|_2^2 \geq \frac{1}{n}$; hence ψ decreases by at least $\frac{1}{4n}$. In general, we cannot give a lower bound for $\|\hat{\chi}\|_2^2$, but again it is larger if $\hat{\chi}$ is smoother.

Gongaza's update (73) can also be extended to the dual case. Then

$$\begin{aligned} &\text{if } \|\sigma^{-1}\mathbf{a}^\top\tilde{\mathbf{d}}\|_\infty < 1, \text{ set} \\ &z_u \leftarrow b^\top y + \frac{(1 + \|\sigma^{-1}\mathbf{a}^\top\tilde{\mathbf{d}}\|_2)}{\tau} (z_u - b^\top y). \end{aligned} \quad (78)$$

Again, this upper bound can be shown to be valid using the primal feasible ξ in (76), and the update can be repeated (a finite number of times) until $\|\sigma^{-1}\mathbf{a}^\top\tilde{\mathbf{d}}\|_\infty \geq 1$. As in the primal case, this update is no stronger than Ye's described above.

We tried these algorithms also on Powell's example, with the starting point $y^0 = (.599, .8)^T$. With $z_u = 10$ and $\tau = 2$, the method with Ye's update improved the upper bound first at the third iteration and terminated in 16 iterations with $y^{16} = (3.29 \times 10^{-4}, -1 + 5.42 \times 10^{-8})$ and $z_u = 1 + 1.10 \times 10^{-4}$. (Termination was for numerical reasons; our usual criterion was not quite met.) Using $\tau = 3$, the method with Gonzaga's update first updated z_u at iteration 4 and then stopped at iteration 17 with $y^{17} = (3.62 \times 10^{-11}, -1 + 1.48 \times 10^{-10})^T$ and $z_u = 1 + 5.15 \times 10^{-5}$. Neither of these methods works as well as the dual projective algorithm, for different reasons. With Ye's update, the upper bound becomes very good quite fast (reaching its final value at iteration 11), but then there is too little centering component in the search direction, and the bound cannot be improved -- there is a close relationship between updating bounds and being close to the central path. With Gonzaga's update, on the other hand, the upper bound is updated too slowly and too many iterations are required, with the upper bound updated at each one (after the first few). Similar behavior is observed over a range of starting points and values for τ and initial z_u , and the projective-scaling method seems to be superior for this problem.

Our final algorithm is the primal-dual potential-reduction method. We suppose we have ξ feasible for (SIP) and (y, σ) feasible for (SID), with $\xi > 0$, $\sigma > 0$. Let $\Delta = \langle \xi, \sigma \rangle$ be the duality gap and $\tau > 1$. Then the search direction is

$$\begin{pmatrix} \pi'_{\text{PRPD}} \\ \tilde{d}'_{\text{PRPD}} \end{pmatrix} := \frac{\tau}{\Delta} \begin{pmatrix} \pi'_{\text{AFF}} \\ \tilde{d}'_{\text{AFF}} \end{pmatrix} + \begin{pmatrix} \pi'_{\text{CEN}} \\ \tilde{d}'_{\text{CEN}} \end{pmatrix}. \quad (77)$$

Once again we can show that this is a descent direction for the appropriate potential function, here the primal dual function

$$\phi_{\text{PD}}(\xi, \sigma) := \tau \ln \langle \xi, \sigma \rangle - \langle \eta, \ln(\xi) \rangle - \langle \eta, \ln(\sigma) \rangle, \quad (78)$$

with $\tau > 1$. However, the analysis appears complicated and the choice of step size not at all obvious.

5. Concluding remarks

We have investigated the application of interior-point algorithms to continuous semi-infinite programming problems in primal and dual form. Our primary motivation comes from our belief that methods that are effective on large-scale problems should have natural extensions to the limiting semi-infinite case. By considering nested discretizations of a given semi-infinite programming problem, we have been led to an invariance principle for methods for finite-dimensional problems. Affine and projective scaling algorithms as well as some path-following and potential-reduction methods satisfy this invariance property, but several methods, including all known $O(\sqrt{n} L)$ -iteration algorithms, do not. Finally, we have considered algorithms for semi-infinite problems that are natural limits of invariant finite-dimensional methods, seen the behavior of three dual methods on a simple semi-infinite problem of Powell [33], and shown that potential reductions can be achieved. The amount of potential reduction appears to be related to the smoothness of the problem, so that the number of iterations may grow with the “roughness” (= inverse smoothness) of the infinite problem, in contrast to the dimension n , which can be viewed as the discreteness of a finite-dimensional problem.

Acknowledgements. I am very indebted to the referees for their helpful suggestions.

References

- [1] I. Adler, M.G.C. Resende, G. Veiga and N. Karmarkar, "An implementation of Karmarkar's algorithm for linear programming," *Mathematical Programming* 44 (1989), 297-335.
- [2] E.L. Allgower, K. Böhmer, F.A. Potra, and W.C. Rheinboldt, "A mesh-independence principle for operator equations and their discretizations," *SIAM Journal on Numerical Analysis* 23 (1986), 160-169.
- [3] E.J. Anderson and P. Nash, *Linear Programming in Infinite-Dimensional Spaces* (Wiley, Chichester, 1987).
- [4] K.M. Anstreicher, "A monotonic projective algorithm for fractional linear programming," *Algorithmica* 1 (1986), 483-498.
- [5] E.R. Barnes, "A variation on Karmarkar's algorithm for solving linear programming problems," *Mathematical Programming* 36 (1986), 174-182.
- [6] R.E. Bixby, J.W. Gregory, I.J. Lustig, R.E. Marsten, and D.F. Shanno, "Very large-scale linear programming: a case study in combining interior point and simplex methods," *Operations Research* 40 (1992), 885-897.
- [7] T.M. Cavalier and A.L. Soyster, "Some computational experience and a modification of the Karmarkar algorithm," Working Paper 85-105, Dept. of Industrial and Management Systems Engineering, Pennsylvania State University, University Park, PA, 1985.
- [8] E. Christiansen and K.O. Kortanek, "Computing material collapse displacement fields on a Cray X-MP/48 by the LP primal affine scaling algorithm," *Annals of Operations Research* 22 (1990), 355-376.
- [9] I.I. Dikin, "Iterative solution of problems of linear and quadratic programming," *Doklady Akademii Nauk SSSR* 174 (1967), 747-748.
- [10] M.C. Ferris and A.B. Philpott, "An interior point algorithm for semi-infinite linear programming," *Mathematical Programming* 43 (1989), 257-276.
- [11] M.C. Ferris and A.B. Philpott, "On affine scaling and semi-infinite programming," *Mathematical Programming* 56 (1992), 361-364.
- [12] R.M. Freund, "An analog of Karmarkar's algorithm for inequality constrained linear programs, with a 'new' class of projective transformations for centering a polytope," *Operations Research Letters* 7 (1988), 9-13.
- [13] R.M. Freund, "Polynomial-time algorithms for linear programming based only on primal scaling and projected gradients of a potential function," *Mathematical Programming* 51 (1991), 203-222.
- [14] D. Gay, "A variant of Karmarkar's linear programming algorithm for problems in standard form," *Mathematical Programming* 37 (1987), 81-90.
- [15] G. de Ghellinck and J.-Ph. Vial, "A polynomial Newton method for linear programming," *Algorithmica* 1 (1986), 425-453.

- [16] K. Glashoff and S.-A. Gustafson, *Linear Optimization and Approximation* (Springer-Verlag, New York, 1983).
- [17] D. Goldfarb and M.J. Todd, "Linear programming," in: G.L. Nemhauser, A.H.G. Rinnooy Kan, and M.J. Todd, eds., *Handbooks in Operations Research and Management Science*, Vol. 1, Optimization (North-Holland, Amsterdam, 1989), pp. 73-170.
- [18] C.C. Gonzaga, "Conical projection algorithms for linear programming," *Mathematical Programming* 43 (1989), 151-173.
- [19] C.C. Gonzaga, "An algorithm for solving linear programming programs in $O(n^3L)$ operations," in: N. Megiddo, ed., *Progress in Mathematical Programming, Interior Point and Related Methods* (Springer-Verlag, New York, 1988), pp. 1-28.
- [20] C.C. Gonzaga, "Search directions for interior linear-programming methods," *Algorithmica* 6 (1991), 153-181.
- [21] C.C. Gonzaga, "Interior point algorithms for linear programming problems with inequality constraints," *Mathematical Programming* 52 (1991), 209-225.
- [22] C.C. Gonzaga, "Polynomial affine algorithms for linear programming," *Mathematical Programming* 49 (1990), 7-21.
- [23] C.C. Gonzaga, "Large-steps path-following methods for linear programming, Part II: Potential reduction method," *SIAM Journal on Optimization* 1 (1991), 280-292.
- [24] D. den Hertog and C. Roos, "A survey of search directions in interior point methods for linear programming," *Mathematical Programming* 52 (1991), 481-509.
- [25] R. Hettich and K.O. Kortanek, "Semi-infinite programming: theory, method and applications," Technical Report, College of Business Administration, The University of Iowa, Iowa City, Iowa (1991), to appear in *Siam Review*.
- [26] N. Karmarkar, "A new polynomial time algorithm for linear programming," *Combinatorica* 4 (1984), 373-395.
- [27] M. Kojima, S. Mizuno and A. Yoshise, "A primal-dual interior point algorithm for linear programming," in: N. Megiddo, ed., *Progress in Mathematical Programming, Interior Point and Related Methods* (Springer-Verlag, New York, 1989), pp. 29-47.
- [28] M. Kojima, S. Mizuno and A. Yoshise, "A polynomial-time algorithm for a class of linear complementarity problems," *Mathematical Programming* 44 (1989), 1-26.
- [29] M. Kojima, S. Mizuno and A. Yoshise, "An $O(\sqrt{n} L)$ iteration potential reduction algorithm for linear complementarity problems," *Mathematical Programming* 50 (1991), 331-342.
- [30] K.O. Kortanek and M. Shi, "Convergence results and numerical experiments on a linear programming hybrid algorithm," *European Journal of Operational Research* 32 (1987), 47-61.
- [31] E. Kranich, "Interior point methods for mathematical programming: a bibliography," Discussion Paper 171, Institute of Economics and Operations Research, Fern University Hagen, Hagen, Germany (also available via netlib).

- [32] I.J. Lustig, R.E. Marston and D.F. Shanno, "The primal-dual interior point method on the Cray supercomputer," in: T.F. Coleman and Y. Li, eds., *Large-Scale Numerical Optimization*, (SIAM, Philadelphia, 1990), pp. 70-80.
- [33] S. Mizuno, M.J. Todd, and Y. Ye, "On adaptive step primal-dual interior-point algorithms for linear programming," Technical Report 944, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY, 1990, to appear in *Mathematics of Operations Research*.
- [34] C. Monma and A. Morton, "Computational experience with a dual affine variant of Karmarkar's method for linear programming," *Operations Research Letters* 6 (1987), 261-267.
- [35] R.C. Monteiro and I. Adler, "Interior path following primal-dual algorithms. Part I: Linear programming," *Mathematical Programming* 44 (1989), 27-42.
- [36] R.C. Monteiro and I. Adler, "Interior path following primal-dual algorithms. Part II: Convex quadratic programming," *Mathematical Programming* 44 (1989), 43-66.
- [37] M.J.D. Powell, "Karmarkar's algorithm: a view from nonlinear programming," *Bulletin of the Institute of Mathematics and its Applications* 26 (1990), 165-181.
- [38] M.J.D. Powell, "The complexity of Karmarkar's algorithm for linear programming," in: D.F. Griffiths and G.A. Watson, eds., *Numerical Analysis 1991*, Pitman Research Notes in Mathematics 260 (Longman, Burnt Hill, UK, 1992), pp. 143-163.
- [39] M.J.D. Powell, "On the number of iterations of Karmarkar's algorithm for linear programming," Technical Report DAMTP 1991/NA23, Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge (1991).
- [40] J. Renegar, "A polynomial-time algorithm based on Newton's method for linear programming," *Mathematical Programming* 40 (1988), 59-94.
- [41] C. Roos and J.-P. Vial, "A polynomial method of approximate centers for linear programming," *Mathematical Programming* 54 (1992), 295-305.
- [42] G. Sonnevend, J. Stoer, and G. Zhao, "On the complexity of following the central path of linear programs by linear extrapolation," *Methods of Operations Research* 63 (1989), 19-31.
- [43] A. Steger, "An extension of Karmarkar's algorithm for bounded linear programming problems," M.S. Thesis, SUNY at Stonybrook, New York (1985).
- [44] K. Tanabe, "Centered Newton method for mathematical programming," in: M. Iri and K. Yajima, eds., *System Modelling and Optimization*, Lecture Notes in Control and Information Sciences 113 (Springer-Verlag, Berlin, 1988), pp. 197-206.
- [45] M.J. Todd, "Recent developments and new directions in linear programming," in: M. Iri and K. Tanabe, editors, *Mathematical Programming: Recent Developments and Applications* (Kluwer Academic Press, Dordrecht, The Netherlands, 1989), pp. 109-157.
- [46] M.J. Todd and B.P. Burrell, "An extension of Karmarkar's algorithm for linear programming using dual variables," *Algorithmica* 1 (1986), 409-424.

- [47] M.J. Todd and Y. Ye, "A centered projective algorithm for linear programming," *Mathematics of Operations Research* 15 (1990), 508-529.
- [48] L. Tuncel and M.J. Todd, "Asymptotic behavior of interior-point methods: a view from semi-infinite programming," Technical Report No. 1031, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY, 1992.
- [49] P.M. Vaidya, "An algorithm for linear programming which requires $O(((m+n)n^2 + (m+n)^{1.5}n)L)$ arithmetic operations," *Mathematical Programming* 47 (1990), 175-202.
- [50] R.J. Vanderbei, M.S. Meketon, and B.A. Freedman, "A modification of Karmarkar's linear programming algorithm," *Algorithmica* 1 (1986), 395-407.
- [51] H. Yamashita, "A polynomially and quadratically convergent method for linear programming," Working paper, Mathematical Systems Institute, Inc., Tokyo, Japan, 1986.
- [52] Y. Ye, "An $O(n^3L)$ potential reduction algorithm for linear programming," *Mathematical Programming* 50 (1991), 239-258.
- [53] Y. Ye and M. Kojima, "Recovering optimal dual solutions in Karmarkar's polynomial algorithm for linear programming," *Mathematical Programming* 39 (1987), 305-317.