Mathematical Programming manuscript No.
(will be inserted by the editor)

Shinji Mizuno · Michael J. Todd

On two homogeneous self-dual approaches to linear programming and its extensions *

Received: date / Revised version: date

Abstract. We investigate the relation between interior-point algorithms applied to two homogeneous self-dual approaches to linear programming, one of which was proposed by Ye, Todd, and Mizuno and the other by Nesterov, Todd, and Ye. We obtain only a partial equivalence of path-following methods (the centering parameter for the first approach needs to be equal to zero or larger than one half), whereas complete equivalence of potential-reduction methods can be shown. The results extend to self-scaled conic programming and to semidefinite programming using the usual search directions.

1. Introduction

Ye, Todd, and Mizuno [24] presented a homogeneous and self-dual interior-point algorithm for solving linear programming (LP) problems. The algorithm can start from arbitrary (infeasible) interior points and achieves the best known com-

Shinji Mizuno: The Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-ku, Tokyo 106, Japan (`mizuno@ism.ac.jp`). Research supported in part by the Ministry of Education, Science, Sports and Culture through the Grant-in-Aid for Scientific Research (C) 08680478.

Michael J. Todd: School of Operations Research and Industrial Engineering, Cornell University, Ithaca, New York 14853, USA (`miketodd@cs.cornell.edu`). Research supported in part by NSF through grant DMS-9505155 and ONR through grant N00014-96-1-0050.

Mathematics Subject Classification (1991): 90C05, 90C22, 90C51

plexity in terms of the number of iterations without using a big initial constant. Recently, Nesterov, Todd, and Ye [18] proposed another type of homogeneous and self-dual interior-point algorithm for solving nonlinear conic problems. (We will explain the sense in which these algorithms are homogeneous in Section 2.) Although the self-dual system treated in [18] resembles that in [24], the algorithm seems rather different from [24], because it generates a sequence of points along the central path such that the parameter of the "duality gap" diverges. Indeed, Nesterov, Todd, and Ye seek a recession direction of a convex set, as in the primal method of de Ghellinck and Vial [2], but from a primal-dual perspective.

Here we investigate the relation of the central paths, their neighborhoods, and the algorithms in [24] and [18]. We mainly consider linear programming, but then extend our results to self-scaled conic programming and to semidefinite programming.

It is easily seen that there exists a bijection from the feasible region of the homogeneous and self-dual LP in [24] to the solution set of the homogeneous and self-dual system in [18]. We show that this map transforms the central paths and their neighborhoods in [24] to the corresponding ones in [18]. However, since the bijection is a projective map (such maps were used in Karmarkar's original method [6]), and hence nonlinear, it is not clear that algorithms for the two approaches will correspond, as they are based on linearizations. We obtain a partial equivalence of the search directions of the path-following algorithms, that is, the set of search directions for path-following algorithms in [18] corresponds to the set of those in [24] only when the centering parameter for the latter is zero or greater than a half. When the parameter is between zero and a half, we can only define a corresponding direction mathematically, but it has no interpretation as the search direction of a path-following method. In contrast to this partial equivalence, we show complete equivalence of the potential-reduction algorithms.

2. Two self-dual systems

We consider the linear programming problem in standard form:

$$\begin{aligned} &\text{minimize} \quad c^T x \\ &\text{subject to } Ax = b, \quad x \geq 0, \end{aligned}$$

where A is an $m \times n$ matrix, $b \in \mathbf{R}^m$, $c \in \mathbf{R}^n$, and $x \in \mathbf{R}^n$. We assume that the rank of A is m. The dual of this primal problem is defined by

$$\begin{aligned} &\text{maximize} \quad b^T y \\ &\text{subject to } A^T y + s = c, \quad s \geq 0, \end{aligned}$$

where $y \in \mathbf{R}^m$ and $s \in \mathbf{R}^n$ are variables.

We consider first the homogeneous and self-dual LP introduced by Ye, Todd, and Mizuno [24]:

$$(HSDP_1) \qquad \begin{array}{lrrrrrrl} \text{minimize} & & & & h^0\theta & & & \\ \text{subject to} & & Ax & -b\tau & +b^0\theta & & & = 0, \\ & -A^T y & & +c\tau & -c^0\theta & -s & & = 0, \\ & b^T y & -c^T x & & -g^0\theta & & -\kappa & = 0, \\ & -(b^0)^T y & +(c^0)^T x & +g^0\tau & & & & = -h^0, \\ & & x \geq 0, & \tau \geq 0, & & s \geq 0, & \kappa \geq 0, & \end{array} \tag{1}$$

where

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

for an initial interior point $(y^0, x^0, \tau^0, \theta^0, s^0, \kappa^0)$. (That is, τ^0, κ^0, and all components of x^0 and s^0 are positive.) We set $\theta^0 := 1$. (In [24], $\tau^0 = 1$ and $\kappa^0 = 1$ are used in addition.) It is easy to see that (1) is self-dual. We call it homogeneous because, with the exception of the final normalizing equation, its constraint system is homogeneous. Also, τ is a homogenizing variable, allowing the right-hand sides b and c to be moved to the left. By multiplying by y^T, x^T, τ, and θ the first, second, third, and fourth equality sets of constraints and summing them up (most of the terms in the left side vanish because of the skew-symmetry of the coefficient matrix), we see that

$$x^T s + \tau\kappa = \theta h^0 = \theta((x^0)^T s^0 + \tau^0 \kappa^0). \tag{2}$$

So the value of the "duality gap" $x^T s + \tau\kappa$ is linear with respect to θ. The optimal value of (HSDP$_1$) is 0, and from (2) the complementarity condition

$$X^* s^* = 0,\ \tau^* \kappa^* = 0$$

holds at any optimal solution $(y^*, x^*, \tau^*, 0, s^*, \kappa^*)$. We are interested in a strictly complementary solution. If $\tau^* > 0$ then x^*/τ^* and $(y^*, s^*)/\tau^*$ are optimal solutions of the original primal and dual LP, respectively. If $\kappa^* > 0$ then we can detect infeasibility of the primal or dual LP. See [24].

Next we consider the self-dual problem introduced by Nesterov, Todd, and Ye [18]:

$$(HSDP_2) \qquad \begin{array}{rrrrrl} & Ax & -b\tau & & & = -b^0, \\ -A^T y & & +c\tau & -s & & = c^0, \\ b^T y & -c^T x & & & -\kappa & = g^0, \\ \multicolumn{6}{c}{x \geq 0,\ \tau \geq 0,\ s \geq 0,\ \kappa \geq 0,} \end{array} \tag{3}$$

where b^0, c^0, and g^0 are as above. We are interested in a recession direction of this problem. Here the linear system is not homogeneous, but again τ is a homogenizing variable; also a recession direction is a solution to the corresponding homogeneous system. Hence we call this approach and the resulting algorithms homogeneous and self-dual too. If $(y^*, x^*, \tau^*, s^*, \kappa^*)$ is such a recession direction, then again skew-symmetry implies that the complementarity condition

$$X^* s^* = 0,\ \tau^* \kappa^* = 0$$

holds. Given such a direction with either $\tau^* > 0$ or $\kappa^* > 0$, we can again extract optimal solutions or a certificate of infeasibility for the original primal or dual LP (see [18]). The fact that we seek a recession direction of a polyhedron gives hope that algorithms for this formulation will be effective, since there appears to be "more room" to find such a direction. However, if the original LP problems both have unique optimal solutions, then the recession direction is unique, and in any case, the complementarity condition must hold, so the set of recession directions is low-dimensional. Indeed, it is easy to see that recession directions of (HSDP_2) suitably scaled are exactly optimal solutions of (HSDP_1) with $\theta = 0$ omitted, and vice versa. Our aim is to extend this analogy between these two homogeneous formulations.

Define the strictly feasible set of the Ye-Todd-Mizuno LP:

$$F_1 := \{(y_1, x_1, \tau_1, \theta_1, s_1, \kappa_1) : \text{strictly feasible solution of } (\mathrm{HSDP}_1)\}$$

and the strictly feasible set of the Nesterov-Todd-Ye problem:

$$F_2 := \{(y_2, x_2, \tau_2, s_2, \kappa_2) : \text{strictly feasible solution of } (\mathrm{HSDP}_2)\},$$

where strictly feasible means that all variables required to be nonnegative must be positive. Then we can define a map $\Phi : F_1 \to F_2$ by

$$\Phi(y_1, x_1, \tau_1, \theta_1, s_1, \kappa_1) := (y_1, x_1, \tau_1, s_1, \kappa_1)/\theta_1$$

for any $(y_1, x_1, \tau_1, \theta_1, s_1, \kappa_1) \in F_1$. Note that $(y^0, x^0, \tau^0, 1, s^0, \kappa^0) \in F_1$, and $\Phi(y^0, x^0, \tau^0, 1, s^0, \kappa^0) = (y^0, x^0, \tau^0, s^0, \kappa^0) \in F_2$.

Proposition 1. *The map Φ is a one-to-one and onto projective transformation from F_1 to F_2. Its inverse is given by*

$$\Phi^{-1}(y_2, x_2, \tau_2, s_2, \kappa_2) = \theta(y_2, x_2, \tau_2, 1, s_2, \kappa_2), \qquad \theta = h^0/(x_2^T s_2 + \tau_2 \kappa_2)$$

for any $(y_2, x_2, \tau_2, s_2, \kappa_2) \in F_2$,

Proof: Straightforward: we only make a few remarks. First, h^0 is positive, so that for any $(y_1, x_1, \tau_1, \theta_1, s_1, \kappa_1) \in F_1$, θ_1 is positive by (2). Conversely, $x_2^T s_2 + \tau_2 \kappa_2$ is clearly positive for any $(y_2, x_2, \tau_2, \theta_2, s_2, \kappa_2) \in F_2$. Note that the value of θ is determined so that (2) holds for $(x, \tau, s, \kappa) = \theta(x_2, \tau_2, s_2, \kappa_2)$, and this implies that the last constraint of (HSDP_1) is satisfied. Note also that the denominator of θ is linear on F_2, despite its appearance. By the skew-symmetry of the constraints of (HSDP_2), $x_2^T s_2 + \tau_2 \kappa_2 = (b^0)^T y_2 - (c^0)^T x_2 - g^0 \tau$. $\square$

We are interested in the relationship between the central paths, their neighborhoods, and especially algorithms for these two self-dual formulations, and whether they correspond under the bijection Φ.

3. The central paths and their neighborhoods

Define the central paths

$$P_1 := \{(y_1, x_1, \tau_1, \theta_1, s_1, \kappa_1) \in F_1 : X_1 s_1 = \mu_1 e, \tau_1 \kappa_1 = \mu_1, \text{ for some } \mu_1 > 0\}$$

and

$$P_2 := \{(y_2, x_2, \tau_2, s_2, \kappa_2) \in F_2 : X_2 s_2 = \mu_2 e, \tau_2 \kappa_2 = \mu_2, \text{ for some } \mu_2 > 0\},$$

where X denotes the diagonal matrix containing the components of x down its diagonal (we use S similarly) and e denotes the vector of ones in R^n. We denote the points on P_1 and P_2 by $(y_1(\mu_1), x_1(\mu_1), \tau_1(\mu_1), \theta_1(\mu_1), s_1(\mu_1), \kappa_1(\mu_1))$ and $(y_2(\mu_2), x_2(\mu_2), \tau_2(\mu_2), s_2(\mu_2), \kappa_2(\mu_2))$ respectively for parameters μ_1 and μ_2. Let

$$\mu^0 := ((x^0)^T s^0 + \tau^0 \kappa^0)/(n + 1).$$

From (2), $\theta_1(\mu_1) = \mu_1/\mu^0$. Hence

Proposition 2. *We have* $P_2 = \Phi(P_1)$*. More precisely*

$$\Phi(y_1(\mu_1), x_1(\mu_1), \tau_1(\mu_1), \theta_1(\mu_1), s_1(\mu_1), \kappa_1(\mu_1)) =$$
$$(y_2(\mu_2), x_2(\mu_2), \tau_2(\mu_2), s_2(\mu_2), \kappa_2(\mu_2))$$

for $\mu_2 = \mu_1/(\theta_1(\mu_1))^2 = \mu^0/\theta_1(\mu_1) = (\mu^0)^2/\mu_1$.

The **proof** is immediate and omitted. □

The path-following interior-point algorithm of [24] (and that of Xu, Hung, and Ye [22]) generates a sequence of points in F_1 approximating the path P_1 so that θ_1 and μ_1 approach 0. On the other hand, the path-following algorithms of [18] generate a sequence of points in F_2 approximating the path P_2 so that μ_2 goes to ∞.

Let $\beta \in (0, 1)$ be a constant. Define the neighborhoods:

$$N_1(\beta) := \{(y_1, x_1, \tau_1, \theta_1, s_1, \kappa_1) \in F_1 : \|(X_1 s_1 - \mu_1 e, \tau_1\kappa_1 - \mu_1)\|_p \le \beta\mu_1,$$
$$\mu_1 = (x_1^T s_1 + \tau_1\kappa_1)/(n + 1), \mu_1 > 0\}.$$

and

$$N_2(\beta) := \{(y_2, x_2, \tau_2, s_2, \kappa_2) \in F_2 : \|(X_2 s_2 - \mu_2 e, \tau_2\kappa_2 - \mu_2)\|_p \le \beta\mu_2,$$
$$\mu_2 = (x_2^T s_2 + \tau_2\kappa_2)/(n + 1), \mu_2 > 0\},$$

where $\|\cdot\|_p$ is the ℓ_p-norm for $p \in [1, \infty]$ (or the so-called one-sided ℓ_∞ norm). It is easy to prove

Proposition 3. $N_2(\beta) = \Phi(N_1(\beta))$.

□

Indeed, the same argument shows that points in the boundary of $N_1(\beta)$ with θ_1 positive correspond one-to-one under Φ to points in the boundary of $N_2(\beta)$ with $x_2^T s_2 + \tau_2\kappa_2$ positive also.

4. Directions for path-following methods

Suppose that we are solving the problem (HSDP$_1$) by a path-following method. Let $(y_1, x_1, \tau_1, \theta_1, s_1, \kappa_1) \in F_1$ be the current iterate. Let $\mu_1 = (x_1^T s_1 + \tau_1\kappa_1)/(n+1)$ and let γ_1 be a constant; typically $\gamma_1 \in [0, 1]$. Consider the nonlinear system defining the center on P_1 corresponding to parameter value $\mu = \gamma_1\mu_1$:

$$\begin{array}{rrrrrrl}
 & Ax & -b\tau & +b^0\theta & & & = 0, \\
-A^T y & & +c\tau & -c^0\theta & -s & & = 0, \\
b^T y & -c^T x & & -g^0\theta & & -\kappa & = 0, \\
-(b^0)^T y & +(c^0)^T x & +g^0\tau & & & & = -h^0, \\
 & & & & Xs & & = \gamma_1\mu_1 e, \\
 & & & & & \tau\kappa & = \gamma_1\mu_1.
\end{array}$$

Then we compute the Newton step for this system at the current iterate:

$$\begin{array}{rrrrrrl}
 & A\Delta x_1 & -b\Delta\tau_1 & +b^0\Delta\theta_1 & & & = 0, \\
-A^T\Delta y_1 & & +c\Delta\tau_1 & -c^0\Delta\theta_1 & -\Delta s_1 & & = 0, \\
b^T\Delta y_1 & -c^T\Delta x_1 & & -g^0\Delta\theta_1 & & -\Delta\kappa_1 & = 0, \\
-(b^0)^T\Delta y_1 & +(c^0)^T\Delta x_1 & +g^0\Delta\tau_1 & & & & = 0, \\
 & S_1\Delta x_1 & & & +X_1\Delta s_1 & & = -X_1 s_1 + \gamma_1\mu_1 e, \\
 & & \kappa_1\Delta\tau_1 & & & +\tau_1\Delta\kappa_1 & = -\tau_1\kappa_1 + \gamma_1\mu_1.
\end{array} \tag{4}$$

We compute the next iterate by

$$(y_1^+, x_1^+, \tau_1^+, \theta_1^+, s_1^+, \kappa_1^+) := (y_1, x_1, \tau_1, \theta_1, s_1, \kappa_1) + \alpha_1(\Delta y_1, \Delta x_1, \Delta\tau_1, \Delta\theta_1, \Delta s_1, \Delta\kappa_1)$$

for some positive α_1. Here α_1 is chosen so that the new iterate either solves (HSDP$_1$) (feasible with $\theta_1 = 0$) or lies in F_1; it is easy to see that α_1 can be chosen to be positive.

The next result is essentially due to Xu, Hung, and Ye [22].

Lemma 1. *We have*

$$\Delta\theta_1 = -(1-\gamma_1)\theta_1.$$

Proof: From (2), we have

$$(x_1 + \alpha\Delta x_1)^T(s_1 + \alpha\Delta s_1) + (\tau_1 + \alpha\Delta\tau_1)(\kappa_1 + \alpha_1\Delta\kappa_1) = (\theta_1 + \alpha\Delta\theta_1)h^0$$

for any $\alpha \in [0, \alpha_1]$. So the coefficients of α in both sides are equal, and we have

$$\begin{aligned}
\Delta\theta_1 h^0 &= s_1^T\Delta x_1 + x_1^T\Delta s_1 + \kappa_1\Delta\tau_1 + \tau_1\Delta\kappa_1 \\
&= -x_1^T s_1 - \tau_1\kappa_1 + (n+1)\gamma_1\mu_1 \\
&= -(n+1)(1-\gamma_1)\mu_1.
\end{aligned}$$

Since $h^0 = (n+1)\mu^0 = (n+1)\mu_1/\theta_1$, we have the result. $\square$

Let $(y_2, x_2, \tau_2, s_2, \kappa_2) = \Phi(y_1, x_1, \tau_1, \theta_1, s_1, \kappa_1) \in F_2$. Define $\mu_2 = (x_2^T s_2 + \tau_2\kappa_2)/(n+1) = \mu^0/\theta_1$. Let γ_2 be a constant; typically but not necessarily we will have $\gamma_2 \geq 1$. Consider the system defining the center on P_2 for $\mu = \gamma_2\mu_2$:

$$\begin{array}{rrrrr} & Ax & -b\tau & & = -b^0, \\ -A^T y & & +c\tau & -s & = c^0, \\ b^T y & -c^T x & & -\kappa & = g^0, \\ & & & Xs & = \gamma_2\mu_2 e, \\ & & & \tau\kappa & = \gamma_2\mu_2. \end{array}$$

Compute the Newton step for this system at $(y_2, x_2, \tau_2, s_2, \kappa_2)$:

$$\begin{array}{rrrrrl} & A\Delta x_2 & -b\Delta\tau_2 & & & = 0, \\ -A^T\Delta y_2 & & +c\Delta\tau_2 & -\Delta s_2 & & = 0, \\ b^T\Delta y_2 & -c^T\Delta x_2 & & & -\Delta\kappa_2 & = 0, \\ & S_2\Delta x_2 & & +X_2\Delta s_2 & & = -X_2 s_2 + \gamma_2\mu_2 e, \\ & & \kappa_2\Delta\tau_2 & & +\tau_2\Delta\kappa_2 & = -\tau_2\kappa_2 + \gamma_2\mu_2. \end{array} \tag{5}$$

We compute the next point:

$$(y_2^+, x_2^+, \tau_2^+, s_2^+, \kappa_2^+) := (y_2, x_2, \tau_2, s_2, \kappa_2) + \alpha_2(\Delta y_2, \Delta x_2, \Delta\tau_2, \Delta s_2, \Delta\kappa_2)$$

for some nonzero (typically, but not necessarily, positive) α_2 so that $(y_2^+, x_2^+, \tau_2^+, s_2^+, \kappa_2^+) \in F_2$.

We make two remarks about (5) in passing. First, suppose we solve this system and the resulting step satisfies the nonnegativity constraints; then it is itself a recession direction for (HSDP_2), and we can terminate the algorithm.

However, this is quite unlikely; more usually, an approximate recession direction is obtained by scaling down the current iterate, e.g. by applying the mapping Φ^{-1}. Secondly, while we are using this system to compute a search direction for (HSDP$_2$), Xu and Ye ([23], Section 4) solve the system (5) to obtain the search direction for their generalized HSD method for solving a linear feasibility system related to (HSDP$_1$). However, a transformation is applied to the solution to obtain their search direction, and it is not hard to see that their search direction is in fact the solution to their system (2)–(3) with parameters $\gamma = \gamma_2/(2\gamma_2 - 1)$ and $\eta := (\gamma_2 - 1)/(2\gamma_2 - 1) = 1 - \gamma$ after scaling. Note that, however large γ_2 is, the resulting γ is at least $1/2$; compare with Lemma 2 below.

Define

$$(\Delta y, \Delta x, \Delta\tau, \Delta s, \Delta\kappa) := \Phi(y_1^+, x_1^+, \tau_1^+, \theta_1^+, s_1^+, \kappa_1^+) - \Phi(y_1, x_1, \tau_1, \theta_1, s_1, \kappa_1).$$

(We assume henceforth that $\theta_1^+ > 0$, i.e., that $\alpha_1(1 - \gamma_1) < 1$. If not, then the algorithm has obtained an exact optimal solution to (HSDP$_1$). Using a limiting argument with α_1 approaching $1/(1 - \gamma_1)$ from below, we can then see using the following argument that the appropriate value for γ_2 gives a search direction that is a recession direction for (HSDP$_2$), so the second algorithm also terminates.) We investigate the relation between $(\Delta y, \Delta x, \Delta\tau, \Delta s, \Delta\kappa)$ and $(\Delta y_2, \Delta x_2, \Delta\tau_2, \Delta s_2, \Delta\kappa_2)$. Note that $(\Delta y, \Delta x, \Delta\tau, \Delta s, \Delta\kappa)$ is nonlinear with respect to γ_1 and α_1, while $(\Delta y_2, \Delta x_2, \Delta\tau_2, \Delta s_2, \Delta\kappa_2)$ is linear with respect to γ_2.

We see that

$$A\Delta x - \Delta\tau b = A(x_1^+/\theta_1^+ - x_1/\theta_1) - (\tau_1^+/\theta_1^+ - \tau_1/\theta_1)b$$

$$
\begin{aligned}
&= (1/\theta_1^+)(Ax_1^+ - \tau_1^+ b) - (1/\theta_1)(Ax_1 - \tau_1 b)\\
&= (1/\theta_1^+)((Ax_1 - \tau_1 b) + \alpha_1(A\Delta x_1 - \Delta\tau_1 b)) - (1/\theta_1)(Ax_1 - \tau_1 b)\\
&= (1/\theta_1^+)(-\theta_1 b^0 - \alpha_1\Delta\theta_1 b^0) + (1/\theta_1)\theta_1 b^0\\
&= -b^0 + b^0\\
&= 0
\end{aligned}
$$

and similarly

$$
\begin{aligned}
-A^T\Delta y + \Delta\tau c - \Delta s &= 0,\\
b^T\Delta y - c^T\Delta x - \Delta\kappa &= 0.
\end{aligned}
$$

Indeed, these equations must be satisfied because $(\Delta y, \Delta x, \Delta\tau, \Delta s, \Delta\kappa)$ is the difference between two points in F_2. We also see that

$$
\begin{aligned}
&S_2\Delta x + X_2\Delta s \qquad (6)\\
&= (1/\theta_1)(S_1\Delta x + X_1\Delta s)\\
&= (1/\theta_1)(S_1(x_1^+/\theta_1^+ - x_1/\theta_1) + X_1(s_1^+/\theta_1^+ - s_1/\theta_1))\\
&= (1/\theta_1)((1/\theta_1^+)(S_1x_1^+ + X_1s_1^+) - (1/\theta_1)(S_1x_1 + X_1s_1))\\
&= (1/\theta_1\theta_1^+)(S_1(x_1 + \alpha_1\Delta x_1) + X_1(s_1 + \alpha_1\Delta s_1)) - (S_1x_1 + X_1s_1)/(\theta_1)^2\\
&= (1/\theta_1\theta_1^+)(2X_1s_1 + \alpha_1(S_1\Delta x_1 + X_1\Delta s_1)) - 2X_1s_1/(\theta_1)^2\\
&= \frac{1}{(1-\alpha_1(1-\gamma_1))(\theta_1)^2}(2X_1s_1 + \alpha_1(-X_1s_1 + \gamma_1\mu_1 e)) - \frac{2X_1s_1}{(\theta_1)^2} \qquad (7)\\
&= \frac{1}{(1-\alpha_1(1-\gamma_1))(\theta_1)^2}((2 - \alpha_1 - 2(1-\alpha_1(1-\gamma_1))X_1s_1 + \alpha_1\gamma_1\mu_1 e)\\
&= \frac{\alpha_1}{(1-\alpha_1(1-\gamma_1))(\theta_1)^2}(-(2\gamma_1 - 1)X_1s_1 + \gamma_1\mu_1 e)\\
&= \frac{\alpha_1(2\gamma_1 - 1)}{1-\alpha_1(1-\gamma_1)}\left(-X_2s_2 + \frac{\gamma_1}{2\gamma_1 - 1}\frac{\mu_1}{(\theta_1)^2}e\right)
\end{aligned}
$$

$$= \frac{\alpha_1(2\gamma_1 - 1)}{1 - \alpha_1(1 - \gamma_1)} \left(-X_2 s_2 + \frac{\gamma_1}{2\gamma_1 - 1} \mu_2 e \right) \tag{8}$$

and similarly

$$\kappa_2 \Delta\tau + \tau_2 \Delta\kappa = \frac{\alpha_1(2\gamma_1 - 1)}{1 - \alpha_1(1 - \gamma_1)} \left(-\tau_2 \kappa_2 + \frac{\gamma_1}{2\gamma_1 - 1} \mu_2 \right). \tag{9}$$

The next result is the key technical tool in our analysis.

Lemma 2. *Suppose* $\gamma_1 \neq 1/2$ *and*

$$\gamma_2 = \frac{\gamma_1}{2\gamma_1 - 1}. \tag{10}$$

Then $(\Delta y, \Delta x, \Delta\tau, \Delta s, \Delta\kappa)$ *is parallel to* $(\Delta y_2, \Delta x_2, \Delta\tau_2, \Delta s_2, \Delta\kappa_2)$*, with the two vectors pointing in the same direction if* $\gamma > 1/2$ *and in opposite directions if* $\gamma_1 < 1/2$*. If in addition*

$$\alpha_2 = \frac{\alpha_1(2\gamma_1 - 1)}{1 - \alpha_1(1 - \gamma_1)} \tag{11}$$

then $(y_2^+, x_2^+, \tau_2^+, s_2^+, \kappa_2^+) = \Phi(y_1^+, x_1^+, \tau_1^+, \theta_1^+, s_1^+, \kappa_1^+)$.

(Note that as γ_1 ranges from 1 down to $1/2$, γ_2 ranges from 1 up to $+\infty$; as γ_1 goes from $1/2$ down to 0, γ_2 goes from $-\infty$ up to 0. Also, α_2 is always nonzero, but negative if $0 \leq \gamma_1 < 1/2$.)

Proof: Since the solution of (5) is unique, when (10) holds, we have by (8) and (9) that

$$(\Delta y, \Delta x, \Delta\tau, \Delta s, \Delta\kappa) = \frac{\alpha_1(2\gamma_1 - 1)}{1 - \alpha_1(1 - \gamma_1)} (\Delta y_2, \Delta x_2, \Delta\tau_2, \Delta s_2, \Delta\kappa_2).$$

If both conditions (10) and (11) hold then

$$(\Delta y, \Delta x, \Delta\tau, \Delta s, \Delta\kappa) = \alpha_2 (\Delta y_2, \Delta x_2, \Delta\tau_2, \Delta s_2, \Delta\kappa_2).$$

So

$$
\begin{aligned}
\Phi(y_1^+, x_1^+, \tau_1^+, \theta_1^+, s_1^+, \kappa_1^+) &= \Phi(y_1, x_1, \tau_1, \theta_1, s_1, \kappa_1) + (\Delta y, \Delta x, \Delta\tau, \Delta s, \Delta\kappa) \\
&= (y_2, x_2, \tau_2, s_2, \kappa_2) + \alpha_2(\Delta y_2, \Delta x_2, \Delta\tau_2, \Delta s_2, \Delta\kappa_2) \\
&= (y_2^+, x_2^+, \tau_2^+, s_2^+, \kappa_2^+).
\end{aligned}
$$

□

Let us now examine the consequences of this lemma for path-following methods. Clearly, if the starting points of two methods for (HSDP_1) and (HSDP_2) correspond via Φ, and at each iteration we choose centering parameters for the two algorithms that are related by (10) and step sizes that are related by (11), then each pair of iterates corresponds via Φ. The initial iterates are usually chosen as $(y^0, x^0, \tau^0, 1, s^0, \kappa^0) \in F_1$ and $(y^0, x^0, \tau^0, s^0, \kappa^0) \in F_2$, and these do indeed correspond under Φ.

For a short-step path-following method driving μ to zero, we choose $\gamma_1 = 1-\omega/\sqrt{n}$ for some ω of order 1. Then, as long as n is above some (usually small) threshold, we have $\gamma_1 > 1/2$, and the corresponding γ_2 is $(1-\omega/\sqrt{n})/(1-2\omega/\sqrt{n}) \approx 1+\omega/\sqrt{n}$, corresponding to a short-step path-following method for the second formulation. However, short-step methods usually employ full Newton steps, and we note that $\alpha_1 = 1$ corresponds to $\alpha_2 = 2-1/\gamma_1 < 1$ for any $\gamma_1 < 1$. Hence such short-step path-following methods do not yield exactly corresponding iterates, although they do in the limit as n tends to ∞.

Since neighborhoods of the two central paths correspond, we might expect that methods that choose step sizes based on the longest step that remains within such a neighborhood would give iterates that correspond. We examine three such methods. The first is the predictor-corrector method of Mizuno-Todd-Ye, which uses two neighborhoods, say $N_1(\beta_S)$ and $N_1(\beta_L)$, $0 < \beta_S < \beta_L < 1$, defined by the ℓ_2-norm. Suppose our current iterates for the two approaches correspond, and lie in the smaller neighborhoods $N_1(\beta_S)$ and $N_2(\beta_S)$. The predictor step (for the first approach) is defined using a centering parameter $\gamma_1 = 0$, and as we have seen, this corresponds to $\gamma_2 = 0$, and, with $\alpha_1 < 1$, to a negative value of α_2. Hence for the second approach, we take the negative of the affine-scaling step, as suggested for a predictor step in Section 5.4 of [18]. In both approaches, we take the longest step along these directions while remaining in the wider of the two neighborhoods, $N_1(\beta_L)$ and $N_2(\beta_L)$. By Proposition 3, the results of these two predictor steps will correspond under Φ. Next we take a single corrector step, with $\gamma_1 = 1$ and $\alpha_1 = 1$; and this corresponds exactly to a single full corrector step in the second approach, since then $\gamma_2 = 1$ and $\alpha_2 = 1$. Hence the two iterates will again correspond under Φ.

The second algorithm is the "largest-step" path-following method of Monteiro and Adler [14] and Mizuno, Yoshise, and Kikuchi [12]. This method (for the first approach) chooses $\alpha_1 = 1$ and the smallest γ_1 so that the resulting iterate lies in a particular neighborhood $N_1(\beta)$ defined by the ℓ_2-norm. For the second approach, we modify this to choose $\alpha_2 = 1$ and the largest γ_2 so that the resulting iterate lies in the corresponding neighborhood $N_2(\beta)$. However, since $\alpha_1 = 1$ and $\alpha_2 = 1$ can never correspond unless $\gamma_1 = 1$, these two largest-step methods do not correspond.

Finally, we consider the long-step method for the first formulation that uses a fixed γ_1 and chooses the longest step size so that the new iterate remains in a fixed neighborhood $N_1(\beta)$ defined by a certain norm. Then, as long as $\gamma_1 > 1/2$, this method does give iterates that correspond under Φ to those generated by a similar method for the second approach using the neighborhood $N_2(\beta)$ defined by the same norm, with fixed $\gamma_2 = \gamma_1/(2\gamma_1 - 1)$. However, such a method is usually implemented for a small value of γ_1, say 1/10, which is less than the critical value 1/2. Note that a value of γ_2 that is large, say 500.5, seems to correspond to a long-step method for the second approach. But the corresponding value for γ_1 is .5005, which is larger than 1/2 and would not generally be viewed as a long-step method for the first approach. Moreover, a step size of 1 for α_1 for these parameters corresponds to the very short step size of $\alpha_2 = .002$, while $\alpha_2 = 1$ corresponds to the unusual choice of $\alpha_1 = 1.998$.

If we do choose $\gamma_1 = 1/10$, the corresponding value for γ_2 is $-1/8$, and the step size α_2 should be negative. This is similar to the predictor step above, but does not seem to correspond to a path-following method for the second approach. We will see later that this is however a reasonable choice for a potential-reduction method.

In summary, we have the following result:

Theorem 1. *Short-step and largest-step path-following algorithms do not give corresponding iterates when applied to (HSDP$_1$) and (HSDP$_2$). Predictor-corrector methods do give corresponding iterates, as do long-step path-following methods* *using parameters* $\gamma_1 > 1/2$ *and* $\gamma_2 = \gamma_1/(2\gamma_1 - 1)$ *and corresponding neighborhoods* $N_1(\beta)$ *and* $N_2(\beta)$.

□

Lemma 2 and Theorem 1 above indicate that γ_1 equal to 1/2 is a critical case. A different parametrization avoids this singularity, but substitutes one at $\gamma_1 = 0$. This reformulation will be useful in the next section. First we note that (10) can be rewritten as

$$\frac{1}{\gamma_1} + \frac{1}{\gamma_2} = 2. \tag{12}$$

Now let us write $\delta_i := 1/\gamma_i$, $i = 1, 2$, and write our directions in terms of the δ's. So in (4) we change the right-hand sides to $-\delta_1 X_1 s_1 + \mu_1 e$ and $-\delta_1 \tau_1 \kappa_1 + \mu_1$, and let the resulting direction be $(\overline{\Delta y}_1, \overline{\Delta x}_1, \overline{\Delta \tau}_1, \overline{\Delta \theta}_1, \overline{\Delta s}_1, \overline{\Delta \kappa}_1)$. Then, with step size $\bar{\alpha}_1 := \gamma_1 \alpha_1$, the next iterate is as before. We make a corresponding change in (5); note that if γ_2 is negative, then the search direction is oppositely directed and correspondingly α_2 and $\bar{\alpha}_2$ have opposite signs. The analysis now proceeds as before: for example, (7) becomes

$$S_2\overline{\Delta x} + X_2\overline{\Delta s} = \frac{1}{(1+\bar{\alpha}_1(1-\delta_1))(\theta_1)^2}(2X_1 s_1 + \bar{\alpha}_1(-\delta_1 X_1 s_1 + \theta_1 \mu^0 e)) - \frac{2X_1 s_1}{(\theta_1)^2}.$$

Thus we obtain

$$S_2\overline{\Delta x} + X_2\overline{\Delta s} = \frac{\bar{\alpha}_1}{1+\bar{\alpha}_1(1-\delta_1)}\left(-(2-\delta_1)X_2 s_2 + \mu_2 e\right)$$

and similarly

$$\kappa_2\overline{\Delta \tau} + \tau_2\overline{\Delta \kappa} = \frac{\bar{\alpha}_1}{1+\bar{\alpha}_1(1-\delta_1)}\left(-(2-\delta_1)\tau_2 \kappa_2 + \mu_2\right).$$

This shows that the iterates correspond as long as

$$\delta_1 + \delta_2 = 2 \tag{13}$$

and

$$\bar{\alpha}_2 = \frac{\bar{\alpha}_1}{1 + \bar{\alpha}_1(1 - \delta_1)}.$$

These are exactly the analogues of conditions (10) and (11).

5. Potential-reduction methods

We note at the outset that the standard primal-dual potential-reduction algorithm of Kojima, Mizuno, and Yoshise [9], which chooses its step sizes to decrease the potential function as much as possible at each iteration, may yield iterates whose limit points are not strictly complementary and hence, if $\tau^* = \kappa^* = 0$, of little use in the self-dual context. However, one can use a variant that guarantees strictly complementary solutions (Güler and Ye [4]) or just use the normal rule and hope for the best.

Define

$$f_1(y_1, x_1, \tau_1, \theta_1, s_1, \kappa_1) = (n + 1 + \eta_1) \ln(x_1^T s_1 + \tau_1 \kappa_1) - \ln(X_1 s_1) - \ln(\tau_1 \kappa_1)$$

for each $(y_1, x_1, \tau_1, \theta_1, s_1, \kappa_1) \in F_1$ and

$$f_2(y_2, x_2, \tau_2, s_2, \kappa_2) = (n + 1 + \eta_2) \ln(x_2^T s_2 + \tau_2 \kappa_2) - \ln(X_2 s_2) - \ln(\tau_2 \kappa_2)$$

for each $(y_2, x_2, \tau_2, s_2, \kappa_2) \in F_2$; here, $\ln(v)$ for a vector v denotes the sum of the logarithms of the components of v. With $\eta_1 > 0$ (say $\eta_1 = \sqrt{n+1}$), f_1 is a potential function used in algorithms for (HSDP$_1$). Similarly, [18] suggests driving to minus infinity the function f_2 in potential-reduction algorithms for (HSDP$_2$), but now with $\eta_2 < 0$ (perhaps $\eta_2 = -\sqrt{n+1}$).

Proposition 4. *For* $(y_1, x_1, \tau_1, \theta_1, s_1, \kappa_1) \in F_1$, *we have*

$$f_2(\Phi(y_1, x_1, \tau_1, \theta_1, s_1, \kappa_1)) = f_1(y_1, x_1, \tau_1, \theta_1, s_1, \kappa_1) + 2\eta_2 \ln h^0 \tag{14}$$

as long as

$$\eta_1 + \eta_2 = 0. \tag{15}$$

Proof: Note that

$$\begin{aligned}
&f_2(\Phi(y_1, x_1, \tau_1, \theta_1, s_1, \kappa_1)) \\
&= (n+1+\eta_2)\ln((x_1^T s_1 + \tau_1\kappa_1)/(\theta_1)^2) - \ln((X_1 s_1)/(\theta_1)^2) - \ln((\tau_1\kappa_1)/(\theta_1)^2) \\
&= (n+1+\eta_2)\ln(x_1^T s_1 + \tau_1\kappa_1) - \ln(X_1 s_1) - \ln(\tau_1\kappa_1) - 2\eta_2 \ln\theta_1 \\
&= (n+1-\eta_2)\ln(x_1^T s_1 + \tau_1\kappa_1) - \ln(X_1 s_1) - \ln(\tau_1\kappa_1) + 2\eta_2 \ln h^0,
\end{aligned}$$

where the last equation uses (2). □

Now let us examine potential-reduction methods for the two formulations. The gradient of f_1 with respect to $(x_1, s_1, \tau_1, \kappa_1)$ at $(y_1, x_1, \tau_1, \theta_1, s_1, \kappa_1) \in F_1$ is

$$\nabla f_1(y_1, x_1, \tau_1, \theta_1, s_1, \kappa_1) = \frac{n+1+\eta_1}{x_1^T s_1 + \tau_1\kappa_1} \begin{pmatrix} s_1 \\ x_1 \\ \kappa_1 \\ \tau_1 \end{pmatrix} - \begin{pmatrix} (x_1)^{-1} \\ (s_1)^{-1} \\ (\tau_1)^{-1} \\ (\kappa_1)^{-1} \end{pmatrix},$$

and similarly that of f_2 with respect to $(x_2, s_2, \tau_2, \kappa_2)$ at $(y_2, x_2, \tau_2, s_2, \kappa_2) \in F_2$ is

$$\nabla f_2(y_2, x_2, \tau_2, s_2, \kappa_2) = \frac{n+1+\eta_2}{x_2^T s_2 + \tau_2 \kappa_2} \begin{pmatrix} s_2 \\ x_2 \\ \kappa_2 \\ \tau_2 \end{pmatrix} - \begin{pmatrix} (x_2)^{-1} \\ (s_2)^{-1} \\ (\tau_2)^{-1} \\ (\kappa_2)^{-1} \end{pmatrix};$$

here $(v)^{-1}$ for a vector v denotes the vector of reciprocals of the components of v. Then the primal-dual-scaled steepest descent direction for f_1 for problem (HSDP_1) is the solution to (4) where we change the right-hand sides to $-\zeta_1 X_1 s_1 + e$ and $-\zeta_1 \tau_1 \kappa_1 + 1$, where

$$\zeta_1 := \frac{n+1+\eta_1}{x_1^T s_1 + \tau_1 \kappa_1}.$$

Let the resulting direction be denoted by $(\widetilde{\Delta y}_1, \widetilde{\Delta x}_1, \widetilde{\Delta \tau}_1, \widetilde{\Delta \theta}_1, \widetilde{\Delta s}_1, \widetilde{\Delta \kappa}_1)$. By scaling up this direction by μ_1, we get the direction $(\overline{\Delta y}_1, \overline{\Delta x}_1, \overline{\Delta \tau}_1, \overline{\Delta \theta}_1, \overline{\Delta s}_1, \overline{\Delta \kappa}_1)$ of the previous section, where

$$\delta_1 = \mu_1 \zeta_1 = \frac{n+1+\eta_1}{n+1} = 1 + \frac{\eta_1}{n+1}.$$

Similarly, the primal-dual-scaled steepest descent direction for f_2 for problem (HSDP_2) is the solution to (5) where we change the right-hand sides to$-\zeta_2 X_2 s_2 + e$ and $-\zeta_2 \tau_2 \kappa_2 + 1$, where

$$\zeta_2 := \frac{n+1+\eta_2}{x_2^T s_2 + \tau_2 \kappa_2}.$$

If the resulting direction is denoted by $(\widetilde{\Delta y}_2, \widetilde{\Delta x}_2, \widetilde{\Delta \tau}_2, \widetilde{\Delta s}_2, \widetilde{\Delta \kappa}_2)$, and we scale it up by μ_2, we get the direction $(\overline{\Delta y}_2, \overline{\Delta x}_2, \overline{\Delta \tau}_2, \overline{\Delta s}_2, \overline{\Delta \kappa}_2)$ of the previous section, where

$$\delta_2 = \mu_2 \zeta_2 = \frac{n+1+\eta_2}{n+1} = 1 + \frac{\eta_2}{n+1}.$$

We note that (13) holds as long as (15) does. Hence, for suitable step sizes, the next iterates of the potential-reduction methods will correspond as long as the present ones do, by the arguments of the previous section. Examples of suitable step sizes are those that minimize the appropriate potential functions, using Proposition 4, or those that minimize these functions subject to remaining within an appropriate neighborhood, using also Proposition 3. (Güler and Ye [4] show that algorithms of the latter type can still achieve a constant reduction in the potential function at each iteration, while having the property that all limit points are strictly complementary solutions.)

Let us look at some special cases. If $\eta_1 = \eta_2 = 0$, then the two potential functions can be viewed as proximity measures (to the respective central paths); in this case $\delta_1 = \delta_2 = 1$, corresponding to $\gamma_1 = \gamma_2 = 1$, and we get centering directions. If $\eta_1 = -\eta_2 = \sqrt{n+1}$, then δ_1, $\delta_2 = 1 \pm (n+1)^{-1/2}$, corresponding to γ_1, $\gamma_2 = 1/(1 \pm (n+1)^{-1/2})$. Thus the directions are those taken in a short-step path-following method, although the iterates are not required to lie in narrow neighborhoods of the central paths, and the step sizes are usually determined by a line search. If $\eta_1 = -\eta_2 = n+1$, then $\delta_1 = 2$ and $\delta_2 = 0$, corresponding to $\gamma_1 = 1/2$ and $\gamma_2 = \infty$. Notice that there is nothing singular about this case in the potential-reduction framework. (Indeed, the case $\eta_2 = -(n+1)$ is of interest, since then the potential function f_2 is exactly the barrier function.) Finally, for $\eta_1 = -\eta_2 > n+1$, $\delta_1 > 2$ and $\delta_2 < 0$, corresponding to $\gamma_1 < 1/2$ and $\gamma_2 < 0$

(with a change in the sense of the direction). If $\eta_1 = -\eta_2$ converges to ∞, γ_1 and γ_2 approach 0, from above and below respectively.

In summary, we have

Theorem 2. *Suppose potential-reduction methods are applied to the two formulations (HSDP*$_1$*) and (HSDP*$_2$*), with*

$$\eta_1 + \eta_2 = 0$$

and step sizes chosen to minimize the corresponding potential functions, either without constraints or subject to the iterates remaining in corresponding neighborhoods. Then if the initial iterates correspond under Φ, so will all subsequent iterates.

□

6. Extensions

Here we show that the results of the previous sections extend to several nonlinear programming problems. We consider self-scaled conic programming problems with the Nesterov-Todd direction [16,17] as well as semidefinite programming problems with several different directions (see, e.g., [20]). Extensions of the first homogeneous self-dual model or a related homogeneous feasibility problem were studied by Potra and Sheng [19], by de Klerk, Roos and Terlaky [7,8], and by Luo, Sturm, and Zhang [11]. The second homogeneous self-dual approach was already considered in a general conic setting in [18]. For these more general problems, x and s lie in more general (finite-dimensional real vector) spaces and are restricted to a closed convex cone and its dual; the inner product of two vectors in $\mathbf{R}^n$ is replaced by a scalar product $\langle s, x\rangle$ defined appropriately; and in (4) and (5) defining the search directions, the equation

$$S\Delta x + X\Delta s = -Xs + \gamma\mu e$$

with appropriate subscripts is replaced by

$$\mathcal{E}\Delta x + \mathcal{F}\Delta s = r_A + \gamma\mu r_C$$

with appropriate subscripts, where $\mathcal{E}$ and $\mathcal{F}$ are suitable operators defined on the appropriate spaces and r_A and r_C are suitable points in the appropriate spaces (the subscripts refer to affine-scaling and centering respectively), all depending on the current iterate (x, s). To express the dependence, we write $\mathcal{E}(x, s)$ etc. where necessary; we also write $\mathcal{E}_1$ for $\mathcal{E}(x_1, s_1)$ etc. Further, the dimension n is replaced with the parameter ν of the appropriate barrier function for the problem; thus

$$\mu := (\langle s, x\rangle + \tau\kappa)/(\nu + 1).$$

(We do not give full details of the extensions or the proofs, but those familiar with self-scaled conic or semidefinite programming should have no difficulty in filling these in.)

As an example, consider the Alizadeh-Haeberly-Overton (AHO) direction [1] for semidefinite programming. Then x and s lie in the space of symmetric matrices of order n, $\nu = n$, and

$$\langle s, x\rangle := s \bullet x := \mathrm{Trace}\,(s^T x) = \mathrm{Trace}\,(sx).$$

Also, the operators $\mathcal{E} = \mathcal{E}(x, s)$ and $\mathcal{F} = \mathcal{F}(x, s)$ are defined by

$$\mathcal{E}u := \frac{1}{2}(su + us), \qquad \mathcal{F}u := \frac{1}{2}(xu + ux).$$

Finally,

$$r_A := -\frac{1}{2}(xs + sx), \qquad r_C := i,$$

where i denotes the identity matrix of order n.

In order for our main results Theorems 1 and 2 to remain true in these more general situations, we only need Lemmas 1 and 2 to hold. Below we give conditions sufficient to assure this:

(i) there is some v such that

$$\mathcal{E}^* v = s, \quad \mathcal{F}^* v = x, \quad \langle v, r_A \rangle = -\langle s, x \rangle, \quad \langle v, r_C \rangle = \nu$$

(here $\mathcal{E}^*$ and $\mathcal{F}^*$ denote the adjoints of $\mathcal{E}$ and $\mathcal{F}$);

(ii) $\mathcal{E}x = \mathcal{F}s = -r_A$; and

(iii) there is some real π so that

$$\mathcal{E}(x/\theta, s/\theta) = \mathcal{E}(x, s)/(\theta)^{\pi},$$

$$r_A(x/\theta, s/\theta) = r_A(x, s)/(\theta)^{\pi+1}, \quad r_C(x/\theta, s/\theta) = r_C(x, s)/(\theta)^{\pi-1}.$$

We now have

Proposition 5. *Under conditions (i)–(iii) above, Lemmas 1 and 2 still hold.*

Proof: We consider first Lemma 1. Note first that the skew-symmetry of the appropriate operator means that we have the analogue

$$\langle s, x \rangle + \tau\kappa = \theta h^0 = \theta(\langle s^0, x^0 \rangle + \tau^0 \kappa^0)$$

of (2). From this we have, as in the proof of Lemma 1,

$$\langle s_1 + \alpha \Delta s_1, x_1 + \alpha \Delta x_1 \rangle + (\tau_1 + \alpha \Delta \tau_1)(\kappa_1 + \alpha_1 \Delta \kappa_1) = (\theta_1 + \alpha \Delta \theta_1) h^0$$

for any $\alpha \in [0, \alpha_1]$. Then the proof can proceed exactly as before if we establish

$$\langle s_1, \Delta x_1 \rangle + \langle \Delta s_1, x_1 \rangle = -\langle s_1, x_1 \rangle + \gamma_1 \mu_1 \nu.$$

However, the equation above follows from

$$\mathcal{E}_1 \Delta x_1 + \mathcal{F}_1 \Delta s_1 = r_{A1} + \gamma_1 \mu_1 r_{C1} \tag{16}$$

after we take the scalar product with v_1 and use (i).

Now we turn to Lemma 2. We need to establish the analogue of (8),

$$\mathcal{E}_2 \Delta x + \mathcal{F}_2 \Delta s = \frac{\alpha_1 (2\gamma_1 - 1)}{1 - \alpha_1 (1 - \gamma_1)} \left(r_{A2} + \frac{\gamma_1}{2\gamma_1 - 1} \mu_2 r_{C2} \right).$$

But this follows from

$$\begin{aligned}
&\mathcal{E}_2 \Delta x + \mathcal{F}_2 \Delta s \\
&= (1/(\theta_1)^\pi)(\mathcal{E}_1 \Delta x + \mathcal{F}_1 \Delta s) \text{ (using (iii))} \\
&= (1/(\theta_1)^\pi)(\mathcal{E}_1 (x_1^+/\theta_1^+ - x_1/\theta_1) + \mathcal{F}_1 (s_1^+/\theta_1^+ - s_1/\theta_1)) \\
&= (1/(\theta_1)^\pi)((1/\theta_1^+)(\mathcal{E}_1 x_1^+ + \mathcal{F}_1 s_1^+) - (1/\theta_1)(\mathcal{E}_1 x_1 + \mathcal{F}_1 s_1)) \\
&= (1/(\theta_1)^\pi \theta_1^+)(\mathcal{E}_1 (x_1 + \alpha_1 \Delta x_1) + \mathcal{F}_1 (s_1 + \alpha_1 \Delta s_1)) - (\mathcal{E}_1 x_1 + \mathcal{F}_1 s_1)/(\theta_1)^{\pi+1} \\
&= (1/(\theta_1)^{\pi+1}(1 - \alpha_1(1-\gamma_1))([1 - (1 - \alpha_1(1-\gamma_1))](\mathcal{E}_1 x_1 + \mathcal{F}_1 s_1) \\
&\quad + \alpha_1 (\mathcal{E}_1 \Delta x_1 + \mathcal{F}_1 \Delta s_1)) \text{ (using Lemma 1)} \\
&= \frac{\alpha_1}{(1 - \alpha_1(1-\gamma_1))(\theta_1)^{\pi+1}} ((1-\gamma_1)(\mathcal{E}_1 x_1 + \mathcal{F}_1 s_1) + r_{A1} + \gamma_1 \mu_1 r_{C1}) \\
&\qquad \text{(using (16))}
\end{aligned}$$

$$= \frac{\alpha_1}{(1-\alpha_1(1-\gamma_1))(\theta_1)^{\pi+1}}\left((2\gamma_1 - 1)r_{A1} + \gamma_1\mu_1 r_{C1}\right) \text{ (using (ii))}$$

$$= \frac{\alpha_1(2\gamma_1 - 1)}{1-\alpha_1(1-\gamma_1)}\left(r_{A2} + \frac{\gamma_1}{2\gamma_1 - 1}\mu_2 r_{C2}\right) \text{ (using (iii))}.$$

□

Corollary 1. *Lemmas 1 and 2 hold for self-scaled conic programming using the Nesterov-Todd direction.*

Proof: In this case we have $\mathcal{E} = F''(w)$, where F is the appropriate self-scaled barrier and w is the scaling point corresponding to the current iterates x and s, so that $F''(w)x = s$, and $\mathcal{F}$ is the identity operator. Also, $r_A = -s$ and $r_C = -F'(x)$. In this case, (i) holds with $v = x$ from basic properties of self-scaled barriers, (ii) clearly holds, and (iii) holds with $\pi = 0$. □

Corollary 2. *Lemmas 1 and 2 hold for semidefinite programming using any of the Alizadeh-Haeberly-Overton [1], Helmberg-Rendl-Vanderbei-Wolkowicz/Kojima-Shindoh-Hara/Monteiro [5,10,13], Nesterov-Todd [16,17], or Gu-Toh [3,21] directions.*

Proof: Indeed, all of the stated directions are members of the Monteiro-Zhang [13,25,15] family and are defined by

$$\mathcal{E} = s \odot m, \quad \mathcal{F} = mx \odot i, \quad r_A = -\frac{1}{2}(mxs + sxm), \text{ and } \quad r_C = m.$$

Here m is a symmetric positive definite matrix (different directions use different matrices m) and $u \odot v$ is defined by $(u \odot v)z := (uzv^T + vzu^T)/2$. See, e.g., [20]. Then it is easy to check that (i) holds for $v = m^{-1}$ (note that $\mathcal{F}^* = (mx)^T \odot i^T =$

$xm \odot i$) and (ii) holds. Moreover, (iii) holds with $\pi = 1$ if m is invariant to changes of scale in x and s (as it is for the AHO, NT, and GT choices), and with $\pi = 2$ if m scales with x and s (as it does for the HRVW/KSH/M choice). (It also holds with $\pi = 0$ if m scales inversely with x and s, as it does for the dual HRVW/KSH/M choice.) □

It is worth pointing out that, although the result above does indeed hold for the Helmberg-Rendl-Vanderbei-Wolkowicz/Kojima-Shindoh-Hara/Monteiro (HRVW/KSH/M) directions as defined by a suitable modification of (4) and (5), the former is *not* the search direction for the extension of the homogeneous self-dual problem (HSDP$_1$) using the HRVW/KSH/M direction! This seeming contradiction arises from the self-dual nature of this problem. Note that the extension of problem (HSDP$_1$) has variables $(y, X, \tau, \theta, S, \kappa)$, and its dual has variables $(\tilde{y}, \tilde{X}, \tilde{\tau}, \tilde{\theta}, \tilde{S}, \tilde{\kappa})$, with X corresponding to $\tilde{S}$, S to $\tilde{X}$, τ to $\tilde{\kappa}$, and κ to $\tilde{\tau}$. A primal-dual algorithm applied to this problem will use a linear system of twice the size of (4) to define the search directions $(\Delta y_1, \Delta X_1, \Delta\tau_1, \Delta\theta_1, \Delta S_1, \Delta\kappa_1)$ for the primal and $(\Delta\tilde{y}_1, \Delta\tilde{X}_1, \Delta\tilde{\tau}_1, \Delta\tilde{\theta}_1, \Delta\tilde{S}_1, \Delta\tilde{\kappa}_1)$ for the dual. However, if the current primal and dual iterates are the same, *and* if the algorithm is self-dual, then the primal and dual directions turn out to be the same and the linear system collapses into one of half the size, i.e., the extension of (4). The next primal and dual iterates will then also be the same. This holds for all the directions above except the HRVW/KSH/M direction (and its dual), which are not self-dual. Hence the convergence theory for the first homogeneous self-dual formulation that follows directly from the results for feasible primal-dual interior-point methods fails to apply automatically to this direction. Of course, corresponding convergence results may hold, but they need to be established separately.

In summary, the results of partial equivalence of path-following algorithms and complete equivalence of potential-reduction methods for the two homogeneous formulations still hold for a wide range of convex programming problems in conic form, with the proviso above.

Acknowledgements

We would like to thank the referees for their very helpful comments.

References

1. Alizadeh, F., Haeberly, J.A., Overton, M.L. (1998): Primal-dual interior-point methods for semidefinite programming: convergence rates, stability and numerical results. SIAM J. Optim. **8,** 746–768
2. de Ghellinck, G., Vial, J.-P. (1986): A polynomial Newton method for linear programming. Algorithmica **1,** 425–454
3. Gu, M. (1997): On primal-dual interior point methods for semidefinite programming. CAM report 97-12, Department of Mathematics, University of California, Los Angeles, California, USA
4. Güler, O., Ye, Y. (1996): Convergence behavior of interior point algorithms. Math. Program. **60,** 215–228
5. Helmberg, C., Rendl, F., Vanderbei, R., Wolkowicz, H. (1996): An interior-point method for semidefinite programming. SIAM J. Optim. **6,** 342–361
6. Karmarkar, N.K. (1984): A new polynomial-time algorithm for linear programming. Combinatorica **4,** 373–395

7. de Klerk, E., Roos, C., Terlaky, T. (1997): Initialization in semidefinite programming via a self-dual skew-symmetric embedding. Oper. Res. Lett. **20,** 213–221

8. de Klerk, E., Roos, C., Terlaky, T. (1998): Infeasible start semidefinite programming algorithms via self-dual embeddings. Fields Inst. Commun. **18,** 215–236

9. Kojima, M., Mizuno, S., Yoshise, A. (1991): An $O(\sqrt{n}L)$ iteration potential reduction algorithm for linear complementarity problems. Math. Program. **50,** 331–342

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

11. Luo, Z.-Q., Sturm, J.F., Zhang, S. (1998): Conic convex programming and self-dual embedding. Optim. Methods Softw. **14,** 169–218

12. Mizuno, S., Yoshise, A., Kikuchi, T. (1989): Practical polynomial time algorithms for linear complementarity problems. J. Oper. Res. Soc. Japan **32,** 75–92

13. Monteiro, R.D.C. (1998): Primal-dual path following algorithms for semidefinite programming. SIAM J. Optim. **7,** 663–678

14. Monteiro, R.D.C., Adler, I. (1989): Interior path following primal-dual algorithms: Part I: Linear programming. Math. Program. **44,** 27–41

15. Monteiro, R.D.C., Zhang, Y. (1998): A unified analysis for a class of path-following primal-dual interior-point algorithms for semidefinite programming. Math. Program. **81,** 281–299

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

17. Nesterov, Yu.E., Todd, M.J. (1998): Primal-dual interior-point methods for self-scaled cones. SIAM J. Optim. **8,** 324–364

18. Nesterov, Yu.E., Todd, M.J., Ye, Y. (1999): Infeasible-start primal-dual methods and infeasibility detectors for nonlinear programming problems. Math. Program. **84,** 227-267

19. Potra, F.A., Sheng, R. (1998): On homogeneous interior-point algorithms for semidefinite programming. Optim. Methods Softw. **9,** 161–184

20. Todd, M.J. (1999): A study of search directions in primal-dual interior-point methods for semidefinite programming. Optim. Methods Softw. **11,** 1–46

21. Toh, K.C. (2000): Some new search directions for primal-dual interior point methods in semidefinite programming. SIAM J. Optim. **11,** 223–242

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

23. Xu, X., Ye, Y. (1995): A generalized homogeneous and self-dual linear programming algorithm. Oper. Res. Lett. **17,** 181–190

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

25. Zhang, Y. (1998): On extending some primal-dual interior-point algorithms from linear programming to semidefinite programming. SIAM J. Optim. **8,** 365–386