AN $O(\sqrt{n} L)$-ITERATION LARGE-STEP PRIMAL-DUAL AFFINE ALGORITHM FOR LINEAR PROGRAMMING$^1$

By

C.C. Gonzaga$^2$ and M.J. Todd$^3$

---

$^1$The work of both authors was supported in part by NSF Grant ECS-8602534 and ONR Contract N00014-87-K0212.

$^2$COPPE-Federal University of Rio de Janeiro, Brazil.

$^3$Center for Applied Mathematics and School of Operations Research and Industrial Engineering, Cornell University.
An $O(\sqrt{n} L)$-iteration large-step primal-dual affine algorithm for linear programming\textsuperscript{1}

C. C. Gonzaga\textsuperscript{2} and M. J. Todd\textsuperscript{3}

Abstract

We describe an algorithm based on reducing a suitable potential function for linear programming. At each iteration, separate scalings are applied to the primal and dual problems, and a step is taken in either the primal or the dual space. We show that a constant reduction can always be achieved, leading to a bound of $O(n^{1/2}L)$ iterations. Moreover, we show that a reduction of $\Omega(n^{1/4})$ can usually be obtained, so that $O(n^{1/4}L)$ iterations are expected to suffice. Finally, we prove that no general algorithm taking either primal or dual steps and guaranteeing the reduction of such a potential function can achieve $R$-order of convergence greater than one.

---

\textsuperscript{1} The work of both authors was supported in part by NSF Grant ECS-8602534 and ONR Contract NOOO14-87-K0212.

\textsuperscript{2} COPPE-Federal University of Rio de Janeiro, Brazil.

\textsuperscript{3} Center for Applied Mathematics and School of Operations Research and Industrial Engineering, Cornell University.
1. Introduction

This paper presents a primal-dual interior-point algorithm for the linear programming problem

$$\min \quad c^T x$$

$$\text{(P)} \quad Ax = b$$

$$x \geq 0.$$ 

The algorithm reduces a primal-dual potential function at each iteration, by applying separate scalings to the primal and dual problems and taking a step in either the primal or the dual space. We show that a constant reduction can always be achieved, leading to a bound of $O(n^{1/2}L)$ iterations, where $n$ is the number of variables and $L$ the size of the input, assuming the data $A$, $b$ and $c$ are all integer. Moreover, we give heuristic arguments showing that a greater reduction can usually be obtained, so that fewer iterations are expected to suffice. Finally, we show that no algorithm that drives this primal-dual potential function to minus infinity by taking either primal or dual steps at each iteration can achieve $R$-order of convergence greater than one, independent of $n$, although superlinear convergence is possible.

Our algorithm uses a primal-dual potential function described in Todd and Ye [15] based on the primal function of Karmarkar [6]. Other algorithms based on reducing this potential function include Ye [16], Kojima, Mizuno and Yoshise [8], Huang and Kortanek [5], and Gonzaga [3].

The conventional dual of (P), written with explicit slacks $s$, is

$$\max \quad b^T y$$

$$\text{(CD)} \quad A^T y + s = c$$

$$s \geq 0.$$ 

It is more convenient to work with a dual problem also in standard form, involving only the nonnegative slacks $s$. Thus let $F$ be a matrix whose rows span the null space of $A$, and let
g = Fc. Then \(ATy + s = c\) for some \(y\) if and only if \(Fs = g\). Let \(d\) be any vector satisfying \(Ad = b\). Then \(b^Ty = d^TATy = d^Tc - d^Ts\). Thus (CD) can be written in terms of \(s\) alone as

\[
\min \quad d^Ts
\]

\[
(D) \quad Fs = g
\]

\[
s \geq 0.
\]

In this form, weak duality takes the form that \(c^Tx + d^Ts \geq c^Td\) for all feasible \(x, s\), and strong duality states that equality holds if and only if \(x\) and \(s\) are optimal. Notice that the duality gap \(c^Tx - b^Ty\) equals \(c^Tx + d^Ts - c^Td\) and also \(x^Ts\). This form of the dual was first presented by Todd and Ye [15].

Let \(F_+(P)\) and \(F_+(D)\) denote the set of strictly positive feasible solutions to (P) and (D) respectively, and \(U_+\) their cartesian product \(F_+(P) \times F_+(D)\). For \((x,s) \in U_+\), we define the primal-dual potential function \(\phi_\rho\) by

\[
\phi_\rho(x,s) := \rho \ln x^Ts - \sum_j \ln x_j s_j
\]

(1)

for \(\rho > n\); the primal penalized function by

\[
f^P_\alpha(x) := \alpha c^Tx - \sum_j \ln x_j
\]

(2)

for \(\alpha > 0\); and the dual penalized function by

\[
f^D_\alpha(s) := \alpha d^Ts - \sum_j \ln s_j
\]

(3)

for \(\alpha > 0\). These functions combine a monotonic function of the objective function or duality gap with a barrier term in primal, dual, or primal-dual space.
We assume that $F_+(P)$ and $F_+(D)$ are nonempty, and that $(x^0, s^0) \in U_+$ is known such that

\[ \phi_+(x^0, s^0) = O(\sqrt{n} L) \]  

where $\rho = n + v\sqrt{n}$ for some constant $v$. Manipulating (P) and (D) to obtain such solutions is described in several papers, for example Ye [16] and Kojima-Mizuno-Yoshise [7,8]. If we can reduce $\phi_+$ by a constant at each iteration, then $O(\sqrt{n} L)$ iterations will yield a pair $(x, s) \in U_+$ such that

\[ (n + v\sqrt{n}) \ln x^T s - \sum \ln x_j s_j \leq -v\sqrt{n} L, \]

or

\[ \ln x^T s \leq -L + \frac{1}{\sqrt{n}} \sum \ln \frac{x_j s_j}{x^T s} \leq -L. \]  

Hence $x^T s \leq 2^{-L}$, and exact solutions to (P) and (D) can then be obtained in $O(n^3)$ additional arithmetic operations. We use a large but constant $v$, as proposed by Gonzaga [2,3], since then, although the complexity bound is larger, it appears that larger steps can be taken and the algorithm will perform better in practice.

Our analysis is based on the relationship between $\phi_+$ and the penalized functions $f_\alpha^P(x)$ and $f_\alpha^D(x)$. Let $P_A$ denote the projection onto the null space of $A$. Then, for any $s \in F_+(D)$, $P_As = P_Ac$ (the rows of $P_A$ span the null space of $A$), so that

\[ P_A \nabla_x \phi_+(x, s) = \frac{\rho}{x^T s} P_A s - P_A X^{-1} e \]

\[ = \frac{\rho}{x^T s} P_A c - P_A X^{-1} e \]

\[ = P_A \nabla f_\alpha^P(x) \]

\[ \]
where \( \alpha = \rho / x^T s \). Here and below, \( X (S) \) denotes the diagonal matrix with the components of \( x (s) \) down its diagonal, and \( e \) denotes the vector of ones in \( \mathbb{R}^n \). Similarly,

\[
P_F \nabla_s \phi_\rho (x,s) = P_F \nabla f^D_\alpha (s)
\]

for the same \( \alpha \). These equations relate the potential function to the penalized functions.

**Remark:** The choice of \( F \) with rows spanning the null space \( N(A) \) of \( A \) has no effect on the projection \( P_F v \) of a given vector \( v \in \mathbb{R}^n \). This is so because by definition the range space \( R(F^T) \) equals \( N(A) \) and hence \( N(F) = R(A^T) \). Since \( R(F^T) = N(A) \) and \( N(F) = R(A^T) \) are orthogonal complements, \( P_F v \) is the projection of \( v \) into the row space of \( A \) and equals \( (I - P_A)v \).

The advantage of working with the penalized functions is that they have been well studied (see Fiacco and McCormick [1]), are strictly convex, and (under our assumptions) achieve their minima for each \( \alpha > 0 \). The minimizer of \( f^P_\alpha \) over \( F_+(P) \) is called the \( \alpha \)-center of \( F_+(P) \) and is denoted \( x(\alpha) \); similarly, \( s(\alpha) \), the \( \alpha \)-center of \( F_+(D) \), minimizes \( f^D_\alpha \) over \( F_+(D) \). Note that if \( x \) is the \( \alpha \)-center of \( F_+(P) \), then

\[
0 = P_A \nabla f^P_\alpha (x) = \alpha P_A c - P_A X^{-1} e = \alpha (P_A c - P_A (\alpha^{-1} X^{-1} e)).
\]

Hence \( s := \alpha^{-1} X^{-1} e \) satisfies \( s > 0, P_A s = P_A c, \) so \( s \in F_+(D) \). Moreover,

\[
P_F \nabla f^D_\alpha (s) = \alpha P_d - P_F S^{-1} e
\]
\[
= \alpha (P_d - P_F x)
\]
\[
= 0,
\]

since \( Ax = Ad = b \); thus \( s \) is the \( \alpha \)-center of \( F_+(D) \). We therefore have the equations

\[
X(\alpha)S(\alpha)e = \alpha^{-1} e, \ x(\alpha)^T s(\alpha) = n/\alpha
\] (7)
relating the $\alpha$-centers of $F_+(P)$ and $F_+(D)$. The path $\{(x(\alpha), s(\alpha))\}$ is called the central path.

Our algorithm makes no mention of the central path, but the convergence proof does. We must prove that the potential function decreases substantially at each iteration. At points where the projected gradients of the potential function are large, this is immediate; we must preclude the possibility of small projected gradients. For the primal or dual penalized functions, small projected gradients characterize nearly central points. We conclude that any reasonable potential restriction method will make good progress far from the central path, but we must pay special attention to its performance near the path. Corollary 2.1 in the next section shows that for our choice of potential function the primal and dual projected gradients can never both be small; its proof uses the contrast between the values $\alpha = n/x(\alpha)^T s(\alpha)$ and $\alpha = \rho/x^T s$ of the last two paragraphs, when $\rho = n + v\sqrt{n}$. Of course, other details such as scaling are necessary to guarantee a suitable step length.

In section 2, we show that, if a projected gradient like that in (6) is small, then $x \in F_+(P)$ is close to the $\alpha$-center $x(\alpha)$ and $c^T x$ is close to $c^T x(\alpha)$. This extends a result in Gonzaga [2]. Hence, we prove that, given $(x, s) \in U_+$, we cannot have both $x$ close to $x(\alpha)$ and $s$ to $s(\alpha)$, for $\alpha = \rho/x^T s$. This result is the basis of our algorithm, described in section 3. We show that constant reduction in $\phi_\rho$ can be achieved at each iteration by taking a step in $x$ or in $s$ (or in both). One reason for taking steps in either the primal or the dual space is that McShane, Monma and Shanno [9], in their empirical study of primal-dual methods, found that taking a step along an appropriate direction in primal-dual space led to poor computational results. They found it necessary to split the primal-dual direction into directions in primal and dual spaces and use separate step lengths (a fixed proportion of the distance to the boundary) in each. We then give heuristic reasons to expect that a reduction of $\Omega(n^{1/4})$ in the potential function can often be achieved. These are based on plausible (but not rigorous) probabilistic assumptions. Rigorous bounds of $O(n^{1/4}L)$ steps have been obtained by Sonnevend, Stoer and Zhao [13] for special classes of problems.
In section 4 we show that no algorithm that drives \( \phi_d \) to \(-\infty\) by taking at each iteration either a primal or a dual step can have R-order of convergence greater than one, independent of \( n \). This result applies to our algorithm as well as to that of Ye [16]. Section 5 contains some concluding remarks and summarizes some preliminary computational results.

2. Preliminaries

2.1 Scaling

Let \( \Lambda \) be a positive definite diagonal matrix. Then we can make the change of variables \( \bar{x} = \Lambda^{-1} x \), so that \( (P) \) becomes

\[
\begin{align*}
\min & \quad \bar{c}^T \bar{x} \\
(\bar{P}) & \quad \bar{A} \bar{x} = b \\
& \quad \bar{x} \geq 0
\end{align*}
\]

with

\[
\bar{A} = \Lambda \Lambda, \bar{c} = \Lambda c.
\]

The dual of \( (\bar{P}) \) can be written as

\[
\begin{align*}
\min & \quad \bar{d}^T \bar{s} \\
(\bar{D}) & \quad \bar{F} \bar{s} = g \\
& \quad \bar{s} \geq 0
\end{align*}
\]

in terms of \( \bar{s} = \Lambda s \), where

\[
\bar{F} = FA^{-1}, \bar{d} = \Lambda^{-1} d.
\]

Note that the rows of \( \bar{F} \) still span the null space of \( \bar{A} \). If \( \bar{r}_\alpha^P, \bar{r}_\alpha^D \), and \( \bar{r}_\alpha \) are defined from \( (\bar{P}) \) and \( (\bar{D}) \) as in (1) - (3), using \( \bar{x}, \bar{s}, \bar{c} \) and \( \bar{d} \), then
\( \tilde{\phi}(\bar{x}, \bar{s}) = \phi_p(x, s), \)

\( \tilde{\Gamma}_\alpha^p(\bar{x}) = \Gamma_\alpha^p(x) + \ln \det \Lambda, \)

and

\( \tilde{\Gamma}_\alpha^d(\bar{s}) = \Gamma_\alpha^d(s) - \ln \det \Lambda, \) \hspace{1cm} (8)

when \( \bar{x} \) and \( x \), and \( \bar{s} \) and \( s \), correspond as above. It follows that

\( \nabla_{\bar{x}} \tilde{\phi}(\bar{x}, \bar{s}) = \Lambda \nabla_x \phi_p(x, s), \)

\( \nabla_{\bar{s}} \tilde{\phi}(\bar{x}, \bar{s}) = \Lambda^{-1} \nabla_s \phi_p(x, s), \)

\( \nabla_{\bar{x}} \tilde{\Gamma}_\alpha^p(\bar{x}) = \Lambda \nabla_x \Gamma_\alpha^p(x), \)

and

\( \nabla_{\bar{s}} \tilde{\Gamma}_\alpha^d(\bar{s}) = \Lambda^{-1} \nabla_s \Gamma_\alpha^d(s). \) \hspace{1cm} (9)

Let \( (\hat{x}, \hat{s}) \in U_+ \). Then we can choose \( \Lambda = \hat{X} = \text{diag}(\hat{x}) \) so that \( \hat{x} \) is transformed into \( e \) in \( \bar{x} \)-space, or \( \Lambda = \hat{S}^{-1} = (\text{diag}(\hat{s}))^{-1} \) to transform \( \hat{s} \) into \( e \). In fact, we can use the first to scale \( x \) when considering changes in \( x \), and the second to scale \( s \) when considering changes in \( s \). This separate scaling contrasts with the symmetric scaling using \( \Lambda = (\hat{X}\hat{S}^{-1})^{1/2} \) which transforms both \( \hat{x} \) and \( \hat{s} \) to \( (\hat{X}\hat{S})^{1/2} e \), used in the primal-dual algorithms of Kojima, Mizuno and Yoshise \[7,8\] and Monteiro and Adler \[11,12\], for example. We only have to be careful when using separate scalings in dealing with \( \phi_p \), since (8) and (9) assume that the same scaling is used on \( x \) and on \( s \).

The effect of scaling is that we can usually assume without loss of generality that our current \( x \) iterate (or our current \( s \) iterate) is \( e \). In this case, any step of length less than one maintains positivity, so that, in the scaled space, steps in the direction of negative (projected) gradients are attractive.

### 2.2 A measure of centrality

Here we refine the analysis of Gonzaga \[3\] to show that we can effectively measure the distance from \( x \in F_+(P) \) to the \( \alpha \)-center of \( F_+(P) \) by means of the norm of the (scaled)
projected gradient of $f^P_{\alpha}$. Recall that $f^P_{\alpha}(x) = \alpha c^T x - \sum_j \ln x_j$ and that $f^P_{\alpha}$ is minimized at the $\alpha$-center $x(\alpha)$.

**Definition 2.1.** The measure of centrality of $x \in F_+(P)$ to the $\alpha$-center $x(\alpha)$ is

$$\delta^P(x, \alpha) := \|P_{A_X}X \nabla f^P_{\alpha}(x)\|_2,$$

where $X = \text{diag}(x)$. Similarly,

$$\delta^D(s, \alpha) := \|P_{F_S}S \nabla f^D_{\alpha}(s)\|_2$$

is the measure of centrality of $s \in F_+(D)$ to the $\alpha$-center $s(\alpha)$.

We can also view $\delta^P(x, \alpha)$ as the length of the Newton step to minimize $f^P_{\alpha}$ from $x$ — see [3]. Note that $\delta^P(x, \alpha) = \bar{\delta}^P(e, \alpha)$, where the scaling is chosen to transform $x$ to $e$, i.e. $\Lambda = X$.

**Proposition 2.1.** Let $x \in F_+(P)$ and

$$\psi := \frac{\|X^{-1}(x(\alpha) - x)\|_2}{\|X^{-1}(x(\alpha) - x)\|_\infty}. \quad (10)$$

If $\delta := \delta^P(x, \alpha) \leq \psi/4$, then

$$\varepsilon := \|X^{-1}(x(\alpha) - x)\|_2 \leq 2\delta \quad (11)$$

and

$$lc^T x(\alpha) - c^T x \leq 3 \sqrt{n} \delta/\alpha. \quad (12)$$

**Proof.** Without loss of generality, we can assume that $x = e$; otherwise, scale so that this is true. Then let

$$h := \frac{x(\alpha) - e}{\|x(\alpha) - e\|_2}$$
so that $\|h\|_2 = 1$ and $x(\alpha) = e + \varepsilon h$. Since $f^P_{\alpha}$ is convex and minimized at $x(\alpha)$, 

$$h^T \nabla f^P_{\alpha}(e + \lambda h) < 0 \text{ for } 0 \leq \lambda < \varepsilon.$$ 

Now

$$\nabla f^P_{\alpha}(e + \lambda h) = \alpha c - ((1 + \lambda h_j)^{-1})$$

$$= \alpha c - e + \lambda \left( \frac{h_j}{1 + \lambda h_j} \right)$$

$$= \nabla f^P_{\alpha}(e) + \lambda h - \lambda^2 \left( \frac{h_j^2}{1 + \lambda h_j} \right).$$

Also,

$$\|h^T \nabla f^P_{\alpha}(e)\| = \|h^T P_A \nabla f^P_{\alpha}(e)\| \leq \|P_A \nabla f^P_{\alpha}(e)\| = \delta,$$ (13)

using $h = P_A h$ and definition 2.1. Hence, for $0 \leq \lambda < \varepsilon$,

$$0 > h^T \nabla f^P_{\alpha}(e + \lambda h)$$

$$= h^T \nabla f^P_{\alpha}(e) + \lambda h^T h - \lambda^2 \sum_j \left( \frac{h_j^3}{1 + \lambda h_j} \right)$$

$$\geq - \delta + \lambda - \lambda^2 \sum_{j: h_j > 0} \left( \frac{h_j^3}{1 + \lambda h_j} \right)$$

$$\geq - \delta + \lambda - \lambda^2 \|h\|_\infty \sum_{j: h_j > 0} h_j^2$$

$$\geq - \delta + \lambda - \lambda^2/\psi,$$ (14)

using $\|h\|_2 = 1$ and $\|h\|_\infty = 1/\psi$. Since $\delta \leq \psi/4$ by hypothesis, (14) gives

$$\frac{(\lambda - \frac{\psi}{2})^2}{\psi} = \frac{\lambda^2}{\psi} - \lambda + \frac{\psi}{4} \geq \frac{\lambda^2}{\psi} - \lambda + \delta > 0$$
for $0 \leq \lambda < \varepsilon$, so $\varepsilon \leq \psi/2$. Thus $\lambda^2/\psi \leq \lambda \varepsilon/\psi \leq \lambda/2$ for $0 \leq \lambda < \varepsilon$, so (14) yields

$$- \delta + \lambda - \lambda/2 < 0 \text{ for } 0 \leq \lambda < \varepsilon,$$

from which $\varepsilon \leq 2\delta$ follows.

To show (12), note that

$$|c^T(x(\alpha) - e)| = |\alpha c^T(\varepsilon h)/\alpha|$$

$$= |(\alpha c - e)^T(\varepsilon h) + e^T(\varepsilon h)/\alpha|$$

$$\leq (\varepsilon/\alpha)(|\nabla f^P_\alpha(e)|_{\|h\|_2} + |e^T h|)$$

$$\leq (2\delta/\alpha)(\delta + |e^T h|)$$

$$\leq 3 \sqrt{n} \delta/\alpha,$$  \hspace{1cm} (15)

using (13), (11), $e^T h \leq \|e\| l_{\|h\|_2} = \sqrt{n}$, and $\delta \leq \psi/4 \leq \sqrt{n}/4$ by definition of $\psi$.

Of course, an analogous result holds for the dual problem (D). An easy consequence is that, for a suitable choice of $\alpha$, we cannot simultaneously have $x$ close to $x(\alpha)$ and $s$ to $s(\alpha)$:

**Corollary 2.1.** Let $(x, s) \in U_+$ and let $\alpha = \rho/x^T s$ where $\rho = n + v\sqrt{n}$, $v > 0$. Let $\Delta > 0$ be such that $\Delta \leq 1/4$ and $\Delta < \psi/6$ (for example, $\Delta = 1/4$ with $v \geq 2$). Then we cannot have both $\delta^P(x, \alpha) \leq \Delta$ and $\delta^D(s, \alpha) \leq \Delta$.

**Proof.** Suppose the contrary. Then proposition (2.1) applies since $\psi \geq 1$, and we deduce that

$$|c^T x(\alpha) - c^T x| \leq 3 \sqrt{n} \delta^P/\alpha,$$

and

$$|d^T s(\alpha) - d^T s| \leq 3 \sqrt{n} \delta^D/\alpha.$$

11
Hence the duality gaps \( x(\alpha)^T s(\alpha) \) and \( x^T s \) satisfy

\[
|x(\alpha)^T s(\alpha) - x^T s| \leq 6\sqrt{n} \Delta / \alpha .
\]

But

\[
x^T s = \frac{\rho}{\alpha} = \frac{n}{\alpha} + \nu \frac{\sqrt{n}}{\alpha} ,
\]

and since \( x(\alpha)^T s(\alpha) = n/\alpha \),

\[
|x(\alpha)^T s(\alpha) - x^T s| = \nu \sqrt{n} / \alpha .
\]

Comparing this expression with (16), we obtain \( \nu \leq 6\Delta \), contrary to the hypothesis.

Now recall (6). The corollary implies that, when \( \rho = n + \nu \sqrt{n} \) with \( \nu \geq 2 \), we cannot have both \( P_{AX}X^T x^T \phi_\rho(x,s) \) and \( P_{FS}S^T s^T \phi_\rho(x,s) \) with norms at most 1/4. Thus it is reasonable that the potential function \( \phi_\rho \) can be reduced substantially by taking either a primal or dual step. This is the basis of our method.

3. **The algorithm**

We can now state our algorithm. To obtain the polynomial bound, \((x^0,s^0)\) must be chosen with \( \phi_\rho(x^0,s^0) = O(\sqrt{n} \ L) \) and \( \eta \) should be \( 2^{-2L} \). For the convergence analysis we choose \( \nu \geq 2 \), but any positive value \( \nu = O(1) \) can be used instead with suitable changes to the constants.

**Algorithm**

Given \((x^0,s^0)\) \( \in U_+ \) and termination parameter \( \eta > 0 \):

- set \( k \leftarrow 0 \), \( \rho \leftarrow n + \nu \sqrt{n} \) for \( 2 \leq \nu = O(1) \).

Repeat until \( (x^k)^T s^k \leq \eta \);

- set \((x,s) \leftarrow (x^k,s^k)\);

**choose** either a primal or a dual step

(a primal step can only be chosen if \( \|P_{AX}X^T x^T \phi_\rho(x,s)\|_2 \geq 1/4 \),
a dual step only if \( \|P_{FS}S\nabla_s \phi_p(x,s)\|_2 \geq 1/4 \);

**primal step:**
\[
x^+ = x - \beta P_{AX}X \nabla_x \phi_p(x,s);
\]
\[
s^+ = s; \text{ or}
\]

**dual step:**
\[
x^+ = x;
\]
\[
s^+ = s - \beta P_{FS}S \nabla_s \phi_p(x,s);
\]

where \( \beta > 0 \) is such that
\[
\phi_p(x^+,s^+) \leq \phi_p(x,s) - \frac{1}{40};
\]

**update:**
\[
k \leftarrow k+1;
\]
\[
(x^k,s^k) \leftarrow (x^+,s^+);
\]

**end.**

Corollary 2.1 implies that the conditions for taking a primal or dual step can always be met. If \( \beta \) can be chosen to assure the decrease of 1/40 in \( \phi_p \), then the argument in the introduction implies that the algorithm will terminate in at most \( O(\phi_p(x^0,s^0) + \sqrt{n \ln(n)}^{-1}) \) iterations.

**Proposition 3.1.** If \( \beta \) in the algorithm is chosen as \((6\|P_{AX}X \nabla_x \phi_p(x,s)\|_2)^{-1}\) or \((6\|P_{FS}S \nabla_s \phi_p(x,s)\|_2)^{-1}\) according as a primal or dual step is taken, then
\[
\phi_p(x^+,s^+) \leq \phi_p(x,s) - \frac{1}{40}.
\]

**Proof.** This follows from standard arguments. We know that
\[
\phi_p(x+\Delta x,x) \leq \phi_p(x,s) + \Delta x^T \nabla_x \phi_p(x,s) - \frac{\|X^{-1} \Delta x\|^2}{2(1 - \|X^{-1} \Delta x\|_\infty)};
\]
see, for instance, Ye [16]. Then if \( \Delta x = -\beta X P_{AX} X \nabla_x \phi_p(x,s) \), we find \( \|X^{-1} \Delta x\|_2 = 1/6 \) and so

\[
\phi_p(x+\Delta x,s) \leq \phi_p(x,s) - \frac{\delta}{6} + \frac{(1/6)^2}{2(1 - 1/6)}
\]

(17)

where \( \delta = \|P_{AX} X \Delta x \phi_p(x,s)\|_2 \). Since \( \delta \) is at least 1/4 if a primal step is taken, we find \( \phi_p \) is reduced by at least 1/40. A similar argument applies if a dual step is taken.

In fact, we could take steps in both primal and dual spaces, exactly as above. Since the step \( \Delta x \) in \( x \) lies in the null space of \( A \), it is orthogonal to the step \( \Delta s \) in \( s \). Hence

\( (x^+)T_s^+ = x^T_s + \Delta x^T s + x^T \Delta s \) is linear in \((\Delta x, \Delta s)\), and the standard argument shows that

\[
\phi_p(x+\Delta x, s+\Delta s) \leq \phi_p(x,s) + \Delta x^T \nabla_x \phi_p(x,s) + \Delta s^T \nabla_s \phi_p(x,s)
\]

\[
\frac{\|X^{-1} \Delta x\|^2_2}{2(1 - \|S^{-1} \Delta s\|_\infty)} - \frac{\|S^{-1} \Delta s\|^2_2}{2(1 - \|S^{-1} \Delta s\|_\infty)}
\]

if we choose step sizes so that \( \|X^{-1} \Delta x\|_2 = \|S^{-1} \Delta s\|_2 = 1/6 \), then we deduce that

\[
\phi_p(x+\Delta x, s+\Delta s) \leq \phi_p(x,s) - \frac{\delta_P}{6} - \frac{\delta_D}{6} + \frac{2(1/6)^2}{2(1 - 1/6)}
\]

Since \( \delta_P \) and \( \delta_D \) are both nonnegative, and one is at least 1/4, we find that \( \phi_p \) is reduced by at least 1/120.

Remarks

a) In most iterations we expect a much greater reduction in potential than guaranteed by the worst case analysis. Indeed, suppose

\[
h = h_p := \frac{X^{-1} (x(\alpha) - x)}{\|X^{-1} (x(\alpha) - x)\|_2}
\]

satisfies \( \|h\|_2 / \|h\|_\infty \geq 4 n^{1/4} \) and \( \|e^T h\| \leq n^{1/4} \) (below we give heuristic reasons for these hypotheses). Then the proof of proposition 2.1 (see especially (15)) shows that if \( \delta_P(x,\alpha) \)
\leq n^{1/4}, \; \|c^T x(\alpha) - e^T x l \| \leq 4\sqrt{n} / \alpha. \) If similar hypotheses hold for the dual direction \( h^D \), we could conclude as in corollary 2.1 that, for \( v > 8 \), we cannot have both \( \delta^P(x, \alpha) \) and \( \delta^D(s, \alpha) \) at most \( n^{1/4} \). This would imply, from (17), that a reduction of potential of \( \Omega(n^{1/4}) \) could be obtained. In the integer model, only \( O(n^{1/4}L) \) iterations of this kind are necessary. This heuristic analysis gives a similar bound to the rigorous results of Sonnevend, Stoor and Zhao [13] for special classes of linear programming problems.

In the paper [4], we show that the hypotheses we have made in the previous paragraph on \( h^P \) and \( h^D \) hold with probability approaching one as \( n \to \infty \) if \( h^P \) and \( h^D \) are uniformly distributed on the unit sphere. Of course, this is not a rigorous analysis of expected behavior, but it gives some justification to the heuristic arguments above.

b) From the description of the algorithm, it seems that two projections must be calculated at each iteration. However, since at most one of \( x \) and \( s \) is changed, one of these projections is the same as at the last iteration. Hence only one new factorization is required at each iteration after the first, and thus one can choose a primal or dual step corresponding to the larger projected gradient. In practice, \( P_{FS} \) can be computed as \( I - P_{AS}^{-1} \) (see the remark below (6)).

4. R-order of convergence

In section 3 we described an algorithm that takes either a primal or a dual step at each iteration and that drives the potential function \( \phi_\rho \) to \( -\infty \), for \( \rho = n + \nu \sqrt{n} \), \( \nu \) a constant. Ye's algorithm [16] is also of this form. In this section, we provide a limit to the convergence rate of such an algorithm. This can be contrasted with the quadratic convergence of Zhang and Tapia's primal-dual algorithm [17].

Recall that, if \( x^k \to x^* \), then \( x^k \) converges to \( x^* \) R-linearly if

\[ \|x^k - x^*\|_2 \leq \alpha \gamma^k \] (18)
for some $\alpha > 0$ and $0 < \gamma < 1$, while the convergence has R-order $q$ ($q > 1$) if for all $1 \leq p < q$, \[ \|x^k - x^*\|_2 \leq \alpha p^k. \] (19)

**Theorem 4.1.** Consider an algorithm that takes either a primal or a dual step at each iteration and that drives $\phi^p$ to $-\infty$, for $p = n + \nu\sqrt{n}$ and $\nu$ a constant. Then this algorithm cannot guarantee that the generated sequences $\{x^k\}$ or $\{s^k\}$ have R-order of convergence greater than one (independent of $n$).

**Proof.** We assume that (P) and (D) have unique nondegenerate optimal solutions $\bar{x}$ and $\bar{s}$ respectively, with
\[ \bar{x}_j > 0 \text{ iff } j \in B, |B| = m \]
\[ \bar{s}_j > 0 \text{ iff } j \in N, |N| = d \]
and $m+d = n$. We indicate the appropriate subvectors of $x$ and $s$ by a subscript $B$ or $N$. Any convergent algorithm has $x \to \bar{x}$, $s \to \bar{s}$, and we will assume henceforth that the iterates have converged sufficiently that
\[ x_B \leq 2 \bar{x}_B, \quad s_N \leq 2 \bar{s}_N. \] (20)

Given, $(x,s) \in U_+$, the duality gap is $x^Ts$, and this is the sum of the primal gap (from optimality), denoted
\[ \pi := s^T x = \bar{s}_N^T \bar{x}_N. \]
and the dual gap
\[ \theta := \bar{x}^T s = \bar{x}_B^T \bar{s}_B. \]
Then, assuming (20) holds, we have for any $\rho \geq n$,
\[ \phi_p(x,s) = \rho \ln(\pi+\theta) - \sum_{j \in B} \ln x js_j - \sum_{j \in N} \ln x js_j \]
\[ \geq \rho \ln(\pi+\theta) - \sum_{j \in B} \ln x js_j - \sum_{j \in N} \ln x js_j - n \ln 2 \]
\[ \geq \rho \ln(\pi+\theta) - m \ln \frac{T}{m} - d \ln \frac{T}{d} - n \ln 2 \]
\[ = \rho \ln(\pi+\theta) - m \ln \theta - d \ln \pi + m \ln 2m + d \ln 2d - n \ln 4 \]
\[ \geq \rho \ln(\pi+\theta) - m \ln \theta - d \ln \pi + m \ln \frac{2m}{n} + d \ln \frac{2d}{n} + n \ln n - n \ln 4 \]
\[ \geq \max\{(\rho - m) \ln \theta - d \ln \pi, (\rho - d) \ln \pi - m \ln \theta\} + n \ln (n/4); \quad (21) \]

the last inequality follows since \(\frac{2m}{n} \ln \frac{2m}{n} + \frac{2d}{n} \ln \frac{2d}{n}\) can be written as \((1+\alpha) \ln (1+\alpha) + (1-\alpha) \ln (1-\alpha)\), which is convex and nonnegative, minimized at \(\alpha = 0\).

Suppose that \(\phi_p\) has been driven below \(n \ln (n/4)\). Then from (21) we can deduce that

\[ d \ln \pi \geq (\rho - m) \ln \theta, \quad m \ln \theta \geq (\rho - d) \ln \pi \]

or

\[ \frac{\rho - m}{\pi} \geq \theta, \quad \frac{\rho - d}{\theta} \geq \pi. \quad (22) \]

Suppose that \(m = d = n/2\), so that \((\rho - m)/d = (\rho - d)/m = 1 + 2v/\sqrt{n}\). Assume that an algorithm of the type considered generates a sequence \((x^k, s^k) \in U_+\) with primal and dual gaps \((\pi_k, \theta_k)\). Then

\[ \min(\pi_{k+1}, \theta_{k+1}) \geq (\min(\pi_k, \theta_k))^{1 + 2v/\sqrt{n}} \quad (23) \]

for all sufficiently large \(k\). Indeed, suppose \(\pi_{k+1} \leq \theta_{k+1}\) (the other case is similar). If a dual step was just taken, then \(\pi_{k+1} = \pi_k\), which is at least the right-hand side of (23) if \(\pi_k < 1\). If a primal step was taken, (22) gives \(\pi_{k+1} \geq \theta_{k+1}^{1 + 2v/\sqrt{n}} = \theta_k^{1 + 2v/\sqrt{n}}\), which is
again greater than or equal to the right-hand side of (23), as long as (20) holds for $(x^{k+1}, s^{k+1})$ and $\phi_p(x^{k+1}, s^{k+1}) \leq n \ln (n/4)$.

Of course, (23) yields

$$\min\{\pi_k, \theta_k\} \geq (\min\{\pi_k, \theta_k\})(1 + 2\nu/\sqrt{n})$$

for any sufficiently large $k$, which implies using (19) that the $R$-order of convergence cannot be greater than $1 + 2\nu/\sqrt{n}$. Since $\nu$ is a constant, this proves the theorem.

Remarks

a) The theorem should not be interpreted as stating that algorithms of this type should not be used. Fast $R$-linear algorithms can be very attractive if $\alpha$ is small and $\gamma$ close to zero in (18). Moreover, our result does not preclude $R$-order of convergence $1 + 2\nu/\sqrt{n}$, which implies $R$-superlinear convergence. However, asymptotic quadratic convergence cannot be achieved, while it is possible for algorithms that move in primal and dual spaces simultaneously [17].

b) From (21) we can establish an interesting property of central pairs $(x, s)$ (i.e., $x = x(\alpha)$, $s = s(\alpha)$ for some $\alpha$). Using (7), we find that $\phi_n(x, s) = n \ln n$, so that (noting that $\rho - m = d, \rho - d = m$ if $\rho = n$) we conclude

$$\pi \geq \left(\frac{1}{4}\right)^{n/d} \theta, \quad \theta \geq \left(\frac{1}{4}\right)^{n/m} \pi.$$  (24)

Hence the primal and dual gaps are of similar order for central pairs sufficiently close to optimal. This result can be contrasted with (22), and provides further motivation for choosing a large value of $\nu$, which gives more freedom in reducing $\pi$ for a given value of $\theta$ or vice versa.