ASYMPTOTIC BEHAVIOR OF INTERIOR-POINT METHODS: A VIEW FROM SEMI-INFINITE PROGRAMMING

by

Levent Tuncel

and

Michael J. Todd

1Research supported in part by Mathematical Sciences Institute at Cornell University, and also by NSF, AFOSR, and ONR through NSF Grant DMS-8920550.

2Research supported in part by NSF, AFOSR, and ONR through NSF Grant DMS-8920550.
ASYMPTOTIC BEHAVIOR OF INTERIOR-POINT METHODS: A VIEW FROM SEMI-INFINITE PROGRAMMING

Levent TUNÇEL*  
Michael J. TODD†

School of Operations Research and Industrial Engineering  
Cornell University  
Ithaca, NY 14853-3801

September 17, 1992

Abstract

We study the asymptotic behavior of interior-point methods for linear programming problems. Attempts to solve larger problems using interior-point methods lead to the question of how these algorithms behave as n (the number of variables) goes to infinity. Here, we take a different point of view and investigate what happens when n is infinite. Motivated by this approach, we study the limits of search directions, potential functions and central paths. We also argue that the complexity of the linear programming problem should depend on the smoothness of the given problem rather than the number of variables. We prove that when n is infinite one can still get a complexity bound on the number of iterations required in terms of the smoothness of the problem and the desired accuracy.

Keywords: Linear programming, interior-point methods, semi-infinite programming.

*Research supported in part by Mathematical Sciences Institute of Cornell University, and also by NSF, AFOSR and ONR through NSF Grant DMS-8920550.  
†Research supported in part by NSF, AFOSR and ONR through NSF Grant DMS-8920550.
1 Introduction

Our aim is to study the asymptotic behavior of interior-point methods. This study
is motivated by Powell’s result [Pow90] providing an example with infinitely many
constraints on which Karmarkar’s algorithm fails to converge to the optimum solution,
by Ferris and Philpott’s paper [FP89] extending the affine-scaling algorithm to semi-
infinite programming problems, and mostly by Todd’s paper [To91] generalizing the
search directions for interior-point methods to the semi-infinite setting and showing
which algorithms can be extended and which cannot. As Todd [To91] points out,
“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 inequalities) becomes
infinite.”

We consider linear programming problems in the following primal ($P$) and dual ($D$)
forms:

\[
(P) \begin{align*}
    \text{minimize} & \quad c^T x \\
    Ax & = b, \\
    x & \geq 0,
\end{align*}
\]

\[
(D) \begin{align*}
    \text{maximize} & \quad b^T y \\
    A^T y + s & = c, \\
    s & \geq 0,
\end{align*}
\]

where $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, and $c \in \mathbb{R}^n$. We assume $A$ has full row rank and that there
exist interior solutions for both problems.

The corresponding semi-infinite problems are denoted by ($SIP$) and ($SID$) respectively:

\[
(SIP) \begin{align*}
    \inf & \quad \int_T \gamma(t) \xi(t) dt \\
    \int_T a(t) \xi(t) dt & = b, \\
    \xi(t) & \geq 0,
\end{align*}
\]

\[
(SID) \begin{align*}
    \max & \quad b^T y \\
    a(t)^T y + \sigma(t) & = \gamma(t) \quad \forall t \in T, \\
    \sigma(t) & \geq 0 \quad \forall t \in T,
\end{align*}
\]

Here, $a(t)$ is an $m$-dimensional vector of functions $a_i(t)$, $b$ is a constant vector in
$\mathbb{R}^m$, and $T$ is a compact subset of some Euclidean space. For simplicity, we assume
that the measure of $T$ is one. We assume that $a(t)$ and $\gamma(t)$ are continuous, hence
the space of functions for the dual problem is clear ($\sigma(t)$ must be continuous). The
choice for the primal problem is not as clear. Anderson and Nash [AN87] point out that the appropriate space should be the space of regular Borel measures on [0,1] (say for $T = [0,1]$), but for simplicity we follow Todd’s treatment [To91], and restrict the primal space to the space of continuous functions on $T$ ($C(T)$) (in this case the optimal value of (SIP) may not be attained even if it is finite). For two excellent surveys on semi-infinite programming see Polak [Pol87] (which provides a view from nondifferentiable optimization) and Hettich and Kortanek [HK91] (which provides a view from continuous optimization).

$F(SIP)$ and $F(SID)$ denote the set of feasible solutions for the primal and dual problems respectively. We will usually deal only with $\sigma(t)$ as a dual feasible solution. So, whenever we say $\sigma(t) \in F(SID)$, we mean that $\sigma(t) \geq 0 \ \forall t \in T$ and there exists a $y \in \mathbb{R}^m$ such that $a^T(t)y + \sigma(t) = \gamma(t) \ \forall t \in T$.

For $\varphi, \vartheta \in C(T)$, the inner product is defined as

$$< \varphi, \vartheta > := \int_T \varphi(t)\vartheta(t)dt.$$ 

The space $C(T)$ with this inner product defines an inner product space (see for instance Rudin [Ru87]). In such a space we have the following properties available to us:

(i) $< \varphi, \vartheta >$ depends linearly on $\vartheta$ for a fixed $\varphi$.

(ii) $< \varphi, \vartheta > = < \vartheta, \varphi >$.

(iii) $< \varphi, \varphi > \geq 0, \ \forall \varphi$.

(iv) $< \varphi, \varphi > = 0 \iff \varphi = 0$.

(v) $||\varphi||_2 := (< \varphi, \varphi >)^{1/2}$ is a norm.

These are all easy to check and (iv) follows from the fact that $\varphi \in C(T)$. We will abuse this notation and write, for instance, $< A, B >$, where $A(t)$ and $B(t)$ are matrices (of appropriate dimensions) of functions. Then the inner product operation implies that we do the matrix multiplication of the functions first and then integrate the resulting matrix entry by entry. We will also write $\xi$ for $\xi(t)$ and $\sigma^{-1}$ for $\frac{1}{\sigma(t)}$ etc. Integrals are always taken over $T$ unless otherwise indicated. Inequalities involving functions are meant to be understood pointwise, i.e. $\xi \geq \sigma$ means $\xi(t) \geq \sigma(t) \ \forall t \in T$. We will usually denote scalar functions by lower case Greek letters; some constants and parameters will also be denoted by lower case Greek letters. Matrices are denoted by upper case, vectors and vector functions are denoted by lower case Roman letters. Logarithms denoted by log will always be base 2.

$F_+(SIP)$ and $F_+(SID)$ denote the “interiors” of $F(SIP)$ and $F(SID)$, i.e.,

$$F_+(SIP) := \{ \xi \in F(SIP) : \xi > 0 \},$$
$$F_+(SID) := \{ \sigma \in F(SID) : \sigma > 0 \}.$$
We also define the interior of the feasible solution set in the primal-dual space:

$$F_0 := \{ (\xi, \sigma) : \xi \in F_+(SIP), \sigma \in F_+(SID) \}.$$  

We assume that both $F_+(SIP)$ and $F_+(SID)$ are nonempty.

One can easily show that the weak duality result holds for the pair $(SIP)$ and $(SID)$. Let $\xi \in F(SIP)$ and $\sigma \in F(SID)$. Then $a^T y + \sigma = \gamma$ and $\int a \xi = b$ imply

$$< \gamma, \xi > - b^T y = < \xi, \sigma > \geq 0,$$

which proves weak duality. Here, $< \xi, \sigma >$ is the duality gap. Clearly, if we have a sequence of primal and dual feasible solutions for which the duality gap tends to zero then the corresponding primal and dual objective values have to tend to their respective optimum values.

Note that if $b = 0$ then $\hat{\xi} := 0$ is a feasible solution in $(SIP)$. Then for any $\sigma \in F(SID)$ we have the duality gap $< \hat{\xi}, \sigma >= 0$. Hence $\hat{\xi}$ is optimal in $(SIP)$ and all feasible solutions are optimal in $(SID)$. Henceforth we assume $b \neq 0$.

2 Projections and Search Directions

We start by generalizing some basic facts for projections from finite dimensional case to the semi-infinite setting. Given $a \in (C(T))^m$, we define an operator $P_a(.)$ on $C(T)$, that projects an element of $C(T)$ onto the null space of $a$:

$$P_a(.) := I(.) - a^T \left[ \int aa^T \right]^{-1} < a, . >,$$

where $I(.)$ is the identity map on $C(T)$. We denote $B := \int aa^T$. A collection of functions $a_i$ for $i \in \{1, 2, \ldots, m\}$ is said to be pseudo-Haar if they are linearly independent on every non-null set. Now we can show that in this case the projection operator is well-defined.

**Proposition 2.1.** If $\{a_i, \ i \in \{1, 2, \ldots, m\}\}$ is a set of continuous functions that is pseudo-Haar, then $B := \int aa^T$ is positive definite, and hence $P_a(.)$ is well defined.

**Proof:** Let $y \in \mathbb{R}^m$. Then we have

$$y^T \left[ \int aa^T \right] y = \int (y^T a)(a^T y) = \int (a^T y)^2 \geq 0. \quad (1)$$
The above inequality holds with equality if and only if $a^Ty = 0$ almost everywhere on $T$. But we assumed that $\{a_i\}$ was pseudo-Haar, so the inequality in (1) holds with equality if and only if $y = 0$. Hence, $B$ is positive definite.

Henceforth we assume the hypothesis of Proposition 2.1. It is easy to see from the definition that $\langle a, P_a(v) \rangle = 0$ (for any $v \in C(T)$). We also have the following results.

**Lemma 2.1.** $P_a^2(v) = P_a(v)$.

**Proof:** From the definition, we have

$$P_a^2(v) = P_a \left( v - a^T B^{-1} \int a v \right)$$

$$= v - a^T B^{-1} \int a v - a^T B^{-1} \int a \left( v - a^T B^{-1} \int a v \right)$$

$$= v - 2a^T B^{-1} \int a v + a^T B^{-1} \int a a^T B^{-1} \int a v$$

$$= v - a^T B^{-1} \int a v$$

$$= P_a(v).$$

\[ \square \]

**Lemma 2.2.** $\langle v_1, P_a(v_2) \rangle = \langle P_a(v_1), v_2 \rangle$.

**Proof:** We find

$$\langle v_1, P_a(v_2) \rangle = \int v_1 \left( v_2 - a^T B^{-1} \int a v_2 \right)$$

$$= \int v_1 v_2 - \int v_1 a^T B^{-1} \int a v_2$$

$$= \int v_1 v_2 - \int v_2 a^T B^{-1} \int a v_1$$

$$= \int v_2 \left( v_1 - a^T B^{-1} \int a v_1 \right)$$

$$= \langle v_2, P_a(v_1) \rangle.$$

\[ \square \]
Lemma 2.3. \( <v_1, P_a(v_2) > = <P_a(v_1), P_a(v_2) >. \)

Proof: We have
\[
< v_1, P_a(v_2) > = < v_1, P_a^2(v_2) > \\
= < v_1, P_a(P_a(v_2)) > \\
= < P_a(v_1), P_a(v_2) >, 
\]
where the first equality follows from Lemma 2.1, and the last from Lemma 2.2.

Corollary 2.1. \( <P_a(v), (I - P_a)(v) > = 0. \)

Note that we easily have
\[
\|P_a(v)\|^2_2 + \|(I - P_a)(v)\|^2_2 = <P_a(v), P_a(v)> + <(I - P_a)(v), (I - P_a)(v)>
\]
\[
= <P_a(v), P_a(v)> - <P_a(v), P_a(v)> + <v, v>
\]
\[
= \|v\|^2_2. \tag{2}
\]

Let \( \eta \) denote the constant function 1 on \( T \). Then we have
\[
\|P_a(\eta)\|^2_2 + \|(I - P_a)(\eta)\|^2_2 = \|\eta\|^2_2 = 1.
\]
So, \( 1 - <P_a(\eta), P_a(\eta)> = 0 \) implies \( P_a(\eta) = \eta \). In this case \( <\eta, (I - P_a)(\eta)> = 0 \), from which
\[
\int a \eta = \int a = 0. \tag{3}
\]

Lemma 2.4. Let \( f \in (C(T))^m, g \in (C(T))^k \), and suppose \( <f, g^T > = 0 \). Then
\[
P \begin{bmatrix} f \\ g \end{bmatrix} (v) = P_f (P_g(v)) = P_g (P_f(v)).
\]

Proof: Indeed,
\[
P \begin{bmatrix} f \\ g \end{bmatrix} (v) = v - [f^T \ g^T] \left[ \begin{bmatrix} f \\ g \end{bmatrix} [f^T \ g^T]^{-1} \right] \begin{bmatrix} f \\ g \end{bmatrix} v
\]
\[
= v - [f^T \ g^T] \begin{bmatrix} (f f^T)^{-1} & 0 \\ 0 & (g g^T)^{-1} \end{bmatrix} \begin{bmatrix} f \ v \\ g \ v \end{bmatrix}
\]
\[ v = f^T \left( \int ff^T \right)^{-1} \int f + g^T \left( \int gg^T \right)^{-1} \int g = P_f(P_g(v)) = P_g(P_f(v)). \]

\[ \square \]

**Lemma 2.5.** Let \( v_p := P_a(v) \). Then we have
\[ v_p = \text{argmin}_\varphi \{ \langle \varphi, \varphi \rangle : \varphi = v - a^Ty, \ y \in \mathbb{R}^m \} \]
and hence \( <P_a(v), P_a(v)> \leq <v, v> \).

*Proof:* Consider the following “least squares” problem:
\[ \min_{y \in \mathbb{R}^m} <v - a^Ty, v - a^Ty> =: f(y). \]

Note that \( f(y) \) is the quadratic function
\[ f(y) = <v, v> - 2<v, a^Ty> + <a^Ty, a^Ty>, \]
so that
\[ \nabla f(y) = -\int av + \int (a^Ty)a, \]
\[ \nabla^2 f(y) = \int aa^T. \]

Thus \( f \) is strictly convex, since \( \int aa^T \) is positive definite. Hence \( f \) is minimized by \( y^* \) with \( \nabla f(y^*) = 0 \), from which
\[ y^* = \left( \int aa^T \right)^{-1} \int av. \]

Hence, we have
\[ v_p = \text{argmin}_\varphi \{ \langle \varphi, \varphi \rangle : \varphi = v - a^Ty, \ y \in \mathbb{R}^m \}. \]

For the last part, note that \( y = 0 \) yields \( <\varphi, \varphi> = <v, v> \), which implies \( <P_a(v), P_a(v)> \leq <v, v> \).

To see the last part directly, one can also use (2).

Now we turn to the local behavior of interior-point algorithms. It is well-known that almost all interior-point algorithms use a linear combination of two directions (affine-scaling and centering) as the search direction (see for instance den Hertog and Roos [HR89]). Todd [To91] shows that the analogous affine-scaling direction and centering...
direction for \((SIP)\) are given by

\[
\pi_{\text{AFF}} := -\xi^2 \left( \gamma - a^T \left[ \int \xi^2 a a^T \right]^{-1} \left[ \int \xi^2 \right] \right),
\]

\[
\pi_{\text{CEN}} := \xi - \xi^2 a^T \left[ \int \xi^2 a a^T \right]^{-1} b.
\]

We note that

\[
\pi_{\text{AFF}} = -\xi P_{\alpha}(\xi \gamma), \quad (4)
\]

\[
\pi_{\text{CEN}} = \xi P_{\alpha}(\eta), \quad (5)
\]

where \(\eta\) is the constant function 1 on \(T\). Henceforth, in primal-only settings \(P\) denotes \(P_{\alpha}\) and in primal-dual settings \(P\) denotes \(P_{\alpha^{1/2} \eta^{-1/2}}\) unless otherwise indicated.

The rest of this section is devoted to proving that some search directions preserve the finite-dimensional property of being a descent direction for an appropriate objective or potential function. We note that Todd [To91] mentioned these facts earlier and gave proofs for some of them. Here we provide the proofs that are not given in Todd’s paper. Namely, we show that the affine-scaling direction is a descent direction for the objective function and that Karmarkar’s algorithm (projective scaling) and primal-dual potential-reduction algorithms give descent directions for the corresponding potential functions in this semi-infinite setting. First, we show that unless all the feasible solutions of \((SIP)\) are optimal, the affine-scaling direction is a descent direction for the objective function \(< \gamma, . >\).

**Proposition 2.2.** Let \(\xi \in F_+(\text{SIP})\) be given. Unless \(< \gamma, \xi >\) is constant on \(F(\text{SIP})\),

\(< \gamma, \pi_{\text{AFF}} > < 0.\)

**Proof:** Using (4) we have

\[
< \gamma, \pi_{\text{AFF}} > = < \gamma, -\xi P(\xi \gamma) >
\]

\[
= -< \xi \gamma, P(\xi \gamma) >
\]

\[
= -< P(\xi \gamma), P(\xi \gamma) >
\]

\[
= -\|P(\xi \gamma)\|_2^2
\]

\[
\leq 0.
\]

Since our space is an inner product space, using the property (iv), \(\|P(\xi \gamma)\|_2 = 0\) implies \(P(\xi \gamma) = 0\). So, there exists a \(y \in \mathbb{R}^m\) such that \(\xi a^T y = \xi \gamma\). We see that \(\sigma := 0\) is feasible in \((SID)\). But for any primal feasible solution \(\hat{\xi}\) we have \(< \hat{\xi}, \sigma > = 0\) and so \(\hat{\xi}\) and \(\sigma\) are optimal; therefore, \(< \gamma, \xi >\) is constant on \(F(\text{SIP})\). \(\square\)
Thus if our next iterate $\xi_+$ is $\xi + \alpha \pi_{AFF}$ for a step size $\alpha < \frac{\inf\{\xi\}}{\inf\{\pi_{AFF}\}}$, then it lies in $F_+(SIP)$ and its objective function value is $\langle \gamma, \xi \rangle + \alpha < \gamma, \pi_{AFF} \rangle = < \gamma, \xi > - \alpha \|P(\xi_\gamma)\|_2^2$, and we have a reduction of $\alpha \|P(\xi_\gamma)\|_2^2$ in the primal objective function value.

**Remark:** The dual affine-scaling algorithm (see [FP89, To91]) has the same monotonicity property. Todd [To91] shows that the affine-scaling direction for $y$ in $(SID)$ is given by $\left[\int \sigma^{-2}aa^T\right]^{-1}b$. Hence the change in the objective function in this direction is $b^T \left[\int \sigma^{-2}aa^T\right]^{-1}b$ which is always strictly positive (since $b \neq 0$ and $\int \sigma^{-2}aa^T$ is positive definite). We can also provide a proof that is more symmetric with the proof for $(SIP)$:

Todd [To91] also shows that the affine-scaling direction for $\sigma$ in $(SIP)$ is given by

$$\delta_{AFF} := -a^T \left[\int \sigma^{-2}aa^T\right]^{-1}b.$$  

We note that

$$\delta_{AFF} := -\sigma(I - P_{aa^{-1}})(\check{\xi}\sigma),$$  

for any $\check{\xi} \in F(SIP)$. We have

$$< \check{\xi}, \delta_{AFF} > = < \check{\xi}, \sigma(I - P_{aa^{-1}})(\check{\xi}\sigma) > = < \check{\xi}\sigma, (I - P_{aa^{-1}})(\check{\xi}\sigma) > = -\| (I - P_{aa^{-1}})(\check{\xi}\sigma) \|_2^2.$$  

This shows that the dual objective function improves by $\alpha \| (I - P_{aa^{-1}})(\check{\xi}\sigma) \|_2^2$ for a suitably chosen step size $\alpha$. Note that $\| (I - P_{aa^{-1}})(\check{\xi}\sigma) \|_2^2 = 0$ implies $\int a\sigma^{-1}\check{\xi}\sigma = \int a\check{\xi} = 0$, which implies $b = \int a\check{\xi} = 0$, a contradiction.

Now we proceed to show that Karmarkar’s search direction, $\pi_{PRO}$, is a descent direction for the potential function

$$\phi(\xi; z_l) := \ln(< \gamma, \xi > - z_l) - < \eta, \ln(\xi) >,$$

where $z_l$ is a suitable valid lower bound on the value of the optimal objective function value in the primal problem. We have

$$\pi_{PRO} := \pi_{AFF} + \nu_{PRO}\pi_{CEN},$$

where

$$\nu_{PRO} := \frac{< \gamma, \xi > - < \gamma, \pi_{CEN} > - z_l}{< \eta, \eta - P(\eta) >}.$$  

9
(see Todd [To91]). Note that if we let \( \Delta := \langle \gamma, \xi > - z_i \) (the estimated optimality gap), then we can write

\[
\nu_{PRO} = \frac{\Delta - \langle P(\eta), P(\xi) \rangle}{1 - \langle P(\eta), P(\eta) \rangle}.
\]

We need to show that \( \nu_{PRO} \) is well defined. Suppose \( 1 - \langle P(\eta), P(\eta) \rangle = 0 \). Then we have \( P(\eta) = \eta \) which implies \( b = 0 \) (see (3)), a contradiction. So, we have \( \nu_{PRO} \) well defined. Next we provide a lemma that will be useful in proving that certain directions are descent directions for corresponding potential functions.

**Lemma 2.6.** Let \( f(\alpha) := -c_1 \alpha + \frac{\alpha^2}{c_2(1-\alpha)}, \quad \alpha \in [0, 1) \), where \( c_1, c_2 > 0 \). Then there exists \( \bar{\alpha} \in (0, 1) \) such that \( f(\bar{\alpha}) < 0 \); in fact

\[
\arg\min_{\alpha \in (0, 1)} f(\alpha) = 1 - \frac{1}{\sqrt{1 + c_1c_2}}.
\]

**Proof:** It is easy to see that \( f(\alpha) \) is strictly convex and differentiable on \((0, 1)\) with

\[
f'(0) = -c_1 < 0.
\]

Solving \( f'(\alpha) = 0 \) for \( \alpha \in (0, 1) \) one gets

\[
\arg\min_{\alpha \in (0, 1)} f(\alpha) = 1 - \frac{1}{\sqrt{1 + c_1c_2}}.
\]

Now, we show that \( \pi_{PRO} \) is a descent direction for the potential function \( \phi \). Note that

\[
\phi(\xi + \alpha \pi_{PRO}; z_i) - \phi(\xi; z_i) = \ln \left( \frac{\langle \gamma, \xi + \alpha \pi_{PRO}, z_i \rangle - \xi_i}{\Delta} \right)
\]

\[
- \langle P(\eta), \ln[\xi^{-1}(\xi + \alpha \pi_{PRO})] \rangle.
\]

Now, we overestimate the first term by its linear approximation; restricting the step size to satisfy \( \alpha |\xi^{-1} \pi_{PRO}| < \eta \) and using the inequality \( \ln(1 + \lambda) \geq \lambda - \frac{\lambda^2}{2(1-\lambda)} \) for \( |\lambda| < 1 \) we have

\[
\phi(\xi + \alpha \pi_{PRO}) - \phi(\xi) \leq \alpha < \frac{\gamma}{\Delta}, \pi_{PRO} > - \alpha < \xi^{-1}, \pi_{PRO} >
\]

\[
+ \frac{1}{2} \xi^{-1} \pi_{PRO} (\eta - \alpha |\xi^{-1} \pi_{PRO}|)^{-1} >
\]

\[
\leq \alpha < \nabla \phi(\xi; z_i), \pi_{PRO} >
\]

\[
+ \frac{\alpha^2}{2(1 - \alpha \sup |\xi^{-1} \pi_{PRO}|)} < \xi^{-1} \pi_{PRO}, \xi^{-1} \pi_{PRO} >.
\]

Now we define \( \nabla \phi(\xi; z_i) \) as

\[
\nabla \phi(\xi; z_i) := \frac{\gamma}{\Delta} - \xi^{-1}.
\]
So, \( \nabla \phi(\xi; z_l) \) is like a “gradient”. Now, it is clear that if we can show \( \langle \nabla \phi(\xi; z_l), \pi_{PRO} \rangle < 0 \) then Lemma 2.6 will imply that \( \pi_{PRO} \) is a descent direction for the potential function \( \phi \). By the definitions we have

\[
\langle \nabla \phi(\xi; z_l), \pi_{PRO} \rangle = \langle \frac{\gamma}{\Delta} - \xi^{-1}, -\xi P(\xi \gamma) + \nu_{PRO} \xi P(\eta) \rangle
\]
\[
= \langle \frac{\xi \gamma}{\Delta} - \eta, -P(\xi \gamma) + \nu_{PRO} P(\eta) \rangle
\]
\[
= -\frac{1}{\Delta} < P(\xi \gamma - \Delta \eta), P(\xi \gamma) - \nu_{PRO} P(\eta) >
\]
\[
= -\frac{1}{\Delta(1 - \|P(\eta)\|_2^2)} < P(\xi \gamma - \Delta \eta), [P(\xi \gamma) - \Delta P(\eta)
-\|P(\eta)\|_2^2 P(\xi \gamma)] + < P(\eta), P(\xi \gamma) > P(\eta) >
\]
\[
= -\frac{1}{\Delta(1 - \|P(\eta)\|_2^2)} [\|P(\xi \gamma)\|_2^2 - 2\Delta < P(\xi \gamma), P(\eta) > + \Delta^2 \|P(\eta)\|_2^2 - \|P(\eta)\|_2^2 \|P(\xi \gamma)\|_2^2 + \Delta \|P(\eta)\|_2^2 < P(\xi \gamma), P(\eta) > + < P(\xi \gamma), P(\eta) >^2
- \Delta \|P(\eta)\|_2^2 < P(\xi \gamma), P(\eta) >].
\]

So,

\[
\langle \nabla \phi(\xi; z_l), \pi_{PRO} \rangle = -\frac{1}{\Delta(1 - \|P(\eta)\|_2^2)} [\|P(\xi \gamma)\|_2^2 - 2\Delta < P(\xi \gamma), P(\eta) > + \Delta^2 \|P(\eta)\|_2^2 - \|P(\eta)\|_2^2 \|P(\xi \gamma)\|_2^2 + < P(\xi \gamma), P(\eta) >^2]
\]

(7)

Note that \( < P(\xi \gamma), P(\eta) >^2 \leq \|P(\xi \gamma)\|_2 \|P(\eta)\|_2 \) by the Cauchy-Schwarz inequality (which is proved exactly as in the finite case); so, we parametrize the related terms as follows:

\[
< P(\xi \gamma), P(\eta) >^2 = \beta \|P(\xi \gamma)\|_2 \|P(\eta)\|_2, \quad \beta \in [0, 1].
\]

Now we have

\[
\langle \nabla \phi(\xi; z_l), \pi_{PRO} \rangle \leq -\frac{1}{\Delta(1 - \|P(\eta)\|_2^2)} [\|P(\xi \gamma)\|_2^2 - 2\Delta \|P(\eta)\|_2 \|P(\xi \gamma)\|_2 + 2\sqrt{\beta} + \Delta^2 \|P(\eta)\|_2^2 + (\beta - 1) \|P(\eta)\|_2 \|P(\xi \gamma)\|_2].
\]

(8)

**Proposition 2.3.** Let \( \xi \in F_+(SIP) \) and assume that \( \xi \) is not optimal. Then either \( \pi_{PRO} \) is a descent direction for the potential function \( \phi \) or the lower bound \( z_l \) can be increased by at least \( \tilde{\alpha} \|I - P_\alpha(\xi)\|_2^2 \), where

\[
\tilde{\alpha} := \sup \{ \alpha : \Delta \xi^{-1} - \alpha (I - P_\alpha)(\xi) \geq 0 \}.
\]
Proof: We use (7) and (8). If $\beta = 1$ we get from (7)

$$<\nabla \phi(\xi; z_1), \pi_{PRO} > = -\frac{\|P(\xi_\gamma - \Delta \eta)/\|_2^2}{\Delta (1 - \|P(\eta)/\|_2^2)}.$$  \hspace{1cm} (9)

The right-hand side of (9) is clearly always non-positive. If it is equal to zero then we have $\|P(\xi_\gamma - \Delta \eta)/\|_2^2 = 0$. This is equivalent to $P(\xi_\gamma) = \Delta P(\eta)$. So, by definition of $P$, there exists some $\hat{\eta} \in \mathbb{R}^m$ such that

$$\xi a^T \hat{\eta} + \Delta \eta = \xi \gamma.$$  \hspace{1cm} (10)

Since $\xi > 0$, equation (10) implies that $\Delta \xi^{-1}$ is feasible in (SID) and the corresponding duality gap is $<\xi, \Delta \xi^{-1}> = \Delta$. But we know that $\Delta \xi^{-1} > 0$, so by moving the dual feasible solution $\Delta \xi^{-1}$ in the direction of $-(I - P_a)(\xi)$, we can easily get a strictly better lower bound. We let

$$\hat{\sigma} := \Delta \xi^{-1} - \hat{\alpha}(I - P_a)(\xi),$$

where

$$\hat{\alpha} := \sup\{\alpha : \Delta \xi^{-1} - \alpha(I - P_a)(\xi) \geq 0\}.$$  

Clearly, $\hat{\sigma} \in P(SID)$ and we have $<\xi, \hat{\sigma}> = <\xi, \Delta \xi^{-1} - \hat{\alpha}(I - P_a)(\xi)> = \Delta - \hat{\alpha}\|P(\xi_\gamma)\|_2^2$. We see that $\|P(\xi_\gamma)\|_2 = 0$ implies $b = \int a\xi = 0$; hence $\hat{\gamma} := \gamma + \hat{\alpha}\|P(\xi_\gamma)\|_2^2 > \gamma$ is an improved lower bound.

If $\beta = 0$, we get from (8)

$$<\nabla \phi(\xi; z_1), \pi_{PRO} > \leq -\frac{\|P(\xi_\gamma)\|_2^2}{\Delta} - \frac{\|P(\eta)\|_2^2}{(1 - \|P(\eta)/\|_2^2)} < 0.$$  

(As in Proposition 2.2, $P(\xi_\gamma) = 0$ implies $\xi$ is optimal.) If $\beta \in (0, 1)$ then the right-hand side of (8) is a quadratic in $\sqrt{\beta}$ (we call the numerator $h(\sqrt{\beta})$, i.e.,

$$h(\sqrt{\beta}) := -\|P(\xi_\gamma)/\|_2^2 - 2\Delta\|P(\eta)/\|_2^2\|P(\xi_\gamma)/\|_2\sqrt{\beta} + \Delta^2\|P(\eta)/\|_2^2 + (\beta - 1)\|P(\eta)/\|_2^2\|P(\xi_\gamma)/\|_2^2);$$

and its maximum value is achieved by

$$\beta_{max} = \frac{\Delta^2}{\|P(\xi_\gamma)/\|_2^2\|P(\eta)/\|_2^2).}$$

If $\beta_{max} \geq 1$ then since $h(0) < 0$ and $h(1) = -(\|P(\xi_\gamma)/\|_2 - \Delta\|P(\eta)/\|_2)^2 \leq 0$, we find $h(\sqrt{\beta}) < 0 \forall \beta \in (0, 1)$. So, we clearly have $<\nabla \phi(\xi; z_1), \pi_{PRO} > < 0$; otherwise ($\beta_{max} \in (0, 1)$) we find $\Delta^2 < \|P(\xi_\gamma)/\|_2^2\|P(\eta)/\|_2^2$. So, by substituting $\beta = \beta_{max}$ and
using the inequality $\Delta < \|P(\xi)\|_2 \|P(\eta)\|_2$, we get

$$
< \nabla \phi(\xi; z_1), \pi_{PRO} > \leq -\frac{1}{\Delta (1 - \|P(\eta)\|_2^2)} \left[ \|P(\xi)\|_2^2 - 2\Delta^2 + \Delta^2 \|P(\eta)\|_2^2 + \Delta^2 \right]
-\|P(\eta)\|_2^2 \|P(\xi)\|_2^2
-\frac{1}{\Delta} \left( (1 - \|P(\eta)\|_2^2) \|P(\xi)\|_2^2 \right)
-\frac{1}{\Delta} \left( (1 - \|P(\eta)\|_2^2) \|P(\xi)\|_2^2 \right)
= -\frac{1}{\Delta} (\|P(\xi)\|_2^2 - \Delta^2).
$$

Using $\Delta^2 < \|P(\xi)\|_2^3 \|P(\eta)\|_2 \leq \|P(\xi)\|_2^3$, we get $< \nabla \phi(\xi; z_1), \pi_{PRO} > < 0$.

So, we established $< \nabla \phi(\xi; z_1), \pi_{PRO} > < 0$. To get an $\alpha > 0$ such that $\phi(\xi + \alpha \pi_{PRO}; z_1) - \phi(\xi; z_1) < 0$, we apply Lemma 2.6 to (6). □

Now we show that the search direction of the primal-dual potential-reduction algorithms (see Kojima, Mizuno, and Yoshise [KMY88b] and Gonzaga and Todd [GT89]) is a descent direction for the potential function

$$
\phi_{PD}(\xi, \sigma) := \tau \ln( < \xi, \sigma >) - < \eta, \ln(\xi) > - < \eta, \ln(\sigma) >,
$$

where $\tau > 1$ is a constant. From now on we will be mostly dealing with the primal-dual set up and hence $P(.)$ will denote $P_{a^{\xi_{1/2}}}\sigma_{1/2}(.)$.

Todd [To91] shows that the affine-scaling and the centering directions for the primal ($\pi'_{AFF}, \pi'_{CEN}$) and the dual ($\delta'_{AFF}, \delta'_{CEN}$) spaces are then given by

$$
\pi'_{AFF} := -\xi + \xi \sigma^{-1} a^T \left[ \int \xi \sigma^{-1} a a^T \right]^{-1} b
$$

$$
\pi'_{CEN} := \sigma^{-1} - \xi \sigma^{-1} a^T \left[ \int \xi \sigma^{-1} a a^T \right]^{-1} \left[ \int \sigma^{-1} a \right]
$$

$$
\delta'_{AFF} := -a^T \left[ \int \xi \sigma^{-1} a a^T \right]^{-1} b
$$

$$
\delta'_{CEN} := a^T \left[ \int \xi \sigma^{-1} a a^T \right]^{-1} \left[ \int \sigma^{-1} a \right].
$$

As Todd [To91] shows, the search directions in the primal ($\pi_{PRPD}$) and the dual ($\delta_{PRPD}$) spaces are given by

$$
\pi_{PRPD} := \frac{\tau}{\Delta} \pi'_{AFF} + \pi'_{CEN}
$$

and

$$
\delta_{PRPD} := \frac{\tau}{\Delta} \delta'_{AFF} + \delta'_{CEN}.
$$
where $\Delta := <\xi, \sigma>$ (the current duality gap). In terms of
\[ v_{P_{RPD}} := \frac{\tau}{\Delta} \xi^{1/2} \sigma^{1/2} - \xi^{-1/2} \sigma^{-1/2}, \]
we find
\begin{align*}
\pi_{P_{RPD}} &= -\xi^{1/2} \sigma^{-1/2} P(v_{P_{RPD}}), \\
\delta_{P_{RPD}} &= -\xi^{-1/2} \sigma^{1/2} (I - P)(v_{P_{RPD}}).
\end{align*}

For a given step size $\alpha \geq 0$ the next iterates in the primal and dual spaces respectively are
\begin{align*}
\xi_{\alpha} &:= \xi + \alpha \pi_{P_{RPD}} = \xi - \alpha \xi^{1/2} \sigma^{-1/2} P(v_{P_{RPD}}), \\
\sigma_{\alpha} &:= \sigma + \alpha \delta_{P_{RPD}} = \sigma - \alpha \xi^{-1/2} \sigma^{1/2} (I - P)(v_{P_{RPD}}).
\end{align*}

Now it is easy to show that at each iteration the duality gap decreases by some amount. Using (13)-(14) we have
\[
<\xi_{\alpha}, \sigma_{\alpha}> = <\xi, \sigma> - \alpha \left[ <\xi^{1/2} \sigma^{1/2}, P(v_{P_{RPD}}) + (I - P)(v_{P_{RPD}})> \right] \\
+ \alpha^2 <P(v_{P_{RPD}}), (I - P)(v_{P_{RPD}})>.
\]

By Corollary 2.1 the coefficient of the quadratic term in $\alpha$ is zero and by simplifying we get
\[
<\xi_{\alpha}, \sigma_{\alpha}> = <\xi, \sigma> - \alpha <\xi^{1/2} \sigma^{1/2}, v_{P_{RPD}}> \\
= <\xi, \sigma> - \alpha \left[ \frac{\tau}{\Delta} \Delta - 1 \right] \\
= <\xi, \sigma> - \alpha (\tau - 1).
\]

For $\tau > 1$ and $\alpha > 0$ we get a reduction in the duality gap. Note that it is also easy to see that
\[
\phi_{PD}(\xi_{\alpha}, \sigma_{\alpha}) - \phi_{PD}(\xi, \sigma) = \tau \ln \left( \frac{<\xi, \sigma> + \alpha <\xi, \pi_{P_{RPD}}> + <\sigma, \pi_{P_{RPD}}>}{\Delta} \right) \\
- \eta, \ln[\xi^{-1}(\xi + \alpha \pi_{P_{RPD}})] > \\
- \eta, \ln[\sigma^{-1}(\sigma + \alpha \delta_{P_{RPD}})] >.
\]

Now, we overestimate the first term by its linear approximation; restricting the step size to satisfy $\alpha|\xi^{-1} \pi_{P_{RPD}}| < \eta$ and $\alpha|\sigma^{-1} \delta_{P_{RPD}}| < \eta$, and then using the inequality $\ln(1 + \lambda) \geq \lambda - \frac{\lambda^2}{2(1 - \lambda)}$ for $|\lambda| < 1$, we have
\[
\phi_{PD}(\xi_0, \sigma_0) - \phi_{PD}(\xi, \sigma) \leq \alpha \tau \left( < \frac{\xi}{\Delta}, \delta_{PRPD} > + < \frac{\sigma}{\Delta}, \pi_{PRPD} > \right)
- \alpha < \xi^{-1}, \pi_{PRPD} > + < \sigma^{-1}, \delta_{PRPD} > \\
+ \alpha^2 < \xi^{-1} \pi_{PRPD}, \frac{1}{2} \xi^{-1} \pi_{PRPD}(\eta - \alpha |\xi^{-1} \pi_{PRPD}|)^{-1} > \\
+ \alpha^2 < \sigma^{-1} \delta_{PRPD}, \frac{1}{2} \sigma^{-1} \delta_{PRPD}(\eta - \alpha |\sigma^{-1} \delta_{PRPD}|)^{-1} > \\
\leq \alpha < \nabla_\xi \phi_{PD}(\xi, \sigma), \pi_{PRPD} > + < \nabla_\sigma \phi_{PD}(\xi, \sigma), \delta_{PRPD} > \\
+ \alpha^2 < \xi^{-1} \pi_{PRPD}, \xi^{-1} \pi_{PRPD} > \\
2(1 - \alpha \sup |\xi^{-1} \pi_{PRPD}|) \\
+ \alpha^2 < \sigma^{-1} \delta_{PRPD}, \sigma^{-1} \delta_{PRPD} > \\
2(1 - \alpha \sup |\sigma^{-1} \delta_{PRPD}|),
\]

where we define \( \nabla_\xi \phi_{PD}(\xi, \sigma) \) and \( \nabla_\sigma \phi_{PD}(\xi, \sigma) \) as

\[
\nabla_\xi \phi_{PD}(\xi, \sigma) := \frac{\tau \sigma - \xi^{-1}}{\Delta} \\
\nabla_\sigma \phi_{PD}(\xi, \sigma) := \frac{\tau \xi - \sigma^{-1}}{\Delta}.
\]

As in Proposition 2.3, \( \nabla_\xi \phi_{PD}(\xi, \sigma) \) and \( \nabla_\sigma \phi_{PD}(\xi, \sigma) \) are like gradients of the potential function. By the definitions and Lemma 2.3, we easily have

\[
< \nabla_\xi \phi_{PD}, \pi_{PRPD} > = -\|P(v_{PRPD})\|^2_2, \\
< \nabla_\sigma \phi_{PD}, \delta_{PRPD} > = -\|(I - P)(v_{PRPD})\|^2_2.
\]

So by (2), we find

\[
< \nabla_\xi \phi_{PD}, \pi_{PRPD} > + < \nabla_\sigma \phi_{PD}, \delta_{PRPD} > = -\|v_{PRPD}\|^2_2.
\]

Hence if \( v_{PRPD} \neq 0 \) then we have a proof that \((\pi_{PRPD}, \delta_{PRPD})\) is a descent direction for \( \phi_{PD} \). If \( v_{PRPD} = 0 \) then \( \xi \sigma = \frac{\Delta}{\tau} \eta \) which implies \( < \xi, \sigma > = < \xi, \sigma > / \tau \), but as long as \( \tau > 1 \) and \( < \xi, \sigma > \neq 0 \) this is impossible. Now, one can use the above discussion with Lemma 2.6 to complete the proof of the following proposition.

**Proposition 2.4.** Let \((\xi, \sigma) \in \mathcal{F}_0\). Then for \( \tau > 1 \), \((\pi_{PRPD}, \delta_{PRPD})\) is a descent direction for the potential function \( \phi_{PD} \).

\( \square \)

## 3 Central Paths, Neighborhoods and Potential Functions

In the semi-infinite setting we define the central path as the set of solutions \((\xi_\mu, \sigma_\mu)\) of the following system of equalities and inequalities, where \( \mu := < \xi, \sigma > \).
\[ <a, \xi> = b, \quad \xi > 0 \]  
\[ a^T y + \sigma = \gamma, \quad \sigma > 0 \]  
\[ \inf \{\xi \sigma\} = \mu. \]  

In this study we are interested in smooth functions only; so, we replace (17) by
\[ \xi \sigma = \mu \eta. \]  

We note that central paths are studied in a similar set up by Sonnevend [So89]. Under suitable conditions, the existence of a solution to this system of equalities and inequalities and the uniqueness of the solutions \((\xi_\mu, \sigma_\mu)\) can be easily extracted from the existing results in semi-infinite programming theory. We use a theorem due to Borwein and Lewis (see [BL91a]). Under our assumptions, Theorem 6.1 of [BL91a] will imply that the system (15)-(17') always has a unique solution. We introduce two problems \((SIPB)\) ((SIP) with barrier) and \((SIDB)\) ((SID) with barrier). These two problems are closely related to the definition of the central path in the semi-infinite setting as they are in the finite setting (for a discussion of the finite (-dimensional) case see for instance Megiddo [Meg88]).

\((SIPB)\)

\[ \inf \frac{1}{\mu} <\gamma, \xi> - <\eta, \ln(\xi)> \]
\[ <a, \xi> = b \]
\[ \xi \geq 0, \]

\((SIDB)\)

\[ \max \frac{1}{\mu} b^T y + <\eta, \ln(\sigma)> \]
\[ a^T y + \sigma = \gamma \]
\[ \sigma \geq 0, \]

From now on we will assume that \(T\) is a finite union of compact intervals on the real line and that \(a_i, \ i \in \{1, 2, \ldots, m\}\) and \(\gamma\) are Lipschitz on \(T\). Then for \(\mu = 1\) using the Integrability Condition stated in [BL91a], and Theorems 6.1 and 6.14 we can conclude
that there exists unique $(\xi, y, \sigma)$ such that $\xi$ is optimal in $\text{(SIPB)}$, $(y, \sigma)$ is optimal in $\text{(SIDB)}$ and that
\[ \xi = (\gamma - a^T y)^{-1} = \sigma^{-1}. \] (18)

It is easy to see that in fact under our assumptions the theorem holds for any $\mu > 0$ with (18) replaced by (18'):
\[ \xi = \mu(\gamma - a^T y)^{-1} = \mu \sigma^{-1}. \] (18')

Hence, we have established the existence and uniqueness of the central path under our assumptions.

Many primal-dual algorithms that follow the central path to the optimal solutions have been developed in the finite case (see for instance Kojima, Mizuno and Yoshise [KMY88a, KMY89], Monteiro and Adler [MA89] and Mizuno, Todd and Ye [MTY90]). To establish convergence, we study a path-following algorithm that keeps its iterates in some neighborhood of the central path similar to the finite case.

Let $\beta \in (0, 1)$ be a constant. We can define some neighborhoods of the central path as in Mizuno, Todd and Ye [MTY90]. We only define the neighborhoods that are based on the infinity norm and only one side of the infinity norm. As Todd [To91] points out the algorithms using two-norm neighborhoods are not invariant (in the sense defined there). Hence let
\[ N_\infty(\beta) := \{ (\xi, \sigma) \in F_0 : (1 - \beta) \eta \leq \frac{\xi \sigma}{\xi, \sigma} \leq (1 + \beta) \eta \}, \]
\[ N_\infty^-(\beta) := \{ (\xi, \sigma) \in F_0 : (1 - \beta) \eta \leq \frac{\xi \sigma}{\xi, \sigma} \}. \]

Note that $N_\infty(\beta) \subset N_\infty^-(\beta)$. As in section 2, the Tanabe-Todd-Ye (see [Ta87, TY90]) potential function extends to
\[ \phi_{PD}(\xi, \sigma) = \tau \ln(<\xi, \sigma>) - <\eta, \ln(\xi) > - <\eta, \ln(\sigma) >, \]
where $\tau > 1$ is a constant. Primal-only or dual-only versions of this potential function can also be defined as extensions of Karmarkar’s potential function [Ka84]. As Powell [Pow90] shows, determining a step size based on a line search on Karmarkar’s potential function could be disastrous.

Example 1: ([Pow90])

\[ (ED) \text{ maximize } \cos(2\pi t) y_1 + \sin(2\pi t) y_2 \leq y_2 \leq 1, \ t \in [0, 1]. \]

So, in the format of (SID), $m = 2$, $b^T = [0, -1]$, $a(t)^T = [\cos(2\pi t), \sin(2\pi t)]$, $\gamma = \eta$ and $T = [0, 1]$. 17
For this example, Powell shows that Karmarkar's potential function is

\[ \phi^{ED}(y_1, y_2) := \ln(1 + y_2) - \ln \left[ 1 + \sqrt{1 - (y_1^2 + y_2^2)} \right] + \text{constant}. \]

Clearly, the barrier part takes values in \([-\ln(2), 0]\) for feasible solutions and is zero on the boundary of the feasible region. So, we no longer have the barrier property \((y^k \rightarrow \partial F(SID) \Rightarrow \phi(y^k) \rightarrow \infty)\). As suggested in [Tu92a], one may try instead the logarithm of the infimum of the function \(\sigma\) which does have the barrier property. In this example it corresponds to the potential function

\[ \rho \ln(1 + y_2) - \ln \left( 1 - \sqrt{y_1^2 + y_2^2} \right), \]

where \(\rho > 1\) is a constant.

Note that Karmarkar's potential function is too smooth in the sense that it treats all of the constraints more uniformly. Even when the current iterate is very close to the boundary of the feasible region, the constraints that are not close to being tight still have considerable effect on the value of the barrier. As a result the potential function fails to keep the barrier property even for Powell's example. However, our barrier always keeps its barrier property and in the case of Powell's example it recognizes that the problem has in effect a single quadratic constraint.

In the primal-dual setting we propose the potential function

\[ \psi_\rho(\xi, \sigma) := (\rho + 1) \ln(\langle \xi, \sigma \rangle) - \ln(\inf\{\xi\sigma\}), \]

where \(\rho > 0\) is a constant.

First of all note that the barrier part and the objective function part balance nicely:

\[ \ln \left( \frac{\int \xi\sigma}{\inf(\xi\sigma)} \right) \geq 0. \]

Equality holds above if and only if \(\int \xi\sigma = \inf\{\xi\sigma\}\) (i.e. \(\xi\sigma\) is constant almost everywhere, meaning \((\xi, \sigma)\) is on the central path). Secondly, \(\psi_\rho(\xi, \sigma)\) is compatible with infinity-norm neighborhoods of the central path in the sense defined in [Tu92b].

Now we give a primal feasible solution for Example 1 and show which neighborhoods of the central path include that solution. The corresponding primal problem for \((ED)\) is

\[
\begin{align*}
(EP) \quad \inf & \quad \int_0^1 \xi(t)dt \\
\text{s.t.} & \quad \int_0^1 \cos(2\pi t)\xi(t)dt = 0 \\
& \quad \int_0^1 \sin(2\pi t)\xi(t)dt = -1 \\
& \quad \xi(t) \geq 0, \quad \forall t \in [0, 1].
\end{align*}
\]
A feasible solution that lies in $F_+ (EP)$ is given by (here $\pi$ denotes the trigonometric constant)

$$\xi(t) := \begin{cases} \frac{\pi^2}{1+\pi} & \text{if } t \in [0, 1/2); \\ \frac{2\pi^2}{1+\pi} t & \text{if } t \in [1/2, 3/4); \\ \frac{2\pi^2}{1+\pi} (\frac{3}{2} - t) & \text{if } t \in [3/4, 1]. \end{cases}$$

Note that $\sigma := \eta$ is in $F_+ (ED)$. We have

$$<\xi, \sigma> = \frac{9}{8} \frac{\pi^2}{1+\pi}.$$ 

So, we find

$$\inf \{<\xi, \sigma> \} = \frac{8}{9} \text{ and } \sup \{<\xi, \sigma> \} = \frac{4}{3}.$$ 

Hence we conclude that $(\xi, \sigma) \in \mathcal{N}_\infty (1/9)$ and $(\xi, \sigma) \in \mathcal{N}_\infty (1/3)$.

4 A Path-Following Algorithm for Semi-Infinite Programming

In this section, we study path-following algorithms. The centering parameter is denoted by $\theta \in [0, 1]$, a constant. Recall from section 2 that the affine-scaling and the centering directions for the primal ($\pi'_{AFF}, \pi'_{CEN}$) and the dual ($\delta'_{AFF}, \delta'_{CEN}$) spaces are given by

$$\begin{align*}
\pi'_{AFF} &:= -\xi + \xi \sigma^{-1} a^T \left[\int \xi \sigma^{-1} aa^T\right]^{-1} b \\
\pi'_{CEN} &:= \sigma^{-1} - \xi \sigma^{-1} a^T \left[\int \xi \sigma^{-1} aa^T\right]^{-1} \left[\int \sigma^{-1} a\right] \\
\delta'_{AFF} &:= -a^T \left[\int \xi \sigma^{-1} aa^T\right]^{-1} b \\
\delta'_{CEN} &:= a^T \left[\int \xi \sigma^{-1} aa^T\right]^{-1} \left[\int \sigma^{-1} a\right].
\end{align*}$$

Then for a given centering parameter $\theta$, the search directions of an algorithm can be written as

$$\begin{align*}
\pi_{PATH} &:= \pi'_{AFF} + \theta <\xi, \sigma> \pi'_{CEN}, \\
\delta_{PATH} &:= \delta'_{AFF} + \theta <\xi, \sigma> \delta'_{CEN}.
\end{align*}$$
We define

\[
\begin{align*}
\mu &:= \langle \xi, \sigma \rangle, \\
v &:= \xi^{1/2}\sigma^{1/2} - \theta \mu \xi^{-1/2}\sigma^{-1/2},
\end{align*}
\]

and observe that

\[
\begin{align}
\pi_{\text{PATH}} &= -\xi^{1/2}\sigma^{-1/2}v_p, \\
\delta_{\text{PATH}} &= -\xi^{-1/2}\sigma^{1/2}v_q,
\end{align}
\]  \quad (19)  \quad (20)

where \(v_p := P(v)\) and \(v_q := (I - P)(v)\). Before describing the algorithm we define the next iterate as a function of the step size:

\[
\begin{align*}
\xi_\alpha &:= \xi + \alpha \pi_{\text{PATH}}, \\
\sigma_\alpha &:= \sigma + \alpha \delta_{\text{PATH}}.
\end{align*}
\]

Suppose \((\bar{\xi}, \bar{\sigma}) \in N_\infty(\beta)\) is given. Then the algorithm can be stated as follows.

**Algorithm I**

Set \((\xi, \sigma) := (\bar{\xi}, \bar{\sigma})\).

While \(\langle \xi, \sigma \rangle > \epsilon^*\) do

compute \(\pi_{\text{PATH}}, \delta_{\text{PATH}}\)

choose the maximum step size \(\bar{\alpha} \in (0, 1)\) such that \((\xi_\alpha, \sigma_\alpha) \in N_\infty(\beta)\)

\((\xi, \sigma) := (\xi_\alpha, \sigma_\alpha)\)

end

In the algorithm above one may instead choose the step size to minimize the potential function \(\psi_p(\xi, \sigma)\) in the neighborhood \(N_\infty(\beta)\) (we will call this Algorithm II). As we show in the next section both versions obtain the same theoretical complexity.
5 Convergence Results

The following lemma was also proved by Todd [To91].

Lemma 5.1.

\[ < \xi, \sigma >= [1 - \alpha(1 - \theta)] < \xi, \sigma >. \]

Proof: Note that by Corollary 2.1 \( < v_p, v_q > = 0 \). This implies \( < \pi_{PATH}, \delta_{PATH} > = 0 \).

So,

\[ < \xi, \sigma > = < \xi, \sigma > + \alpha(< \sigma, \pi_{PATH} > + < \xi, \delta_{PATH} >) + \alpha^2 < \pi_{PATH}, \delta_{PATH} > \]

\[ = < \xi, \sigma > - \alpha < \xi^{1/2} \sigma^{1/2}, v_p > + < \xi^{1/2} \sigma^{1/2}, v_q > \]

\[ = < \xi, \sigma > - \alpha < \xi^{1/2} \sigma^{1/2}, v > \]

\[ = [1 - \alpha(1 - \theta)] < \xi, \sigma >. \]

\[ \square \]

Now we need to show a lower bound on \( \bar{\alpha} \). We have the following lemma.

Lemma 5.2.

(a) Let \((\xi, \sigma) \in \mathcal{N}_\infty(\beta)\) and

\[ \alpha \leq \min \left\{ \frac{\beta \theta \mu}{\inf\{v_p v_q\}}, 1 \right\}. \]  

(21)

Then \((\xi_\alpha, \sigma_\alpha) \in \mathcal{N}_\infty(\beta)\).

(b) Let \((\xi, \sigma) \in \mathcal{N}_\infty(\beta)\) and

\[ \alpha \leq \min \left\{ \frac{\beta \theta \mu}{\sup\{|v_p v_q|\}}, 1 \right\}. \]  

(22)

Then \((\xi_\alpha, \sigma_\alpha) \in \mathcal{N}_\infty(\beta)\). In other words,

\[ \bar{\alpha} \geq \min \left\{ \frac{\beta \theta \mu}{\sup\{|v_p v_q|\}}, 1 \right\}. \]
Proof:
(a) Using (19) and (20) and assuming \( \alpha \) satisfies (21),

\[
\xi_\alpha \sigma_\alpha = \xi \sigma - \alpha \xi^{1/2} \sigma^{1/2} (v_p + v_q) + \alpha^2 v_p v_q
\]

\[
= \xi \sigma - \alpha \xi^{1/2} \sigma^{1/2} v + \alpha^2 v_p v_q
\]

\[
\geq (1 - \alpha)(1 - \beta) \mu + \alpha \theta \mu + \alpha^2 v_p v_q
\]

\[
\geq (1 - \alpha)(1 - \beta) \mu + \alpha \theta \mu - \alpha \beta \theta \mu
\]

\[
= (1 - \alpha + \alpha \theta) \mu (1 - \beta)
\]

\[
< \xi_\alpha, \sigma_\alpha > (1 - \beta).
\]

Now we show that for such an \( \alpha (\xi_\alpha, \sigma_\alpha) \) is feasible. Note that \( \alpha \leq 1 \) and \( \theta < 1 \) imply \(< \xi_\alpha, \sigma_\alpha > > 0 \). Then by the continuity of \( \xi_\alpha \sigma_\alpha \) in \( \alpha \) we must have \( \xi_\alpha \sigma_\alpha > 0 \) for all \( \hat{\alpha} \in [0, \alpha] \). Hence by continuity of \( \xi_\alpha \) and \( \sigma_\alpha \) we see that \( \xi_\hat{\alpha} > 0 \) and \( \sigma_\hat{\alpha} > 0 \) for all \( \hat{\alpha} \in [0, \alpha] \). Part (b) can also be easily proven in a similar way. \( \Box \)

If we can show that (for all iterations)

\[
\sup \{|v_p v_q|\} \leq C \mu, \quad (23)
\]

where \( C \geq 1 \) is a constant, then we have a convergence proof for the algorithm. The number of iterations to attain a given improvement in precision is then constant, but depends on \( C \).

Theorem 5.1. Let \((\xi, \sigma) \in \mathcal{N}_\alpha(\beta)\) such that < \( \xi, \sigma > \leq \frac{1}{\epsilon^*}\), where \( \epsilon^* \) is the desired accuracy (i.e. we want to find a solution \((\xi^*, \sigma^*)\) such that < \( \xi^*, \sigma^* > \leq \epsilon^*\)). Suppose \( \sup \{|v_p v_q|\} \leq C \mu \) for all iterations. Then Algorithm I finds such a solution in at most

\[
\frac{2C}{\beta \theta (1 - \theta) \log \left( \frac{1}{\epsilon^*} \right)}
\]

iterations. Under the same assumptions and a suitable value for \( \rho \), Algorithm II terminates in at most

\[
\frac{4C}{\beta \theta (1 - \theta) \ln \left( \frac{1}{\epsilon^*} \right)} + 1
\]

iterations.

Proof: By Lemma 5.2 we have \( \hat{\alpha}_k \geq \frac{\beta \theta}{C} \) for all \( k \); here \( \hat{\alpha}_k \) denotes \( \hat{\alpha} \) at the \( k^{th} \) iteration. Let \( \mu_0 := < \xi, \sigma > \) and \( \mu_k \) be the duality gap after the \( k^{th} \) iteration. Then by Lemma
5.1 we have
\[
\frac{\mu_k}{\mu_0} = \prod_{i=1}^{k} [1 - \bar{\alpha}_i(1 - \theta)].
\]

So,
\[
\mu_k \leq \frac{1}{\epsilon^*} \left[ 1 - \frac{\beta \theta (1 - \theta)}{C} \right]^k.
\]

Therefore, to get \( \mu_k \leq \epsilon^* \) it suffices to have
\[
k \geq \frac{2C}{\beta \theta (1 - \theta)} \log \left( \frac{1}{\epsilon^*} \right).
\]

Now, we turn to Algorithm II. We have
\[
\psi_p(\xi_\alpha, \sigma_\alpha) - \psi_p(\xi, \sigma) = \rho \ln \left( \frac{\mu_1 - \alpha(1 - \theta)}{\mu} \right) - \ln \left( \inf \{ \xi_\alpha \sigma_\alpha \} \right).
\]
\[
+ \ln \left( \inf \{ \xi \sigma \} \right)
\]
\[
\leq -\rho (1 - \theta) - \ln (1 - \beta).
\]

We simply wrote down the change in the potential function and then overestimated the logarithm by its first-order Taylor approximation and used the fact that the iterates lie in \( N_{\infty}(\beta) \). Choosing
\[
\rho = \frac{2C}{\beta \theta (1 - \theta)} \ln \left( \frac{1}{1 - \beta} \right)
\]
gives
\[
\psi_p(\xi_\alpha, \sigma_\alpha) - \psi_p(\xi, \sigma) \leq \psi_p(\xi_\alpha, \sigma_\alpha) - \psi_p(\xi, \sigma) \leq -\ln \left( \frac{1}{1 - \beta} \right).
\]
i.e. we get a constant reduction in the potential function at each iteration. We know that \( \langle \xi, \bar{\sigma} \rangle \leq 1/\epsilon^* \) and that \((\xi, \bar{\sigma}) \in N_{\infty}(\beta)\). So,
\[
\psi_p(\xi, \bar{\sigma}) \leq \rho \ln \left( \frac{1}{\epsilon^*} \right) - \ln (1 - \beta)
\]
\[
\leq \left[ \frac{2C}{\beta \theta (1 - \theta)} \ln \left( \frac{1}{\epsilon^*} \right) + 1 \right] \ln \left( \frac{1}{1 - \beta} \right).
\]

We want \((\xi^*, \sigma^*) \in N_{\infty}(\beta)\) such that
\[
\psi_p(\xi^*, \sigma^*) \leq -\frac{2C}{\beta \theta (1 - \theta)} \ln \left( \frac{1}{1 - \beta} \right) \ln \left( \frac{1}{\epsilon^*} \right).
\]
Since $\psi_p$ decreases by at least $\ln\left(\frac{1}{1-\beta}\right)$ at each iteration the bound follows.

The rest of this section is dedicated to finding bounds on $\sup\{|v_p v_q|\}$. In the finite case, one uses the following inequalities to bound this term:

$$\|v_p\|_2 \leq \|v\|_2 \quad \text{and} \quad \|v_q\|_2 \leq \|v\|_2.$$  \hfill (24)

Then for any $x \in \mathbb{R}^n$,

$$\|x\|_\infty \leq \|x\|_2.$$  \hfill (25)

Hence,

$$\sup\{|v_p v_q|\} \leq \|v_p\|_\infty \|v_q\|_\infty \leq \|v_p\|_2 \|v_q\|_2 \leq \|v\|_2^2.$$  \hfill (26)

In the semi-infinite case (as above) we define

$$\|v\|_2 := \langle v, v \rangle^{1/2},$$

and

$$\|v\|_\infty := \sup |v|.$$  

Then (24) holds (by Lemma 2.5) but the inequality (25) (and hence (26)) may not. This is the only critical relation that does not hold in the semi-infinite setting. It is not hard to see that we may have

$$\frac{\|v_p\|_\infty}{\|v\|_2} \to \infty$$

even on the central path. One of the most important reasons for this behavior is getting too close to the boundary of the feasible set. Since we stop at a desired precision, in certain cases it might be possible to prove that for all iterates the following inequalities hold:

$$\frac{\|v_p\|_\infty}{\|v_p\|_2} \leq C_{2,\infty}, \quad \frac{\|v_q\|_\infty}{\|v_q\|_2} \leq C_{2,\infty}. \hfill (27)$$

Then (23) will hold with $C = C_{2,\infty}$.

This analysis makes the arguments of Todd [To91] more rigorous. As Todd [To91] comments, the complexity of these problems should depend on their smoothness and the smoothness can be quantified as the smallest constant satisfying the inequalities of (27). So, in the case of path-following algorithms an input problem is said to be $(\epsilon^*, \beta, C_{2,\infty})$-smooth if (27) holds for the resulting $v$ for all $(\xi, \sigma) \in \mathcal{N}_\infty(\beta)$ such that
\( \epsilon^* \leq < \xi, \sigma > \leq 1/\epsilon^* \). So, by Theorem 5.1 we have a complexity bound on the number of iterations for all \((\epsilon^*, \beta, C_{2,\infty})\)-smooth problems.

Note that we still have nice properties for \( v \) if the current iterate lies in the neighborhood \( \mathcal{N}_\infty(\beta) \) of the central path:

\[
\|v\|_2^2 = \mu - 2\theta\mu + \theta^2\mu^2 < \xi^{-1}, \sigma^{-1} > \\
\leq \mu \left( 1 - 2\theta + \frac{\theta^2}{1 - \beta} \right) \\
\leq \mu. \tag{29}
\]

For simplicity we assumed here (and will assume from now on) that \( \theta \leq \min\{\frac{1}{4}, 2(1 - \beta)\} \). Since all the iterates lie in \( \mathcal{N}_\infty(\beta) \) throughout the algorithm we can show that \( v \) is smooth and that the smoothness constant is independent of the iteration number and the current duality gap. We have

\[
\|v\|_\infty \leq \|\xi^{1/2}\sigma^{1/2}\|_\infty + \theta\mu\|\xi^{-1/2}\sigma^{-1/2}\|_\infty \\
\leq \left( \sqrt{1 + \beta} + \frac{\theta}{\sqrt{1 - \beta}} \right) \sqrt{\mu} \tag{31} \\
\leq \left( \sqrt{1 + \beta} + 2\sqrt{1 - \beta} \right) \sqrt{\mu} \\
\leq 3\sqrt{\mu}. \tag{32}
\]

Note that to go from (30) to (31) we needed to provide an upperbound on \( \|\xi^{1/2}\sigma^{1/2}\|_\infty \). This is the main reason for using the neighborhood \( \mathcal{N}_\infty(\beta) \) rather than \( \mathcal{N}_\infty^{-\infty}(\beta) \). We also know from (28) that

\[
\|v\|_2 = \sqrt{\mu} \left( 1 - 2\theta + \theta^2\mu < \xi^{-1}, \sigma^{-1} > \right)^{1/2} \\
\geq \frac{\sqrt{\mu}}{2}.
\]

The last inequality follows from our assumption that \( \theta \leq \frac{1}{4} \). So,

\[
\|v\|_\infty \leq \left( \sqrt{2(1 + \beta)} + 2\sqrt{2(1 - \beta)} \right) \|v\|_2 \\
\leq 5\|v\|_2. \tag{33}
\]

Now, we attempt to relate smoothness more directly to the input data. We will use some results due to Amir and Ziegler [AZ76] (Theorem 5.2) and Ziegler [Zi77] (Theorem 5.3). Also see Borwein and Lewis [BL91b] for a discussion of some related issues.
Theorem 5.2. Let $S_n$ be the set of polynomials on $[0, 1]$ of degree at most $n$. Define

$$k_n(p) := \min\{ \frac{\| \varphi \|_p}{\| \varphi \|_{\infty}} : \varphi \in S_n \}.$$ 

Then for all $p > 1$ we have

$$k_n(p) \geq \frac{1}{(n + 1)^2},$$

and for $n$ even

$$k_n(p) \leq \left[ \frac{4}{(n + 2)^2} \right]^{1/p},$$

while for $n$ odd

$$k_n(p) \leq \left[ \frac{4}{(n + 1)(n + 3)} \right]^{1/p}.$$ 

Moreover, the lower bound can be improved to $(\frac{1}{n+1})^{2/p}$ for $p \leq 2$.

\[ \Box \]

Theorem 5.3. Let $S_n$ be the set of trigonometric polynomials on $[0, 1]$ of degree at most $n$. Define $k_n(p)$ as before. Then for $1 \leq p \leq 2$ we have

$$\frac{1}{(2n + 1)^{1/p}} \leq k_n(p) \leq \frac{1}{n^{1/p}}.$$ 

For $p > 2$ we have

$$\frac{1}{(2np + 1)^{1/p}} \leq k_n(p) \leq \frac{1}{(np/2 + 1)^{1/p}}.$$

\[ \Box \]

Unfortunately, these results do not directly imply that when the input data is smooth, $C_{2,\infty}$ can be bounded, because even though the original data is smooth the scaling and projection operations may be very rough, i.e., destroy the smoothness. Now we show how to use theorems 5.2 and 5.3 to get complexity bounds.

Lemma 5.3. Suppose $(\xi, \sigma) \in \mathcal{N}_{\infty}^{\infty}(\beta)$ and $\xi \leq M_\xi \eta, \sigma \leq M_\sigma \eta$. Then

$$\| \xi^{1/2} \sigma^{-1/2} \|_{\infty} \| \xi^{-1/2} \sigma^{1/2} \|_{\infty} \leq \frac{M_\xi M_\sigma}{(1 - \beta)} < \xi, \sigma >.$$ 

Proof: Since $(\xi, \sigma) \in \mathcal{N}_{\infty}^{\infty}(\beta)$ we have

$$\xi \sigma \geq (1 - \beta) < \xi, \sigma >$$

26
which implies
\[ \xi \sigma^{-1} \leq \frac{\xi^2}{(1 - \beta)} < \xi, \sigma > \]
and
\[ \xi^{-1} \sigma \leq \frac{\sigma^2}{(1 - \beta)} < \xi, \sigma >. \]
Therefore,
\[ \|\xi^{1/2} \sigma^{-1/2}\|_\infty \|\xi^{-1/2} \sigma^{1/2}\|_\infty \leq \frac{M_\xi M_\sigma}{(1 - \beta)} < \xi, \sigma >. \]
\[ \square \]

The following lemma provides a way of relating the smoothness of \( v_p \) and \( v_q \) to that of the original data.

**Lemma 5.4.** Suppose \((\xi, \sigma) \in \mathcal{N}_\infty(\beta)\) and \(\xi \leq M_\xi \eta, \sigma \leq M_\sigma \eta\). Let \( S \subset \mathcal{C}(T) \) be a subspace of functions such that
\[ \varphi \in S \Rightarrow \frac{\|\varphi\|_\infty}{\|\varphi\|_2} \leq K(S), \]
where \( K(S) \) is a constant (we say \( S \) is \( K(S) \)-smooth). If \( \{a_i, i \in \{1, \ldots, m\}\} \subset S \) then
\[ \|v_q\|_\infty \leq \frac{K(S) M_\xi M_\sigma}{(1 - \beta) \mu} \|v\|_2, \]
and
\[ \|v_p\|_\infty \leq \left( 5 + \frac{K(S) M_\xi M_\sigma}{(1 - \beta) \mu} \right) \|v\|_2. \]

Hence, assuming \( \frac{K(S) M_\xi M_\sigma}{(1 - \beta) \mu} \geq 3 \), we have
\[ \|v_p v_q\|_\infty \leq \frac{2}{\mu} \left( \frac{K(S) M_\xi M_\sigma}{(1 - \beta)} \right)^2. \]

**Proof:** By definition we have
\[ \frac{\|v_q\|_\infty}{\|v_q\|_2} = \frac{\|\xi^{1/2} \sigma^{-1/2} a^T \left[ \int \xi \sigma^{-1} a a^T \right]^{-1} \int a \xi^{1/2} \sigma^{-1/2} v\|_\infty}{\|v_q\|_2} \]
\[ \leq \frac{\|\xi^{1/2} \sigma^{-1/2}\|_\infty}{\|v_q\|_2} \frac{\|a^T \left[ \int \xi \sigma^{-1} a a^T \right]^{-1} \int a \xi^{1/2} \sigma^{-1/2} v\|_\infty}{\|v_q\|_2}. \] (34)
Note that the function in the numerator of (34) belongs to \( S \). Hence,

\[
\frac{\|v_q\|_\infty}{\|v_q\|_2} \leq K(S)\|\xi^{1/2}\sigma^{-1/2}\|_\infty \frac{\|a^T[a\sigma^{-1}a]^T\|^{-1}\|\xi^{1/2}\sigma^{-1/2}\|_2}{\|v_q\|_2}
\]

\[
= K(S)\|\xi^{1/2}\sigma^{-1/2}\|_\infty \frac{\|\xi^{1/2}\sigma^{-1/2}\|_2}{\|v_q\|_2}
\]

\[
\leq K(S)\|\xi^{1/2}\sigma^{-1/2}\|_\infty \|\xi^{-1/2}\sigma^{1/2}\|_\infty.
\]

Using Lemma 5.3, we get

\[
\frac{\|v_q\|_\infty}{\|v_q\|_2} \leq \frac{K(S)M_\xi M_\sigma}{(1 - \beta)\mu}.
\]

So,

\[
\|v_q\|_\infty \leq \frac{K(S)M_\xi M_\sigma}{(1 - \beta)\mu} \|v\|_2.
\]

By definition \( v_p = v - v_q \); and using (33) we find

\[
\|v_p\|_\infty \leq \|v\|_\infty + \|v_q\|_\infty \leq \left( 5 + \frac{K(S)M_\xi M_\sigma}{(1 - \beta)\mu} \right) \|v\|_2.
\]

Finally, the inequalities above together with (29) and (33) give the last result.

\[\square\]

**Corollary 5.1.** Suppose the assumptions of Lemma 5.4 hold. Then

\[
\tilde{\alpha} \geq \min \left\{ \frac{\theta\beta(1 - \beta)^2\mu^2}{2M_\xi^2 M_\sigma^2[K(S)]^2}, 1 \right\}.
\]

**Proof:** Follows from Lemma 5.2 and Lemma 5.4. \(\square\)
Corollary 5.2. Suppose the assumptions of Lemma 5.4 hold. Then

(a) If the functions \( a_i, i \in \{1, 2, \ldots, m\} \) are trigonometric polynomials of degree \( n \), then

\[
\bar{\alpha} \geq \min \left\{ \frac{\theta (1 - \beta)^2 \mu^2}{2(M_\xi)^2(M_\sigma)^2(2n + 1)^1} \right\}.
\]

(b) If the functions \( a_i, i \in \{1, 2, \ldots, m\} \) are polynomials of degree \( n \), then

\[
\bar{\alpha} \geq \min \left\{ \frac{\theta (1 - \beta)^2 \mu^2}{2(M_\xi)^2(M_\sigma)^2(n + 1)^2} \right\}.
\]

Proof: Part (a) follows from Theorem 5.3 and Corollary 5.1. Part (b) follows from Theorem 5.2 and Corollary 5.1.

Now we can establish some complexity bounds.

Theorem 5.4. Suppose we are given \((\xi, \sigma) \in N_\infty(\beta)\) such that \( < \xi, \sigma > = \mu_0 \). Let \( \epsilon^* \) be the desired accuracy. Suppose \( \xi \leq M_\xi \eta \) and \( \sigma \leq M_\sigma \eta \) (with \( M_\xi \geq 1, M_\sigma \geq 1 \)) for all \((\xi, \sigma) \in N_\infty(\beta)\) such that \( < \xi, \sigma > \geq [\epsilon^*, \mu^*_0] \). If \( \{a_i, i \in \{1, \ldots, m\}\} \subset S \subset C(T) \) where \( S \) is a subspace of \( K(S) \)-smooth functions \( (K(S) \geq 1) \) and \( M_\xi, M_\sigma \) are \( O(1) \), then the algorithm terminates in \( O\left(\frac{(K(S))^2}{\epsilon^*} \right) \) iterations.

Proof: Define

\[
R := \frac{3\theta (1 - \beta)^2}{8M_\xi^2M_\sigma^2K(S)^2}.
\]

So, by Lemma 5.1 and Corollary 5.1,

\[
\mu_k \leq \mu_0 \prod_{i=1}^{k} (1 - R\mu_{i-1}^2).
\]

(35)

To estimate the required number of iterations, we treat the sequence \( \{\mu_k\} \) in chunks. Suppose the current duality gap is \( \mu_j \). Then to get \( \mu_k / \mu_j \leq 1/2 \) one needs at most \( 4/R\mu_0^2 \) iterations, because we have \( \mu_i \geq \mu_j/2 \forall i \in \{j, j+1, \ldots, k-1\} \) and (35). Therefore, the number of iterations required to get \( \mu_k \leq \epsilon^* \) is at most

\[
\sum_{j=1}^{\lceil \log(\mu_0 / \epsilon^*) \rceil} \frac{4^j}{R\mu_0^2} \leq \frac{4}{R\mu_0^2} 2^\log((2\mu_0 / \epsilon^*)^2) = \frac{16}{R(\epsilon^*)^2}.
\]
Hence, $O \left( \frac{K(S) v}{\epsilon^2} \right)^2$ iterations suffice.

**Corollary 5.3.** Suppose the assumptions of Lemma 5.4 hold. Then

(a) If the functions $a_i, i \in \{1, 2, \ldots, m\}$ are trigonometric polynomials of degree $n$, then the algorithm converges in $O\left( \frac{n}{\epsilon^{1/2}} \right)$ iterations.

(b) If the functions $a_i, i \in \{1, 2, \ldots, m\}$ are polynomials of degree $n$, then the algorithm converges in $O\left( \frac{n^2}{\epsilon^{1/2}} \right)$ iterations.

**Proof:** Follows from Theorem 5.4 and Corollary 5.2.

Now we consider a more general setting in which the functions $\{a_i, i \in \{1, \ldots, m\}\}$ do not necessarily belong to a subspace $S$ which is $K(S)$-smooth. We let

$$t^* := \text{argsup}\{t \in T : \xi(t)\sigma^{-1}(t)\},$$

and to define the measure of smoothness we also define the set $WIM_\kappa$ (Where It Matters):

$$WIM_\kappa := \left\{ t \in T : \xi(t)\sigma^{-1}(t) \geq \frac{1}{\kappa} \xi(t^*)\sigma^{-1}(t^*) \right\},$$

where $\kappa \geq 1$ is a parameter. Note that if $\kappa$ is fixed, as the iterates converge to an optimal solution the measure of the set $WIM_\kappa$, $M(WIM_\kappa)$ will probably converge to zero. First we give a lemma that will be useful.

**Lemma 5.5.** Suppose $(\xi, \sigma) \in \mathcal{N}_\infty(\beta)$ and $\xi \leq M\xi, \sigma \leq M\sigma$. Then

$$\left\| \int [\xi\sigma^{-1}aa^T]^{-1} \right\|_\infty \leq \sqrt{m}\kappa\xi^{-1}(t^*)\sigma(t^*) \left\| \left[ \int_{WIM_\kappa} aa^T \right]^{-1} \right\|_2.$$

**Proof:** The following inequalities are matrix inequalities with respect to the cone of symmetric positive semi-definite matrices ($A \geq B$ iff $A - B$ is positive semi-definite). We find

$$\int [\xi\sigma^{-1}aa^T] \geq \int_{WIM_\kappa} [\xi\sigma^{-1}aa^T] \geq \frac{1}{\kappa} \xi(t^*)\sigma^{-1}(t^*) \int_{WIM_\kappa} aa^T.$$
So,

\[
\left[ \int \xi \sigma^{-1} a a^T \right]^{-1} \leq \kappa \xi^{-1}(t^*) \sigma(t^*) \left[ \int_{WIM} a a^T \right]^{-1},
\]

which implies

\[
\left\| \left[ \int \xi \sigma^{-1} a a^T \right]^{-1} \right\|_2 \leq \kappa \xi^{-1}(t^*) \sigma(t^*) \left\| \left[ \int_{WIM} a a^T \right]^{-1} \right\|_2.
\]

Hence, we have

\[
\left\| \left[ \int \xi \sigma^{-1} a a^T \right]^{-1} \right\|_\infty \leq \sqrt{m} \kappa \xi^{-1}(t^*) \sigma(t^*) \left\| \left[ \int_{WIM} a a^T \right]^{-1} \right\|_2.
\]

Clearly, Lemma 5.5 provides a gross overestimate, yet it suffices to provide a lower bound on the maximum step size. Given a vector of functions \( a \), \( \sup(a) \) will denote the vector of the individual supremums of the functions.

**Lemma 5.6.** Suppose \( (\xi, \sigma) \in N_\infty(\beta) \) and \( \xi \leq M \eta, \sigma \leq M \eta \). Then

\[
\tilde{\alpha} \geq \min \left\{ \frac{\beta \theta}{18 \kappa^2 m^3 \sup(|a|) \| \| \left[ \int_{WIM} a a^T \right]^{-1} \|_2} \right\}.
\]

**Proof:** We have

\[
\| u \|_\infty = \left\| \xi^{1/2} \sigma^{-1/2} a^T \left[ \int \xi \sigma^{-1} a a^T \right]^{-1} \int \xi^{1/2} \sigma^{-1/2} a^T \right\|_\infty \\
\leq m \xi(t^*) \sigma^{-1}(t^*) \| \sup(|a|) \| \| \left[ \int \xi \sigma^{-1} a a^T \right]^{-1} \|_\infty \\
\leq 3 m^{3/2} \kappa \| \sup(|a|) \| \| \left[ \int_{WIM} a a^T \right]^{-1} \|_2 \sqrt{\mu}.
\]

The last inequality follows from Lemma 5.5 and (32). Note that

\[
\left\| \left[ \int_{WIM} a a^T \right]^{-1} \right\|_2 \geq \left\| \left[ I a a^T \right]^{-1} \right\|_2
\]

31
\[
\text{So, in a similar way to the proof of Lemma 5.4 we get}
\]
\[
\|v_p\|_\infty \leq 6m^{3/2}\kappa \|\text{sup}(|a|)\|_\infty^2 \left\| \int_{WIM_\kappa} aa^T \right\|_2^{-1} \sqrt{\mu}.
\]

So, by Lemma 5.2,
\[
\bar{\alpha} \geq \min \left\{ \frac{\beta \theta}{18\kappa^2 m^3 \|\text{sup}(|a|)\|_\infty^4 \left\| \int_{WIM_\kappa} aa^T \right\|_2^{-1} \| \right\}, 1 \right\}.
\]

\[\square\]

If one lets
\[
\kappa := \frac{M_\kappa^2 M_\sigma^2}{(1 - \beta)^2 \mu^2},
\]
then by Lemma 5.3 we get \(WIM_\kappa = T\) so that the expression providing a lower bound on the maximum step size is independent of the measure of \(WIM_\kappa\). Following a similar proof to that of Theorem 5.4, it is not hard to see that the number of iterations grows with the following expression:
\[
\left( \frac{1}{\epsilon^2} \right)^4 m^3 \|\text{sup}(|a|)\|_\infty^4 \left\| \int_{WIM_\kappa} aa^T \right\|_2^{-1} \|.
\]

6 Conclusion

We showed that the many of the ingredients used in interior-point methods for linear programming problems hold in the semi-infinite setting. One of the most important properties that fail in the limit is a norm inequality. In a finite-dimensional Euclidean space it is clear that the infinity norm of an element is at most equal to the two norm of the same element which clearly is not true in general for a pre-Hilbert space. Along these lines, one future possibility is to search for algorithms that do not use the norm inequality (as described above) in the finite case. Since we showed that in the limit it is possible to relate different measures to the iteration complexity of the algorithms, this might also help to explain the slow growth of the number of iterations taken by interior-point algorithms as the number of variables grows.

We finally note that the algorithms discussed here could also be considered for implementation on semi-infinite programming problems. However, the derivation and/or evaluation of the functions defined here would be a very hard problem in practice.
References


