Mathematical Programming 47 (1990) 175-201
North-Holland

AN ALGORITHM FOR LINEAR PROGRAMMING WHICH REQUIRES $O(((m+n)n^2+(m+n)^{1.5}n)L)$ ARITHMETIC OPERATIONS

Pravin M. VAIDYA
AT&T Bell Laboratories, Murray Hill, NJ 07974, USA

Received 10 December 1987
Revised manuscript received 29 August 1988

We present an algorithm for linear programming which requires $O(((m+n)n^2+(m+n)^{1.5}n)L)$ arithmetic operations where m is the number of constraints, and n is the number of variables. Each operation is performed to a precision of $O(L)$ bits. L is bounded by the number of bits in the input. The worst-case running time of the algorithm is better than that of Karmarkar's algorithm by a factor of $\sqrt{m+n}$.

Key words: Optimization, linear programming, complexity, polynomial time algorithms.

1. Introduction

We study the linear programming problem

$$\begin{array}{ll} \max & c^{\mathrm{T}}x \\ \text{s.t.} & Ax \geqslant b \end{array}$$

where $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$ and $c \in \mathbb{R}^n$. We assume that the polytope defined by $Ax \geqslant b$ is bounded and has a non-zero interior. As the polytope is bounded we can assume that $m \geqslant n$, and that the columns of A are linearly independent.

A polynomial time algorithm for the linear programming problem was first presented by Khachian [6] using the ellipsoid method. Khachian's algorithm requires $O(mn^3L)$ arithmetic operations in the worst case, and each operation is performed to a precision of $O(L)$ bits where

$$L = \log_2(\text{largest absolute value of the determinant of any square submatrix of } A)$$
$$+\log_2\left(\max_i c_i\right)+\log_2\left(\max_i b_i\right)+\log_2(m+n).$$

In [5], Karmarkar presents an interior point algorithm which requires $O((m^{1.5}n^2+m^2n)L)$ arithmetic operations, each operation being performed to a precision of $O(L)$ bits. We present an algorithm for the linear programming problem which requires $O((mn^2+m^{1.5}n)L)$ arithmetic operations in the worst case, and it is adequate to perform each arithmetic operation to a precision of $O(L)$ bits. The algorithm presented in this paper is thus faster than Karmarkar's algorithm [5] by a factor of $\sqrt{m}$ for all values of m and n. It is also faster than Khachian's ellipsoid algorithm by a factor of n for $m \leqslant n^2$, and continues to be faster than his algorithm for $m \leqslant n^4$. (Typically, m and n are of the same order.)

In [8], Renegar gives an algorithm for linear programming which requires $O(m^{1.5}n^2L)$ arithmetic operations. We build on the ideas in [8], and obtain a faster algorithm. In our algorithm we construct a sequence $P^0, P^1, \ldots, P^k, \ldots,$ of smaller and smaller polytopes which shrink towards the optimal vertex (facet). During the kth iteration, we move from the center of polytope P^{k-1} to the center of polytope P^k by performing a local optimization which consists of minimizing a linear function over an ellipsoid. The total number of iterations performed by the algorithm is $O(\sqrt{m}\,L)$, and on the average each iteration requires $O(\sqrt{m}\,n^2+mn)$ arithmetic operations where the average is taken over all the iterations. Thus the total number of arithmetic operations is $O((mn^2+m^{1.5}n)L)$. When the algorithm terminates we have a feasible point that is sufficiently close in objective function value to the optimum over the original polytope, and we may then construct an optimal vertex as described in Section 9 in $O(mn^2)$ arithmetic operations.

In Section 2 we give an overview, and describe the relationship with related work [1, 8]. In Section 3 we give the algorithm. In Section 4 we give some properties of the potentials used to measure convergence, and in Section 5 we describe the local optimizations. In Section 6 we show how to amortize the number of arithmetic operations. In Section 7 we show how to reduce any linear program to the format required by the algorithm, and in Section 8 we show that $O(L)$ precision is adequate for arithmetic operations. In Section 9 we describe how an optimal vertex may be obtained from a feasible point which is sufficiently close in objective function value to the optimum.

At this point we note that a condensed version [11] of this paper appeared in the Proceedings of the 19th Annual ACM Symposium on Theory of Computing, May 1987. We would also like to point out that a bound of $O((m+n)^3L)$ arithmetic operations for linear programming has been independently obtained by Gonzaga [4].

2. An overview

Let P be the given polytope

$$P=\{x\colon Ax \geqslant b\}$$

and let $\beta^{\max}$ be the maximum value of the objective function $c^{\mathrm{T}}x$ over P. Let $\beta^0, \beta^1, \ldots, \beta^k, \ldots,$ be a strictly increasing sequence such that $\lim_{k\to\infty}\beta^k=\beta^{\max}$. Let π^k denote the set of linear inequalities $\{Ax\geqslant b, c^{\mathrm{T}}x\geqslant\beta^k\}$, and P^k denote the polytope $P^k=\{x\colon Ax\geqslant b, c^{\mathrm{T}}x\geqslant\beta^k\}$. Let a_i^{T} denote the ith row of the constraint matrix A. The center of the system of linear inequalities π^k is defined to be the unique point that maximizes the potential function

$$F^k(x)=\sum_{i=1}^{m}\ln(a_i^{\mathrm{T}}x-b_i)+m\ln(c^{\mathrm{T}}x-\beta^k)$$

over the interior of the polytope P^k. Since the columns of A are linearly independent, $F^k(x)$ is a strictly concave function over the interior of P^k, and so the center of π^k is indeed a unique point. We let ω^k denote the center of the system π^k. Let $f^k(x)$ be the normalized potential function given by $f^k(x) = F^k(\omega^k) - F^k(x)$. $f^k(x)$ is a strictly convex function. Furthermore, the value of $F^k(x)$ at ω^k is not required for evaluating the derivatives of $f^k(x)$, and for computing $f^k(y) - f^k(z)$ for a pair of points y and z.

The algorithm generates a sequence of points $x^0, x^1, \ldots, x^k, \ldots$, such that x^k is in the interior of polytope P^k, and x^k is a good approximation to the center ω^k of the system π^k. Specifically, each x^k satisfies the condition $f^k(x^k) \leq 0.04$. We shall assume that β^0 satisfies the condition $\beta^{\max} - \beta^0 = 2^{O(L)}$, and that we have a point x^0 such that $f^0(x^0) \leq 0.04$. In Section 7 we show how to transform the given linear program so that a required starting point x^0 is available. In the kth iteration, β^k is computed as

$$\beta^k = \beta^{k-1} + (\alpha/\sqrt{m})(c^{\mathrm{T}}x^{k-1} - \beta^{k-1})$$

for some constant α. As the bounding objective function hyperplane shifts due to the change in β, x^{k-1} may no longer be a good approximation to the new center ω^k, and so a local optimization is performed to decrease the potential $f^k(x)$ and obtain a better approximation x^k to the center ω^k. The local optimization consists of minimizing a linear function over an ellipsoid. The sequence $\beta^0, \beta^1, \ldots, \beta^k, \ldots$, satisfies the condition

$$\beta^{\max} - \beta^k \leq (1 - \alpha'/\sqrt{m})(\beta^{\max} - \beta^{k-1})$$

for some constant α' dependent on α. So the corresponding sequence of systems $\pi^0, \pi^1, \ldots, \pi^k, \ldots$, induces a sequence of smaller and smaller polytopes $P^0, P^1, \ldots, P^k, \ldots$, which shrink geometrically towards the optimal vertex (facet). In $O(\sqrt{m}\, L)$ iterations we obtain a point where the objective function value is $2^{-O(L)}$ away from the optimal value over the original polytope P, and we may then isolate a set of constraints which define an optimal vertex of P as described in Section 9.

We shall now describe how to find x^k from x^{k-1}. It may be shown that if $f^{k-1}(x^{k-1}) \leq 0.04$ and $\alpha \leq \frac{1}{30}$ then $f^k(x^{k-1}) \leq 0.05$. Thus at the start of the kth iteration we have a point x^{k-1} satisfying the condition $f^k(x^{k-1}) \leq 0.05$, and we have to obtain a point x^k such that $f^k(x^k) \leq 0.04$. Let D be an $m \times m$ diagonal matrix such that the ith diagonal entry D_{ii} satisfies the condition $1/(1.1(a_i^{\mathrm{T}}x^{k-1} - b_i)^2) \leq D_{ii} \leq 1.1/(a_i^{\mathrm{T}}x^{k-1} - b_i)^2$. Let $E^k(r)$ be the ellipsoid defined by

$$E^k(r) = \left\{x\colon (x - x^{k-1})^{\mathrm{T}}\left(A^{\mathrm{T}}DA + \frac{m}{(c^{\mathrm{T}}x^{k-1} - \beta^k)^2}cc^{\mathrm{T}}\right)(x - x^{k-1}) \leq r^2 f^k(x^{k-1})\right\}$$

where r is a suitable parameter between 0 and 1. We note that $E^k(r)$ is contained in the polytope P^k. Let η^k be the gradient of $f^k(x)$ evaluated at x^{k-1}. Consider the power series expansion of $f^k(x)$ at x^{k-1}. Within the ellipsoid $E^k(r)$, the magnitude of the second order term in this series is bounded by $0.55r^2 f^k(x^{k-1})$, and the sum of magnitudes of the higher order terms is at most $0.1r^3 f^k(x^{k-1})/(1-0.27r)$, whereas the minimum value attained by the first order (linear) term within the ellipsoid is less than $-0.9rf^k(x^{k-1})$. So minimizing $(\eta^k)^T(x-x^{k-1})$, the linear term in the power series, over the ellipsoid $E^k(r)$ will decrease $f^k(x)$ by approximately a factor of $(1-0.9r+0.55r^2)$. Let $z^k(r)$ be the point that minimizes the linear function $(\eta^k)^T x$ over the ellipsoid $E^k(r)$. If $0.5 \leq r \leq 0.8$ then $z^k(r)$ reduces the potential $f^k(x)$ by at least 25% and $f^k(z^k(r)) \leq 0.75 f^k(x^{k-1})$.

x^k is computed as follows. From the theory of convex functions [13], $z^k(r)-x^{k-1}$ satisfies the system of linear equations

$$\left(A^T D A + \frac{m}{(c^T x^{k-1} - \beta^k)^2} cc^T\right)(z^k(r) - x^{k-1}) = -t(r)\eta^k$$

for some scalar $t(r) > 0$. We first compute ξ^k, a vector in the direction of $z^k(r) - x^{k-1}$, by solving the system of linear equations

$$\left(A^T D A + \frac{m}{(c^T x^{k-1} - \beta^k)^2} cc^T\right)\xi^k = -\eta^k.$$

It is worth noting that the vector ξ^k could be in a direction that is different from the direction in Newton's method [8]. Next, we find a scalar $t^k > 0$ such that

$$0.018 \leq (t^k)^2(\xi^k)^T\left(A^T D A + \frac{m}{(c^T x^{k-1} - \beta^k)^2} cc^T\right)\xi^k \leq 0.0196.$$

If $0.04 \leq f^k(x^{k-1}) \leq 0.05$ then $x^{k-1} + t^k\xi^k$ minimizes the function $(\eta^k)^T x$ over the ellipsoid $E^k(r_0)$, for some r_0 in the range $[0.6, 0.7]$, and $f^k(x^{k-1} + t^k\xi^k) \leq 0.75 f^k(x^{k-1}) \leq 0.04$. Thus either $f^k(x^{k-1}) \leq 0.04$ or $f^k(x^{k-1} + t^k\xi^k) \leq 0.04$. It suffices to let x^k be that point where the potential $f^k(x)$ is lower among the two points x^{k-1} and $x^{k-1} + t^k\xi^k$.

An algorithm based on the idea of producing a sequence of shrinking polytopes together with a sequence of approximate centers was given by Renegar [8]. An approach based on centers is also suggested in [1, 9] but without any analysis. In developing our algorithm we follow Renegar's approach, but there are two critical differences which enable us to obtain a faster algorithm. First, closeness to the center of a polytope is measured in a different manner. Renegar [8] measures closeness to the center in terms of euclidean distance in a transformed domain, and shows that a local optimization decreases the distance to the current center. We measure closeness in terms of the potentials $f^k(x)$. Second, the local optimizations described in this paper are quite different from Renegar's [8], and can actually increase the distance metric used by him in [8] to measure closeness to the center. Finally, Renegar's algorithm requires $O(m^{1.5}n^2L)$ arithmetic operations. Measuring local convergence in terms of the potentials $f^k(x)$ allows us to amortize the number of arithmetic operations, and obtain a bound of $O((mn^2 + m^{1.5}n)L)$ on the total number of arithmetic operations performed by our algorithm.

In [1] Bayer and Lagarias analyze an infinitesimal version of Karmarkar's algorithm, and study trajectories leading from each point to the optimum defined by taking infinitesimal steps. It is interesting to note that the points ω^0, $\omega^1, \ldots, \omega^k, \ldots$, lie on such a trajectory that would be generated by starting the infinitesimal version of Karmarkar's algorithm at ω^0. Thus our algorithm could also be viewed as efficiently following this trajectory to the optimum.

3. The algorithm

In this section we give the actual algorithm. We assume that we are given a β^0 such that $\beta^{\max} - \beta^0 = 2^{O(L)}$, and an x^0 which is close enough to the center ω^0 of the system of linear inequalities $\pi^0 = \{Ax \geq b, c^T x \geq \beta^0\}$. Specifically, $f^0(x^0) \leq 0.04$. In Section 7 we shall describe how to suitably transform the given linear program so that a required β^0 and x^0 are available.

At the beginning of the kth iteration we have a parameter β^{k-1}, and a feasible point x^{k-1} such that $c^T x^{k-1} > \beta^{k-1}$ and $f^{k-1}(x^{k-1}) \leq 0.04$. We also have a diagonal matrix D such that $1/(1.1(a_i^T x^{k-1} - b_i)^2) \leq D_{ii} \leq 1.1/(a_i^T x^{k-1} - b_i)^2$ for $i = 1, \ldots, m$. During the kth iteration we perform the following computations in sequence.

(1) $\beta^k := \beta^{k-1} + (1/30\sqrt{m})(c^T x^{k-1} - \beta^{k-1})$.

(2) Determine a direction ξ^k by solving

$$\left(A^T D A + \frac{m}{(c^T x^{k-1} - \beta^k)^2} c c^T\right) \xi^k = -\eta^k,$$

where η^k is the gradient of $f^k(x)$ evaluated at x^{k-1}.

(3) Compute a scalar $t^k > 0$ such that

$$0.018 \leq (t^k)^2 (\xi^k)^T \left(A^T D A + \frac{m}{(c^T x^{k-1} - \beta^k)^2} c c^T\right) \xi^k \leq 0.0196.$$

(4) If $f^k(x^{k-1} + t^k \xi^k) < f^k(x^{k-1})$ then $x^k := x^{k-1} + t^k \xi^k$ else $x^k := x^{k-1}$.

(5) For each i, $1 \leq i \leq m$, if

$$D_{ii} < 1/(1.1(a_i^T x^k - b_i)^2) \quad \text{or} \quad D_{ii} > 1.1/(a_i^T x^k - b_i)^2$$

then

$$D_{ii} := 1/(a_i^T x^k - b_i)^2.$$

The algorithm halts when $c^T x^k - \beta^k \leq 2^{-13L}$. Each iterate x^k is a good approximation to the center ω^k, and satisfies the condition $f^k(x^k) \leq 0.04$. It is this property of the algorithm which leads to a bound of $O(\sqrt{m}\, L)$ on the number of iterations. Using this property we can show the following theorem.

Theorem 1. *If* $m \geqslant 16$ *then for all k,* $c^{\mathrm{T}}x^k - \beta^k \geqslant 0.4(\beta^{\max} - \beta^k)$. □

From Theorem 1, it follows that

$$\beta^{\max} - \beta^k \leqslant (1 - 0.4/(30\sqrt{m}))(\beta^{\max} - \beta^{k-1}),$$

and since $\beta^{\max} - \beta^0 = 2^{O(L)}$, $c^{\mathrm{T}}x^k - \beta^k$ must fall below 2^{-13L} in $O(\sqrt{m}\, L)$ iterations. Thus the algorithm halts in $O(\sqrt{m}\, L)$ iterations. From Theorem 1, it also follows that when the algorithm halts, $\beta^{\max} - c^{\mathrm{T}}x^k \leqslant 1.25 \times 2^{-13L}$. Then, as described in Section 9, using the final point generated by the algorithm an optimal vertex may be computed in $O(mn^2)$ arithmetic operations, each operation being performed to a precision of $O(L)$ bits.

The proof of Theorem 1 is based on the following lemmas. For the lemmas below we assume that $m \geqslant 16$.

Lemma 1. *For all k,*

$$c^{\mathrm{T}}\omega^k - \beta^k \geqslant 0.5(\beta^{\max} - \beta^k),$$

and if $f^k(x) \leqslant 0.04$ *then*

$$|c^{\mathrm{T}}\omega^k - c^{\mathrm{T}}x|/(c^{\mathrm{T}}\omega^k - \beta^k) \leqslant 0.4/\sqrt{2m}. \qquad \square$$

Lemma 2. *For all k, if* $f^{k-1}(x^{k-1}) \leqslant 0.04$ *then* $f^k(x^{k-1}) \leqslant 0.05$. □

Lemma 3. *For all k, if* $0.04 \leqslant f^k(x^{k-1}) \leqslant 0.05$ *then* $f^k(x^{k-1} + t^k\xi^k) \leqslant 0.04$.

Lemma 1 follows from Lemmas 4 and 5 which are proved in Section 4. Lemma 2 follows from Lemma 7 that is also proved in Section 4. A proof of Lemma 3 is given in Section 5. From Lemmas 2 and 3, it inductively follows that for all k, $f^k(x^k) \leqslant 0.04$. Theorem 1 then follows from Lemma 1.

4. Potential functions

We shall study some properties of the potential

$$F(x) = \sum_{i=1}^{m} \ln(a_i^{\mathrm{T}}x - b_i) + m \ln(c^{\mathrm{T}}x - \beta)$$

and the point ω that maximizes $F(x)$ over the polytope $P^{\beta} = \{x\colon Ax \geqslant b,\ c^{\mathrm{T}}x \geqslant \beta\}$. We note that ω is a unique point since $F(x)$ is strictly concave. ($F(x)$ is strictly concave since the polytope is bounded and has a non-empty interior.) Let $f(x)$ be the normalized potential given by $f(x) = F(\omega) - F(x)$.

Lemma 4. *For any point x in the polytope P^{β},*

$$\sum_{i=1}^{m} \frac{(a_i^{\mathrm{T}}x - b_i)}{(a_i^{\mathrm{T}}\omega - b_i)} + m\frac{c^{\mathrm{T}}x - \beta}{c^{\mathrm{T}}\omega - \beta} = 2m$$

and

$$c^{\mathrm{T}}\omega - \beta \geqslant 0.5(c^{\mathrm{T}}x - \beta).$$

Proof. A proof is given in [8] but we sketch one for completeness. Since the gradient of $F(x)$ vanishes at ω, taking the dot product of the gradient of $F(x)$ at ω with $x - \omega$ gives

$$\sum_{i=1}^{m} \frac{a_i^{\mathrm{T}}(x-\omega)}{a_i^{\mathrm{T}}\omega - b_i} + m\frac{c^{\mathrm{T}}(x-\omega)}{c^{\mathrm{T}}\omega - \beta} = 0.$$

The lemma then follows. □

The next lemma states that if x and ω are close in potential then the value of $a_i^{\mathrm{T}}x$ (as well as $c^{\mathrm{T}}x$) at x and at ω can differ only by a small amount.

Lemma 5. *If $f(x) \leqslant \frac{1}{2}\delta^2 - \frac{1}{3}\delta^3$, where $0 \leqslant \delta < 1$, then*

$$|a_i^{\mathrm{T}}x - a_i^{\mathrm{T}}\omega|/(a_i^{\mathrm{T}}\omega - b_i) \leqslant \delta, \quad 1 \leqslant i \leqslant m,$$

and

$$|c^{\mathrm{T}}\omega - c^{\mathrm{T}}x|/(c^{\mathrm{T}}\omega - \beta) \leqslant \delta/\sqrt{2m}.$$

Proof. Suppose that $f(x) \leqslant \frac{1}{2}\delta^2 - \frac{1}{3}\delta^3$, where $0 \leqslant \delta < 1$. Let $\psi_i(x) = (a_i^{\mathrm{T}}x - a_i^{\mathrm{T}}\omega)/(a_i^{\mathrm{T}}\omega - b_i)$ for $i = 1, 2, \ldots, m$, and let $\psi_i(x) = (c^{\mathrm{T}}x - c^{\mathrm{T}}\omega)/(c^{\mathrm{T}}\omega - \beta)$ for $i = m+1, \ldots, 2m$. We note that the coordinates $\psi_i(x)$ were introduced in [8, 9]. First, we shall show that

$$\sum_{i=1}^{2m} \psi_i(x)^2 \leqslant \delta^2. \tag{4.1}$$

From (4.1) the first inequality of the lemma is immediate. As $f(x)$ is strictly convex, the minimum value of $f(x)$ over the region $\{x\colon \sum_{i=1}^{2m} \psi_i(x)^2 \geqslant \delta^2,\ Ax \geqslant b,\ c^{\mathrm{T}}x \geqslant \beta\}$ occurs on the boundary of the ellipsoid $\Sigma(\delta) = \{x\colon \sum_{i=1}^{2m} \psi_i(x)^2 \leqslant \delta^2\}$. We shall lower bound the value of $f(x)$ on the boundary of $\Sigma(\delta)$. We have that

$$f(x) = -\sum_{i=1}^{2m} \ln(1 + \psi_i(x)).$$

Using the Taylor series expansion for $\ln(1 + \psi_i(x))$, within the ellipsoid $\Sigma(\delta)$, $f(x)$ may be expressed as

$$f(x) = \sum_{i=1}^{2m} \sum_{j=1}^{\infty} \frac{(-1)^j \psi_i(x)^j}{j}.$$

From Lemma 4, $\sum_{i=1}^{2m} \psi_i(x)=0$. Also, on the boundary of the region $\Sigma(\delta)$,

$$\sum_{i=1}^{2m} \sum_{j=4}^{\infty} \frac{(-1)^j \psi_i(x)^j}{j} > 0.$$

Thus on the boundary of $\Sigma(\delta)$,

$$f(x) > \sum_{i=1}^{2m} \tfrac{1}{2}\psi_i(x)^2 - \tfrac{1}{3}\psi_i(x)^3$$

$$> \sum_{i=1}^{2m} \psi_i(x)^2(\tfrac{1}{2}-\tfrac{1}{3}\delta) \quad (\text{since } |\psi_i(x)| \leqslant \delta)$$

$$> \delta^2(\tfrac{1}{2}-\tfrac{1}{3}\delta).$$

(4.1) above then follows. Then

$$\sum_{i=1}^{2m} |\psi_i(x)| \leqslant \delta\sqrt{2m},$$

and since $\sum_{i=1}^{2m} \psi_i(x)=0$ and $\psi_{m+1}(x)=\cdots=\psi_{2m}(x)$, we get that

$$\sum_{i=m+1}^{2m} |\psi_i(x)| = m\frac{|c^{\mathrm{T}}x - c^{\mathrm{T}}\omega|}{c^{\mathrm{T}}\omega - \beta} \leqslant \delta\sqrt{\tfrac{1}{2}m}. \qquad \square$$

We note that Lemma 1 in Section 3 follows from Lemmas 4 and 5 above.

Let $F'(x)$ be the potential defined as

$$F'(x) = \sum_{i=1}^{m} \ln(a_i^{\mathrm{T}}x - b_i) + m\ln(c^{\mathrm{T}}x - \beta'),$$

and let ω' be the point that maximizes $F'(x)$ over the region $\{x\colon Ax \geqslant b,\ c^{\mathrm{T}}x \geqslant \beta'\}$. Let $f'(x) = F(\omega') - F'(x)$ be the normalized potential corresponding to $F'(x)$. The next lemma bounds the change in objective function value at the center due to change in the parameter β.

Lemma 6. *Let $\beta' \geqslant \beta$. Then $c^{\mathrm{T}}\omega' \geqslant c^{\mathrm{T}}\omega$ and $c^{\mathrm{T}}\omega' - c^{\mathrm{T}}\omega \leqslant \beta' - \beta$.*

Proof. A proof is given in [8] but we shall give it here for completeness. Let $\chi(t) = \beta + t(\beta' - \beta)$, and let $\omega(t)$ be the point that maximizes the function

$$F_t(x) = \sum_{i=1}^{m} \ln(a_i^{\mathrm{T}}x - b_i) + m\ln(c^{\mathrm{T}}x - \chi(t)).$$

Since

$$c^{\mathrm{T}}\omega' - c^{\mathrm{T}}\omega = \int_0^1 c^{\mathrm{T}}\left(\frac{\mathrm{d}}{\mathrm{d}t}\,\omega(t)\right)\mathrm{d}t$$

it suffices to show that

$$0 \leqslant c^{\mathrm{T}}\left(\frac{\mathrm{d}}{\mathrm{d}t}\,\omega(t)\right) \leqslant \beta' - \beta.$$

As the gradient of $F_t(x)$ vanishes at $\omega(t)$,

$$\sum_{i=1}^{m} \frac{1}{a_i^T\omega(t)-b_i} a_i + \frac{m}{c^T\omega(t)-\chi(t)} c = 0.$$

Differentiating the above expression w.r.t. t, and rearranging gives

$$\left(-\sum_{i=1}^{m} \frac{1}{(a_i^T\omega(t)-b_i)^2} a_i a_i^T\right)\left(\frac{d}{dt}\omega(t)\right)$$
$$=\left(m\left(c^T\left(\frac{d}{dt}\omega(t)\right)-\frac{d}{dt}\chi(t)\right)\Big/(c^T\omega(t)-\chi(t))^2\right)c.$$

The dot product of the left hand side of this equation with $(d/dt)\omega(t)$ is negative since the matrix on the left side is negative-definite, and so from the right hand side of this equation we get

$$\left(c^T\left(\frac{d}{dt}\omega(t)\right)\right)^2 \leqslant \left(c^T\left(\frac{d}{dt}\omega(t)\right)\right)\left(\frac{d}{dt}\chi(t)\right).$$

Since $(d/dt)\chi(t)=\beta'-\beta \geqslant 0$, we may conclude that $0 \leqslant c^T((d/dt)\omega(t)) \leqslant \beta'-\beta$, and the lemma then follows. □

The next lemma bounds the change in potential at a point due to change in the parameter β.

Lemma 7. *Let x be a point in the interior of the polytope $P^\beta=\{x\colon Ax \geqslant b,\ c^Tx \geqslant \beta\}$ such that $f(x) \leqslant \frac{1}{2}\delta^2-\frac{1}{3}\delta^3$, where $0 \leqslant \delta < 1$. Let*

$$\beta'=\beta+(\alpha/\sqrt{m})(c^Tx-\beta),$$

where $\alpha \geqslant 0$. Then

$$f'(x) \leqslant f(x)+\frac{\alpha\delta}{\sqrt{2}(1-\alpha/\sqrt{m})}+\frac{\alpha^2(1+\delta/\sqrt{2m})^2}{(1-\alpha/\sqrt{m}-\alpha\delta/(\sqrt{2}\,m))}.$$

Proof. We may write $f'(x)$ as

$$f'(x)=f(x)+f'(\omega)+m\ln\left(\frac{(c^Tx-\beta)(c^T\omega-\beta')}{(c^Tx-\beta')(c^T\omega-\beta)}\right).$$

We have

$$\frac{(c^Tx-\beta)(c^T\omega-\beta')}{(c^Tx-\beta')(c^T\omega-\beta)} = 1+\frac{(\beta'-\beta)(c^T\omega-c^Tx)}{(c^Tx-\beta')(c^T\omega-\beta)}$$
$$\leqslant 1+\frac{\alpha\delta}{m}\frac{1}{\sqrt{2}(1-\alpha/\sqrt{m})} \quad \text{(by Lemma 5).}$$

Thus

$$m\ln\left(\frac{(c^Tx-\beta)(c^T\omega-\beta')}{(c^Tx-\beta')(c^T\omega-\beta)}\right) \leqslant \frac{\alpha\delta}{\sqrt{2}(1-\alpha/\sqrt{m})}.$$

Next,

$$\frac{c^{\mathrm{T}}\omega'-\beta'}{c^{\mathrm{T}}\omega-\beta'}=\frac{c^{\mathrm{T}}\omega'-\beta}{c^{\mathrm{T}}\omega-\beta}\left(1+\frac{(\beta'-\beta)(c^{\mathrm{T}}\omega'-c^{\mathrm{T}}\omega)}{(c^{\mathrm{T}}\omega'-\beta)(c^{\mathrm{T}}\omega-\beta')}\right)$$

$$\leqslant\frac{c^{\mathrm{T}}\omega'-\beta}{c^{\mathrm{T}}\omega-\beta}+\frac{(\beta'-\beta)^2}{(c^{\mathrm{T}}\omega-\beta)(c^{\mathrm{T}}\omega-\beta')}\quad\text{(by Lemma 6)}$$

$$\leqslant\frac{c^{\mathrm{T}}\omega'-\beta}{c^{\mathrm{T}}\omega-\beta}+\frac{\alpha^2(1+\delta/\sqrt{2m})^2}{m(1-\alpha/\sqrt{m}-\alpha\delta/(\sqrt{2}\,m))}$$

(by Lemma 5 and definition of β').

Then

$$f'(\omega)=\sum_{i=1}^{m}\ln\left(\frac{a_i^{\mathrm{T}}\omega'-b_i}{a_i^{\mathrm{T}}\omega-b_i}\right)+m\ln\left(\frac{c^{\mathrm{T}}\omega'-\beta'}{c^{\mathrm{T}}\omega-\beta'}\right)$$

$$\leqslant\sum_{i=1}^{m}\left(\frac{a_i^{\mathrm{T}}\omega'-b_i}{a_i^{\mathrm{T}}\omega-b_i}-1\right)+m\left(\frac{c^{\mathrm{T}}\omega'-\beta'}{c^{\mathrm{T}}\omega-\beta'}-1\right)$$

$$\leqslant\sum_{i=1}^{m}\frac{a_i^{\mathrm{T}}\omega'-b_i}{a_i^{\mathrm{T}}\omega-b_i}+m\frac{c^{\mathrm{T}}\omega'-\beta}{c^{\mathrm{T}}\omega-\beta}-2m$$

$$+\alpha^2(1+\delta/\sqrt{2m})^2(1-\alpha/\sqrt{m}-\alpha\delta/(\sqrt{2}\,m))^{-1}.$$

From Lemma 4,

$$\sum_{i=1}^{m}\frac{a_i^{\mathrm{T}}\omega'-b_i}{a_i^{\mathrm{T}}\omega-b_i}+m\frac{c^{\mathrm{T}}\omega'-\beta}{c^{\mathrm{T}}\omega-\beta}-2m=0.$$

Thus

$$f'(\omega)\leqslant\frac{\alpha^2(1+\delta/\sqrt{2m})^2}{(1-\alpha/\sqrt{m}-\alpha\delta/(\sqrt{2}\,m))}.\qquad\square$$

We note that Lemma 2 in Section 3 follows from Lemma 7 above.

5. Local optimizations

In this section we describe the local optimizations, and show that x^k, the point computed during the kth iteration, sufficiently reduces the potential $f^k(x)$. We shall first collect together a few definitions. Let $F^k(x)$ be the potential defined as

$$F^k(x)=\sum_{i=1}^{m}\ln(a_i^{\mathrm{T}}x-b_i)+m\ln(c^{\mathrm{T}}x-\beta^k)$$

and ω^k be the point that maximizes $F^k(x)$ over the region $\{x\colon Ax\geqslant b,\ c^{\mathrm{T}}x\geqslant\beta^k\}$. Let $f^k(x)=F^k(\omega^k)-F^k(x)$ be the normalized potential corresponding to $F^k(x)$.

Let x^{k-1} be the point at the beginning of the kth iteration. Let η^k be the gradient of $f^k(x)$ evaluated at x^{k-1}, and let D be a diagonal matrix such that $1/(1.1(a_i^{\mathrm{T}}x^{k-1}-b_i)^2) \leqslant D_{ii} \leqslant 1.1/(a_i^{\mathrm{T}}x^{k-1}-b_i)^2$. x^k is obtained from x^{k-1} as follows.

(1) Determine a direction ξ^k by solving the system of linear equations

$$\left(A^{\mathrm{T}}DA+\frac{m}{(c^{\mathrm{T}}x^{k-1}-\beta^k)^2}cc^{\mathrm{T}}\right)\xi^k=-\eta^k.$$

(2) Next, compute a scalar $t^k>0$ satisfying the condition

$$0.018 \leqslant (t^k)^2(\xi^k)^{\mathrm{T}}\left(A^{\mathrm{T}}DA+\frac{m}{(c^{\mathrm{T}}x^{k-1}-\beta^k)^2}cc^{\mathrm{T}}\right)\xi^k \leqslant 0.0196.$$

(3) If $f^k(x^{k-1}+t^k\xi^k))<f^k(x^{k-1})$ then $x^k:=x^{k-1}+t^k\xi^k$ else $x^k:=x^{k-1}$.

It is adequate to show that if $f^k(x^{k-1})$ exceeds 0.04 then $x^{k-1}+t^k\xi^k$ sufficiently reduces the potential $f^k(x)$.

In this section we shall prove an alternate lemma for Lemma 3 which was introduced in Section 3, i.e. Lemma 8 below, and Lemma 3 will follow as a consequence of Lemma 8. We shall require some additional notation. Let $E^k(r)$ be the ellipsoid

$$E^k(r)=\left\{x\colon (x-x^{k-1})^{\mathrm{T}}\left(A^{\mathrm{T}}DA+\frac{m}{(c^{\mathrm{T}}x^{k-1}-\beta^k)^2}cc^{\mathrm{T}}\right)(x-x^{k-1})\leqslant r^2f^k(x^{k-1})\right\}$$

where r lies between 0 and 1. Let $z^k(r)$ be the point that minimizes the linear function $(\eta^k)^{\mathrm{T}}x$ over the ellipsoid $E^k(r)$. Let H^k denote the Hessian of $f^k(x)$ evaluated at x^{k-1}. Note that H^k may be written as

$$H^k=\sum_{i=1}^{m}\frac{1}{(a_i^{\mathrm{T}}x^{k-1}-b_i)^2}a_ia_i^{\mathrm{T}}+\frac{m}{(c^{\mathrm{T}}x^{k-1}-\beta^k)^2}cc^{\mathrm{T}},$$

and η^k may be written as

$$\eta^k=-\left(\sum_{i=1}^{m}\frac{1}{a_i^{\mathrm{T}}x^{k-1}-b_i}a_i+\frac{m}{c^{\mathrm{T}}x^{k-1}-\beta^k}c\right).$$

Lemma 8. *Suppose that* $0<f^k(x^{k-1})\leqslant\frac{1}{2}\delta^2-\frac{1}{3}\delta^3$, *where* $0\leqslant\delta<1$. *Then*

$$f^k(z^k(r))\leqslant(1-\mu r+0.55r^2+\nu r^3)f^k(x^{k-1}),$$

where

$$\mu=\sqrt{(1-\delta)}\Bigg/\left(1.1\sum_{j=0}^{\infty}\frac{\delta^j}{(j+1)(j+2)}\right)^{1/2}$$

and

$$\nu=\tfrac{1}{3}(1.1)^{3/2}\sqrt{f^k(x^{k-1})}/(1-\sqrt{1.1r^2f^k(x^{k-1})}).$$

Before proving Lemma 8, we shall show how Lemma 3 follows from Lemma 8. As $z^k(r)$ minimizes $(\eta^k)^{\mathrm{T}}x$ over the ellipsoid $E^k(r)$, from the theory of convex optimization [13], it follows that $z^k(r)-x^{k-1}$ satisfies the system of linear equations (given by the Karush–Kuhn–Tucker conditions),

$$\left(A^{\mathrm{T}}DA+\frac{m}{(c^{\mathrm{T}}x^{k-1}-\beta^k)^2}cc^{\mathrm{T}}\right)(z^k(r)-x^{k-1})=-t(r)\eta^k,$$

for some scalar $t(r)>0$. So ξ^k, the direction computed during the kth iteration, and $z^k(r)-x^{k-1}$ are in the same direction. Furthermore, when $f^k(x^{k-1})$ is in the range $[0.04, 0.05]$, $x^{k-1}+t^k\xi^k$ equals $z^k(r_0)$, for some r_0 in the range $[0.6, 0.7]$. So from Lemma 8 we may conclude that if $0.04\leqslant f^k(x^{k-1})\leqslant 0.05$ then $f^k(x^{k-1}+t^k\xi^k)\leqslant 0.75f^k(x^{k-1})\leqslant 0.04$.

We shall now give a proof of Lemma 8.

Proof of Lemma 8. Let x be a point in the ellipsoid $E^k(r)$. Write x as $x=x^{k-1}+t\xi$ where t is a scalar. Using the power series expansion at x^{k-1}, $f^k(x^{k-1}+t\xi)$ may be written as

$$\begin{aligned}f^k(x^{k-1}+t\xi)&=f^k(x^{k-1})+t(\eta^k)^{\mathrm{T}}\xi+\tfrac{1}{2}t^2\xi^{\mathrm{T}}H^k\xi\\&\quad+\sum_{j=3}^{\infty}\frac{(-1)^jt^j}{j}\left(\sum_{i=1}^{m}\frac{(a_i^{\mathrm{T}}\xi)^j}{(a_i^{\mathrm{T}}x^{k-1}-b_i)^j}+\frac{m(c^{\mathrm{T}}\xi)^j}{(c^{\mathrm{T}}x^{k-1}-\beta^k)^j}\right).\end{aligned}$$

Since $1/(1.1(a_i^{\mathrm{T}}x^{k-1}-b_i)^2)\leqslant D_{ii}\leqslant 1.1/(a_i^{\mathrm{T}}x^{k-1}-b_i)^2$,

$$\begin{aligned}\tfrac{1}{2}t^2\xi^{\mathrm{T}}H^k\xi&=\tfrac{1}{2}t^2\left(\sum_{i=1}^{m}\frac{(a_i^{\mathrm{T}}\xi)^2}{(a_i^{\mathrm{T}}x^{k-1}-b_i)^2}+\frac{m(c^{\mathrm{T}}\xi)^2}{(c^{\mathrm{T}}x^{k-1}-\beta^k)^2}\right)\\&\leqslant\tfrac{1}{2}(1.1)t^2\xi^{\mathrm{T}}\left(A^{\mathrm{T}}DA+\frac{m}{(c^{\mathrm{T}}x^{k-1}-\beta^k)^2}cc^{\mathrm{T}}\right)\xi\\&\leqslant 0.55r^2f^k(x^{k-1}).\end{aligned}\tag{5.1}$$

Next, note that for $j\geqslant 3$,

$$\sum_i\theta_i^2\leqslant\sigma^2\Rightarrow\sum_i|\theta_i|^j\leqslant|\sigma|^j.$$

So from (5.1) we can conclude that for $j\geqslant 3$,

$$\left|t^j\left(\sum_{i=1}^{m}\frac{(a_i^{\mathrm{T}}\xi)^j}{(a_i^{\mathrm{T}}x^{k-1}-b_i)^j}+\frac{m(c^{\mathrm{T}}\xi)^j}{(c^{\mathrm{T}}x^{k-1}-\beta^k)^j}\right)\right|\leqslant(1.1r^2f^k(x^{k-1}))^{j/2}.$$

Thus

$$\begin{aligned}&\left|\sum_{j=3}^{\infty}\frac{(-1)^jt^j}{j}\left(\sum_{i=1}^{m}\frac{(a_i^{\mathrm{T}}\xi)^j}{(a_i^{\mathrm{T}}x^{k-1}-b_i)^j}+\frac{m(c^{\mathrm{T}}\xi)^j}{(c^{\mathrm{T}}x^{k-1}-\beta^k)^j}\right)\right|\\&\leqslant\sum_{j=3}^{\infty}\frac{1}{j}(1.1r^2f^k(x^{k-1}))^{j/2}\\&\leqslant\tfrac{1}{3}(1.1)^{3/2}\frac{r^3\sqrt{f^k(x^{k-1})}}{1-\sqrt{1.1r^2f^k(x^{k-1})}}f^k(x^{k-1}).\end{aligned}\tag{5.2}$$

From (5.1) it follows that within the ellipsoid $E^k(r)$ the magnitude of the second order term in the above series is upper bounded by $0.55r^2f^k(x^{k-1})$. (5.2) gives an upper bound on the sum of the magnitudes of the third and higher order terms in the above power series for a point in $E^k(r)$. Next, we shall lower bound the maximum change in the linear (first order) term that is possible within the ellipsoid $E^k(r)$. Such a bound is provided by Lemma 9. Let x' be the point where the line joining x^{k-1} and ω^k intersects the boundary of the ellipsoid $E^k(r)$. By Lemma 9,

$$(\eta^k)^T(x'-x^{k-1}) \leq -\mu r f^k(x^{k-1}).$$

Since $(\eta^k)^T z^k(r) \leq (\eta^k)^T x'$, the lemma follows from Lemma 9 and the upper bounds given by (5.1) and (5.2) on sum of the second and higher order terms in the above power series. □

Lemma 9. *Let x' be the point where the line joining x^{k-1} and ω^k intersects the boundary of the ellipsoid $E^k(r)$. If $0 < f^k(x^{k-1}) \leq \frac{1}{2}\delta^2 - \frac{1}{3}\delta^3$, where $0 \leq \delta < 1$, then*

$$(\eta^k)^T(x'-x^{k-1}) \leq -\mu r f^k(x^{k-1})$$

where

$$\mu = \sqrt{(1-\delta)} \Big/ \left(1.1 \sum_{j=0}^{\infty} \frac{\delta^j}{(j+1)(j+2)}\right)^{1/2}.$$

Proof. Let $x^{k-1} - x' = \lambda u$, where u is the unit vector in the direction of $x^{k-1} - x'$, and $\lambda = \|x^{k-1} - x'\|_2$. We have

$$\lambda^2 u^T \left(A^T D A + \frac{m}{(c^T x^{k-1} - \beta^k)^2} c c^T \right) u = r^2 f^k(x^{k-1}).$$

Since $1/(1.1(a_i^T x^{k-1} - b_i)^2) \leq D_{ii} \leq 1.1/(a_i^T x^{k-1} - b_i)^2$,

$$\lambda^2 u^T H^k u \geq r^2 f^k(x^{k-1})/1.1.$$

Thus

$$\lambda \geq r(f^k(x^{k-1})/(1.1 u^T H^k u))^{1/2}.$$

Hence

$$(\eta^k)^T(x^{k-1} - x') \geq \frac{r\sqrt{f^k(x^{k-1})}}{\sqrt{1.1}} \frac{(\eta^k)^T u}{\sqrt{u^T H^k u}}$$

$$\geq \frac{r f^k(x^{k-1})}{\sqrt{1.1}} \frac{(\eta^k)^T(x^{k-1} - \omega^k)}{\sqrt{f^k(x^{k-1})}\sqrt{(x^{k-1}-\omega^k)^T H^k (x^{k-1} - \omega^k)}}. \qquad (5.3)$$

Let $y_i = (a_i^{\mathrm{T}} x^{k-1} - a_i^{\mathrm{T}} \omega^k)/(a_i^{\mathrm{T}} \omega^k - b_i)$ for $i = 1, \ldots, m$, and let $y_i = (c^{\mathrm{T}} x^{k-1} - c^{\mathrm{T}} \omega^k)/(c^{\mathrm{T}} \omega^k - \beta^k)$ for $i = m+1, \ldots, 2m$. Then

$$(\eta^k)^{\mathrm{T}}(x^{k-1} - \omega^k) = \sum_{i=1}^{2m} \left(\frac{1}{1+y_i} - 1\right),$$

$$(x^{k-1} - \omega^k)^{\mathrm{T}} H^k (x^{k-1} - \omega^k) = \sum_{i=1}^{2m} \left(\frac{1}{1+y_i} - 1\right)^2$$

and

$$f^k(x^{k-1}) = \sum_{i=1}^{2m} \ln\left(\frac{1}{1+y_i}\right).$$

From Lemma 4 in Section 4,

$$\sum_{i=1}^{2m} y_i = 0.$$

Since $f^k(x^{k-1}) \leqslant \frac{1}{2}\delta^2 - \frac{1}{3}\delta^3$, where $0 \leqslant \delta < 1$, from Lemma 5 in Section 4 we get that

$$|y_i| \leqslant \delta, \quad i = 1, 2, \ldots, 2m.$$

We can thus apply Lemma 10 below, and from (5.3) above conclude that

$$(\eta^k)^{\mathrm{T}}(x^{k-1} - x') \geqslant r f^k(x^{k-1}) \sqrt{(1-\delta)} \Big/ \left(1.1 \sum_{j=0}^{\infty} \frac{\delta^j}{(j+1)(j+2)}\right)^{1/2}. \qquad \square$$

Lemma 10. *Suppose that* $\sum_{i=1}^{2m} y_i = 0$, $\sum_{i=1}^{2m} \ln(1+y_i) < 0$ *and* $|y_i| \leqslant \delta < 1$ *for* $i = 1, 2, \ldots, 2m$. *Then*

$$\sum_{i=1}^{2m} \left(\frac{1}{1+y_i} - 1\right) \geqslant (1-\delta) \sum_{i=1}^{2m} \left(\frac{1}{1+y_i} - 1\right)^2$$

and

$$\sum_{i=1}^{2m} \left(\frac{1}{1+y_i} - 1\right) \geqslant \sum_{i=1}^{2m} \ln\left(\frac{1}{1+y_i}\right) \Big/ \sum_{j=0}^{\infty} \frac{\delta^j}{(j+1)(j+2)}.$$

Proof. Since $\sum_{i=1}^{2m} y_i = 0$,

$$\sum_{i=1}^{2m} \left(\frac{1}{1+y_i} - 1\right) = \sum_{i=1}^{2m} \left(\frac{1}{1+y_i} + y_i - 1\right) = \sum_{i=1}^{2m} \left(\frac{y_i^2}{1+y_i}\right).$$

So as $|y_i| \leqslant \delta$,

$$\sum_{i=1}^{2m} \left(\frac{1}{1+y_i} - 1\right) = \sum_{i=1}^{2m} \left(\frac{y_i^2}{1+y_i}\right) \geqslant (1-\delta) \sum_{i=1}^{2m} \frac{y_i^2}{(1+y_i)^2}$$

$$\geqslant (1-\delta) \sum_{i=1}^{2m} \left(\frac{1}{1+y_i} - 1\right)^2.$$

Next, as $\sum_{i=1}^{2m} y_i = 0$,

$$\sum_{i=1}^{2m} \ln\left(\frac{1}{1+y_i}\right) = \sum_{i=1}^{2m} (y_i - \ln(1+y_i))$$

$$= \sum_{y_i \neq 0, 1 \leqslant i \leqslant m} \frac{y_i^2}{1+y_i} \frac{(1+y_i)(y_i - \ln(1+y_i))}{y_i^2}.$$

Using the Taylor series expansion for $\ln(1+y_i)$, for $y_i \neq 0$ we get

$$\frac{(1+y_i)(y_i-\ln(1+y_i))}{y_i^2} = \tfrac{1}{2} + \sum_{j=1}^{\infty} \frac{(-1)^{j+1} y_i^j}{(j+1)(j+2)}$$

$$\leqslant \sum_{j=0}^{\infty} \frac{\delta^j}{(j+1)(j+2)} \quad \text{as } |y_i| \leqslant \delta.$$

Thus

$$\sum_{i=1}^{2m} \ln\left(\frac{1}{1+y_i}\right) \leqslant \left(\sum_{j=0}^{\infty} \frac{\delta^j}{(j+1)(j+2)}\right) \sum_{i=1}^{2m} \left(\frac{y_i^2}{1+y_i}\right).$$

The lemma then follows. □

6. Amortizing the number of arithmetic operations

In this section we show that the total number of arithmetic operations performed by the algorithm is $O((mn^2 + m^{1.5}n)L)$. The amortization of the number of arithmetic operations is similar to the one in [5]. The total number of arithmetic operations is determined by the number of operations required for the following computations.

(1) Solving systems of linear equations to determine the directions ξ^k.

(2) Computing the gradients η^k of the potentials $f^k(x)$ and computing the scalars t^k.

In the kth iteration we determine a direction ξ^k by solving the system of linear of equations

$$\left(A^{\mathrm{T}}DA + \frac{m}{(c^{\mathrm{T}}x^{k-1} - \beta^k)^2} cc^{\mathrm{T}}\right)\xi^k = -\eta^k$$

where η^k is the gradient of $f^k(x)$ evaluated at x^{k-1}, and we find a scalar t^k such that

$$0.018 \leqslant (t^k)^2(\xi^k)^{\mathrm{T}}\left(A^{\mathrm{T}}DA + \frac{m}{(c^{\mathrm{T}}x^{k-1} - \beta^k)^2} cc^{\mathrm{T}}\right)\xi^k \leqslant 0.0196.$$

The gradient η^k can be computed in $O(mn)$ operations, and once we have ξ^k, a required t^k may also be obtained in $O(mn)$ operations. The total number of iterations is $O(\sqrt{m}\,L)$. So the total number of operations required to compute η^k and scalar t^k over all the iterations is $O(m^{1.5}nL)$. We maintain $(A^{\mathrm{T}}DA)^{-1}$ and update it whenever the matrix D changes. Once $(A^{\mathrm{T}}DA)^{-1}$ is available, $(A^{\mathrm{T}}DA + (m/(c^{\mathrm{T}}x^{k-1} - \beta^k)^2)cc^{\mathrm{T}})^{-1}$ may be computed in $O(n^2)$ operations as the two matrices differ by a rank one matrix, and then ξ^k may be obtained in $O(n^2)$ extra operations. We shall show that the total number of operations required to maintain $(A^TDA)^{-1}$ during the entire execution of the algorithm is $O(mn^2L)$, and then the desired bound on the total number of arithmetic operations performed by the algorithm follows.

At the end of the kth iteration D is updated as follows. For $i=1,\ldots,m$, the ith diagonal element D_{ii} is reset to $1/(a_i^Tx^k-b_i)^2$ if $D_{ii}\notin[1/(1.1(a_i^Tx^k-b_i)^2), 1.1/(a_i^Tx^k-b_i)^2]$. Suppose that D' is the matrix obtained by changing the ith diagonal element of D to d'. Then

$$A^TD'A=A^TDA+(d'-D_{ii})a_ia_i^T.$$

Thus whenever an element of D is changed, A^TDA changes by a rank one matrix, and hence $(A^TDA)^{-1}$ changes by a rank one matrix. Therefore when an element of D is changed, $(A^TDA)^{-1}$ may be updated in $O(n^2)$ operations, using the rank one update formula

$$(B+uv^T)^{-1}=B^{-1}-B^{-1}uv^TB^{-1}/(1+v^TB^{-1}u).$$

So to obtain a bound of $O(mn^2L)$ on the total number of operations required to maintain $(A^TDA)^{-1}$, it is sufficient to show that the total number of changes to the matrix D during the entire execution of the algorithm is $O(mL)$.

Let

$$\phi_i^k=|\ln((a_i^Tx^k-b_i)/(a_i^Tx^{k-1}-b_i))|.$$

Suppose D_{ii} was reset at the lth iteration and at the jth iteration but was not reset between the lth and jth iterations. Then $\sum_{k=l+1}^{j}\phi_i^k\geq\ln(1.1)$, and the total number of times D_{ii} is changed during the execution of the algorithm is $O(\sum_{k=1}^{I}\phi_i^k)$, where I is the number of iterations performed by the algorithm. Thus, the total number of changes to D is $O(\sum_{i=1}^{m}\sum_{k=1}^{I}\phi_i^k)$.

A bound on $\sum_{i=1}^{m}\sum_{k=1}^{I}\phi_i^k$ may be obtained as follows. x^k lies within an ellipsoid around x^{k-1}, and x^k-x^{k-1} satisfies the condition

$$(x^k-x^{k-1})^T\left(A^TDA+\frac{m}{(c^Tx^{k-1}-\beta^k)^2}cc^T\right)(x^k-x^{k-1})\leq 0.0196.$$

Since at the start of the kth iteration, for $i=1,\ldots,m$, $D_{ii}\in[1/(1.1(a_i^Tx^{k-1}-b_i)^2), 1.1/(a_i^Tx^{k-1}-b_i)^2]$, it follows that

$$\sum_{i=1}^{m}\left(\frac{a_i^Tx^k-b_i}{a_i^Tx^{k-1}-b_i}-1\right)^2\leq 1.1\times 0.0196,$$

and thus

$$\sum_{i=1}^{m}\left|\left(\frac{a_i^Tx^k-b_i}{a_i^Tx^{k-1}-b_i}-1\right)\right|\leq 1.1\times 0.0196\times\sqrt{m}.$$

Then using the Taylor series expansion for the natural logarithm, it is easily shown that $\sum_{i=1}^{m}\phi_i^k=O(\sqrt{m})$, and since I, the number of iterations, is $O(\sqrt{m}\,L)$ we may conclude that $\sum_{i=1}^{m}\sum_{k=1}^{I}\phi_i^k=O(mL)$.

7. Recasting a linear program into the required format

In this section we show how to transform the given linear program so that a suitable starting point for the transformed program is available. There are several ways to carry out such a transformation. The one we give is similar to the one in [8]. However, there is one critical difference. The transformation given in [8] can increase the parameter L by a factor of n in the worst case, whereas the transformation given in this section can increase L by at most a constant factor. The given linear program is

$$\begin{aligned} \max \quad & p^{\mathrm{T}} z \\ \text{s.t.} \quad & Hz \geqslant q \end{aligned}$$

where $z \in \mathbb{R}^{n_1}$, $p \in \mathbb{R}^{n_1}$, $q \in \mathbb{R}^{m_1}$, and $H \in \mathbb{R}^{m_1 \times n_1}$. We reserve A, x, b and c to refer to a linear program that is already in the required format. Let

$$L_1 = \log_2(\text{largest absolute value of the determinant of any square submatrix of } H) + \log_2\left(\max_i p_i\right) + \log_2\left(\max_i q_i\right) + \log_2(m_1 + n_1).$$

Note that:

(i) If the given linear program has an optimal solution then every optimal vertex z^{opt} satisfies the condition $\|z^{\mathrm{opt}}\|_\infty \leqslant 2^{L_1}$.

(ii) If the polytope $\{z\colon Hz \geqslant q\}$ is unbounded then there is a feasible solution z^{f} such that $\|z^{\mathrm{f}}\|_\infty \leqslant 2^{L_1}$.

Let $t \in \mathbb{R}$, let $e \in \mathbb{R}^{m_1}$ be a vector given by $e^{\mathrm{T}} = [1, 1, \ldots, 1]$, let $\lambda = m_1 n_1 2^{2L_1}$, and let $\mu = 2^{30L_1}$. Let h_i^{T} denote the ith row of H. The transformed linear program is as follows.

$$\begin{aligned} \max \quad & p^{\mathrm{T}} z + \mu t \\ \text{s.t.} \quad & h_i^{\mathrm{T}} z - (\lambda + q_i) t \geqslant -\lambda, \quad i = 1, 2, \ldots, m_1, \\ & -e^{\mathrm{T}} H z \geqslant -\lambda, \\ & z_j \geqslant -\lambda, \quad j = 1, 2, \ldots, n_1, \\ & -z_j \geqslant -\lambda, \quad j = 1, 2, \ldots, n_1, \\ & -\lambda t \geqslant -\lambda, \\ & ((m+1)\lambda + e^{\mathrm{T}} q) t \geqslant -\lambda. \end{aligned}$$

Let $m = m_1 + 2n_1 + 3$, and let $n = n_1 + 1$. Let $A \in \mathbb{R}^{m \times n}$ denote the constraint matrix, $b \in \mathbb{R}^m$ denote the right hand side of the constraints, $c \in \mathbb{R}^n$ denote the objective function vector, and $x \in \mathbb{R}^n$ denote the variables in the transformed linear program.

Note that $c^{\mathrm{T}}=[p^{\mathrm{T}}, \mu]$ and $x^{\mathrm{T}}=[z^{\mathrm{T}}, t]$. The transformed linear program may be written as

$$\max \quad c^{\mathrm{T}}x$$
$$\text{s.t.} \quad Ax \geqslant b.$$

The transformed problem has the following properties.

(a) Since $((m+1)\lambda+e^{\mathrm{T}}q)>0$, t is bounded, and thus the polytope $\{x\colon Ax \geqslant b\}$ is bounded.

(b) Let

$$L=\log_2(\text{largest absolute value of the determinant of any square submatrix of } A)$$
$$+\log_2\left(\max_i c_i\right)+\log_2\left(\max_i b_i\right)+\log_2(m+n).$$

Then $L \leqslant 40L_1$. The bound on L follows from the observations that the largest absolute value of the determinant of any square submatrix of A is at most $(m_1+n_1)^6 2^{6L_1}$, and that $\max_i b_i \leqslant m_1 n_1 2^{2L_1}$ and $\max_i c_i \leqslant 2^{30L_1}$.

(c) 0 is feasible. Since the sum of the rows of A is the zero vector and all the coordinates of b have the same value, the gradient of the function $\sum_{i=1}^m \ln(a_i^{\mathrm{T}}x-b_i)$ vanishes at 0. Furthermore, as the polytope $\{x\colon Ax \geqslant b\}$ is bounded, 0 is the unique point that maximizes the function $\sum_{i=1}^m \ln(a_i^{\mathrm{T}}x-b_i)$.

(d) A point

$$\begin{bmatrix} z \\ 1 \end{bmatrix}$$

is feasible for the transformed linear program iff $Hz \geqslant q$, $e^{\mathrm{T}}Hz \leqslant \lambda$ and $-\lambda \leqslant z_j \leqslant \lambda$, $j=1,2,\ldots,n_1$.

(e) μ is large enough so that if there exists a feasible point with $t=1$ then every optimal solution has $t=1$. This is because the minimum vertex to vertex variation of the function μt exceeds the maximum change in the function $p^{\mathrm{T}}z$ over the entire polytope $\{x\colon Ax \geqslant b\}$.

We shall now show that 0 is an adequate starting point for running the algorithm in Section 3 on the transformed problem. Let $\beta^0=-m^3 2^{2L}$. Let

$$F^0(x)=\sum_{i=1}^m \ln(a_i^{\mathrm{T}}x-b_i)+m\ln(c^{\mathrm{T}}x-\beta^0),$$

and ω^0 be the point that maximizes $F^0(x)$. Since the magnitude of β^0 is large enough,

$$|m\ln((c^{\mathrm{T}}\omega^0-\beta^0)/(c^{\mathrm{T}}x-\beta^0))| \leqslant 1/m$$

for any point x in the transformed polytope. Thus, $F^0(\omega^0)-F^0(0) \leqslant 0.04$ as required, for $m \geqslant 4$.

Finally, we have the following easily shown lemma.

Lemma 11. *Let*

$$\begin{bmatrix} z^{\text{opt}} \\ t^{\text{opt}} \end{bmatrix}$$

be an optimal vertex in the transformed linear program.

(i) *If* $t^{\text{opt}} < 1$ *then the original linear program is infeasible.*

(ii) *If* $t^{\text{opt}} = 1$, $e^{\mathrm{T}} H z^{\text{opt}} < \lambda$ *and* $|z_j^{\text{opt}}| < \lambda$, $j = 1, 2, \ldots, n_1$, *then* z^{opt} *is an optimal vertex in the original linear program.*

(iii) *If* $t^{\text{opt}} = 1$, *and either* $|z_j^{\text{opt}}| = \lambda$, *for some* j, *or* $e^{\mathrm{T}} H z^{\text{opt}} = \lambda$, *then either the original problem is unbounded or* z *is an optimal solution. In this case the transformed problem is solved again with* λ *replaced by* 2λ *to obtain a new optimal point*

$$\begin{bmatrix} z^* \\ t^* \end{bmatrix}.$$

If $p^{\mathrm{T}} z^* > p^{\mathrm{T}} z^{\text{opt}}$ *then the original problem is unbounded, otherwise* z^{opt} *is an optimal solution.* □

8. Precision of arithmetic operations

During each iteration of the algorithm, a direction is obtained by solving a system of linear equations whose matrix is symmetric positive-definite. The error in the solution of such a system of linear equations is directly related to the extremal eigenvalues and the condition number of the matrix describing the system, and the precision used for arithmetic operations [10, 12]. In the first part of this section we shall bound the entries in the matrix D, and also the extremal eigenvalues and the condition number of the matrices arising during local optimization at each iteration. The bounds obtained will be valid during the entire execution of the algorithm. Using these bounds we shall argue that it is adequate to maintain $(A^{\mathrm{T}} D A)^{-1}$ to an accuracy of γL bits for some constant γ. Then in the second part of the section we shall describe how sufficient accuracy in $(A^{\mathrm{T}} D A)^{-1}$ may be maintained during rank one changes using $\mathrm{O}(L)$ bits of precision.

8.1. Extremal eigenvalues and condition number of local optimization matrices

As before let x^{k-1} be the point at the beginning of the kth iteration. During the kth iteration we determine a direction ξ^k by solving the system of linear equations

$$\left(A^{\mathrm{T}} D A + \frac{m}{(c^{\mathrm{T}} x^{k-1} - \beta^k)^2} c c^{\mathrm{T}} \right) \xi^k = -\eta^k$$

where D is a diagonal matrix such that $D_{ii} \in [1/(1.1(a_i^{\mathrm{T}} x^{k-1} - b_i)), 1.1/(a_i^{\mathrm{T}} x^{k-1} - b_i)]$, and

$$\eta^k = -\left(\sum_{i=1}^{m} \frac{1}{a_i^{\mathrm{T}} x^{k-1} - b_i} a_i + \frac{m}{c^{\mathrm{T}} x^{k-1} - \beta^k} c \right)$$

is the gradient of $f^k(x)$ evaluated at x^{k-1}. We maintain $(A^TDA)^{-1}$ by performing rank one changes, and compute $(A^TDA+(m/(c^Tx^{k-1}-\beta^k)^2)cc^T)^{-1}$ by a rank one change to $(A^TDA)^{-1}$.

We shall first bound the entries in D. Note that the absolute value of $a_i^Tx-b_i$ and c^Tx for all feasible x is upper bounded by 2^{4L}. Thus

$$a_i^Tx^{k-1}-b_i\leq 2^{4L}.$$

Moreover, we may assume that $c^Tx^{k-1}-\beta^{k-1}>2^{-13L}$ (since the algorithm halts when $c^Tx^k-\beta^k$ is less than 2^{-13L}). Next, we shall lower bound the value of $a_i^Tx^{k-1}-b_i$, for $1\leq i\leq m$. Let $\overline{\beta^{k-1}}$ be obtained by rounding β^{k-1} to $15L$ bits. A vertex of the polytope $\{x: Ax\geq b, c^Tx\geq\overline{\beta^{k-1}}\}$ has rational coordinates with a common denominator which is at most 2^{16L}, and so the maximum change in the value of $a_i^Tx-b_i$ over this polytope is at least 2^{-16L}. Thus, the maximum value of $a_i^Tx-b_i$ over the polytope $\{x: Ax\geq b, c^Tx\geq\beta^{k-1}\}$ is at least 2^{-16L}. Then from Lemma 4 in Section 3,

$$a_i^T\omega^{k-1}-b_i\geq 2^{-16L}/(2m),$$

where ω^{k-1} is the point that minimizes $f^{k-1}(x)$. Furthermore, as $f^{k-1}(x^{k-1})\leq 0.04$, from Lemma 5 in Section 3 we get that

$$a_i^Tx^{k-1}-b_i\geq 2^{-16L}/(4m),$$

for $m\geq 16$. Thus from the bounds on $a_i^Tx^{k-1}-b_i$ and the manner in which D is updated, we may conclude that the entries in D are upper and lower bounded by 2^{36L} and 2^{-8L} respectively.

Before obtaining bounds on condition numbers and eigenvalues, we shall define some standard notation [10, 12] that will be used throughout this section. For a matrix B, let $\lambda_{\max}(B)$ and $\lambda_{\min}(B)$ denote the largest and smallest eigenvalues of B. The 2-norm of B, denoted $\|B\|_2$, is defined as

$$\|B\|_2=\sqrt{\lambda_{\max}(B^TB)}.$$

The frobenius norm of B, denoted $\|B\|_F$, is defined as

$$\|B\|_F=\left(\sum_{i=1}^{p}\sum_{j=1}^{q}B_{ij}^2\right)^{1/2},$$

where p and q denote the number of rows and columns of B respectively, and B_{ij} denotes the ijth element of B. Note that for a $p\times p$ matrix B, $\|B\|_2\leq\|B\|_F$, and $\|B\|_F\leq\sqrt{p}\|B\|_2$. For a list of the properties of these norms the reader may refer to [10, 12]. The condition number of B, denoted $\kappa(B)$, is defined as

$$\kappa(B)=\|B\|_2\|B^{-1}\|_2.$$

Note that if B is symmetric positive definite then

$$\kappa(B)=\lambda_{\max}(B)/\lambda_{\min}(B), \tag{8.1.1}$$

$$\|B\|_2=\lambda_{\max}(B) \tag{8.1.2}$$

and

$$\|B^{-1}\|_2 = 1/\lambda_{\min}(B). \tag{8.1.3}$$

We shall now obtain bounds on the extremal eigenvalues and the condition number of the matrices involved in the computation at each iteration in the algorithm. Note that D, $A^{\mathrm{T}}DA$ and $A^{\mathrm{T}}DA + (m/(c^{\mathrm{T}}x^{k-1} - \beta^k)^2)cc^{\mathrm{T}}$ are symmetric positive definite matrices. From the bounds on the entries in D it follows that

$$2^{-8L} \leqslant \lambda_{\min}(D) \leqslant \lambda_{\max}(D) \leqslant 2^{36L}. \tag{8.1.4}$$

An elementary argument using rayleigh quotients [10, 11] shows that

$$\lambda_{\min}(A^{\mathrm{T}}DA) \geqslant \lambda_{\min}(A^{\mathrm{T}}A)\lambda_{\min}(D) \tag{8.1.5}$$

and

$$\lambda_{\max}(A^{\mathrm{T}}DA) \leqslant \lambda_{\max}(A^{\mathrm{T}}A)\lambda_{\max}(D). \tag{8.1.6}$$

Next, we bound the eigenvalues of $A^{\mathrm{T}}A$. We have

$$\|Ax\|_2 \leqslant \|A\|_{\mathrm{F}}\|x\|_2 \leqslant 2^L\|x\|_2.$$

Also,

$$\|x\|_2 \leqslant \|A_1^{-1}\|_{\mathrm{F}}\|A_1x\|_2$$

where A_1 is an $n \times n$ submatrix of A of rank n. By definition of L each entry in A_1^{-1} is at most 2^L, and hence

$$\|x\|_2 \leqslant 2^{2L}\|Ax\|_2.$$

Thus

$$\lambda_{\max}(A^{\mathrm{T}}A) = \max_x \frac{x^{\mathrm{T}}A^{\mathrm{T}}Ax}{x^{\mathrm{T}}x} \leqslant 2^{2L} \tag{8.1.7}$$

and

$$\lambda_{\min}(A^{\mathrm{T}}A) = \min_x \frac{x^{\mathrm{T}}A^{\mathrm{T}}Ax}{x^{\mathrm{T}}x} \geqslant 2^{-4L}. \tag{8.1.8}$$

From (8.1.4)-(8.1.8) we may conclude that

$$2^{-12L} \leqslant \lambda_{\min}(A^{\mathrm{T}}DA) \leqslant \lambda_{\max}(A^{\mathrm{T}}DA) \leqslant 2^{38L}. \tag{8.1.9}$$

Furthermore,

$$\lambda_{\min}\left(A^{\mathrm{T}}DA + \frac{m}{(c^{\mathrm{T}}x^{k-1} - \beta^k)^2}cc^{\mathrm{T}}\right) \geqslant \lambda_{\min}(A^{\mathrm{T}}DA) \geqslant 2^{-12L} \tag{8.1.10}$$

and

$$\lambda_{\max}\left(A^{\mathrm{T}}DA + \frac{m}{(c^{\mathrm{T}}x^{k-1} - \beta^k)^2}cc^{\mathrm{T}}\right)$$

$$\leqslant \lambda_{\max}(A^{\mathrm{T}}DA) + \lambda_{\max}\left(\frac{m}{(c^{\mathrm{T}}x^{k-1} - \beta^k)^2}cc^{\mathrm{T}}\right) \leqslant 2^{39L}. \tag{8.1.11}$$

From (8.1.1), (8.1.9), (8.1.10) and (8.1.11) we get that

$$\kappa(A^{\mathrm{T}}DA) \leq 2^{50L}$$

and

$$\kappa\left(A^{\mathrm{T}}DA + \frac{m}{(c^{\mathrm{T}}x^{k-1} - \beta^k)^2} cc^{\mathrm{T}}\right) \leq 2^{51L}.$$

Let γ be a fixed constant greater than 200. We maintain an approximate inverse $(A^{\mathrm{T}}DA)_{\mathrm{a}}^{-1}$ of $A^{\mathrm{T}}DA$ such that

$$(A^{\mathrm{T}}DA)_{\mathrm{a}}^{-1} = (A^{\mathrm{T}}DA)^{-1} + \Delta$$

where $\|\Delta\|_2 \leq 2^{-\gamma L}$. Then, from the above bounds on condition numbers and eigenvalues, and noting that $c^{\mathrm{T}}x^{k-1} - \beta^k \geq 2^{-14L}$, it is easily shown that

$$(A^{\mathrm{T}}DA)_{\mathrm{a}}^{-1} - \frac{m(A^{\mathrm{T}}DA)_{\mathrm{a}}^{-1}cc^{\mathrm{T}}(A^{\mathrm{T}}DA)_{\mathrm{a}}^{-1}}{(c^{\mathrm{T}}x^{k-1} - \beta^k)^2 + mc^{\mathrm{T}}(A^{\mathrm{T}}DA)_{\mathrm{a}}^{-1}c} = \left(A^{\mathrm{T}}DA + \frac{m}{(c^{\mathrm{T}}x^{k-1} - \beta^k)^2} cc^{\mathrm{T}}\right)^{-1} + \Delta'$$

where $\|\Delta'\|_2 \leq 2^{-(\gamma-80)L}$. Once the approximate inverse $(A^{\mathrm{T}}DA)_{\mathrm{a}}^{-1}$ is available, $x^{k-1} + t^k\xi^k$ is obtained as follows. First, the gradient η^k, the approximation to $(A^{\mathrm{T}}DA + (m/(c^{\mathrm{T}}x^{k-1} - \beta^k)^2)cc^{\mathrm{T}})^{-1}$ given by the above formula, the direction ξ^k, the scalar t^k, and $x^{k-1} + t^k\xi^k$ are all computed using $2\gamma L$ bits of precision. Then each coordinate of the point $x^{k-1} + t^k\xi^k$ is rounded off to $40L$ bits.

Finally, as $\|\eta^k\|_2 \leq 2^{22L}$, and $t^k \leq 2^{2L}$, it is easily shown that the error in $x^{k-1} + t^k\xi^k$ is $\mathrm{O}(2^{-40L})$. Then the potential difference between the computed and the exact value of $x^{k-1} + t^k\xi^k$ is neglible.

8.2. Maintaining accuracy in the inverse

Here we shall briefly describe how to maintain accuracy in $(A^{\mathrm{T}}DA)^{-1}$ during rank one changes. Let B denote the matrix $A^{\mathrm{T}}DA$ before a rank one change, and let $B + uv^{\mathrm{T}}$, where $u \in \mathbb{R}^n$, $v \in \mathbb{R}^n$, denote the matrix $A^{\mathrm{T}}DA$ after the rank one change. Given an approximate inverse B' of B, we must compute an approximate inverse of $B + uv^{\mathrm{T}}$. We may restrict ourselves to the case where v is either u or $-u$, because whenever the ith element of D is updated, $A^{\mathrm{T}}DA$ changes by a rank one matrix of the form $d'a_ia_i^{\mathrm{T}}$, where d' is some scalar. Let us assume that an approximate inverse B' of B, such that

$$BB' = I + E_1 \quad \text{and} \quad \|E_1\|_2 \leq 2^{-\gamma_1 L}.$$

for some constant γ_1, is available. Furthermore, let us also assume that each entry of B' is represented using $\gamma_2 L$ bits for some constant γ_2. A good approximation to $(B + uv^{\mathrm{T}})^{-1}$ is computed as follows.

(1) Compute an initial approximation $B'' = B' - B'uv^{\mathrm{T}}B'/(1 + v^{\mathrm{T}}B'u)$.

(2) $(B + uv^{\mathrm{T}})(B'') = I + E_1 + E_2$, where E_2 is a constant rank matrix computable in $\mathrm{O}(n^2)$ operations. An adequate approximate inverse of $B + uv^{\mathrm{T}}$ is obtained by rounding the entries in $B'' - B''E_2$ to multiples of $2^{-\gamma_2 L}$.

Let $B''-B''E_2+E_3$ be the computed approximate inverse of $B+uv^{\mathrm{T}}$. Then

$$(B+uv^{\mathrm{T}})(B''-B''E_2+E_3)=I+E_1-(E_1+E_2)E_2+(B+uv^{\mathrm{T}})E_3.$$

We shall obtain an upper bound for $\|E_2\|_2$ in terms of $\|E_1\|_2$. Using this bound we shall show that it is possible to choose γ_1 and γ_2 so that

$$\|E_1-(E_1+E_2)E_2+(B+uv^{\mathrm{T}})E_3\|_2\leqslant(1+2^{-L})\|E_1\|_2.$$

Note that from (8.1.2), (8.1.3) and (8.1.9) in Section 8.1 it follows that

$$2^{-12L}\leqslant 1/\|B^{-1}\|_2\leqslant\|B\|_2\leqslant 2^{38L} \tag{8.2.1}$$

and

$$2^{-12L}\leqslant 1/\|(B+uv^{\mathrm{T}})^{-1}\|_2\leqslant\|B+uv^{\mathrm{T}}\|_2\leqslant 2^{38L}. \tag{8.2.2}$$

Next, we obtain a lower bound for $1+v^{\mathrm{T}}B'u$. Using the rank one correction formula for the inverse of a matrix, we may write

$$(B^{-1}u)(B^{-1}v)^{\mathrm{T}}/(1+v^{\mathrm{T}}B^{-1}u)=B^{-1}-(B+uv^{\mathrm{T}})^{-1},$$

and so

$$\begin{aligned}|1+v^{\mathrm{T}}B^{-1}u|&\geqslant\frac{\|(B^{-1}u)(B^{-1}v)^{\mathrm{T}}\|_{\mathrm{F}}}{\|B^{-1}\|_{\mathrm{F}}+\|(B+uv^{\mathrm{T}})^{-1}\|_{\mathrm{F}}}\\&\geqslant\frac{\|B^{-1}u\|_2\|B^{-1}v\|_2}{\sqrt{n}(\|B^{-1}\|_2+\|(B+uv^{\mathrm{T}})^{-1}\|_2)}\\&\geqslant 2^{-90L}\|u\|_2\|v\|_2\quad\text{(from 8.2.1, 8.2.2).}\end{aligned}$$

Noting that $B'=B^{-1}+B^{-1}E_1$ we get

$$\begin{aligned}|1+v^{\mathrm{T}}B'u|&\geqslant|1+v^{\mathrm{T}}B^{-1}u|-|v^{\mathrm{T}}B^{-1}E_1u|\\&\geqslant(2^{-90L}-\|B^{-1}\|_2\|E_1\|_2)\|u\|_2\|v\|_2.\end{aligned} \tag{8.2.3}$$

Next, observe that

$$E_2=-E_1uv^{\mathrm{T}}B'/(1+v^{\mathrm{T}}B'u)$$

and hence

$$\begin{aligned}\|E_2\|_2&\leqslant\|E_1\|_2\|B'\|_2\|u\|_2\|v\|_2/|1+v^{\mathrm{T}}B'u|\\&\leqslant\frac{2^{13L}}{2^{-90L}-2^{12L}\|E_1\|_2}\|E_1\|_2\quad\text{(from 8.2.1, 8.2.3).}\end{aligned}$$

From the bounds on $\|E_2\|_2$, and on $\|B+uv^{\mathrm{T}}\|_2$, it follows that we may choose suitably large values for the constants γ_1 and γ_2 so that $\|E_2^2\|_2$, $\|E_1E_2\|_2$, and $\|(B+uv^{\mathrm{T}})E_3\|_2$ are each less than $2^{-(L+2)}\|E_1\|_2$. Choosing $\gamma_1\geqslant 300$ and $\gamma_2\geqslant\gamma_1+50$ suffices. Then

$$\|E_1-(E_1+E_2)E_2+(B+uv^{\mathrm{T}})E_3\|_2\leqslant(1+2^{-L})\|E_1\|_2.$$

Thus the error in the maintained approximate inverse of $A^{\mathrm{T}}DA$ grows very slowly.

As before let $(A^{\mathrm{T}}DA)_{\mathrm{a}}^{-1}$ be the maintained approximate inverse of $A^{\mathrm{T}}DA$, and let

$$(A^{\mathrm{T}}DA)(A^{\mathrm{T}}DA)_{\mathrm{a}}^{-1} = I + E.$$

Suppose initially $\|E\|_2 \leqslant 2^{-(\gamma_1+1)L}$. Then from the above bound on the growth of the error in $(A^{\mathrm{T}}DA)_{\mathrm{a}}^{-1}$ we may conclude that after j rank one changes to $A^{\mathrm{T}}DA$, $j \leqslant m+n$, $\|E\|_2 \leqslant (1+2^{-L})^j 2^{-(\gamma_1+1)L}$. Thus for $j \leqslant m+n$, $\|E\|_2 \leqslant 2^{-\gamma_1 L}$. After $m+n$ rank one changes to $A^{\mathrm{T}}DA$ we can recompute $(A^{\mathrm{T}}DA)_{\mathrm{a}}^{-1}$ by directly inverting $A^{\mathrm{T}}DA$. (This recomputation is performed at most $\mathrm{O}(L)$ times, and so does not increase the running time of the algorithm.)

As mentioned in Section 8.1, we require an $(A^{\mathrm{T}}DA)_{\mathrm{a}}^{-1}$ such that

$$(A^{\mathrm{T}}DA)_{\mathrm{a}}^{-1} = (A^{\mathrm{T}}DA)^{-1} + \Delta$$

where $\|\Delta\|_2 \leqslant 2^{-\gamma L}$. It is easily seen that $\Delta = (A^{\mathrm{T}}DA)^{-1}E$. So from (8.1.2) and (8.1.9) in Section 8.1 it follows that if $\gamma_1 \geqslant \gamma + 50$ then $\|\Delta\|_2 \leqslant 2^{-\gamma L}$. So from the other requirement on γ_1, we can conclude that choosing $\gamma_1 \geqslant \max\{\gamma+50, 300\}$ is adequate.

Finally, since each entry in B' is represented using $\gamma_2 L$ bits, the intermediate matrices B'', E_2 and $B''-B''E_2$, in steps (1) and (2) above may be computed exactly using rational arithmetic and $\mathrm{O}(L)$ bits of precision.

9. Finding an optimal vertex

In this section we describe how to find an optimal vertex by performing at most $\mathrm{O}(mn^2)$ arithmetic operations to a precision of $\mathrm{O}(L)$ bits, once a feasible point that is sufficiently close in objective function value to the optimum is available. Let $\hat{x}$ be a point such that

$$A\hat{x} \geqslant b,$$

$$c^{\mathrm{T}}\hat{x} \geqslant \beta^{\max} - \varepsilon$$

where $\varepsilon = 2^{-12L}$ and $\beta^{\max}$ is the maximum value of $c^{\mathrm{T}}x$ over the polytope $\{x\colon Ax \geqslant b\}$. Given $\hat{x}$, there are several ways of finding an optimal vertex x^{opt}, the procedure we shall describe is similar to the one in [7, pp. 173–174]. First, note that $\beta^{\max}$ is a rational number with numerator and denominator at most 2^{2L}. So $\beta^{\max}$ is the unique rational, with denominator less than or equal to 2^{2L}, closest to $c^{\mathrm{T}}\hat{x}$, and may be computed using the method of continued fractions. Let S denote the set of column vectors of A^{T}. W.r.t. a point x define $V(\delta, x)$ to be a subset of $S \cup \{c\}$ such that:

(i) $|c^{\mathrm{T}}x - \beta^{\max}| \leqslant \delta$ iff $c \in V(\delta, x)$.

(ii) For each $a_i \in S$, $|a_i^{\mathrm{T}}x - b_i| \leqslant \delta$ iff $a_i \in V(\delta, x)$.

$V(\delta, x)$ may be thought of as the set of constraints that are almost satisfied with equality (with an error of at most δ). Let $\dim(V(\delta, x))$ denote the dimension of the vector space spanned by the elements of $V(\delta, x)$. Using $\hat{x}$ we obtain a point x^* such that:

(i) $|c^{\mathrm{T}}x^* - \beta^{\max}| \leqslant 2\varepsilon$.

(ii) $a_i^{\mathrm{T}}x^* \geqslant b_i - 2\varepsilon$, $a_i \in S$.

(iii) Every vector in $S \cup \{c\}$ can be expressed as a linear combination of the vectors in $V(2\varepsilon, x^*)$, i.e. $\dim(V(2\varepsilon, x^*)) = n$.

Then the point x^{opt} which satisfies the system of linear equations

$$a_i^{\mathrm{T}} x^{\mathrm{opt}} = b_i, \quad a_i \in V(2\varepsilon, x^*),$$

$$c^{\mathrm{T}} x^{\mathrm{opt}} = \beta^{\max},$$

is feasible, and is an optimal vertex (see [7, pp. 173–174]) (note that even though the $V(2\varepsilon, x^*)$ may contain more than n vectors the above system of linear equations does have a solution and so x^{opt} is a well-defined point). Once V is available, x^{opt} may be computed by using a version of gaussian elimination given in [2, 3], and performing rational arithmetic with $O(L)$ bits of precision.

We shall give a procedure to obtain x^*. The procedure is almost identical to the one given in [4, pp. 173–174]. The idea is to stay close to the plane $c^{\mathrm{T}} x = \beta^{\max}$, remain almost feasible, and at the same time increase the rank of the set of constraints that are almost satisfied with equality. Suppose that we have a point x' such that:

(i) $|c^{\mathrm{T}} x' - \beta^{\max}| \leq \delta$.

(ii) $a_i^{\mathrm{T}} x' \geq b_i - \delta$, $a_i \in S$.

(iii) There is a column a_μ of A^{T} which is linearly independent of the vectors in $V(\delta, x')$.

We may then compute a u such that

$$|c^{\mathrm{T}} u| \leq 2^{-16L} \varepsilon,$$

$$|a_i^{\mathrm{T}} u| \leq 2^{-16L} \varepsilon, \quad a_i \in V(\delta, x'),$$

$$a_\mu^{\mathrm{T}} u \leq -1.$$

Thus by moving in the direction of u from x' we can substantially decrease the distance from the constraining plane $a_\mu^{\mathrm{T}} x = b_\mu$ without appreciably changing the value of the objective function or the distance from any of the constraining planes $a_i^{\mathrm{T}} x = b_i$, $a_i \in V(\delta, x')$. Specifically, we can find a scalar $0 < \lambda \leq 2^{6L}$ such that:

(i) $|c^{\mathrm{T}}(x' + \lambda u) - \beta^{\max}| \leq \delta + 2^{-2L} \varepsilon$.

(ii) $a_i^{\mathrm{T}}(x' + \lambda u) \geq b_i - \delta - 2^{-2L} \varepsilon$, $a_i \in S$.

(iii) $\dim(V(\delta + 2^{-2L}\varepsilon, x' + \lambda u)) \geq \dim(V(\delta, x')) + 1$.

The existence of such a λ follows from two observations. First, the coordinates of any feasible point are bounded by 2^L. Second, if a_j can be expressed as a linear combination of the elements of $V(\delta, x')$ then the coefficients in the linear combination are rationals with numerators and denominators bounded by 2^{2L}.

Thus we can construct a sequence $x^0 = \hat{x}, x^1, \ldots, x^k, \ldots, x^*$ such that $\dim(V(\varepsilon^k, x^k)) > \dim(V(\varepsilon^{k-1}, x^{k-1}))$, where $\varepsilon^k = (1 + k2^{-L})\varepsilon$. We shall show that computing x^* from $\hat{x}$ requires $O(mn^2)$ arithmetic operations, and it is adequate to perform each operation to a precision of $O(L)$ bits. Let M_k be a matrix such that:

(i) Each row of M_k is an element of $V(\varepsilon^k, x^k)$.

(ii) The rows of M_k are linearly independent.

(iii) The number of rows of M_k equals $\dim(V(\varepsilon^k, x^k))$.

Let a_j be a vector in the set $S - V(\varepsilon^k, x^k)$. Note that a_j is linearly independent of the rows of M_k iff the orthogonal projection of a_j onto the orthogonal complement of the row space of M_k is non-zero. Since entries in A are integers, the orthogonal complement of the row space of M_k (i.e. the subspace $\{x\colon M_k x = 0\}$) has a basis such that the coordinates of each vector in the basis are rationals with a common denominator which is at most 2^L. Thus if the orthogonal projection of a_j onto this subspace is non-zero then the 2-norm of the orthogonal projection is at least 2^{-6L}. Furthermore, the said orthogonal projection of a_j is the vector

$$(I - M_k^{\mathrm{T}}(M_k M_k^{\mathrm{T}})^{-1} M_k) a_j.$$

To be able to compute the orthogonal projection to a sufficient degree of accuracy, we maintain an approximate inverse $(M_k M_k^{\mathrm{T}})_{\mathrm{a}}^{-1}$ of $M_k M_k^{\mathrm{T}}$ such that

$$(M_k M_k^{\mathrm{T}})_{\mathrm{a}}^{-1} = (M_k M_k^{\mathrm{T}})^{-1} + E_k$$

where E_k is an error matrix such that $\|E_k\|_2 \leq 2^{-\gamma_3 L}$ for a suitable constant $\gamma_3 > 100$. (Alternately, the LU decomposition of $M_k M_k^{\mathrm{T}}$ could be maintained.) In a manner similar to Section 8.1, it is easily shown that

$$2^{-4L} \leq \lambda_{\min}(M_k M_k^{\mathrm{T}}) \leq \lambda_{\max}(M_k M_k^{\mathrm{T}}) \leq 2^{2L}.$$

Then from the bounds on the eigenvalues of $M_k M_k^{\mathrm{T}}$, it is easily seen that using the approximate inverse of $M_k M_k^{\mathrm{T}}$ (instead of the exact inverse) will produce an error of at most $2^{-(\gamma_3 - 40L)}$ in computing the 2-norm of the orthogonal projection of a_j. So from the computed orthogonal projection we may correctly determine whether a_j is or is not linearly dependent on the rows of M_k. Once the approximate inverse of $M_k M_k^{\mathrm{T}}$ is available, finding the orthogonal projection of a_j requires $O(n^2)$ operations. Furthermore, a suitable approximate inverse of $M_{k+1} M_{k+1}^{\mathrm{T}}$ is obtained from the approximate inverse of $M_k M_k^{\mathrm{T}}$ in a manner similar to Section 8.2, using the updating formula [5, 12],

$$\begin{bmatrix} M & v \\ v^{\mathrm{T}} & d \end{bmatrix}^{-1} = \frac{1}{d - v^{\mathrm{T}} M^{-1} v} \begin{bmatrix} (d - v^{\mathrm{T}} M^{-1} v) M^{-1} + (M^{-1} v)(M^{-1} v)^{\mathrm{T}} & -M^{-1} v \\ -(M^{-1} v)^{\mathrm{T}} & 1 \end{bmatrix},$$

and this requires $O(n^2)$ arithmetic operations performed to precision of $O(L)$ bits. Thus given $\hat{x}$, x^* may be computed in $O(mn^2)$ arithmetic operations, with each operation being performed to a precision of $O(L)$ bits.

Acknowledgements

The author would like to thank Sanjiv Kapoor, Narendra Karmarkar, and Jeff Lagarias for helpful discussions and suggestions. The author would like to thank M.R. Garey and D.S. Johnson for useful suggestions concerning the organization of this paper.

References

[1] D.A. Bayer and J.C. Lagarias, "The non-linear geometry of linear programming I. Affine and projective scaling trajectories," *Transactions of the American Mathematical Society* (1989), to appear.

[2] J. Edmonds, "Systems of distinct representatives and linear algebra," *Journal of Research of the National Bureau of Standards* 71B (1967) 241-245.

[3] F.R. Gantmacher, *Matrix Theory, Vol. 1* (Chelsea, London, 1959) Chapter 2.

[4] C. Gonzaga, "An algorithm for solving linear programming problems in $O(n^3L)$ operations," Memorandum UCB/ERL M87/10, Electronics Research Laboratory, University of Berkeley (Berkeley, CA, 1987).

[5] N. Karmarkar, "A new polynomial time algorithm for linear programming," *Combinatorica* 4 (1984) 373-395.

[6] L.G. Khachian, "Polynomial algorithms in linear programming," *Žurnal Vyčislitel'noĭ Matematiki i Matematičeskoĭ Fiziki* 20 (1980) 53-72.

[7] C. Papadimitriou and K. Steiglitz, *Combinatorial Optimization: Algorithms and Complexity* (Prentice-Hall, Englewood Cliffs, NJ, 1982).

[8] J. Renegar, "A polynomial-time algorithm, based on Newton's method, for linear programming," *Mathematical Programming*, 40 (1988) 59-93.

[9] Gy. Sonnevand, "An analytical center for polyhedrons and new classes of global algorithms for linear (smooth, convex) programming," preprint, Department of Numerical Analysis, Institute of Mathematics, Eötvös University (Budapest, 1989) 6-8.

[10] G.W. Stewart, *Introduction to Matrix Computations* (Academic Press, New York, 1973).

[11] P.M. Vaidya, "An algorithm for linear programming which requires $O(((m+n)n^2+(m+n)^{1.5}n)L)$ arithmetic operations," *Proceedings 19th Annual ACM Symposium Theory of Computing* (1987) 29-38.

[12] J.H. Wilkinson, *The Algebraic Eigenvalue Problem* (Oxford University Press (Clarendon), London and New York, 1965).

[13] G. Zoutendijk, *Mathematical Programming Methods* (North-Holland, New York, 1976).