Mathematical Programming 73 (1996) 291–341

A new algorithm for minimizing convex functions over convex sets

Pravin M. Vaidya

AT & T Bell Laboratories, Murray Hill, NJ 07974, USA

Received 12 October 1989; revised manuscript received 11 June 1991

Abstract

Let $S \subseteq \mathbb{R}^n$ be a convex set for which there is an oracle with the following property. Given any point $z \in \mathbb{R}^n$ the oracle returns a "Yes" if $z \in S$; whereas if $z \notin S$ then the oracle returns a "No" together with a hyperplane that separates z from S. The feasibility problem is the problem of finding a point in S; the convex optimization problem is the problem of minimizing a convex function over S. We present a new algorithm for the feasibility problem. The notion of a volumetric center of a polytope and a related ellipsoid of maximum volume inscribable in the polytope are central to the algorithm. Our algorithm has a significantly better global convergence rate and time complexity than the ellipsoid algorithm. The algorithm for the feasibility problem easily adapts to the convex optimization problem.

Keywords: Convex programming; Linear programming; Optimization; Complexity

1. Introduction

Let $S \subseteq \mathbb{R}^n$ be a convex set for which there is an oracle with the following property. The oracle accepts as input any point in $\mathbb{R}^n$. If the input $z \in S$ then the oracle returns a "Yes"; whereas if $z \notin S$ then the oracle returns a "No" along with a vector $c \in \mathbb{R}^n$ such that $S \subseteq \{x\colon c^{\mathrm{T}}x \geqslant c^{\mathrm{T}}z\}$. (Typically, S is of the form $S = \{x\colon g_i(x) \leqslant 0, 1 \leqslant i \leqslant p\}$ where $g_i(x)$ are differentiable convex functions.) The *feasibility problem* is the problem of finding a point in S given an oracle for S. Let $g(x)$ be a convex function such that given a point $z \in \mathbb{R}^n$ in the domain of g we can compute a vector c such that $\{x\colon g(x) \leqslant g(z)\} \subseteq \{x\colon c^{\mathrm{T}}x \geqslant c^{\mathrm{T}}z\}$. The *convex optimization problem* is the problem of minimizing $g(x)$ over S. In this paper we shall describe a new algorithm for the feasibility problem. An easy modification to the algorithm for the feasibility problem

SSDI 0025-5610(92)00021-K

will lead to an algorithm for the convex optimization problem. We note that a preliminary version of this paper appeared in [9].

In the case of the feasibility problem we shall assume that S is contained in a ball of radius 2^L centered at the origin and that if S is nonempty then it contains a ball of radius 2^{-L}. Let x^{opt} be the point that minimizes $g(x)$ over S and let γ be a given parameter. In the case of the optimization problem we shall assume that S is contained in a ball of radius 2^L centered at the origin and that the set $\{x\colon x \in S,\ g(x) - g(x^{\text{opt}}) \leqslant \gamma\}$ contains a ball of radius 2^{-L}; the output of the algorithm is a point $x^* \in S$ such that $g(x^*) - g(x^{\text{opt}}) \leqslant \gamma$. We note that our algorithm easily adapts to the different versions of the feasibility and the optimization problems described in [4].

A generic iterative algorithm for the feasibility problem is as follows. We maintain a region R such that $S \subseteq R$. At each iteration we choose a test point z in R and call the oracle with z as input. We halt if $z \in S$. So suppose $z \notin S$. Then the oracle returns a vector c such that $\forall x \in S$, $c^{\mathrm{T}}x \geqslant c^{\mathrm{T}}z$. Let $\beta \leqslant c^{\mathrm{T}}z$. Then $S \subseteq (R \cap \{x\colon c^{\mathrm{T}}x \geqslant \beta\})$ and R is reset to be the region $(R \cap \{x\colon c^{\mathrm{T}}x \geqslant \beta\})$. As the algorithm proceeds R shrinks and its volume decreases at a certain rate. If S is nonempty then it contains a ball of radius 2^{-L} and the algorithm halts with a point in S before the volume of R falls below 2^{-nL}. If S is empty then the algorithm halts the first time the volume of R falls below 2^{-nL} and since R contains S this gives a proof that S is empty. During the course of the algorithm the description of R can become complicated and choosing the test point can become expensive; so if the region R becomes too complicated we replace R by a simpler region that contains R; such a replacement trades volume for computational efficiency and the algorithm still converges.

The well-known ellipsoid algorithm [4,5] falls in this generic scheme; in the ellipsoid algorithm the region R is an ellipsoid and the test point used is the center of the ellipsoid. Another algorithm due to Levin [5] uses simplices instead of ellipsoids. Our algorithm also follows the above scheme. In our case the region R is a bounded full-dimensional polytope $P = \{x\colon Ax \geqslant b\}$ where $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$. The test point we use is the point that minimizes the determinant of the Hessian of the logarithmic barrier for P. Specifically, the logarithmic barrier is the function $-\sum_{i=1}^{m} \ln(a_i^{\mathrm{T}}x - b_i)$ and its Hessian evaluated at x, denoted by $H(x)$, is given by

$$H(x) = \sum_{i=1}^{m} \frac{a_i a_i^{\mathrm{T}}}{\left(a_i^{\mathrm{T}}x - b_i\right)^2},$$

where a_i^{T} denotes the ith row of A. Let $F(x) = \frac{1}{2}\ln(\det(H(x)))$ where $\det(H(x))$ denotes the determinant of $H(x)$, and let ω be the point that minimizes $F(x)$ over P. The point ω will be called the *volumetric center* of P. We use ω (or a good approximation to ω) as our test point. The function $F(x)$ is strictly convex and a Newton-type method can be used to compute a good approximation to ω efficiently. The polytope P is also trimmed from time to time (i.e., some of the planes defining P are dropped) so that the number of planes in the description of P does not grow beyond $O(n)$.

The volume of P decreases by a fixed constant factor (independent of the dimension n) at each iteration on the average, and our algorithm halts with a point in S (or with the conclusion that S is empty) in $O(nL)$ iterations. During each iteration we have to invert an $n \times n$ matrix (and solve a system of linear equations), and possibly query the oracle once. Let T be the cost (in terms of number of arithmetic operations) of one query to the oracle. Then the total number of arithmetic operations performed by our algorithm is $O(TnL + n^4L)$, and the total number of calls to the oracle is $O(nL)$. If we use fast matrix multiplication for performing the matrix inversion the total number of arithmetic operations reduces to $O(TnL + M(n)nL)$, where $M(n)$ is the number of operations for multiplying two $n \times n$ matrices. (It is known that $M(n) = O(n^{2.38})$, see [3].) The ellipsoid algorithm was previously the best known algorithm for the feasibility problem. In the ellipsoid algorithm the volume falls by a factor of about $(1 - 1/n)$ at each iteration, and the number of iterations is $O(n^2L)$. During each iteration we have to perform one query to the oracle together with a rank one correction to an $n \times n$ matrix and a matrix–vector multiplication. So the total number of arithmetic operations in the ellipsoid algorithm is $O(Tn^2L + n^4L)$, and the total number of calls to the oracle is $O(n^2L)$. Thus our algorithm performs asymptotically fewer operations as well as fewer calls to the oracle. The reason for stressing the number of calls to the oracle is that in many cases the cost of querying the oracle far exceeds the other costs in the algorithm [4]. It is worth noting that using fast matrix multiplication does not reduce the number of operations performed by the ellipsoid algorithm.

Next, we mention some recent related work. The analytic center of a polytope has been widely used in interior point methods for linear programming, and algorithms for convex optimization based on the analytic center have been suggested in [10–12] but without any convergence analysis. An algorithm based on a maximal inscribed ellipsoid is given in [13]; this algorithm has a rate of convergence (i.e., volume decrease) comparable to the one in this paper, but computing a maximal inscribed ellipsoid is much more expensive (at least by a factor of $\sqrt{n}$) than computing the volumetric center.

A natural question that arises is: Is there a simple but intuitive explanation for why is the volumetric center ω a good test point? The question may be answered as follows. Let $E(H(x), x, r)$ denote the ellipsoid given by

$$E(H(x), x, r) = \{y: (y-x)^{\mathrm{T}} H(x)(y-x) \leqslant r^2\}.$$

$E(H(x), x, 1) \subseteq P$ and may be thought of as a local quadratic approximation to P. $E(H(\omega), \omega, 1)$ has the largest volume among all such ellipsoids $E(H(x), x, 1)$ and is hence a maximum volume quadratic approximation to P. A plane through ω divides $E(H(\omega), \omega, 1)$ into two parts of equal volume; so there is a good chance that a plane through ω divides P into two parts with approximately equal volume (loosely speaking). So if the process of cutting P through ω is iterated the volume would be expected to decrease at a good rate.

There is also a simple intuitive reason for why our algorithm has a faster rate of convergence than the ellipsoid algorithm. In the ellipsoid algorithm the half-ellipsoid to which the set S is localized after an oracle query is immediately enclosed in another smaller ellipsoid and the vector c generated by the oracle is not used in subsequent steps; as a result a considerable amount of information is given up at each step. Since our algorithm works with polytopes instead of ellipsoids the cutting planes generated by the oracle are maintained for several steps after they are generated and continue to directly influence the choice of the test point. Furthermore, hyperplanes are dropped and the polytope P is trimmed not at each step but whenever necessary. As a consequence more of the information generated by the oracle gets utilized and the volume of P shrinks at a geometric rate independent of n.

A byproduct of our algorithm is an algorithm for solving linear programming problems which performs a total of $O(mn^2L + M(n)nL)$ arithmetic operations in the worst case, where m is the number of constraints and n is the number of variables; this gives an improvement in the time complexity of linear programming for $m > n^2$, see [8]. We also note that if the polytope P is not trimmed in our algorithm (i.e., we do not discard any plane generated by the oracle) we still get a convergent algorithm that halts in $O(n^2L^2)$ iterations. Finally, note that any problem that can be solved by the ellipsoid algorithm can be solved with a better time complexity by our algorithm.

In Section 2 we give an overview and describe the algorithm for the feasibility problem. In Section 3 we discuss three theorems which summarize the behaviour of a Newton-type method for minimizing $F(x)$ and which characterize how $F(\omega)$ changes when we add (remove) a plane to (from) the polytope P. In Section 4 we briefly describe how an easy modification to the algorithm for the feasibility problem leads to an algorithm for the convex optimization problem. In Section 5 we discuss variants of the basic algorithm in Section 2. In Section 6 we propose algorithms for solving linear programming problems that follow a path of volumetric centers or hybrid centers. In Section 7 we state and prove various properties of $F(x)$. In Section 8 we prove the theorems introduced in Section 3.

2. An overview

In this section we shall describe the algorithm for the feasibility problem. But first we shall introduce some notation. Let P be the bounded full-dimensional polytope

$$P = \{x:\ Ax \geqslant b\},$$

where $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$ and $x \in \mathbb{R}^n$. Let $H(x)$ be defined as

$$H(x) = \sum_{i=1}^{m} \frac{a_i a_i^{\mathrm{T}}}{\left(a_i^{\mathrm{T}} x - b_i\right)^2},$$

where a_i^{T} denotes the ith row of A. $H(x)$ is the Hessian of the logarithmic barrier function $\sum_{i=1}^{m} -\ln(a_i^{\mathrm{T}}x - b_i)$ and is positive definite for all x in the interior of P. Let $F(x)$ be defined as

$$F(x) = \tfrac{1}{2}\ln(\det(H(x))),$$

where $\det(H(x))$ denotes the determinant of $H(x)$, and let ω be the point that minimizes $F(x)$ over the polytope P. The point ω will be called the volumetric center of P. Let $\nabla F(x)$ ($\nabla^2 F(x)$) denote the gradient (Hessian) of $F(x)$ evaluated at x. Let

$$\sigma_i(x) = \frac{a_i^{\mathrm{T}} H(x)^{-1} a_i}{\left(a_i^{\mathrm{T}} x - b_i\right)^2}, \quad 1 \leqslant i \leqslant m.$$

By Lemma 1 (Section 7), the gradient $\nabla F(x)$ may be written as

$$\nabla F(x) = -\sum_{i=1}^{m} \sigma_i(x) \frac{a_i}{a_i^{\mathrm{T}} x - b_i}.$$

Let $Q(x)$ be defined as

$$Q(x) = \sum_{i=1}^{m} \sigma_i(x) \frac{a_i a_i^{\mathrm{T}}}{\left(a_i^{\mathrm{T}} x - b_i\right)^2}.$$

Note that $Q(x)$ is positive definite over the interior of P. $Q(x)$ is a good approximation to $\nabla^2 F(x)$; specifically, the quadratic forms $\xi^{\mathrm{T}} \nabla^2 F(x) \xi$ and $\xi^{\mathrm{T}} Q(x) \xi$ satisfy the condition

$$\forall \xi \in \mathbb{R}^n, \quad 5\xi^{\mathrm{T}} Q(x) \xi \geqslant \xi^{\mathrm{T}} \nabla^2 F(x) \xi \geqslant \xi^{\mathrm{T}} Q(x) \xi.$$

Since $Q(x)$ is positive definite this condition implies that $F(x)$ is a strictly convex function over the interior of P. Let $\mu(x)$ be the largest number λ satisfying the condition that

$$\forall \xi \in \mathbb{R}^n, \quad \xi^{\mathrm{T}} Q(x) \xi \geqslant \lambda \xi^{\mathrm{T}} H(x) \xi.$$

Note that $\mu(x) \geqslant \min_{1 \leqslant i \leqslant m}\{\sigma_i(x)\}$.

We shall now describe the algorithm for the feasibility problem. The algorithm starts out with the simplex $P = \{x\colon x_j \geqslant -2^L, 1 \leqslant j \leqslant n, \sum_{j=1}^{n} x_j \leqslant n2^L\}$. (Note that the algorithm could start with any polytope whose volumetric center is easy to compute, say for example a box.) Since S is contained in a ball of radius 2^L centered at the origin, initially $S \subseteq P$. Throughout the algorithm S and P satisfy the relation $S \subseteq P$. Let δ and ε be small constants such that $\delta \leqslant 10^{-4}$, and $\varepsilon \leqslant 10^{-3}\delta$. At the beginning of each iteration we have a point $z \in P$ such that

$$F(z) - F(\omega) \leqslant \varepsilon^4 \mu(\omega).$$

(Note that when the algorithm starts the polytope P is just a simplex and we can explicitly solve the equation $\nabla F(x) = 0$ to obtain the volumetric center of a simplex; in this case the analytic center [6] is also the volumetric center.) The computation performed during an iteration falls into two cases depending on the value of $\min_{1 \leqslant i \leqslant m}\{\sigma_i(z)\}$.

Case 1. $\min_{1 \leqslant i \leqslant m}\{\sigma_i(z)\} \geqslant \varepsilon$. In this case we add a plane to the polytope P. First, the oracle is called with the current point z as input. The algorithm halts if $z \in S$; otherwise the oracle returns a vector c such that

$$\forall x \in S, \quad c^{\mathrm{T}} x \geqslant c^{\mathrm{T}} z.$$

We choose β such that $c^{\mathrm{T}}z \geqslant \beta$ and

$$\frac{c^{\mathrm{T}}H(z)^{-1}c}{(c^{\mathrm{T}}z-\beta)^2} = \tfrac{1}{2}(\delta\varepsilon)^{1/2}.$$

Let $\tilde{A} = (A\ c)^{\mathrm{T}}$ and let $\tilde{b} = (b\ \beta)^{\mathrm{T}}$. A and B are reset as

$$A \leftarrow \tilde{A}, \quad b \leftarrow \tilde{b}.$$

Since ω shifts due to the addition of a plane to P, we use a Newton-type method to move closer to ω as follows.

For $j = 1$ to $\lceil 30 \ln(2\varepsilon^{-4.5}) \rceil$ do $z \leftarrow z - 0.18Q(z)^{-1}\nabla F(z)$.

Case 2. $\min_{1 \leqslant i \leqslant m}\{\sigma_i(z)\} < \varepsilon$. In this case we remove a plane from the polytope P. W.l.o.g. suppose that $\sigma_m(z) = \min_{1 \leqslant i \leqslant m}\{\sigma_i(z)\}$. Let $a_m = c$, $b_m = \beta$, $A = (\tilde{A}\ c)^{\mathrm{T}}$, and $b = (\tilde{b}\ \beta)^{\mathrm{T}}$. A and b are reset as

$$A \leftarrow \tilde{A}, \quad b \leftarrow \tilde{b}.$$

Since ω shifts due to the removal of a plane, we use a Newton-type method to move closer to ω as follows.

For $j = 1$ to $\lceil 30 \ln(4\varepsilon^{-3}) \rceil$ do $z \leftarrow z - 0.18Q(z)^{-1}\nabla F(z)$.

The convergence lemma below summarizes the behaviour of the algorithm.

Convergence Lemma. *Let $\delta \leqslant 10^{-4}$, let $\varepsilon \leqslant 10^{-3}\delta$, and let ρ^k denote the value of $F(\omega)$ at the beginning of the kth iteration. Then at the beginning of each iteration z satisfies the condition*

$$F(z) - F(\omega) \leqslant \varepsilon^4\mu(\omega).$$

Furthermore, the following statements hold.

(1) *If Case 1 occurs during the kth iteration then*

$$\rho^{k+1} - \rho^k \geqslant \frac{(\delta\varepsilon)^{1/2}}{5}.$$

(2) *If Case 2 occurs during the kth iteration then*

$$\rho^k - \rho^{k+1} \leqslant 5\varepsilon.$$

The proof of the Convergence Lemma is based on Theorems 1, 2 and 3 in Section 3 and will be given in Section 3.

Bounding the number of iterations

Let π^k denote the volume of the polytope P at the beginning of the kth iteration. Using the Convergence Lemma we shall next obtain an upper bound on π^k, and show that the algorithm halts in $O(nL)$ iterations. First, we shall show that

$$\rho^k \geqslant \rho^0 + \tfrac{1}{2}k\varepsilon.$$

Since P is bounded, the number of bounding planes of P is at least $n+1$, and to start with this number is exactly $n+1$. Thus by the kth iteration Case 1 must have occurred at least as often as Case 2; otherwise the number of planes would have fallen below $n+1$ which is not possible. So by the kth iteration Case 1 must have occurred at least $\frac{1}{2}k$ times, and Case 2 could have occurred at most $\frac{1}{2}k$ times. Hence

$$\rho^k - \rho^0 \geqslant \tfrac{1}{2}\left(\tfrac{1}{5}k(\delta\varepsilon)^{1/2} - 5k\varepsilon\right) \geqslant \tfrac{1}{2}k\varepsilon, \quad \text{since } \varepsilon \leqslant 10^{-3}\delta.$$

We shall next bound π^k. It is well-known (see [6]) that if x^* is the point that maximizes the logarithmic barrier over P, then

$$P \subseteq \left\{x\colon (x-x^*)^{\mathrm{T}} H(x^*)(x-x^*) \leqslant m^2\right\}.$$

Then from the relation between determinants and volume (see [4]) it follows that

$$\begin{aligned}\text{volume}(P) &\leqslant (2m)^n\left(\det(H(x^*))\right)^{-1/2} \leqslant (2m)^n\left(\det(H(\omega))\right)^{-1/2}\\ &\leqslant (2m)^n\, \mathrm{e}^{-F(\omega)}.\end{aligned}$$

Since $\sum_{i=1}^m \sigma_i(x) = n$ (see Claim 3, Section 7.1), Case 2 is forced to occur at an iteration if the number of planes defining P is greater than n/ε, and hence m never exceeds n/ε. Then since $\rho^0 \geqslant -(n(L+1) + \ln(n+1))$, we get that

$$\begin{aligned}\ln(\pi^k) &\leqslant n\ln(2m) - \rho^k\\ &\leqslant n\ln(2n/\varepsilon) - \rho^0 - \tfrac{1}{2}k\varepsilon\\ &\leqslant n(L + \ln(2n/\varepsilon) + 1) + \ln(n+1) - \tfrac{1}{2}k\varepsilon.\end{aligned}$$

Thus the volume of P must fall below 2^{-nL} in $O(nL)$ iterations. Hence the algorithm must halt in $O(nL)$ iterations since $S \subseteq P$ and S contains a ball of radius 2^{-L} if it is nonempty.

Bounding the number of arithmetic operations

Note that since $\sum_{i=1}^m \sigma_i(x) = n$ (see Claim 3, Section 7.1), Case 2 is forced to occur in the algorithm if the number of planes grows beyond n/ε. So $m = O(n)$. The number of operations per iteration may be accounted for as follows.

1. Since $m = O(n)$, the quantities $\sigma_i(z)$, $1 \leqslant i \leqslant m$, may be evaluated in $O(n^3)$ operations.

2. $O(1)$ steps of the Newton-type method are executed per iteration, and so $Q(z)^{-1}$ and $\nabla F(z)$ are computed $O(1)$ times. $\nabla F(z)$ may be expressed as

$$\nabla F(z) = -\sum_{i=1}^m \sigma_i(z)\frac{a_i}{a_i^{\mathrm{T}} z - b_i}.$$

So once $\sigma_i(z)$, $1 \leqslant i \leqslant m$, are available, $Q(z)$, $Q(z)^{-1}$ and $\nabla F(z)$ can be evaluated in $O(n^3)$ operations.

3. The oracle is called at most once and one such call costs T operations.

4. Computing β such that $c^{\mathrm{T}} H(z)^{-1} c/(c^{\mathrm{T}} z - \beta)^2 = \frac{1}{2}(\delta\varepsilon)^{1/2}$ also requires $O(n^3)$ operations. (It suffices to compute β approximately.)

Thus the number of operations per iteration is $O(T + n^3)$. Using fast matrix multiplication the number of operations per iteration may be reduced to $O(T + M(n))$ where

$M(n)$ is the number of operations for multiplying two $n \times n$ matrices. (It is known that $M(n) = O(n^{2.38})$, see [3].) Since the number of iterations is $O(nL)$, the total number of operations is $O(TnL + n^4L)$ without fast matrix multiplication and $O(TnL + M(n)nL)$ with fast matrix multiplication. The total number of calls to the oracle is $O(nL)$.

3. Adding / deleting a plane and moving closer to the volumetric center ω

In this section we shall discuss three theorems on which the proof of the Convergence Lemma in Section 2 is based.

Theorem 1. *Let $\delta \leqslant 10^{-4}$, let $z \in P$, let $\eta = Q(z)^{-1}\nabla F(z)$, and let r be a scalar such that $0 \leqslant r \leqslant 0.2$. Let $z' = z - \lambda\eta$ where λ is defined as follows.*

$$\text{If } F(z) - F(\omega) \leqslant \delta\sqrt{\mu(\omega)} \text{ then } \lambda = r \text{ else } \lambda = \frac{r\delta^{1/2}(\mu(z))^{1/4}}{\left(\nabla F(z)^{\mathrm{T}}\eta\right)^{1/2}}.$$

Then the following statements hold.

1. *If $F(z) - F(\omega) \leqslant \delta\sqrt{\mu(\omega)}$ then*
$$F(z') - F(\omega) \leqslant (1 - 0.71r + 1.9r^2)(F(z) - F(\omega)).$$
2. *If $F(z) - F(\omega) > \delta\sqrt{\mu(\omega)}$ then*
$$F(z) - F(z') \geqslant \frac{(r - 2.65r^2)\delta}{2\sqrt{m}}.$$

Theorem 1 states that a Newton-type algorithm for minimizing $F(x)$ will converge linearly if started from a point z such that $F(z) - F(\omega) \leqslant \delta\sqrt{\mu(\omega)}$, $\delta \leqslant 10^{-4}$. It also states that taking a Newton-like step from a point z such that $F(z) - F(\omega) > \delta\sqrt{\mu(\omega)}$ will decrease F by at least $\Omega(1/\sqrt{m})$.

The next two theorems address the following question: by how much does the minimum value of $F(\omega)$ increase (decrease) when we add (remove) the constraint $c^{\mathrm{T}}x \geqslant \beta$ to (from) the set of constraints defining the polytope P? We shall require some additional notation to denote the polytope obtained by adding (removing) a plane to (from) P, and the related functions and matrices. Let $\tilde{A} \in \mathbb{R}^{\tilde{m}\times n}$, $\tilde{b} \in \mathbb{R}^{\tilde{m}}$, and let $\tilde{P}$ be the polytope $\tilde{P} = \{x\colon \tilde{A}x \geqslant \tilde{b}\}$. Let $\tilde{a}_i^{\mathrm{T}}$ denote the ith row of $\tilde{A}$, let $\tilde{H}(x)$ be defined as

$$\tilde{H}(x) = \sum_{i=1}^{\tilde{m}} \frac{\tilde{a}_i\tilde{a}_i^{\mathrm{T}}}{\left(\tilde{a}_i^{\mathrm{T}}x - \tilde{b}_i\right)^2},$$

let $\tilde{F}(x) = \frac{1}{2}\ln(\det(\tilde{H}(x)))$, and let $\tilde{\omega}$ be the point that minimizes $\tilde{F}(x)$ over the polytope $\tilde{P}$. Let $\tilde{\sigma}_i(x) = \tilde{a}_i^{\mathrm{T}}\tilde{H}(x)^{-1}\tilde{a}_i/(\tilde{a}_i^{\mathrm{T}}x - \tilde{b}_i)^2$, $1 \leqslant i \leqslant \tilde{m}$, let

$$\tilde{Q}(x) = \sum_{i=1}^{\tilde{m}} \tilde{\sigma}_i(x)\frac{\tilde{a}_i\tilde{a}_i^{\mathrm{T}}}{\left(\tilde{a}_i^{\mathrm{T}}x - \tilde{b}_i\right)^2},$$

and let $\tilde{\mu}(x)$ be the largest number λ such that $\forall\xi \in \mathbb{R}^n$, $\xi^{\mathrm{T}}\tilde{Q}(x)\xi \geqslant \lambda\xi^{\mathrm{T}}\tilde{H}(x)\xi$.

Theorem 2. *Let* $\tilde{A} = (A\ c)^{\mathrm{T}}$, $\tilde{b} = (b\ \beta)^{\mathrm{T}}$ *and* $\tilde{P} = P \cap \{x\colon c^{\mathrm{T}}x \geqslant \beta\}$. *Let* $z \in \tilde{P}$, *let* $\alpha \leqslant \delta \leqslant 10^{-4}$ *and let* $\alpha \leqslant \delta\mu(z)$. *Suppose that* $F(z) - F(\omega) \leqslant \alpha^2\mu(\omega)$ *and that* $c^{\mathrm{T}}H(z)^{-1}c/(c^{\mathrm{T}}z - \beta)^2 = \frac{1}{2}\alpha^{1/2}$. *Then*

$$\tilde{F}(z) - \tilde{F}(\tilde{\omega}) \leqslant 0.33(\delta\alpha)^{1/2} \leqslant 0.43\delta\sqrt{\tilde{\mu}(\tilde{\omega})}$$

and

$$\tilde{F}(\tilde{\omega}) - F(\omega) \geqslant \tfrac{1}{5}\alpha^{1/2}.$$

Theorem 3. *Let* $A = (\tilde{A}\ c)^{\mathrm{T}}$, $b = (\tilde{b}\ \beta)^{\mathrm{T}}$ *and* $P = \tilde{P} \cap \{x\colon c^{\mathrm{T}}x \geqslant \beta\}$. *Let* $z \in P$, *let* $\alpha \leqslant \delta \leqslant 10^{-4}$, *and let* $F(z) - F(\omega) \leqslant \alpha^2\mu(\omega)$. *Suppose that* $c^{\mathrm{T}}H(z)^{-1}c/(c^{\mathrm{T}}z - \beta)^2 \leqslant \min\{\alpha, \mu(z)\}$. *Then the polytope* $\tilde{P}$ *is bounded,*

$$\tilde{F}(z) - \tilde{F}(\tilde{\omega}) \leqslant \min\left\{4\alpha\tilde{\mu}(\tilde{\omega}),\ \delta\sqrt{\tilde{\mu}(\tilde{\omega})}\right\},$$

and

$$F(\omega) - \tilde{F}(\tilde{\omega}) \leqslant 5\alpha.$$

The proofs of the three theorems will be given in Section 8. We shall next give a proof of the Convergence Lemma introduced in Section 2 based on these theorems.

Convergence Lemma. *Let* $\delta \leqslant 10^{-4}$, *let* $\varepsilon \leqslant 10^{-3}\delta$, *and let* ρ^k *denote the value of* $F(\omega)$ *at the beginning of the kth iteration. Then at the beginning of each iteration* z *satisfies the condition*

$$F(z) - F(\omega) \leqslant \varepsilon^4\mu(\omega).$$

Furthermore, the following statements hold.

1. *If Case* 1 *occurs during the kth iteration then*

$$\rho^{k+1} - \rho^k \geqslant \tfrac{1}{5}(\delta\varepsilon)^{1/2}.$$

2. *If Case* 2 *occurs during the kth iteration then*

$$\rho^k - \rho^{k+1} \leqslant 5\varepsilon.$$

Proof. We shall prove the Lemma by induction. Suppose at the beginning of the kth iteration we have a point z such that $F(z) - F(\omega) \leqslant \varepsilon^4\mu(\omega)$. The computation performed during the iteration depends on the value of $\min_{1 \leqslant i \leqslant m}\{\sigma_i(z)\}$.

Case 1. $\min_{1 \leqslant i \leqslant m}\{\sigma_i(z)\} \geqslant \varepsilon$. In this case the algorithm halts if z is feasible; otherwise the oracle returns a vector $c \in \mathbb{R}^n$ and we compute a β such that

$$\frac{c^{\mathrm{T}}H(z)^{-1}c}{\left(c^{\mathrm{T}}z - \beta\right)^2} = \tfrac{1}{2}(\delta\varepsilon)^{1/2}.$$

Let $\tilde{A} = (A\ c)^{\mathrm{T}}$ and $\tilde{b} = (b\ \beta)^{\mathrm{T}}$. The polytope $\tilde{P} = P \cap \{x\colon c^{\mathrm{T}}x \geqslant \beta\}$. From the definition of $\mu(z)$ we have that

$$\min_{1 \leqslant i \leqslant m}\{\sigma_i(z)\} \leqslant \mu(z).$$

So in this case

$$\varepsilon \leqslant \min_{1 \leqslant i \leqslant m} \{\sigma_i(z)\} \leqslant \mu(z).$$

and

$$\delta\varepsilon \leqslant \delta\mu(z).$$

Thus the conditions of Theorem 2 are satisfied with $\alpha = \delta\varepsilon$ and we may conclude that

$$\tilde{F}(\tilde{\omega}) - F(\omega) \geqslant \tfrac{1}{5}(\delta\varepsilon)^{1/2}$$

and

$$\tilde{F}(z) - \tilde{F}(\tilde{\omega}) \leqslant 0.33(\delta^2\varepsilon)^{1/2} \leqslant 0.43\delta\sqrt{\tilde{\mu}(\tilde{\omega})}\,.$$

During the iteration A and b are reset to $\tilde{A}$ and $\tilde{b}$, a few Newton steps are then performed to move closer to the new volumetric center and the iteration ends. So from the definition of ρ^k we can conclude that

$$\rho^{k+1} - \rho^k \geqslant \tfrac{1}{5}(\delta\varepsilon)^{1/2}.$$

After A and b have been reset the point z satisfies the condition

$$F(z) - F(\omega) \leqslant 0.33(\delta^2\varepsilon)^{1/2} \leqslant 0.43\delta\sqrt{\mu(\omega)}\,,$$

and from this condition we have that

$$\varepsilon \leqslant 1.8\mu(\omega).$$

The Newton steps to move closer to the new volumetric center are as follows.

For $j = 1$ to $\lceil 30 \ln(2\varepsilon^{-4.5}) \rceil$ do $z \leftarrow z - 0.18Q(z)^{-1}\nabla F(z)$.

By Theorem 1, it follows that if $F(z) - F(\omega) \leqslant \delta\sqrt{\mu(\omega)}$ then

$$F\big(z - 0.18Q(z)^{-1}\nabla F(z)\big) - F(\omega) \leqslant 0.94(F(z) - F(\omega)).$$

Hence after the Newton steps the resulting point z satisfies

$$F(z) - F(\omega) \leqslant 0.5\varepsilon^5 \leqslant \varepsilon^4\mu(\omega)$$

as required.

Case 2. $\min_{1 \leqslant i \leqslant m}\{\sigma_i(z)\} < \varepsilon$. In this case we remove a plane from the polytope P. W.l.o.g. suppose that $\sigma_m(z) = \min_{1 \leqslant i \leqslant m}\{\sigma_i(z)\}$. Let $a_m = c$, $b_m = \beta$, $A = (\tilde{A}\ c)^{\mathrm{T}}$, and $b = (\tilde{b}\ \beta)^{\mathrm{T}}$. The polytope $P = \tilde{P} \cap \{x\colon c^{\mathrm{T}}x \geqslant \beta\}$. Note that by the definition of $\mu(z)$

$$\min_{1 \leqslant i \leqslant m} \{\sigma_i(z)\} \leqslant \mu(z).$$

So in this case we have that

$$\frac{c^{\mathrm{T}}H(z)^{-1}c}{(c^{\mathrm{T}}z - \beta)^2} = \sigma_m(z) = \min_{1 \leqslant i \leqslant m} \{\sigma_i(z)\} \leqslant \min\{\varepsilon, \mu(z)\}.$$

Thus the conditions of Theorem 3 are satisfied with $\alpha = \varepsilon$ and we may conclude that

$$F(\omega) - \tilde{F}(\tilde{\omega}) \leqslant 5\varepsilon,$$

and that

$$\tilde{F}(z) - \tilde{F}(\tilde{\omega}) \leqslant \min\left\{4\varepsilon\tilde{\mu}(\nabla\omega),\ \delta\sqrt{\tilde{\mu}(\tilde{\omega})}\right\}.$$

During the iteration A and b are reset to $\tilde{A}$ and $\tilde{b}$ respectively, a few Newton steps are then performed to move closer to the new volumetric center and the iteration ends. So by the definition of ρ^k it follows that

$$\rho^k - \rho^{k-1} \leqslant 5\varepsilon.$$

After A and b have been reset z satisfies the condition

$$F(z) - F(\omega) \leqslant \min\left\{4\varepsilon\mu(\omega),\ \delta\sqrt{\mu(\omega)}\right\}.$$

The Newton steps for moving closer to the new volumetric center are as follows.

For $j = 1$ to $\lceil 30\ \ln(4\varepsilon^{-3})\rceil$ do $z \leftarrow z - 0.18Q(z)^{-1}\nabla F(z)$.

From Theorem 1, it follows that if z is a point such that $F(z) - F(\omega) \leqslant \delta\sqrt{\mu(\omega)}$ then

$$F\left(z - 0.18Q(z)^{-1}\nabla F(z)\right) - F(\omega) \leqslant 0.94(F(z) - F(\omega)).$$

Hence after the Newton steps have been performed the resulting point z satisfies the condition

$$F(z) - F(\omega) \leqslant \varepsilon^4\mu(\omega)$$

as required. □

4. Modifying the algorithm to solve the convex optimization problem

We shall briefly discuss how to modify the algorithm for the feasibility problem and thereby obtain an algorithm for the solution of the convex optimization problem

$$\min\ g(x)$$
$$\text{s.t. } x \in S,$$

where $g(x)$ is convex. Furthermore, g is such that given a point z in the domain of g we can compute a vector c such that $\{x\colon g(x) \leqslant g(z)\} \subseteq \{x\colon c^{\mathrm{T}}x \geqslant c^{\mathrm{T}}z\}$. Let x^{opt} minimize $g(x)$ over S. We shall assume that s is contained in a ball of radius 2^L centered at the origin and that the set $\{x\colon x \in S,\ g(x) - g(x^{\text{opt}}) \leqslant \gamma\}$ contains a ball of radius 2^{-L}. The output of the algorithm will be a point $x^* \in S$ such that $g(x^*) - g(x^{\text{opt}}) \leqslant \gamma$.

The modification is as follows. The algorithm for the optimization problem proceeds exactly as the algorithm for the feasibility problem in Section 2 except when the current point z is found to be feasible in Case 1 during an iteration. Then instead of halting, we compute a vector c such that $\{x\colon g(x) \leqslant g(z)\} \subseteq \{x\colon c^{\mathrm{T}}x \geqslant c^{\mathrm{T}}z\}$ and use this vector in place of the vector that would have been returned by the oracle had z not been feasible. Once c is available, the remaining computations during the iteration are exactly the same as in the feasibility algorithm. Moreover, among all the feasible points z encountered thus far we maintain the one with the lowest value of the objective function $g(x)$; let z^* be this point.

The modified algorithm halts when the volume of P falls below 2^{-nL}. Suppose none of the points z encountered in the algorithm are feasible or if z^* exists then it is such that $g(z^*) - g(x^{\text{opt}}) > \gamma$. Then it is easily seen that each constraint $c^{\mathrm{T}}x \geqslant \beta$ that is added to the polytope P during the course of the modified algorithm is such that

$$\{x\colon\ x \in S,\ g(x) - g(x^{\text{opt}}) \leqslant \gamma\} \subseteq \{x\colon\ c^{\mathrm{T}}x \geqslant c^{\mathrm{T}}z\} \subseteq \{x\colon\ c^{\mathrm{T}}x \geqslant \beta\}$$

and so when the modified algorithm halts

$$\{x\colon\ x \in S,\ g(x) - g(x^{\text{opt}}) \leqslant \gamma\} \subseteq P.$$

This means that when the modified algorithm halts P contains a ball of radius 2^{-L} and this cannot happen. So z^* exists and serves as the required output point x^*.

5. Variants of the algorithm

The algorithm in Section 2 is designed to obtain the best worst case time complexity. But an algorithm that has the best worst case running time may not necessarily be the one that gives the best performance in practice. Moreover, in many cases extra information may be available about the oracle and the set S, and this information may be utilized to obtain a better algorithm. Building on the ideas in the basic algorithm we can construct a wide variety of algorithms for the solution of convex programming problems (or related feasibility problems.) This will give us the flexibility of being able to design algorithms that suit the given problem and possibly the given computational facilities. It will also enable us to exploit any special structure in the set of constraints if any.

Several variants of the basic algorithm are possible. One possibility is to cut the current polytope very close to the volumetric center ω during each iteration instead of taking shallow cuts as in the algorithm in Section 2. Another possibility is to keep on adding planes generated by the oracle without ever removing any plane (i.e., discard Case 2 from the algorithm); such an algorithm would converge in $O(n^2L^2)$ iterations since by Theorem 2 (Section 3, with $\alpha = \delta/4m$) the value of $F(\omega)$ would increase by $\Omega(1/\sqrt{m})$ at each iteration. In the remainder of the section, as a sample we shall describe two more ways of obtaining variants of the basic algorithm.

The volumetric center as a weighted analytic center

Variants of the algorithm in Section 2 can be constructed by interpreting the volumetric center as a weighted analytic center. The weighted analytic center $\pi(w)$ of the polytope P is the point that minimizes the weighted logarithmic barrier function

$$\text{logbar}(w, x) = -\sum_{i=1}^{m} w_i \ln\left(a_i^{\mathrm{T}}x - b_i\right)$$

over P where w_i, $1 \leqslant i \leqslant m$, are positive weights. (w_i is the weight on the plane $a_i^{\mathrm{T}}x = b_i$.) The gradient of the weighted logarithmic barrier is given by

$$\nabla\,\text{logbar}(w, x) = -\sum_{i=1}^{m} w_i \frac{a_i}{a_i^{\mathrm{T}}x - b_i}.$$

Comparing this with

$$\nabla F(x) = -\sum_{i=1}^{m} \sigma_i(x) \frac{a_i}{a_i^{\mathrm{T}} x - b_i},$$

we get that the volumetric center ω is the minimizer of the weighted logarithmic barrier $-\sum_{i=1}^{m} \sigma_i(\omega) \ln(a_i^{\mathrm{T}} x - b_i)$.

In the basic algorithm in Section 2 the volumetric center of P is used as a test point. The idea is to use a weighted analytic center as a test point instead of the volumetric center. The weights $\sigma_i(x)$ would guide the choice of the weights w_i. One possibility is to use a weighted analytic center $\pi(w)$ such that the weights w_i satisfy the condition $\alpha_1 \sigma_i(\pi(w)) \leqslant w_i \leqslant \alpha_2 \sigma_i(\pi(w))$, $1 \leqslant i \leqslant m$, where α_1, α_2 are some constants. The key point is to ensure that the weighted analytic center $\pi(w)$ does not lie close to a plane with a small weight on it. One important reason for looking for variants along these lines is as follows. The main computational effort in the basic algorithm (except for querying the oracle) is in computing the weights $\sigma_i(x)$; so if one can design an algorithm where it suffices to compute coarse approximations to the weights $\sigma_i(x)$ then it could lead to a better running time in theory and/or practice. Even better would be an algorithm that somehow uses these weights implicitly and does not require their explicit computation.

Combination of determinant barrier and logarithmic barrier

Here we consider a special convex programming problem; specifically, that of minimizing a convex function $g(x)$ over the polytope P. We want to solve

$$\begin{aligned} &\min\ g(x) \\ &\text{s.t. } x \in P, \end{aligned}$$

where $g(x)$ is a differentiable convex function. An iterative algorithm for the solution of this problem is as follows. During the kth iteration we choose a test point $z(k)$ in P and compute a vector $c(k)$ (by differentiating $g(x)$ at $z(k)$) such that

$$\{x: g(x) \leqslant g(z(k))\} \subseteq \{x: c(k)^{\mathrm{T}} x \geqslant c(k)^{\mathrm{T}} z(k)\},$$

and compute a suitable $\beta(k)$ such that $c(k)^{\mathrm{T}} z(k) > \beta(k)$. Let

$$B(k, x)) = rI + \sum_{i=1}^{k-1} \frac{c(k) c(k)^{\mathrm{T}}}{\left(c(k)^{\mathrm{T}} x - \beta(k)\right)^2},$$

where $r > 0$ is a suitable fixed scale factor, and let

$$\psi(k, x) = \ln(\det(B(k, x))) - \sum_{i=1}^{m} \ln\left(a_i^{\mathrm{T}} x - b_i\right).$$

The test point $z(k)$ is chosen to be the minimizer (or a good approximation to the minimizer) of $\psi(k, x)$ over the polytope $P \cap \{x: c(j)^{\mathrm{T}} x \geqslant \beta(j), 1 \leqslant j \leqslant k-1\}$. $\psi(k, x)$ consists of a determinant barrier together with the logarithmic barrier for P; the determinant component pushes $z(k)$ towards decreasing values of $g(x)$ and the logarithmic barrier keeps $z(k)$ away from the boundaries of P.

6. Linear programming via path of volumetric centers

Consider the linear programming problem

$$\max\ c^{\mathrm{T}}x$$
$$\text{s.t. } x \in P.$$

Various known interior point algorithms for this problem follow the path of analytic centers to the optimum [6,7]. (The analytic center is the weighted analytic center with each of the weights w_i equal to 1.) Instead we can design an algorithm that follows the path of volumetric centers. The path of volumetric centers is defined by the equation

$$\nabla F(x) = tc, \quad t \in \mathbb{R},\ t \geqslant 0.$$

It is the set of all points in the polytope P where the gradient of $F(x)$ is a non-negative multiple of the cost vector c. Such an algorithm would start from the volumetric center and follow the path of volumetric centers using Newton–Raphson steps in a manner similar to the algorithms that follow the path of analytic centers [6,7].

Another possibility is to follow a path of hybrid centers. The path of hybrid centers is defined by

$$\nabla F(x) + r\nabla \operatorname{logbar}(e, x) = tc, \quad t \in \mathbb{R},\ t \geqslant 0.$$

where $e \in \mathbb{R}^m$ is the vector of all ones, and r is a fixed positive constant. Note that logbar(e, x) is just the logarithmic barrier for P; so the hybrid center may be thought of as a combination of the analytic center and the volumetric center. The author has obtained an algorithm that follows a path of hybrid centers (with $r = n/m$) and solves linear programming problems in $O((mn)^{1/4}L)$ iterations; each iteration is a Newton–Raphson step and involves inverting a matrix and solving a system of linear equations. (Here L is a standard parameter; for a definition of L see [7].) This improves on the previously best known bound of $O(\sqrt{m}\,L)$ iterations [6] when $n = O(m)$. Details and a complete presentation will be given in a subsequent paper.

7. Properties of $F(x)$

In this section we shall study the function $F(x)$. We shall first collect together some notation. Let P be the polytope

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

where $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$ and $x \in \mathbb{R}^n$. We assume that P is full dimensional and bounded; we will be interested only in points that lie in the interior of P, denoted Interior(P). Let a_i^{T} denote the ith row of A. Let $H(x)$ be defined as

$$H(x) = \sum_{i=1}^{m} \frac{a_i a_i^{\mathrm{T}}}{\left(a_i^{\mathrm{T}}x - b_i\right)^2}.$$

$H(x)$ is the Hessian of the logarithmic barrier function $\sum_{i=1}^{m} -\ln(a_i^{\mathrm{T}}x - b_i)$ and is positive definite for all x in the interior of P. Let $F(x)$ be the function

$$F(x) = \tfrac{1}{2}\ln(\det(H(x))),$$

where $\det(H(x))$ denotes the determinant of $H(x)$, and let the volumetric center ω be the point that minimizes $F(x)$ over the polytope P.

For a function $\Psi(x)$ we let $\nabla\Psi(x)$ denote the gradient of $\Psi(x)$ evaluated at x, and let $\nabla^2\Psi(x)$ denote the Hessian of $\Psi(x)$ evaluated at x. So $\nabla F(x)$ ($\nabla^2 F(x)$) denotes the gradient (Hessian) of $F(x)$ evaluated at x. Let

$$\sigma_i(x) = \frac{a_i^{\mathrm{T}} H(x)^{-1} a_i}{\left(a_i^{\mathrm{T}}x - b_i\right)^2}, \quad 1 \leqslant i \leqslant m,$$

and let $Q(x)$ be defined as

$$Q(x) = \sum_{i=1}^{m} \sigma_i(x) \frac{a_i a_i^{\mathrm{T}}}{\left(a_i^{\mathrm{T}}x - b_i\right)^2}.$$

Note that $Q(x)$ is positive definite over the interior of P. We shall show that $Q(x)$ is a good approximation to $\nabla^2 F(x)$; specifically, the quadratic forms $\xi^{\mathrm{T}}\nabla^2 F(x)\xi$ and $\xi^{\mathrm{T}}Q(x)\xi$ will be shown to satisfy the condition

$$\forall\xi \in \mathbb{R}^n, \quad 5\xi^{\mathrm{T}}Q(x)\xi \geqslant \xi^{\mathrm{T}}\nabla^2 F(x)\xi \geqslant \xi^{\mathrm{T}}Q(x)\xi.$$

Since $Q(x)$ is positive definite this condition implies that $F(x)$ is a strictly convex function over the interior of P.

Let μ be the largest number λ satisfying the condition that

$$\forall\xi \in \mathbb{R}^n, \quad \xi^{\mathrm{T}}Q(x)\xi \geqslant \lambda\xi^{\mathrm{T}}H(x)\xi.$$

We shall show that

$$1 \geqslant \mu(x) \geqslant 1/(4m).$$

Let $\Sigma(x, r)$ be the region

$$\Sigma(x, r) = \left\{ y: \forall i,\ 1 \leqslant i \leqslant m,\ 1 - r \leqslant \frac{a_i^{\mathrm{T}}y - b_i}{a_i^{\mathrm{T}}x - b_i} \leqslant 1 + r \right\}.$$

Note that if $r \leqslant 1$ then $\Sigma(x, r) \subseteq P$.

For a symmetric positive definite $n \times n$ matrix B, we shall let $E(B, x, r)$ denote the ellipsoid given by

$$E(B, x, r) = \left\{ y: (y - x)^{\mathrm{T}}B(y - x) \leqslant r^2 \right\}.$$

We shall show that

$$E\left(Q(x), x, (\mu(x))^{1/4} r\right) \subseteq \Sigma(x, r).$$

Lemmas 1 through 10 below summarize some of the properties of $F(x)$. We shall prove these lemmas in Section 7.2. Lemmas 1 and 2 give explicit formulae for the gradient and the Hessian of $F(x)$ respectively.

Lemma 1.

$$\nabla F(x) = -\sum_{i=1}^{m} \frac{a_i^{\mathrm{T}} H(x)^{-1} a_i}{\left(a_i^{\mathrm{T}} x - b_i\right)^2} \frac{a_i}{a_i^{\mathrm{T}} x - b_i} = -\sum_{i=1}^{m} \sigma_i(x) \frac{a_i}{a_i^{\mathrm{T}} x - b_i}.$$

Lemma 2. *Let* $u_{ij} = a_i/(a_i^{\mathrm{T}} x - b_i) - a_j/(a_j^{\mathrm{T}} x - b_j)$. *Then*

$$\nabla^2 F(x) = Q(x) + 2 \sum_{1 \leqslant i \leqslant j \leqslant m} \frac{\left(a_i^{\mathrm{T}} H(x)^{-1} a_j\right)^2}{\left(a_i^{\mathrm{T}} x - b_i\right)^2 \left(a_j^{\mathrm{T}} x - b_j\right)^2} u_{ij} u_{ij}^{\mathrm{T}}.$$

Lemma 3 states that $Q(x)$ serves as a good approximation to $\nabla^2 F(x)$.

Lemma 3. *The matrices* $Q(x)$ *and* $\nabla^2 F(x)$ *satisfy the condition*

$$\forall \xi \in \mathbb{R}^n, \quad 5\xi^{\mathrm{T}} Q(x) \xi \geqslant \xi^{\mathrm{T}} \nabla^2 F(x) \xi \geqslant \xi^{\mathrm{T}} Q(x) \xi.$$

Hence, $\nabla^2 F(x)$ *is positive definite and* $F(x)$ *is strictly convex over the interior of* P.

Lemma 4 states that the value of the quadratic form $\xi^{\mathrm{T}} Q(x) \xi$ does not deviate too far from the value of the quadratic form $\xi^{\mathrm{T}} H(x) \xi$ and gives bounds on $\mu(x)$.

Lemma 4.

$$\forall \xi \in \mathbb{R}^n, \quad \xi^{\mathrm{T}} H(x) \xi \geqslant \xi^{\mathrm{T}} Q(x) \xi \geqslant \frac{1}{4m} \xi^{\mathrm{T}} H(x) \xi,$$

and thus

$$1 \geqslant \mu(x) \geqslant 1/(4m).$$

Lemma 5 formalizes the observation that for all x in $\Sigma(\hat{x}, r)$ the quadratic form $\xi^{\mathrm{T}} Q(x) \xi$ does not deviate too far from the quadratic form $\xi^{\mathrm{T}} Q(\hat{x}) \xi$ if r is less than some small constant.

Lemma 5. *Suppose that* $r < 1$ *and that* $x \in \Sigma(\hat{x}, r)$. *Then for all* $\xi \in \mathbb{R}^n$,

$$\frac{(1-r)^2}{(1+r)^4} \xi^{\mathrm{T}} Q(\hat{x}) \xi \leqslant \xi^{\mathrm{T}} Q(x) \xi \leqslant \frac{(1+r)^2}{(1-r)^4} \xi^{\mathrm{T}} Q(\hat{x}) \xi$$

and

$$\frac{(1-r)^4}{(1+r)^4} \mu(\hat{x}) \leqslant \mu(x) \leqslant \frac{(1+r)^4}{(1-r)^4} \mu(\hat{x}).$$

Consider the equation $\nabla F(x) = tw$ where t is a scalar and w is a fixed n-dimensional vector. This equation implicitly defines x as a function of t and Lemma 6 summarizes some of the properties of this implicitly defined function that can be derived from the implicit function theorem [1,2].

Lemma 6. *Let $\Phi(x, t) = \nabla F(x) - tw$ where $t \in \mathbb{R}$ and w is a fixed vector in $\mathbb{R}^n$. Then the equation $\Phi(x, t) = 0$ implicitly defines x as a function of t, and we may write $x = x(t)$. Moreover, $x(t)$ is an analytic function of t, and $\mathrm{d}x(t)/\mathrm{d}t$, the derivative of $x(t)$ w.r.t. t evaluated at t, may be written as*

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = \nabla^2 F(x)^{-1} w,$$

and if $0 \leqslant t_1 \leqslant t_2$ then

$$\tfrac{1}{5}\int_{t_1}^{t_2} tw^{\mathrm{T}} Q(x)^{-1} w \,\mathrm{d}t \leqslant F(x(t_2)) - F(x(t_1)) \leqslant \int_{t_1}^{t_2} tw^{\mathrm{T}} Q(x)^{-1} w \,\mathrm{d}t.$$

Consider the trajectory $\nabla F(x) = tw$, where $t \in \mathbb{R}$ and w is fixed, that passes through $\hat{x}$. Lemma 7 gives an upper bound on the derivative of $\ln(a_i^{\mathrm{T}} x(t) - b_i)$ w.r.t. t for the portion of this trajectory in $\Sigma(\hat{x}, r)$ in terms of quantities evaluated at $\hat{x}$. Lemma 8 gives a lower bound on how much t must change before the trajectory reaches the boundary of $\Sigma(\hat{x}, r)$.

Lemma 7. *Let w be a fixed vector in $\mathbb{R}^n$, and let $\hat{x}$ be such that $\nabla F(\hat{x}) = \hat{t}w$ for some scalar $\hat{t}$. Let $t \in \mathbb{R}$ and let $x = x(t)$ be a point on the trajectory $\nabla F(x) = tw$ such that $x \in \Sigma(\hat{x}, r)$, $r < 1$. Then for $1 \leqslant i \leqslant m$,*

$$\left| \frac{a_i^{\mathrm{T}}(\mathrm{d}x(t)/\mathrm{d}t)}{a_i^{\mathrm{T}} x - b_i} \right| \leqslant \frac{(1+r)^3}{(1-r)^2} \frac{\left(w^{\mathrm{T}} Q(\hat{x})^{-1} w\right)^{1/2}}{(\mu(\hat{x}))^{1/4}}.$$

Lemma 8. *Let $r < 1$, let w be a fixed vector in $\mathbb{R}^n$, and let $\hat{x}$ be such that $\nabla F(\hat{x}) = \hat{t}w$ for some scalar $\hat{t}$. Let $x(\bar{t})$ be a point on the trajectory $\nabla F(x) = tw$, $t \in \mathbb{R}$, such that $x(\bar{t})$ does not lies in the interior of $\Sigma(\hat{x}, r)$. Then*

$$|\hat{t} - \bar{t}| \geqslant \frac{\left(r - \frac{1}{2}r^2\right)(1-r)^2(\mu(\hat{x}))^{1/4}}{(1+r)^3 \sqrt{w^{\mathrm{T}} Q(\hat{x})^{-1} w}}.$$

Lemma 9 gives a sufficient condition in terms of $\nabla F(z)^{\mathrm{T}} Q(z)^{-1} \nabla F(z)$ and $\mu(z)$ for the point z to be in the region $\Sigma(\omega, r)$.

Lemma 9. *Let $\delta \leqslant 10^{-4}$, let $z \in P$ and suppose that $\nabla F(z)^{\mathrm{T}} Q(z)^{-1} \nabla F(z) \leqslant \delta\sqrt{\mu(z)}$. Then $\omega \in \Sigma(z, 1.1\sqrt{\delta})$, $\mu(z) \leqslant 1.1\mu(\omega)$, and*

$$F(z) - F(\omega) \leqslant 0.55\, \nabla F(z)^{\mathrm{T}} Q(z)^{-1} \nabla F(z).$$

Lemma 10 states that if $F(z) - F(\omega)$ is small then the quantities $F(z) - F(\omega)$ and $\nabla F(z)^{\mathrm{T}} Q(z)^{-1} \nabla F(z)$ are closely related.

Lemma 10. *Let $\delta \leqslant 10^{-4}$ and let z be a point in P such that $F(z) - F(\omega) \leqslant \delta\sqrt{\mu(\omega)}$. Then $z \in \Sigma(\omega, 5\sqrt{\delta})$, $\mu(\omega) \leqslant 1.5\mu(z)$, and*

$$0.14\, \nabla F(z)^{\mathrm{T}} Q(z)^{-1} \nabla F(z) \leqslant F(z) - F(\omega) \leqslant 1.4\, \nabla F(z)^{\mathrm{T}} Q(z)^{-1} \nabla F(z).$$

7.1. Some useful claims

In this section we shall state and prove four claims which will be used in the proofs of various lemmas and theorems.

Claim 1. *Let B be an $n \times n$ symmetric positive definite matrix and let $E(B, x, r) = \{y: (y-x)^{\mathrm{T}}B(y-x) \leqslant r^2\}$. Let w be an arbitrary fixed vector in $\mathbb{R}^n$. Then*

$$\max_{y \in E(B,x,r)} \left\{ \left(w^{\mathrm{T}}(y-x)\right)^2 \right\} = r^2 w^{\mathrm{T}} B^{-1} w.$$

Claim 2. *Let $\theta > 0$, and let B_1, B_2 be $n \times n$ positive definite matrices. Then*

$$\forall \xi \in \mathbb{R}^n,\ \xi^{\mathrm{T}} B_1 \xi \geqslant \theta \xi^{\mathrm{T}} B_2 \xi \quad \Rightarrow \quad \forall \xi \in \mathbb{R}^n,\ \xi^{\mathrm{T}} B_1^{-1} \xi \leqslant \frac{1}{\theta} \xi^{\mathrm{T}} B_2^{-1} \xi.$$

Claim 3.

$$\sigma_i(x) = \sum_{j=1}^{m} \frac{\left(a_i^{\mathrm{T}} H(x)^{-1} a_j\right)^2}{\left(a_i^{\mathrm{T}} x - b_i\right)^2 \left(a_j^{\mathrm{T}} x - b_j\right)^2},$$

and $\sigma_i(x) \leqslant 1$, $1 \leqslant i \leqslant m$. Moreover, $\sum_{i=1}^{m} \sigma_i(x) = n$.

Claim 4.

$$\frac{a_i^{\mathrm{T}} Q(x)^{-1} a_i}{\left(a_i^{\mathrm{T}} x - b_i\right)^2} \leqslant \frac{1}{\sqrt{\mu(x)}}, \qquad 1 \leqslant i \leqslant m,$$

and thus

$$E\left(Q(x), x, (\mu(x))^{1/4} r\right) \subseteq \Sigma(x, r).$$

Proof of Claim 1. Let x^{opt} be the point that maximizes the linear function $w^{\mathrm{T}} y$ over the ellipsoid $E(B, x, r)$. From the Karush–Kuhn–Tucker conditions [5] it follows that

$$B(x^{\mathrm{opt}} - x) = tw,$$

where t is a scalar, and since x^{opt} lies on the boundary of $E(B, x, r)$ we can write x^{opt} as

$$x_{\mathrm{opt}} = x + \frac{r}{\sqrt{w^{\mathrm{T}} B^{-1} w}} B^{-1} w.$$

Thus

$$w^{\mathrm{T}}(x^{\mathrm{opt}} - x) = r\sqrt{w^{\mathrm{T}} B^{-1} w}$$

and the claim then follows. □

Proof of Claim 2. Suppose that $\forall \xi \in \mathbb{R}^n$, $\xi^{\mathrm{T}} B_1 \xi \geqslant \theta \xi^{\mathrm{T}} B_2 \xi$. Then

$$\begin{aligned} \xi^{\mathrm{T}} B_1 \xi \leqslant 1 \quad &\Rightarrow \quad \theta \xi^{\mathrm{T}} B_2 \xi \leqslant 1 \\ &\Rightarrow \quad \xi^{\mathrm{T}} B_2 \xi \leqslant 1/\theta. \end{aligned}$$

Thus

$$E(B_1, 0, 1) \subseteq E\left(B_2, 0, 1/\sqrt{\theta}\right)$$

and hence for $w \in \mathbb{R}^n$,

$$\max_{\xi \in E(B_1, 0, 1)} \left\{ \left(w^{\mathrm{T}} \xi \right)^2 \right\} \leqslant \max_{\xi \in E(B_2, 0, 1/\sqrt{\theta})} \left\{ \left(w^{\mathrm{T}} \xi \right)^2 \right\}.$$

So by Claim 1,

$$w^{\mathrm{T}} B_1^{-1} w \leqslant \frac{1}{\theta} w^{\mathrm{T}} B_2^{-1} w.$$

Claim 2 then follows. □

Proof of Claim 3. Note that $H(x)^{-1}$ may be written as

$$H(x)^{-1} = H(x)^{-1} H(x) H(x)^{-1} = \sum_{j=1}^{m} \frac{H(x)^{-1} a_j a_j^{\mathrm{T}} H(x)^{-1}}{\left(a_j^{\mathrm{T}} x - b_j\right)^2}.$$

Thus we may write $\sigma_i(x)$ as

$$\sigma_i(x) = \frac{a_i^{\mathrm{T}} H(x)^{-1} a_i}{\left(a_i^{\mathrm{T}} x - b_i\right)^2} = \sum_{j=1}^{m} \frac{\left(a_i^{\mathrm{T}} H(x)^{-1} a_j\right)^2}{\left(a_i^{\mathrm{T}} x - b_i\right)^2 \left(a_j^{\mathrm{T}} x - b_j\right)^2}.$$

Then $\sigma_i(x) \leqslant 1$ is seen as follows. First note that the ellipsoid $E(H(x), x, 1)$ is contained in the polytope P. So, by Claim 1,

$$a_i^{\mathrm{T}} H(x)^{-1} a_i = \max_{y \in E(H(x), x, 1)} \left\{ \left(a_i^{\mathrm{T}} (y - x) \right)^2 \right\} \leqslant \left(a_i^{\mathrm{T}} x - b_i \right)^2.$$

Finally,

$$\sum_{i=1}^{m} \sigma_i(x) = \operatorname{trace}\left(\sum_{i=1}^{m} \frac{H(x)^{-1/2} a_i a_i^{\mathrm{T}} H(x)^{-1/2}}{\left(a_i^{\mathrm{T}} x - b_i\right)^2} \right) = n,$$

since $H(x)^{-1/2} H(x) H(x)^{-1/2}$ is the identity. □

Proof of Claim 4. From the definition of $\mu(x)$ and Claim 2 it follows that

$$a_i^{\mathrm{T}} Q(x)^{-1} a_i \leqslant \frac{1}{\mu(x)} a_i^{\mathrm{T}} H(x)^{-1} a_i$$

and thus

$$\frac{a_i^{\mathrm{T}} Q(x)^{-1} a_i}{\left(a_i^{\mathrm{T}} x - b_i\right)^2} \leqslant \frac{1}{\mu(x)} \frac{a_i^{\mathrm{T}} H(x)^{-1} a_i}{\left(a_i^{\mathrm{T}} x - b_i\right)^2} = \frac{\sigma_i(x)}{\mu(x)}.$$

Note that if B may be written as

$$B = \hat{B} + w w^{\mathrm{T}},$$

where $\hat{B}$ is a symmetric positive semi-definite matrix then

$$w^{\mathrm{T}}B^{-1}w \leqslant 1.$$

This is seen as follows.

$$w^{\mathrm{T}}B^{-1}w = w^{\mathrm{T}}B^{-1}BB^{-1}w = w^{\mathrm{T}}B^{-1}\hat{B}B^{-1}w + \left(w^{\mathrm{T}}B^{-1}w\right)^2$$

and hence

$$w^{\mathrm{T}}B^{-1}w(1 - w^{\mathrm{T}}B^{-1}w) = w^{\mathrm{T}}B^{-1}\hat{B}B^{-1}w \geqslant 0.$$

So $w^{\mathrm{T}}B^{-1}w \leqslant 1$. Since $Q(x)$ may be expressed as

$$Q(x) = \hat{Q}(x) + \sigma_i(x)\frac{a_i a_i^{\mathrm{T}}}{\left(a_i^{\mathrm{T}}x - b_i\right)^2},$$

where $\hat{Q}(x)$ is positive semi-definite we can conclude that

$$\sigma_i(x)\frac{a_i^{\mathrm{T}}Q(x)^{-1}a_i}{\left(a_i^{\mathrm{T}}x - b_i\right)^2} \leqslant 1.$$

It then follows that

$$\frac{a_i^{\mathrm{T}}Q(x)^{-1}a_i}{\left(a_i^{\mathrm{T}}x - b_i\right)^2} \leqslant \min\left\{\frac{1}{\sigma_i(x)}, \frac{\sigma_i(x)}{\mu(x)}\right\}$$

and hence

$$\frac{a_i^{\mathrm{T}}Q(x)^{-1}a_i}{\left(a_i^{\mathrm{T}}x - b_i\right)^2} \leqslant \frac{1}{\sqrt{\mu(x)}}.$$

By Claim 2 we get that

$$\max_{y \in E(Q(x), x, (\mu(x))^{1/4}r)} \left\{\left(a_i^{\mathrm{T}}(y - x)\right)^2\right\} = r^2\sqrt{\mu(x)}\, a_i^{\mathrm{T}}Q(x)^{-1}a_i \leqslant r^2\left(a_i^{\mathrm{T}}x - b_i\right)^2.$$

Thus if $y \in E(Q(x), x, (\mu(x))^{1/4}r)$, then $|a_i^{\mathrm{T}}(y - x)/(a_i^{\mathrm{T}}x - b_i)| \leqslant r$ and $y \in \Sigma(x, r)$. □

7.2. Proofs of Lemmas 1 to 10

In this section we shall prove Lemmas 1 through 10 introduced in Section 7.

Proof of Lemma 1. To prove the lemma it will suffice to show that for all directions

$$\lim_{t \to 0} \frac{F(x + t\xi) - F(x)}{t} = -\sum_{i=1}^{m} \sigma_i(x)\frac{a_i^{\mathrm{T}}\xi}{a_i^{\mathrm{T}}x - b_i}.$$

For small t, we have that

$$\frac{1}{\left(a_i^{\mathrm{T}}(x + t\xi) - b_i\right)^2} = \frac{1}{\left(a_i^{\mathrm{T}}x - b_i\right)^2} - \frac{2ta_i^{\mathrm{T}}\xi}{\left(a_i^{\mathrm{T}}x - b_i\right)^3} + \mathrm{O}(t^2).$$

Let $\Delta = H(x+t\xi) - H(x)$. For small t, we may write Δ as

$$\Delta = -2t\sum_{i=1}^{m} \frac{a_i a_i^{\mathrm{T}}}{\left(a_i^{\mathrm{T}}x - b_i\right)^2}\frac{a_i^{\mathrm{T}}\xi}{a_i^{\mathrm{T}}x - b_i} + \mathrm{O}(t^2), \tag{7.1}$$

and $H(x+t\xi)$ may be written as

$$H(x+t\xi) = H(x) + \Delta = H(x)^{1/2}\left(I + H(x)^{-1/2}\Delta H(x)^{-1/2}\right)H(x)^{1/2}.$$

Thus

$$\begin{aligned} F(x+t\xi) &= \tfrac{1}{2}\ln(\det(H(x+t\xi))) \\ &= F(x) + \tfrac{1}{2}\ln\left(\det\left(I + H(x)^{-1/2}\Delta H(x)^{-1/2}\right)\right). \end{aligned} \tag{7.2}$$

Let $\varepsilon_1, \ldots, \varepsilon_n$ be the eigenvalues of $H(x)^{-1/2}\Delta H(x)^{-1/2}$. Then $|\varepsilon_j| = \mathrm{O}(t)$, $1 \leqslant j \leqslant n$. Moreover, the eigenvalues of $I + H(x)^- 1/2\,\Delta H(x)^{-1/2}$ are $1+\varepsilon_1, \ldots, 1+\varepsilon_n$, and

$$\begin{aligned} \ln\left(\det\left(I + H(x)^{-1/2}\Delta H(x)^{-1/2}\right)\right) &= \sum_{j=1}^{n}\ln(1+\varepsilon_j) \\ &= \sum_{j=1}^{n}\varepsilon_j + \mathrm{O}(t^2) \\ &= \mathrm{trace}\left(H(x)^{-1/2}\Delta H(x)^{-1/2}\right) + \mathrm{O}(t^2). \end{aligned} \tag{7.3}$$

From (7.1) we get that

$$\begin{aligned} &\mathrm{trace}\left(H(x)^{-1/2}\Delta H(x)^{-1/2}\right) \\ &\quad = \mathrm{trace}\left(-2t\sum_{i=1}^{m}\frac{H(x)^{-1/2}a_i a_i^{\mathrm{T}}H(x)^{-1/2}}{\left(a_i^{\mathrm{T}}x - b_i\right)^2}\frac{a_i^{\mathrm{T}}\xi}{a_i^{\mathrm{T}}x - b_i} + \mathrm{O}(t^2)\right) \\ &\quad = -2t\sum_{i=1}^{m}\frac{a_i^{\mathrm{T}}H(x)^{-1}a_i}{\left(a_i^{\mathrm{T}}x - b_i\right)^2}\frac{a_i^{\mathrm{T}}\xi}{a_i^{\mathrm{T}}x - b_i} + \mathrm{O}(t^2) \\ &\quad = -2t\sum_{i=1}^{m}\sigma_i(x)\frac{a_i^{\mathrm{T}}\xi}{a_i^{\mathrm{T}}x - b_i} + \mathrm{O}(t^2). \end{aligned} \tag{7.4}$$

Thus from (7.2) and (7.3) we get that

$$\begin{aligned} F(x+t\xi) &= F(x) + \tfrac{1}{2}\,\mathrm{trace}\left(H(x)^{-1/2}\Delta H(x)^{-1/2}\right) + \mathrm{O}(t^2) \\ &= F(x) - t\sum_{i=1}^{m}\sigma_i(x)\frac{a_i^{\mathrm{T}}\xi}{a_i^{\mathrm{T}}x - b_i} + \mathrm{O}(t^2) \quad (\text{by } 7.4). \end{aligned}$$

Thus

$$\lim_{t\to 0}\frac{F(x+t\xi) - F(x)}{t} = -\sum_{i=1}^{m}\sigma_i(x)\frac{a_i^{\mathrm{T}}\xi}{a_i^{\mathrm{T}}x - b_i}. \quad \square$$

Proof of Lemma 2. From Lemma 1 it follows that

$$\begin{aligned}
&\frac{\partial}{\partial x_l}\frac{\partial}{\partial x_k}F(x)\\
&\quad= -\left[\sum_{i=1}^{m} a_i^{\mathrm{T}}H(x)^{-1}a_i\frac{\partial}{\partial x_l}\left(\frac{a_{ik}}{\left(a_i^{\mathrm{T}}x-b_i\right)^3}\right)+\frac{a_{ik}}{\left(a_i^{\mathrm{T}}x-b_i\right)^3}\frac{\partial}{\partial x_l}a_i^{\mathrm{T}}H(x)^{-1}a_i\right]\\
&\quad= 3\sum_{i=1}^{m}\frac{a_i^{\mathrm{T}}H(x)^{-1}a_i}{\left(a_i^{\mathrm{T}}x-b_i\right)^2}\frac{a_{ik}a_{il}}{\left(a_i^{\mathrm{T}}x-b_i\right)^2}-\sum_{i=1}^{m}\frac{a_{ik}}{\left(a_i^{\mathrm{T}}x-b_i\right)^3}\frac{\partial}{\partial x_l}a_i^{\mathrm{T}}H(x)^{-1}a_i.
\end{aligned}\tag{7.5}$$

We shall show that

$$\frac{\partial}{\partial x_l}a_i^{\mathrm{T}}H(x)^{-1}a_i = 2\sum_{j=1}^{m}\frac{\left(a_i^{\mathrm{T}}H(x)^{-1}a_j\right)^2}{\left(a_j^{\mathrm{T}}x-b_j\right)^2}\frac{a_{jl}}{a_j^{\mathrm{T}}x-b_j}.\tag{7.6}$$

From (7.5) and (7.6) and the definition of $\sigma_i(x)$ we get that

$$\begin{aligned}
\frac{\partial}{\partial x_l}\frac{\partial}{\partial x_k}F(x) &= 3\sum_{i=1}^{m}\sigma_i(x)\frac{a_{ik}a_{il}}{\left(a_i^{\mathrm{T}}x-b_i\right)^2}\\
&\quad-2\sum_{i=1}^{m}\sum_{j=1}^{m}\frac{\left(a_i^{\mathrm{T}}H(x)^{-1}a_j\right)^2}{\left(a_i^{\mathrm{T}}x-b_i\right)^2\left(a_j^{\mathrm{T}}x-b_j\right)^2}\frac{a_{ik}a_{jl}}{\left(a_i^{\mathrm{T}}x-b_i\right)\left(a_j^{\mathrm{T}}x-b_j\right)},
\end{aligned}$$

and thus

$$\begin{aligned}
\nabla^2F(x) &= 3\sum_{i=1}^{m}\sigma_i(x)\frac{a_ia_i^{\mathrm{T}}}{\left(a_i^{\mathrm{T}}x-b_i\right)^2}\\
&\quad-2\sum_{i=1}^{m}\sum_{j=1}^{m}\frac{\left(a_i^{\mathrm{T}}H(x)^{-1}a_j\right)^2}{\left(a_i^{\mathrm{T}}x-b_i\right)^2\left(a_j^{\mathrm{T}}x-b_j\right)^2}\frac{a_ia_j^{\mathrm{T}}}{\left(a_i^{\mathrm{T}}x-b_i\right)\left(a_j^{\mathrm{T}}x-b_j\right)}.
\end{aligned}$$

By Claim 3,

$$\sigma_i(x) = \sum_{j=1}^{m}\frac{\left(a_i^{\mathrm{T}}H(x)^{-1}a_j\right)^2}{\left(a_i^{\mathrm{T}}x-b_i\right)^2\left(a_j^{\mathrm{T}}x-b_j\right)^2},$$

and so we may rewrite $\nabla^2F(x)$ as

$$\nabla^2F(x) = \sum_{i=1}^{m}\sigma_i(x)\frac{a_ia_i^{\mathrm{T}}}{\left(a_i^{\mathrm{T}}x-b_i\right)^2}+2\sum_{i=1}^{m}\sum_{j=1}^{m}\frac{\left(a_i^{\mathrm{T}}H(x)^{-1}a_j\right)^2}{\left(a_i^{\mathrm{T}}x-b_i\right)^2\left(a_j^{\mathrm{T}}x-b_j\right)^2}$$

$$\times\left(\frac{a_i a_i^{\mathrm T}}{\left(a_i^{\mathrm T}x-b_i\right)^2}-\frac{a_i a_j^{\mathrm T}}{\left(a_i^{\mathrm T}x-b_i\right)\left(a_j^{\mathrm T}x-b_j\right)}\right)$$

$$=\sum_{i=1}^{m}\sigma_i(x)\frac{a_i a_i^{\mathrm T}}{\left(a_i^{\mathrm T}x-b_i\right)^2}$$

$$+2\sum_{1\leqslant i<j\leqslant m}\frac{\left(a_i^{\mathrm T}H(x)^{-1}a_j\right)^2}{\left(a_i^{\mathrm T}x-b_i\right)^2\left(a_j^{\mathrm T}x-b_j\right)^2}\left(\frac{a_i}{a_i^{\mathrm T}x-b_i}-\frac{a_j}{a_j^{\mathrm T}x-b_j}\right)$$

$$\times\left(\frac{a_i}{a_i^{\mathrm T}x-b_i}-\frac{a_j}{a_j^{\mathrm T}x-b_j}\right)^{\mathrm T}.$$

Lemma 2 then follows from the definitions of Q and u_{ij}'s.

We shall now prove (7.6). To show (7.6) it will suffice to show that all directions ξ,

$$\lim_{t\to 0}\frac{a_i^{\mathrm T}H(x+t\xi)^{-1}a_i-a_i^{\mathrm T}H(x)^{-1}a_i}{t}=2\sum_{j=1}^{m}\frac{\left(a_i^{\mathrm T}H(x)^{-1}a_j\right)^2}{\left(a_j^{\mathrm T}x-b_j\right)^2}\frac{a_j^{\mathrm T}\xi}{a_j^{\mathrm T}x-b_j}. \tag{7.7}$$

Let $\Delta=H(x+t\xi)-H(x)$. For small t, we have that

$$\frac{1}{\left(a_j^{\mathrm T}(x+t\xi)-b_j\right)^2}=\frac{1}{\left(a_j^{\mathrm T}x-b_j\right)^2}-\frac{2ta_j^{\mathrm T}\xi}{\left(a_j^{\mathrm T}x-b_j\right)^3}+\mathrm{O}(t^2),$$

and Δ may be written as

$$\Delta=-2t\sum_{j=1}^{m}\frac{a_j a_j^{\mathrm T}}{\left(a_j^{\mathrm T}x-b_j\right)^2}\frac{a_j^{\mathrm T}\xi}{a_j^{\mathrm T}x-b_j}+\mathrm{O}(t^2), \tag{7.8}$$

and $H(x+t\xi)^{-1}$ may be expressed as

$$H(x+t\xi)^{-1}=\left(H(x)^{1/2}\left(I+H(x)^{-1/2}\Delta H(x)^{-1/2}\right)H(x)^{1/2}\right)^{-1}$$

$$=H(x)^{-1/2}\left(I+H(x)^{-1/2}\Delta H(x)^{-1/2}\right)^{-1}H(x)^{-1/2}.$$

Since $\| H(x)^{-1/2}\Delta H(x)^{-1/2}\|=\mathrm{O}(t)$ we have that

$$\left(I+H(x)^{-1/2}\Delta H(x)^{-1/2}\right)^{-1}=I-H(x)^{-1/2}\Delta H(x)^{-1/2}+\mathrm{O}(t^2)$$

and hence

$$H(x+t\xi)^{-1}=H(x)^{-1}-H(x)^{-1}\Delta H(x)^{-1}+\mathrm{O}(t^2).$$

Thus

$$a_i^{\mathrm{T}}H(x+t\xi)^{-1}a_i - a_i^{\mathrm{T}}H(x)^{-1}a_i = -a_i^{\mathrm{T}}H(x)^{-1}\Delta H(x)^{-1}a_i + \mathrm{O}(t^2)$$
$$= 2t\sum_{j=1}^{m}\frac{\left(a_i^{\mathrm{T}}H(x)^{-1}a_j\right)^2}{\left(a_j^{\mathrm{T}}x-b_j\right)^2}\,\frac{a_j^{\mathrm{T}}\xi}{a_j^{\mathrm{T}}x-b_j} + \mathrm{O}(t^2) \quad \text{(by 7.8)}.$$

(7.7) then follows. □

Proof of Lemma 3. From Lemma 2 we get that

$$\xi^{\mathrm{T}}\nabla^2F(x)\xi = \xi^{\mathrm{T}}Q(x)\xi + 2\sum_{1\leqslant i<j\leqslant m}\frac{\left(a_i^{\mathrm{T}}H(x)^{-1}a_j\right)^2}{\left(a_i^{\mathrm{T}}x-b_i\right)^2\left(a_j^{\mathrm{T}}x-b_j\right)^2}\left(\xi^{\mathrm{T}}u_{ij}\right)^2, \tag{7.9}$$

where $u_{ij} = a_i/(a_i^{\mathrm{T}}x-b_i) - a_j/(a_j^{\mathrm{T}}x-b_j)$. Thus

$$\xi^{\mathrm{T}}\nabla^2F(x)\xi \geqslant \xi^{\mathrm{T}}Q(x)\xi. \tag{7.10}$$

Also,

$$\left(\xi^{\mathrm{T}}u_{ij}\right)^2 = \frac{\left(a_i^{\mathrm{T}}\xi\right)^2}{\left(a_i^{\mathrm{T}}x-b_i\right)^2} + \frac{\left(a_j^{\mathrm{T}}\xi\right)^2}{\left(a_j^{\mathrm{T}}x-b_j\right)^2} - \frac{2\left(a_i^{\mathrm{T}}\xi\right)\left(a_j^{\mathrm{T}}\xi\right)}{\left(a_i^{\mathrm{T}}x-b_i\right)\left(a_j^{\mathrm{T}}x-b_j\right)}$$
$$\leqslant 2\left(\frac{\left(a_i^{\mathrm{T}}\xi\right)^2}{\left(a_i^{\mathrm{T}}x-b_i\right)^2} + \frac{\left(a_j^{\mathrm{T}}\xi\right)^2}{\left(a_j^{\mathrm{T}}x-b_j\right)^2}\right). \tag{7.11}$$

We have that

$$4\xi^{\mathrm{T}}Q(x)\xi = 4\sum_{i=1}^{m}\sigma_i(x)\frac{\left(a_i^{\mathrm{T}}\xi\right)^2}{\left(a_i^{\mathrm{T}}x-b_i\right)^2}$$
$$= 4\sum_{i=1}^{m}\sum_{j=1}^{m}\frac{\left(a_i^{\mathrm{T}}H(x)^{-1}a_j\right)^2}{\left(a_i^{\mathrm{T}}x-b_i\right)^2\left(a_j^{\mathrm{T}}x-b_j\right)^2}\,\frac{\left(a_i^{\mathrm{T}}\xi\right)^2}{\left(a_i^{\mathrm{T}}x-b_i\right)^2} \quad \text{(by Claim 3)}$$
$$\geqslant 4\sum_{1\leqslant i<j\leqslant m}\frac{\left(a_i^{\mathrm{T}}H(x)^{-1}a_j\right)^2}{\left(a_i^{\mathrm{T}}x-b_i\right)^2\left(a_j^{\mathrm{T}}x-b_j\right)^2}\left(\frac{\left(a_i^{\mathrm{T}}\xi\right)^2}{\left(a_i^{\mathrm{T}}x-b_i\right)^2} + \frac{\left(a_j^{\mathrm{T}}\xi\right)^2}{\left(a_j^{\mathrm{T}}x-b_j\right)^2}\right)$$
$$\geqslant 2\sum_{1\leqslant i<j\leqslant m}\frac{\left(a_i^{\mathrm{T}}H(x)^{-1}a_j\right)^2}{\left(a_i^{\mathrm{T}}x-b_i\right)^2\left(a_j^{\mathrm{T}}x-b_j\right)^2}\left(\xi^{\mathrm{T}}u_{ij}\right)^2 \quad \text{(by (7.11))}.$$

Then from (7.9) we may conclude that $\xi^{\mathrm{T}}\nabla^2F(x)\xi \leqslant 5\xi^{\mathrm{T}}Q(x)\xi$. Lemma 3 then follows from (7.10). □

Proof of Lemma 4. By the definition of $Q(x)$ we have that

$$\xi^{\mathrm{T}} Q(x) \xi = \sum_{i=1}^{m} \sigma_i(x) \frac{\left(a_i^{\mathrm{T}} \xi\right)^2}{\left(a_i^{\mathrm{T}} x - b_i\right)^2}.$$

By Claim 3, $\sigma_i(x) \leqslant 1$, $1 \leqslant i \leqslant m$, and hence

$$\xi^{\mathrm{T}} Q(x) \xi \leqslant \sum_{i=1}^{m} \frac{\left(a_i^{\mathrm{T}} \xi\right)^2}{\left(a_i^{\mathrm{T}} x - b_i\right)^2} = \xi^{\mathrm{T}} H(x) \xi.$$

We shall show that

$$\xi^{\mathrm{T}} Q(x) \xi \geqslant \frac{1}{4m} \xi^{\mathrm{T}} H(x) \xi. \tag{7.12}$$

Since $\mu(x)$ is the largest number λ such that $\forall \xi \in \mathbb{R}^n$, $\xi^{\mathrm{T}} Q(x) \xi \geqslant \lambda \xi^{\mathrm{T}} H(x) \xi$, it will then follow that

$$1 \geqslant \mu(x) \geqslant 1/(4m).$$

We shall next show (7.12). Let

$$\hat{H}(x) = \sum_{\sigma_i(x) \geqslant 1/(2m)} \frac{a_i a_i^{\mathrm{T}}}{\left(a_i^{\mathrm{T}} x - b_i\right)^2}.$$

We have that

$$\begin{aligned} \xi^{\mathrm{T}} Q(x) \xi &\geqslant \sum_{\sigma_i(x) \geqslant 1/(2m)} \sigma_i(x) \frac{\left(a_i^{\mathrm{T}} \xi\right)^2}{\left(a_i^{\mathrm{T}} x - b_i\right)^2} \\ &\geqslant \frac{1}{2m} \sum_{\sigma_i(x) \geqslant 1/(2m)} \frac{\left(a_i^{\mathrm{T}} \xi\right)^2}{\left(a_i^{\mathrm{T}} x - b_i\right)^2} \\ &\geqslant \frac{1}{2m} \xi^{\mathrm{T}} \hat{H}(x) \xi. \end{aligned} \tag{7.13}$$

Moreover,

$$\begin{aligned} \xi^{\mathrm{T}} \hat{H}(x) \xi &= \xi^{\mathrm{T}} H(x) \xi - \xi^{\mathrm{T}} \left(\sum_{\sigma_i(x) < 1/(2m)} \frac{a_i a_i^{\mathrm{T}}}{\left(a_i^{\mathrm{T}} x - b_i\right)^2} \right) \xi \\ &= \xi^{\mathrm{T}} H(x) \xi - \xi^{\mathrm{T}} H(x)^{1/2} \left(\sum_{\sigma_i(x) < 1/(2m)} \frac{H(x)^{-1/2} a_i a_i^{\mathrm{T}} H(x)^{-1/2}}{\left(a_i^{\mathrm{T}} x - b_i\right)^2} \right) \\ &\quad \times H(x)^{1/2} \xi. \end{aligned} \tag{7.14}$$

The matrix $\sum_{\sigma_i(x) < 1/(2m)} (H(x)^{-1/2} a_i a_i^{\mathrm{T}} H(x)^{-1/2} / (a_i^{\mathrm{T}} x - b_i)^2)$ is symmetric positive semidefinite and its largest eigenvalue is bounded by its trace which equals $\sum_{\sigma_i(x)<1/(2m)}\sigma_i(x)$. Thus from (7.14) it follows that

$$\xi^{\mathrm{T}}\hat{H}(x)\xi \geqslant \xi^{\mathrm{T}}H(x)\xi\left(1-\sum_{\sigma_i(x)<1/(2m)}\sigma_i(x)\right) \geqslant \tfrac{1}{2}\xi^{\mathrm{T}}H(x)\xi.$$

Then from (7.13) we can conclude that

$$\xi^{\mathrm{T}}Q(x)\xi \geqslant \frac{1}{4m}\xi^{\mathrm{T}}H(x)\xi. \quad \square$$

Proof of Lemma 5. Suppose $x \in \Sigma(\hat{x}, r)$. We have that

$$\begin{aligned}\xi^{\mathrm{T}}H(x)\xi &= \sum_{i=1}^{m}\frac{(a_i^{\mathrm{T}}\xi)^2}{(a_i^{\mathrm{T}}x-b_i)^2}\\ &= \sum_{i=1}^{m}\frac{(a_i^{\mathrm{T}}\hat{x}-b_i)^2}{(a_i^{\mathrm{T}}x-b_i)^2}\frac{(a_i^{\mathrm{T}}\xi)^2}{(a_i^{\mathrm{T}}\hat{x}-b_i)^2}.\end{aligned}$$

Hence

$$\begin{aligned}\min_{1\leqslant i\leqslant m}\left\{\frac{(a_i^{\mathrm{T}}\hat{x}-b_i)^2}{(a_i^{\mathrm{T}}x-b_i)^2}\right\}\xi^{\mathrm{T}}H(\hat{x})\xi &\leqslant \xi^{\mathrm{T}}H(x)\xi\\ &\leqslant \max_{1\leqslant i\leqslant m}\left\{\frac{(a_i^{\mathrm{T}}\hat{x}-b_i)^2}{(a_i^{\mathrm{T}}x-b_i)^2}\right\}\xi^{\mathrm{T}}H(\hat{x})\xi\end{aligned}$$

and hence

$$\frac{\xi^{\mathrm{T}}H(\hat{x})\xi}{(1+r)^2} \leqslant \xi^{\mathrm{T}}H(x)\xi \leqslant \frac{\xi^{\mathrm{T}}H(\hat{x})\xi}{(1-r)^2}. \tag{7.15}$$

So from Claim 2 we may conclude that

$$(1+r)^2 a_i^{\mathrm{T}}H(\hat{x})^{-1}a_i \geqslant a_i^{\mathrm{T}}H(x)^{-1}a_i \geqslant (1-r)^2 a_i^{\mathrm{T}}H(\hat{x})^{-1}a_i.$$

Then noting that $\sigma_i(x) = a_i^{\mathrm{T}}H(x)^{-1}a_i/(a_i^{\mathrm{T}}x-b_i)^2$ it follows that for $1 \leqslant i \leqslant m$,

$$\frac{(1-r)^2}{(1+r)^2}\sigma_i(\hat{x}) \leqslant \sigma_i(x) \leqslant \frac{(1+r)^2}{(1-r)^2}\sigma_i(\hat{x}), \tag{7.16}$$

and hence

$$\frac{(1-r)^2}{(1+r)^4} \leqslant \frac{\sigma_i(x)}{\sigma_i(\hat{x})}\frac{(a_i^{\mathrm{T}}\hat{x}-b_i)^2}{(a_i^{\mathrm{T}}x-b_i)^2} \leqslant \frac{(1+r)^2}{(1-r)^4}. \tag{7.17}$$

We can write $\xi^{\mathrm{T}}Q(x)\xi$ as

$$\xi^{\mathrm{T}}Q(x)\xi=\sum_{i=1}^{m}\sigma_i(x)\frac{\left(a_i^{\mathrm{T}}\xi\right)^2}{\left(a_i^{\mathrm{T}}x-b_i\right)^2}$$
$$=\sum_{i=1}^{m}\frac{\sigma_i(x)\left(a_i^{\mathrm{T}}\hat{x}-b_i\right)^2}{\sigma_i(\hat{x})\left(a_i^{\mathrm{T}}x-b_i\right)^2}\sigma_i(\hat{x})\frac{\left(a_i^{\mathrm{T}}\xi\right)^2}{\left(a_i^{\mathrm{T}}\hat{x}-b_i\right)^2}$$

and thus from (7.17) we can conclude that for all $\xi\in\mathbb{R}^n$,

$$\frac{(1-r)^2}{(1+r)^4}\xi^{\mathrm{T}}Q(\hat{x})\xi\leqslant\xi^{\mathrm{T}}Q(x)\xi\leqslant\frac{(1+r)^2}{(1-r)^4}\xi^{\mathrm{T}}Q(\hat{x})\xi. \tag{7.18}$$

From (7.18) and (7.15) it follows that

$$\forall\xi\in\mathbb{R}^n,\quad \xi^{\mathrm{T}}Q(x)\xi\geqslant\frac{(1-r)^2}{(1+r)^4}\xi^{\mathrm{T}}Q(\hat{x})\xi$$
$$\geqslant\frac{(1-r)^2\mu(\hat{x})}{(1+r)^4}\xi^{\mathrm{T}}H(\hat{x})\xi\quad(\text{by def. of }\mu(\hat{x}))$$
$$\geqslant\frac{(1-r)^4\mu(\hat{x})}{(1+r)^4}\xi^{\mathrm{T}}H(x)\xi.$$

Thus

$$\mu(x)\geqslant\frac{(1-r)^4\mu(\hat{x})}{(1+r)^4}.$$

From (7.18) and (7.15) it also follows that

$$\forall\xi\in\mathbb{R}^n,\quad \xi^{\mathrm{T}}Q(\hat{x})\xi\geqslant\frac{(1-r)^4}{(1+r)^2}\xi^{\mathrm{T}}Q(x)\xi$$
$$\geqslant\frac{\mu(x)(1-r)^4}{(1+r)^2}\xi^{\mathrm{T}}H(x)\xi\quad(\text{by def. of }\mu(x))$$
$$\geqslant\frac{\mu(x)(1-r)^4}{(1+r)^4}\xi^{\mathrm{T}}H(\hat{x})\xi.$$

Thus

$$\mu(\hat{x})\geqslant\frac{\mu(x)(1-r)^4}{(1+r)^4}.\quad\square$$

Proof of Lemma 6. Note that from the strict convexity of $F(x)$ it follows that for each value of t there is a unique x which satisfies $\Phi(x,t)=0$. Let $\Phi_j(x,t)$ denote the jth coordinate function of $\Phi(x,t)$. Note that

$$\Phi_j(x,t)=\sum_{i=1}^{m}\frac{a_i^{\mathrm{T}}H(x)^{-1}a_i}{\left(a_i^{\mathrm{T}}x-b_i\right)^2}\frac{a_{ij}}{a_i^{\mathrm{T}}x-b_i}-tw_j.$$

$\Phi_j(x, t)$ is an analytic function which is easily seen as follows. First, note that the sum of two analytic functions is analytic, the product of two analytic functions is analytic, and the ratio of two analytic functions is analytic provided the denominator function is different from zero at each point in the domain of interest. Also, note that each of the functions $a_i^T x - b_i$, $1 \leqslant i \leqslant m$ is analytic. It then follows that each entry of the matrix $H(x) = \sum_{i=1}^m (a_i a_i^T / (a_i^T x - b_i)^2)$ is an analytic function in the interior of the polytope P. Moreover, since the determinant of $H(x)$ is different from zero at each point in the interior of P, each element of $H(x)^{-1}$ is also analytic in the interior of P. We can then conclude that the coordinate functions $\Phi_j(x, t)$ are analytic in the interior of P. Then from the implicit function theorem [1,2] we can conclude that the equation $\Phi(x, t) = 0$ implicitly defines x as a function of t, and that the function $x(t): \mathbb{R} \to \text{Interior}(P)$ is an analytic function of t. Then using the chain rule for differentiation gives

$$\Phi_t(x(t), t) + \Phi_x(x(t), t) \frac{\mathrm{d}x(t)}{\mathrm{d}t} = 0,$$

where Φ_t is computed with x held fixed and Φ_x is computed with t held fixed. Observing that $\Phi_t(x(t), t) = -w$ and $\Phi_x(x(t), t) = \nabla^2 F(x)$ we get that

$$-w + \nabla^2 F(x) \frac{\mathrm{d}x(t)}{\mathrm{d}t} = 0,$$

and since $\nabla^2 F(x)$ is nonsingular

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = \nabla^2 F(x)^{-1} w.$$

We can write

$$F(x(t_2)) - F(x(t_1)) = \int_{t_1}^{t_2} \nabla F(x)^T \frac{\mathrm{d}x(t)}{\mathrm{d}t}$$

$$= \int_{t_1}^{t_2} t w^T \nabla^2 F(x)^{-1} w \, \mathrm{d}t$$

From Lemma 3 and Claim 2 we get that

$$\tfrac{1}{5} w^T Q(x)^{-1} w \leqslant w^T \nabla^2 F(x)^{-1} w \leqslant w^T Q(x)^{-1} w,$$

and since $0 \leqslant t_1 \leqslant t_2$, for $t_1 \leqslant t \leqslant t_2$ we have that

$$\tfrac{1}{5} t w^T Q(x)^{-1} w \leqslant t w^T \nabla^2 F(x)^{-1} w \leqslant t w^T Q(x)^{-1} w.$$

Hence

$$\tfrac{1}{5} \int_{t_1}^{t_2} t w^T Q(x)^{-1} w \, \mathrm{d}t \leqslant F(x(t_2)) - F(x(t_1)) \leqslant \int_{t_1}^{t_2} t w^T Q(x)^{-1} w \, \mathrm{d}t. \quad \square$$

Proof of Lemma 7. We have that

$$\left(\frac{a_i^{\mathrm{T}}(\mathrm{d}x(t)/\mathrm{d}t)}{a_i^{\mathrm{T}}x-b_i}\right)^2 = \frac{\left(a_i^{\mathrm{T}}\nabla^2F(x)^{-1}w\right)^2}{\left(a_i^{\mathrm{T}}x-b_i\right)^2} \quad \text{(by Lemma 6)}$$

$$\leqslant \frac{\left(a_i^{\mathrm{T}}\nabla^2F(x)^{-1}a_i\right)\left(w^{\mathrm{T}}\nabla^2F(x)^{-1}w\right)}{\left(a_i^{\mathrm{T}}x-b_i\right)^2}$$

since $u^{\mathrm{T}}v \leqslant \|u\|_2\|v\|_2$.

By Lemma 3 and Claim 2,

$$\forall \xi \in \mathbb{R}^n, \quad \xi^{\mathrm{T}}\nabla^2F(x)^{-1}\xi \leqslant \xi^{\mathrm{T}}Q(x)^{-1}\xi.$$

Thus

$$\left(\frac{a_i^{\mathrm{T}}(\mathrm{d}x(t)/\mathrm{d}t)}{a_i^{\mathrm{T}}x-b_i}\right)^2 \leqslant \frac{\left(a_i^{\mathrm{T}}Q(x)^{-1}a_i\right)\left(w^{\mathrm{T}}Q(x)^{-1}w\right)}{\left(a_i^{\mathrm{T}}x-b_i\right)^2}.$$

By Claim 4,

$$\frac{a_i^{\mathrm{T}}Q(x)^{-1}a_i}{\left(a_i^{\mathrm{T}}x-b_i\right)^2} \leqslant \frac{1}{\sqrt{\mu(x)}}$$

and hence

$$\left(\frac{a_i^{\mathrm{T}}(\mathrm{d}x(t)/\mathrm{d}t)}{a_i^{\mathrm{T}}x-b_i}\right)^2 \leqslant \frac{w^{\mathrm{T}}Q(x)^{-1}w}{\sqrt{\mu(x)}}.$$

Since $x \in \Sigma(\hat{x}, r)$, from Lemma 5 and Claim 2 we get that

$$w^{\mathrm{T}}Q(x)^{-1}w \leqslant \frac{(1+r)^4}{(1-r)^2}w^{\mathrm{T}}Q(\hat{x})^{-1}w$$

and that

$$\frac{1}{\sqrt{\mu(x)}} \leqslant \frac{(1+r)^2}{(1-r)^2\sqrt{\mu(\hat{x})}}$$

Thus

$$\left(\frac{a_i^{\mathrm{T}}(\mathrm{d}x(t)/\mathrm{d}t)}{a_i^{\mathrm{T}}x-b_i}\right)^2 \leqslant \frac{(1+r)^6}{(1-r)^4}\frac{w^{\mathrm{T}}Q(\hat{x})^{-1}w}{\sqrt{\mu(\hat{x})}}. \quad \square$$

Proof of Lemma 8. Let $x(t^*)$ be the first point on the boundary of $\Sigma(\hat{x}, r)$ as we move from $\hat{x}$ to $x(\bar{t})$ on the trajectory $\nabla F(x) = tw$. ($x(t^*)$ exists since by Lemma 6 the trajectory is continuous.) Note that all points on the trajectory between $\hat{x}$ and $x(t^*)$ lie in $\Sigma(\hat{x}, r)$. There exists an index j, $1 \leqslant j \leqslant m$ such that $|a_j^{\mathrm{T}}(x(t^*)-\hat{x})/(a_j^{\mathrm{T}}\hat{x}-b_j)| = r$. Hence

$$\left|\ln\left(\frac{a_j^{\mathrm{T}}x(t^*)-b_j}{a_j^{\mathrm{T}}\hat{x}-b_j}\right)\right| \geqslant r - \tfrac{1}{2}r^2.$$

We have that

$$\left|\ln\left(\frac{a_j^{\mathrm{T}}x(t^*)-b_j}{a_j^{\mathrm{T}}\hat{x}-b_j}\right)\right| = \left|\int_{t^*}^{\hat{t}} \frac{a_j^{\mathrm{T}}(\mathrm{d}x(t)/\mathrm{d}t)}{a_j^{\mathrm{T}}x-b_j}\,\mathrm{d}t\right|,$$

and hence

$$\left|\int_{t^*}^{\hat{t}} \frac{a_j^{\mathrm{T}}(\mathrm{d}x(t)/\mathrm{d}t)}{a_j^{\mathrm{T}}x-b_j}\,\mathrm{d}t\right| \geqslant r-\tfrac{1}{2}r^2. \tag{7.19}$$

By Lemma 7, for all $x\in\Sigma(\hat{x},r)$,

$$\left|\frac{a_j^{\mathrm{T}}(\mathrm{d}x(t)/\mathrm{d}t)}{a_j^{\mathrm{T}}x-b_j}\right| \leqslant \frac{(1+r)^3}{(1-r)^2}\,\frac{\left(w^{\mathrm{T}}Q(\hat{x})^{-1}w\right)^{1/2}}{(\mu(\hat{x}))^{1/4}},$$

and hence

$$\left|\int_{t^*}^{\hat{t}} \frac{a_j^{\mathrm{T}}(\mathrm{d}x(t)/\mathrm{d}t)}{a_j^{\mathrm{T}}x-b_j}\,\mathrm{d}t\right| \leqslant |\hat{t}-t^*|\;\frac{(1+r)^3}{(1-r)^2}\,\frac{\left(w^{\mathrm{T}}Q(\hat{x})^{-1}w\right)^{1/2}}{(\mu(\hat{x}))^{1/4}}.$$

Thus from (7.19) it follows that

$$|\hat{t}-t^*| \geqslant \frac{\left(r-\frac{1}{2}r^2\right)(1-r)^2}{(1+r)^3}\,\frac{(\mu(\hat{x}))^{1/4}}{\left(w^{\mathrm{T}}Q(\hat{x})^{-1}w\right)^{1/2}}.$$

Lemma 8 then follow from the observation that $|\hat{t}-\bar{t}| \geqslant |\hat{t}-t^*|$. □

Proof of Lemma 9. Let $r=1.1\sqrt{\delta}$. Suppose that the $\omega\notin\Sigma(z,r)$. Let $w=\nabla F(z)$. Note that both z and ω lie on the trajectory $\nabla F(x)=tw$, $t\in\mathbb{R}$. Thus applying Lemma 8 with $\hat{x}=z$, $\hat{t}=1$, $x(\bar{t})=\omega$ and $\bar{t}=0$ we get that

$$1 \geqslant \frac{\left(r-\frac{1}{2}r^2\right)(1-r)^2(\mu(z))^{1/4}}{(1+r)^3\sqrt{w^{\mathrm{T}}Q(z)^{-1}w}},$$

and noting that $w=\nabla F(z)$ and $r=1.1\sqrt{\delta}\leqslant 0.011$, we get that

$$\nabla F(z)^{\mathrm{T}}Q(z)^{-1}\nabla F(z) \geqslant 1.02\sqrt{\mu(z)} > \delta\sqrt{\mu(z)}.$$

By assumption, this cannot happen. So $\omega\in\Sigma(z,r)$.

Next, we shall show that

$$F(z)-F(\omega) \leqslant 0.55\nabla F(z)^{\mathrm{T}}Q(z)^{-1}\nabla F(z).$$

From Lemma 5 and Claim 2, we have that for $x\in\Sigma(z,r)$,

$$w^{\mathrm{T}}Q(x)^{-1}w \leqslant \frac{(1+r)^4}{(1-r)^2}w^{\mathrm{T}}Q(z)^{-1}w.$$

Thus we have that

$$
\begin{aligned}
F(z) - F(\omega) &\leqslant \int_0^1 t w^{\mathrm{T}} Q(x)^{-1} w \, \mathrm{d}t \quad \text{(by Lemma 6)} \\
&\leqslant \int_0^1 t \frac{(1+r)^4}{(1-r)^2} w^{\mathrm{T}} Q(z)^{-1} w \, \mathrm{d}t \\
&= \frac{(1+r)^4}{2(1-r)^2} w^{\mathrm{T}} Q(z)^{-1} w \\
&= \frac{(1+r)^4}{2(1-r)^2} \nabla F(z)^{\mathrm{T}} Q(z)^{-1} \nabla F(z) \\
&\leqslant 0.55\, \nabla F(z)^{\mathrm{T}} Q(z)^{-1} \nabla F(z), \quad \text{since } r = 1.1\sqrt{\delta} \leqslant 0.011.
\end{aligned}
$$

Finally, since $\omega \in \Sigma(z, r)$, from Lemma 5 we get that

$$
\begin{aligned}
\mu(z) &\leqslant \frac{(1+r)^4}{(1-r)^4} \mu(\omega) \\
&\leqslant 1.1 \mu(\omega), \quad \text{since } r \leqslant 0.011.
\end{aligned}
$$

That concludes the proof of the Lemma. □

Proof of Lemma 10. Let $r = 5\sqrt{\delta}$. We shall first show that $z \in \Sigma(\omega, r)$. Suppose $z \notin \Sigma(\omega, r)$. Let $w = \nabla F(z)$. Note that the trajectory $\nabla F(x) = tw$, $t \in \mathbb{R}$, passes through both z and ω. Let $x(\bar{t})$ be the first point on the boundary of $\Sigma(\omega, r)$ as we move on the trajectory $\nabla F(x) = tw$ from ω to z. Note that $\nabla F(x(\bar{t})) = \bar{t}w$, and that $\forall t$, $0 \leqslant t \leqslant \bar{t}$, $x(t) \in \Sigma(\omega, r)$. We can then apply Lemma 8 with $\hat{x} = \omega$ and $\hat{t} = 0$ and conclude that

$$
\bar{t} \geqslant \frac{\left(r - \frac{1}{2}r^2\right)(1-r)^2(\mu(\omega))^{1/4}}{(1+r)^3\sqrt{w^{\mathrm{T}} Q(\omega)^{-1} w}}
$$

and hence

$$
(\bar{t})^2 w^{\mathrm{T}} Q(\omega)^{-1} w \geqslant \frac{\left(r - \frac{1}{2}r^2\right)^2 (1-r)^4}{(1+r)^6} \sqrt{\mu(\omega)}. \tag{7.20}
$$

By Lemma 6,

$$
F(x(\bar{t})) - F(\omega) \geqslant \tfrac{1}{5} \int_0^{\bar{t}} t w^{\mathrm{T}} Q(x)^{-1} w \, \mathrm{d}t.
$$

From Lemma 5 and Claim 2 we get that for all $x \in \Sigma(\omega, r)$

$$
w^{\mathrm{T}} Q(x)^{-1} w \geqslant \frac{(1-r)^4}{(1+r)^2} w^{\mathrm{T}} Q(\omega)^{-1} w
$$

and hence

$$\begin{aligned} F(x(\bar{t})) - F(\omega) &\geqslant \frac{(1-r)^4 w^{\mathrm{T}} Q(\omega)^{-1} w}{5(1+r)^2} \int_0^{\bar{t}} t\, \mathrm{d}t. \\ &\geqslant \frac{(1-r)^4 (\bar{t})^2}{10(1+r)^2} w^{\mathrm{T}} Q(\omega)^{-1} w \\ &\geqslant \frac{(1-r)^8 \left(r - \frac{1}{2} r^2\right)^2}{10(1+r)^8} \sqrt{\mu(\omega)}. \quad (\text{by } (7.20)). \end{aligned}$$

Then noting that $r = 5\sqrt{\delta} \leqslant 0.05$, we get that

$$F(x(\bar{t})) - F(\omega) \geqslant 1.01\, \delta \sqrt{\mu(\omega)}. \tag{7.21}$$

On the other hand since $F(x)$ increases monotonically on the trajectory $\nabla F(x) = tw$ from ω to z, we have that

$$F(x(\bar{t})) - F(\omega) \leqslant F(z) - F(\omega) \leqslant \delta \sqrt{\mu(\omega)}. \tag{7.22}$$

But (7.21) and (7.22) cannot be simultaneously true. So z must be in $\Sigma(\omega, r)$.

We shall now show that

$$0.14\, \nabla F(z)^{\mathrm{T}} Q(z)^{-1} \nabla F(z) \leqslant F(z) - F(\omega) \leqslant 1.4\, \nabla F(z)^{\mathrm{T}} Q(z)^{-1} \nabla F(z). \tag{7.23}$$

Since $z \epsilon \Sigma(\omega, r)$, it follows that $\omega \epsilon \Sigma(z, r/(1-r))$. Let $\bar{r} = r/(1-r)$. So from Lemma 5 and Claim 2, we can conclude that for all $x \epsilon \Sigma(z, \bar{r})$,

$$\frac{(1-\bar{r})^4 w^{\mathrm{T}} Q(z)^{-1} w}{(1+\bar{r})^2} \leqslant w^{\mathrm{T}} Q(x)^{-1} w \leqslant \frac{(1+\bar{r})^4 w^{\mathrm{T}} Q(z)^{-1} w}{(1-\bar{r})^2}. \tag{7.24}$$

From Lemma 6,

$$\tfrac{1}{5} \int_0^1 t w^{\mathrm{T}} Q(x)^{-1} w\, \mathrm{d}t \leqslant F(z) - F(\omega) \leqslant \int_0^1 t w^{\mathrm{T}} Q(x)^{-1} w\, \mathrm{d}t$$

and hence from (7.24) it follows that

$$\frac{(1-\bar{r})^4 w^{\mathrm{T}} Q(z)^{-1} w}{5(1+\bar{r})^2} \int_0^1 t\, \mathrm{d}t \leqslant F(z) - F(\omega) \leqslant \frac{(1+\bar{r})^4 w^{\mathrm{T}} Q(z)^{-1} w}{(1-\bar{r})^2} \int_0^1 t\, \mathrm{d}t.$$

(7.23) then follows by carrying out the integration, and noting that $\bar{r} = r/(1-r)$ and $r = 5\sqrt{\delta} \geqslant 0.05$. Furthermore, since $z \epsilon \Sigma(\omega, r)$, by Lemma 5 we get that

$$\begin{aligned} \mu(\omega) &\leqslant \frac{(1+r)^4}{(1-r)^4} \mu(z) \\ &\leqslant 1.5 \mu(z), \quad \text{since } r \leqslant 0.05. \end{aligned}$$

That concludes the proof of the Lemma. □

8. Proofs of theorems

In this section we shall prove Theorems 1, 2 and 3 introduced in Section 3. But first we shall collect together some notation. Recall that P is the polytope $P=\{x: Ax \geqslant b\}$, where $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, and $x \in \mathbb{R}^n$. Also, note that a_i^{T} denotes the ih row of A, $H(x)=\sum_{i=1}^m (a_i a_i^{\mathrm{T}}/(a_i^{\mathrm{T}}x-b_i)^2)$, $F(x)=\frac{1}{2}\ln(\det(H(x)))$, and the volumetric center ω is the point that minimizes $F(x)$ over the polytope P. Moreover, $\sigma_i(x)=a_i^{\mathrm{T}}H(x)^{-1}a_i/(a_i^{\mathrm{T}}x-b_i)^2$, $1 \leqslant i \leqslant m$, $Q(x)=\sum_{i=1}^m (\sigma_i(x)a_i a_i^{\mathrm{T}}/(a_i^{\mathrm{T}}x-b_i)^2)$, and $\mu(x)$ is the largest number λ such that $\forall \xi \in \mathbb{R}^n$, $\xi^{\mathrm{T}}Q(x)\xi \geqslant \lambda \xi^{\mathrm{T}}H(x)\xi$.

Similarly, $\tilde{P}$ is the polytope $P=\{x: \tilde{A}x \geqslant \tilde{b}\}$ where $\tilde{A} \in \mathbb{R}^{\tilde{m}\times n}$ and $\tilde{b} \in \mathbb{R}^{\tilde{m}}$. Note that $\tilde{a}_i^{\mathrm{T}}$ denotes the ith row of $\tilde{A}$, $\tilde{H}(x)=\sum_{i=1}^{\tilde{m}}(\tilde{a}_i\tilde{a}_i^{\mathrm{T}}/(\tilde{a}_i^{\mathrm{T}}x-\tilde{b}_i)^2)$, $\tilde{F}(x)=\frac{1}{2}\ln(\det(\tilde{H}(x)))$, and the volumentric center $\tilde{\omega}$ is the point that minimizes $\tilde{F}(x)$ over $\tilde{P}$. Furthermore, $\tilde{\sigma}_i(x)=\tilde{a}_i^{\mathrm{T}}\tilde{H}(x)^{-1}\tilde{a}_i/(\tilde{a}_i^{\mathrm{T}}x-\tilde{b}_i)^2$, $1 \leqslant i \leqslant \tilde{m}$, $\tilde{Q}(x)=\sum_{i=j}^{\tilde{m}}(\tilde{\sigma}_i(x)\tilde{a}_i\tilde{a}_i^{\mathrm{T}}/(\tilde{a}_i^{\mathrm{T}}x-\tilde{b}_i)^2)$ and $\tilde{\mu}(x)$ is the largest number λ such that $\forall\xi \in \mathbb{R}^n$, $\xi^{\mathrm{T}}Q(x)\xi \geqslant \lambda\xi^{\mathrm{T}}\tilde{H}(x)\xi$.

Remark. Note that as long as the polytope $\tilde{P}$ is full dimensional and bounded Lemmas 1 through 10 in Section 7 are valid with the relevant symbols without tilde replaced by the corresponding symbols with tilde, e.g., with P replaced by $\tilde{P}$; $H(x)$, $Q(x)$ replaced by $\tilde{H}(x)$, $\tilde{Q}(x)$; m replaced by $\tilde{m}$ etc.

Instead of proving Theorem 3 we shall prove Theorem 4 below and Theorem 3 will follow from Theorem 4 by interchanging the roles of the tilded and the untilded quantities (i.e., by interchanging $\tilde{P}$ and P, $\tilde{A}$ and A, etc.)

Theorem 4. *Let* $\tilde{A}=(A\ c)^{\mathrm{T}}$, $\tilde{b}=(b\ \beta)^{\mathrm{T}}$ *and* $\tilde{P}=P\cap\{x: c^{\mathrm{T}}x \geqslant \beta\}$. *Let* $z \in \tilde{P}$, *let* $\alpha \leqslant \delta \leqslant 10^{-4}$, *and let* $\tilde{F}(z)-\tilde{F}(\tilde{w}) \leqslant \alpha^2\tilde{\mu}(\tilde{\omega})$. *Suppose that* $c^{\mathrm{T}}\tilde{H}(z)^{-1}c/(c^{\mathrm{T}}z-\beta)^2 \leqslant \min\{\alpha, \tilde{\mu}(z)\}$. *Then the polytope P is bounded,*

$$F(z)-F(\omega) \leqslant \min\left\{4\alpha\mu(\omega),\ \delta\sqrt{\mu(\omega)}\right\},$$

and

$$\tilde{F}(\tilde{\omega})-F(\omega) \leqslant 5\alpha.$$

In addition to the lemmas in Section 7 and the claims in Section 7.1, the proofs of the theorems make use of Lemmas 11–14 stated below. Lemmas 11–14 will be proved in Section 8.1.

Lemma 11. *Let* $\tilde{A}=(A\ c)^{\mathrm{T}}$, $\tilde{b}=(b\ \beta)^{\mathrm{T}}$, $\tilde{P}=P\cap\{x: c^{\mathrm{T}}x \geqslant B\}$, *and let* $z \in \tilde{P}$. *Then*

(a) $$\forall\xi \in \mathbb{R}^n,\quad \xi^{\mathrm{T}}\tilde{H}(z)\xi\left(1-\frac{c^{\mathrm{T}}\tilde{H}(z)^{-1}c}{(c^{\mathrm{T}}z-\beta)^2}\right) \leqslant \xi^{\mathrm{T}}H(z)\xi \leqslant \xi^{\mathrm{T}}\tilde{H}(z)\xi.$$

(b) $$c^{\mathrm{T}}\tilde{H}(z)^{-1}c=\left(c^{\mathrm{T}}H(z)^{-1}c\right)\frac{(c^{\mathrm{T}}z-\beta)^2}{(c^{\mathrm{T}}z-\beta)^2+c^{\mathrm{T}}H(z)^{-1}c}.$$

(c) *For* $1 \leqslant i \leqslant m$,

$$\sigma_i(z) - \tilde{\sigma}_i(z) = \frac{\left(a_i^{\mathrm{T}} H(z)^{-1} c\right)^2}{\left((c^{\mathrm{T}} z - \beta)^2 + c^{\mathrm{T}} H(z)^{-1} c\right)\left(a_i^{\mathrm{T}} z - b_i\right)^2}.$$

(d) For $1 \leqslant i \leqslant m$,

$$\sigma_i(z)\left(1 - \frac{c^{\mathrm{T}} H(z)^{-1} c}{(c^{\mathrm{T}} z - \beta)^2}\right) \leqslant \tilde{\sigma}_i(z) \leqslant \sigma_i(z).$$

(e) $$\tilde{F}(z) - F(z) = \frac{1}{2}\ln\left(1 + \frac{c^{\mathrm{T}} H(z)^{-1} c}{(c^{\mathrm{T}} z - \beta)^2}\right) = -\frac{1}{2}\ln\left(1 - \frac{c^{\mathrm{T}} \tilde{H}(z)^{-1} c}{(c^{\mathrm{T}} z - \beta)^2}\right).$$

Lemma 12. *Let* $\tilde{A} = (A\ c)^{\mathrm{T}}$, $\tilde{b} = (b\ \beta)^{\mathrm{T}}$, *and let* $z \in \tilde{P}$. *Let* $u = \sum_{i=1}^{m} (\tilde{\sigma}_i(z) - \sigma_i(z)) a_i / (a_i^{\mathrm{T}} z - b_i)$. *Then*

$$u^{\mathrm{T}} Q(z)^{-1} u \leqslant \frac{\left(c^{\mathrm{T}} H(z)^{-1} c\right)^2}{(c^{\mathrm{T}} z - \beta)^4}.$$

Lemma 13. *Let* $\tilde{A} = (A\ c)^{\mathrm{T}}$, $\tilde{b} = (b\ \beta)^{\mathrm{T}}$, *and let* $z \in \tilde{P}$. *Suppose that* $c^{\mathrm{T}} H(z)^{-1} c / (c^{\mathrm{T}} z - \beta)^2 \leqslant \tau < 1$. *Then*

$$\left(1 + \frac{\tau^2}{\mu(z)}\right)\mu(z) \geqslant \tilde{\mu}(z) \geqslant \frac{1-\tau}{1+\tau}\mu(z)$$

and for all $\xi \in \mathbb{R}^n$,

$$\frac{1}{(1 + \tau^2/\mu(z))}\xi^{\mathrm{T}} Q(z)^{-1} \xi \leqslant \xi^{\mathrm{T}} \tilde{Q}(z)^{-1} \xi \leqslant \frac{1}{1-\tau}\xi^{\mathrm{T}} Q(z)^{-1} \xi.$$

Lemma 14. *Let* $z \in P$ *and let* $\delta \leqslant 10^{-4}$. *Suppose that* A *has linearly independent columns and that* $\nabla F(z)^{\mathrm{T}} Q(z)^{-1} \nabla F(z) \leqslant \delta\sqrt{\mu(z)}$. *Then the polytope* P *is bounded.*

It is worth noting that as long as A has linearly independent columns the required inverses in Lemmas 11, 12 and 13 exist and the lemmas remain valid; the boundedness of P (or $\tilde{P}$) is not required for their validity. We shall now prove Theorems 1, 2 and 4.

Proof of Theorem 1. Let $y = z - t\eta$ where $t \in \mathbb{R}$. Define $\lambda_{\max}$ as follows.

$$\text{If } F(z) - F(\omega) \leqslant \delta\sqrt{\mu(\omega)} \text{ then } \lambda_{\max} = 0.2 \text{ else } \lambda_{\max} = \frac{0.2\delta^{1/2}\mu(z)^{1/4}}{\left(\nabla F(z)^{\mathrm{T}}\eta\right)^{1/2}}.$$

We shall show that

$$\forall t,\ 0 \leqslant t \leqslant \lambda_{\max},\quad \frac{\mathrm{d}F(y(t))}{\mathrm{d}t} \leqslant -(1 - 5.3t)\nabla F(z)^{\mathrm{T}}\eta, \tag{8.1}$$

where $\mathrm{d}F(y(t))/\mathrm{d}t$ is the derivative of $F(y(t))$ w.r.t. t evaluated at t. Before proving (8.1) we shall complete the proof of the theorem using (8.1). From (8.1) it follows that

$$\begin{aligned} F(z) - F(z') &= F(z) - F(z - \lambda\eta) \\ &= -\int_0^\lambda \frac{\mathrm{d}F(y(t))}{\mathrm{d}t}\,\mathrm{d}t \\ &\geqslant \nabla F(z)^{\mathrm{T}}\eta\int_0^\lambda (1 - 5.3t)\,\mathrm{d}t \\ &= (\lambda - 2.65\lambda^2)\nabla F(z)^{\mathrm{T}}\eta. \end{aligned} \tag{8.2}$$

There are two cases depending on the value of $F(z) - F(\omega)$. Suppose that $F(z) - F(\omega) \leqslant \delta\sqrt{\mu(\omega)}$. Then by Lemma 10 and the definition of η it follows that

$$1.4\,\nabla F(z)^{\mathrm{T}}\eta \geqslant F(z) - F(\omega).$$

So by (8.2) and noting that in this case $\lambda = r$ we may conclude that

$$\begin{aligned} F(z) - F(z') &\geqslant (r - 2.65r^2)\nabla F(z)^{\mathrm{T}}\eta \\ &\geqslant \frac{r - 2.65r^2}{1.4}(F(z) - F(\omega)) \\ &\geqslant (0.71r - 1.9r^2)(F(z) - F(\omega)). \end{aligned}$$

Thus if $F(z) - F(\omega) \leqslant \delta\sqrt{\mu(\omega)}$ then

$$F(z') - F(\omega) \leqslant (1 - 0.71r + 1.9r^2)(F(z) - F(\omega)).$$

Next, suppose that $F(z) - F(\omega) > \delta\sqrt{\mu(\omega)}$. By Lemma 9 and the definition of η it follows that

$$\nabla F(z)^{\mathrm{T}}\eta \leqslant \delta\sqrt{\mu(z)} \quad \Rightarrow \quad F(z) - F(\omega) \leqslant 0.55\delta\sqrt{\mu(z)} \leqslant 0.61\,\delta\sqrt{\mu(\omega)}.$$

So in this case

$$\nabla F(z)^{\mathrm{T}}\eta > \delta\sqrt{\mu(z)}.$$

Note that in this case $\lambda = r\delta^{1/2}\mu(z)^{1/4}/(\nabla F(z)^{\mathrm{T}}\eta)^{1/2}$. So from (8.2) we get that

$$\begin{aligned} F(z) - F(z') &\geqslant (\lambda - 2.65\lambda^2)\nabla F(z)^{\mathrm{T}}\eta \\ &= r\delta^{1/2}\mu(z)^{1/4}\left(\nabla F(z)^{\mathrm{T}}\eta\right)^{1/2} - 2.65r^2\delta\sqrt{\mu(z)} \\ &\geqslant (r - 2.65r^2)\,\delta\sqrt{\mu(z)} \\ &\geqslant \frac{(r - 2.65r^2)\,\delta}{2\sqrt{m}}, \quad \text{since } \mu(z) \geqslant \frac{1}{4m} \text{ by Lemma 4.} \end{aligned}$$

Thus if $F(z) - F(\omega) > \delta\sqrt{\mu(z)}$ then

$$F(z) - F(z') \geqslant \frac{(r - 2.65r^2)\,\delta}{2\sqrt{m}}.$$

That completes the proof of Theorem 1.

We shall now prove (8.1). We shall restrict ourselves to values of t in the range 0 to λ_{max}. To show (8.1) we shall require an upper bound on $d^2F(y(t))/dt^2$, the second derivative of $F(y(t))$ w.r.t. t evaluated at t. We have that

$$\frac{d^2F(y(t))}{dt^2} = \eta^T \nabla^2 F(y(t))\eta.$$

By Lemma 3,

$$\eta^T \nabla^2 F(y(t))\eta \leqslant 5\eta^T Q(y(t))\eta$$

and hence

$$\frac{d^2F(y(t))}{dt^2} \leqslant 5\eta^T Q(y(t))\eta. \tag{8.3}$$

We shall show that

$$y(t) \in \Sigma(z, 0.6\delta^{1/2}). \tag{8.4}$$

From (8.4) and Lemma 5 we get that

$$\begin{aligned}\eta^T Q(y(t))\eta &\leqslant \frac{(1+0.6\delta^{1/2})^2}{(1-0.6\delta^{1/2})^4}\eta^T Q(z)\eta \\ &\leqslant 1.06\eta^T Q(z)\eta \quad (\text{since } \delta \leqslant 10^{-4}) \\ &\leqslant 1.06\,\nabla F(z)^T\eta \quad (\text{by def. of } \eta).\end{aligned}$$

So from (8.3) we can conclude that

$$\frac{d^2F(y(t))}{dt^2} \leqslant 5.3\,\nabla F(z)^T\eta.$$

We can write

$$\begin{aligned}\frac{dF(y(t))}{dt} &= \frac{dF(y(0))}{dt} + \int_0^t \frac{d^2F(y(t))}{dt^2}\,dt \\ &\leqslant -\nabla F(z)^T\eta + \int_0^t 5.3\,\nabla F(z)^T\eta\,dt \\ &= -(1-5.3t)\nabla F(z)^T\eta.\end{aligned}$$

(8.2) then follows.

We shall now show (8.4). We have that

$$\begin{aligned}(y(t)-z)^T Q(z)(y(t)-z) &= t^2\eta^T Q(z)\eta \\ &\leqslant \lambda_{max}^2 \nabla F(z)^T\eta.\end{aligned}$$

Thus

$$(y(t)-z)^T Q(z)(y(t)-z) \leqslant \lambda_{max}^2 \nabla F(z)^T\eta. \tag{8.5}$$

There are two cases depending on the value of $F(z)-F(\omega)$. Suppose that $F(z)-F(\omega) \leqslant \delta\sqrt{\mu(\omega)}$. By Lemma 10 and the definition of η we get that

$$\nabla F(z)^T\eta \leqslant \frac{1}{0.14}(F(z)-F(\omega)) \leqslant 7.2\delta\sqrt{\mu(\omega)}$$

and that

$$\mu(\omega) \leqslant 1.5\mu(z).$$

It then follows that

$$\nabla F(z)^{\mathrm{T}}\eta \leqslant 7.2\,\delta\sqrt{\mu(\omega)} \leqslant 9\delta\sqrt{\mu(z)}.$$

Then by (8.5) and noting that in this case $\lambda_{\max} = 0.2$ we get that

$$(y(t) - z)^{\mathrm{T}}Q(z)(y(t) - z) \leqslant 0.36\delta\sqrt{\mu(z)}.$$

Next, suppose that $F(z) - F(\omega) > \delta\sqrt{\mu(\omega)}$. In this case $\lambda_{\max} = 0.2\,\delta^{1/2} \times (\mu(z))^{1/4}/(\nabla F(z)^{\mathrm{T}}\eta)^{1/2}$. Hence

$$\lambda_{\max}^2 \nabla F(z)^{\mathrm{T}}\eta = 0.04\delta\sqrt{\mu(z)} \leqslant 0.36\delta\sqrt{\mu(z)},$$

and so by (8.5),

$$(y(t) - z)^{\mathrm{T}}Q(z)(y(t) - z) \leqslant 0.36\delta\sqrt{\mu(z)}.$$

Thus in both the cases $y(t) \in E(Q(z),\ z,\ 0.6\delta^{1/2}(\mu(z))^{1/4})$ where

$$E\big(Q(z),\ z,\ \theta(\mu(z))^{1/4}\big) = \Big\{x:\ (x - z)^{\mathrm{T}}Q(z)(x - z) \leqslant \theta^2\sqrt{\mu(z)}\Big\}.$$

By Claim 4,

$$E\big(Q(z),\ z,\ \theta(\mu(z))^{1/4}\big) \subseteq \Sigma(z,\ \theta)$$

and hence $y(t) \in \Sigma(z, 0.6\delta^{1/2})$. This proves (8.4). □

Proof of Theorem 2. We shall show below that

$$\nabla\tilde{F}(z)^{\mathrm{T}}\tilde{Q}(z)^{-1}\nabla\tilde{F}(z) \leqslant 0.6(\delta\alpha)^{1/2} \leqslant 0.66\delta\sqrt{\tilde{\mu}(z)}. \tag{8.6}$$

Before proving (8.6) we shall complete the proof of the theorem using (8.6). By (8.6) and Lemma 9 we get that

$$\tilde{F}(z) - \tilde{F}(\tilde{\omega}) \leqslant 0.55\nabla\tilde{F}(z)^{\mathrm{T}}\tilde{Q}(z)^{-1}\nabla\tilde{F}(z)$$

and that

$$\tilde{\mu}(z) \leqslant 1.1\tilde{\mu}(\tilde{\omega}).$$

So from (8.6) it follows that

$$\tilde{F}(z) - \tilde{F}(\tilde{\omega}) \leqslant 0.33(\delta\alpha)^{1/2} \leqslant 0.43\delta\sqrt{\tilde{\mu}(\tilde{\omega})}. \tag{8.7}$$

From Lemma 11(e), we have that

$$\tilde{F}(z) - F(z) = \tfrac{1}{2}\ln\left(1 + \frac{c^{\mathrm{T}}H(z)^{-1}c}{(c^{\mathrm{T}}z - \beta)^2}\right) \geqslant \tfrac{1}{4}\alpha^{1/2} - \tfrac{1}{8}\alpha. \tag{8.8}$$

Then from (8.7) and (8.8) we may conclude that

$$\begin{aligned}\tilde F(\tilde\omega) - F(\omega) &= \big(\tilde F(z) - F(z)\big) + \big(F(z) - F(\omega)\big) + \big(\tilde F(\tilde\omega) - \tilde F(z)\big)\\ &\geqslant \big(\tilde F(z) - F(z)\big) + \big(\tilde F(\tilde\omega) - \tilde F(z)\big), \text{ since } F(z) - F(\omega) \geqslant 0\\ &\geqslant \tfrac{1}{4}\alpha^{1/2} - \tfrac{1}{8}\alpha - 0.33(\delta\alpha)^{1/2}\\ &\geqslant \tfrac{1}{5}\alpha^{1/2} \quad \text{since } \alpha \leqslant \delta \leqslant 10^{-4}.\end{aligned}$$

Theorem 2 then follows from (8.7).

We shall now show (8.6). Note that $\tilde m = m + 1$, $\tilde a_i = a_i$, $1 \leqslant i \leqslant m$, $a_{\tilde m} = c$, $\tilde b_i = b_i$, $1 \leqslant i \leqslant m$, and $b_{\tilde m} = \beta$. By Lemma 1,

$$\begin{aligned}-\nabla\tilde F(z) &= \sum_{i=1}^{\tilde m} \tilde\sigma_i(z)\frac{\tilde a_i}{\tilde a_i^{\mathrm T} z - \tilde b_i}\\ &= \sum_{i=1}^{m} \tilde\sigma_i(z)\frac{a_i}{a_i^{\mathrm T} z - b_i} + \frac{c^{\mathrm T}\tilde H(z)^{-1}c}{(c^{\mathrm T}z-\beta)^2}\frac{c}{c^{\mathrm T}z-\beta}\\ &= \sum_{i=1}^{m} \sigma_i(z)\frac{a_i}{a_i^{\mathrm T} z - b_i} + \sum_{i=1}^{m}\big(\tilde\sigma_i(z) - \sigma_i(z)\big)\frac{a_i}{a_i^{\mathrm T} z - b_i}\\ &\quad + \frac{c^{\mathrm T}\tilde H(z)^{-1}c}{(c^{\mathrm T}z-\beta)^2}\frac{c}{c^{\mathrm T}z-\beta}\\ &= -\nabla F(z) + u + \nu,\end{aligned}$$

where

$$u = \sum_{i=1}^{m}\big(\tilde\sigma_i(z) - \sigma_i(z)\big)\frac{a_i}{a_i^{\mathrm T} z - b_i} \quad \text{and} \quad \nu = \frac{c^{\mathrm T}\tilde H(z)^{-1}c}{(c^{\mathrm T}z-\beta)^2}\frac{c}{c^{\mathrm T}z-\beta}.$$

We shall obtain an upper bound on each of $\nabla F(z)^{\mathrm T}Q(z)^{-1}\nabla F(z)$, $u^{\mathrm T}Q(z)^{-1}u$ and $\nu^{\mathrm T}Q(z)^{-1}\nu$; these bounds will lead to a bound on $\nabla\tilde F(z)^{\mathrm T}Q(z)^{-1}\tilde F(z)$ which in turn will give a bound on $\nabla\tilde F(z)^{\mathrm T}\tilde Q(z)^{-1}\nabla\tilde F(z)$. Since $\alpha \leqslant \delta \leqslant 10^{-4}$ and $\mu(\omega) \leqslant 1$ we have that

$$F(z) - F(\omega) \leqslant \alpha^2\mu(\omega) \leqslant \delta\sqrt{\mu(\omega)}\,.$$

So by Lemma 10 we get that

$$\nabla F(z)^{\mathrm T}Q(z)^{-1}\nabla F(z) \leqslant \frac{1}{0.14}\big(F(z) - F(\omega)\big) \leqslant 7.2\,\alpha^2\mu(\omega)$$

and that

$$\mu(\omega) \leqslant 1.5\mu(z).$$

So

$$\begin{aligned}\nabla F(z)^{\mathrm T}Q(z)^{-1}\nabla F(z) &\leqslant 7.2\,\alpha^2\mu(\omega)\\ &\leqslant 11\,\alpha^2\mu(z)\\ &\leqslant 11\,\delta^{3/2}\alpha^{1/2}, \quad \text{since } \alpha \leqslant \delta \text{ and } \mu(z) \leqslant 1.\end{aligned}$$

Thus

$$\nabla F(z)^{\mathrm T}Q(z)^{-1}\nabla F(z) \leqslant 11\,\delta^{3/2}\alpha^{1/2}. \tag{8.9}$$

From Lemma 12 we get that

$$u^{\mathrm{T}}Q(z)^{-1}\mu \leqslant \frac{\left(c^{\mathrm{T}}H(z)^{-1}c\right)^2}{\left(c^{\mathrm{T}}z-\beta\right)^4} \leqslant \tfrac{1}{4}\alpha \leqslant \tfrac{1}{4}(\delta\alpha)^{1/2}, \quad \text{since } \alpha \leqslant \delta.$$

Thus

$$u^{\mathrm{T}}Q(z)^{-1}u \leqslant \tfrac{1}{4}(\delta\alpha)^{1/2}, \tag{8.10}$$

We have that

$$\nu^{\mathrm{T}}Q(z)^{-1}\nu = \frac{\left(c^{\mathrm{T}}\tilde{H}(z)^{-1}c\right)^2}{\left(c^{\mathrm{T}}z-\beta\right)^4}\frac{c^{\mathrm{T}}Q(z)^{-1}c}{\left(c^{\mathrm{T}}z-\beta\right)^2}. \tag{8.11}$$

By Lemma 11(b), we get that

$$c^{\mathrm{T}}\tilde{H}(z)^{-1}c \leqslant c^{\mathrm{T}}H(z)^{-1}c.$$

Furthermore, from the definition of $\mu(z)$ and Claim 2 it follows that

$$c^{\mathrm{T}}Q(z)^{-1}c \leqslant \frac{1}{\mu(z)}c^{\mathrm{T}}H(z)^{-1}c.$$

Thus from (8.11) we can conclude that

$$\nu^{\mathrm{T}}Q(z)^{-1}\nu \leqslant \frac{1}{\mu(z)}\frac{\left(c^{\mathrm{T}}H(z)^{-1}c\right)^3}{\left(c^{\mathrm{T}}z-\beta\right)^6} \leqslant \frac{\alpha^{3/2}}{8\mu(z)} \leqslant \tfrac{1}{8}\delta\alpha^{1/2}, \quad \text{since } \alpha \leqslant \delta\mu(z).$$

Thus

$$\nu^{\mathrm{T}}Q(z)^{-1}\nu \leqslant \tfrac{1}{8}\delta\alpha^{1/2}. \tag{8.12}$$

Using the relation $x^{\mathrm{T}}y \leqslant \|x\|_2\|y\|_2$, we get that

$$\begin{aligned} &\nabla\tilde{F}(z)^{\mathrm{T}}Q(z)^{-1}\nabla\tilde{F}(z) \\ &\quad = (-\nabla F(z)+u+\nu)^{\mathrm{T}}Q(z)^{-1}(-\nabla F(z)+u+\nu) \\ &\quad \leqslant \left(\left(\nabla F(z)^{\mathrm{T}}Q(z)^{-1}\nabla F(z)\right)^{1/2}+\left(u^{\mathrm{T}}Q(z)^{-1}u\right)^{1/2}+\left(\nu^{\mathrm{T}}Q(z)^{-1}\nu\right)^{1/2}\right)^2. \end{aligned} \tag{8.13}$$

From (8.9), (8.10), (8.12), (8.13), and the assumption that $\delta \leqslant 10^{-4}$ we can conclude that

$$\nabla\tilde{F}(z)^{\mathrm{T}}Q(z)^{-1}\nabla\tilde{F}(z) \leqslant 0.5(\delta\alpha)^{1/2}. \tag{8.14}$$

By Lemma 13, (with $\tau = 0.5\alpha^{1}/2$), we get that

$$\nabla\tilde{F}(z)^{\mathrm{T}}\tilde{Q}(z)^{-1}\nabla\tilde{F}(z) \leqslant \frac{\nabla\tilde{F}(z)^{\mathrm{T}}Q(z)^{-1}\nabla\tilde{F}(z)}{(1-0.5\alpha^{1/2})}, \tag{8.15}$$

and that

$$\alpha \leqslant \delta\mu(z) \leqslant \frac{(1+0.5\alpha^{1/2})\delta\tilde{\mu}(z)}{(1-0.5\alpha^{1/2})}. \tag{8.16}$$

From (8.14)–(8.16) and the assumption that $\alpha \leqslant 10^{-4}$, it follows that

$$\nabla\tilde{F}(z)^{\mathrm{T}}\tilde{Q}(z)^{-1}\nabla\tilde{F}(z) \leqslant 0.6(\delta\alpha)^{1/2} \leqslant 0.66\delta\sqrt{\tilde{\mu}(z)}\,.$$

That concludes the proof of (8.6). □

Proof of Theorem 4. By Lemma 11(a), we get that

$$\forall\xi\in\mathbb{R}^{n},\quad \xi^{\mathrm{T}}H(z)\xi \geqslant \left(1-\frac{c^{\mathrm{T}}\tilde{H}(z)^{-1}c}{(c^{\mathrm{T}}z-\beta)^{2}}\right)\xi^{\mathrm{T}}\tilde{H}(z)\xi.$$

Then since $c^{\mathrm{T}}\tilde{H}(z)^{-1}c/(c^{\mathrm{T}}z-\beta)^{2} < 1$, it follows that $H(z)$ is nonsingular and that A has linearly independent columns. We shall show below that

$$\nabla F(z)^{\mathrm{T}}Q(z)^{-1}\nabla F(z) \leqslant \min\left\{5\alpha\mu(z),\ \delta\sqrt{\mu(z)}\right\}. \tag{8.17}$$

Before proving (8.17) we shall complete the proof of the theorem using (8.17). Since A has linearly independent columns, from (8.17) and Lemma 14 we can conclude that the polytope P is bounded. Then by (8.17) and Lemma 9 we get that

$$F(z) - F(\omega) \leqslant 0.55\,\nabla F(z)^{\mathrm{T}}Q(z)^{-1}\nabla F(z)$$

and that

$$\mu(z) \leqslant 1.1\mu(\omega).$$

So from (8.17) we can conclude that

$$F(z) - F(\omega) \leqslant \min\left\{4\alpha\mu(\omega),\ \delta\sqrt{\mu(\omega)}\right\}. \tag{8.18}$$

From Lemma 11(e), we have that

$$\tilde{F}(z) - F(z) = -\tfrac{1}{2}\ln\left(1-\frac{c^{\mathrm{T}}\tilde{H}(z)^{-1}c}{(c^{\mathrm{T}}z-\beta)^{2}}\right) \leqslant \frac{\alpha}{2(1-\alpha)}. \tag{8.19}$$

Then from (8.18) and (8.19) it follows that

$$\begin{aligned}\tilde{F}(\tilde{\omega}) - F(\omega) &= \left(\tilde{F}(z)-F(z)\right) + (F(z)-F(\omega)) + \left(\tilde{F}(\tilde{\omega})-\tilde{F}(z)\right)\\ &\leqslant \left(\tilde{F}(z)-F(z)\right) + (F(z)-F(\omega)) \quad \left(\text{since } \tilde{F}(\tilde{\omega})-\tilde{F}(z)\leqslant 0\right)\\ &\leqslant \frac{\alpha}{2(1-\alpha)} + 4\alpha\mu(\omega)\\ &\leqslant 5\alpha \quad \left(\text{since } \alpha\leqslant 10^{-4},\ \mu(\omega)\leqslant 1\right).\end{aligned}$$

Theorem 4 then follows from (8.18).

Next, we shall relate $\mu(z)$ and $\tilde{\mu}(z)$, $\xi^{\mathrm{T}}Q(z)^{-1}\xi$ and $\xi^{\mathrm{T}}\tilde{Q}(z)^{-1}\xi$; these relations will be useful in proving (8.17). Let $\tau = (1/(1-\alpha))\min\{\alpha, \tilde{\mu}(z)\}$. By Lemma 11(b), we have that

$$\begin{aligned}\frac{c^{\mathrm{T}}H(z)^{-1}c}{(c^{\mathrm{T}}z-\beta)^2} &= \frac{(c^{\mathrm{T}}z-\beta)^2}{(c^{\mathrm{T}}z-\beta)^2 - c^{\mathrm{T}}\tilde{H}(z)^{-1}c}\,\frac{c^{\mathrm{T}}\tilde{H}(z)^{-1}c}{(c^{\mathrm{T}}z-\beta)^2}\\ &= \left(1-\frac{c^{\mathrm{T}}\tilde{H}(z)^{-1}c}{(c^{\mathrm{T}}z-\beta)^2}\right)^{-1}\frac{c^{\mathrm{T}}\tilde{H}(z)^{-1}c}{(c^{\mathrm{T}}z-\beta)^2}\\ &\leqslant \frac{1}{1-\alpha}\min\{\alpha, \tilde{\mu}(z)\}\\ &\leqslant \tau.\end{aligned}$$

Thus

$$\frac{c^{\mathrm{T}}H(z)^{-1}c}{(c^{\mathrm{T}}z-\beta)^2} \leqslant \tau. \tag{8.20}$$

Then by Lemma 13 we get that

$$\begin{aligned}\tilde{\mu}(z) &\leqslant \mu(z) + \tau^2\\ &\leqslant \mu(z) + \frac{\tau\tilde{\mu}(z)}{1-\alpha} \quad (\text{by def. of } \tau).\end{aligned}$$

Thus

$$\tilde{\mu}(z) \leqslant \frac{1-\alpha}{1-\alpha-\tau}\, m(z). \tag{8.21}$$

Also, by Lemma 13, for all $\xi \in \mathbb{R}^n$

$$\begin{aligned}\xi^{\mathrm{T}}Q(z)^{-1}\xi &\leqslant \left(1+\frac{\tau^2}{\mu(z)}\right)\xi^{\mathrm{T}}\tilde{Q}(z)^{-1}\xi\\ &\leqslant \left(1+\frac{\tau^2(1-\alpha)}{(1-\alpha-\tau)\tilde{\mu}(z)}\right)\xi^{\mathrm{T}}\tilde{Q}(z)^{-1}\xi \quad (\text{by } 8.21)\\ &\leqslant \left(1+\frac{\tau}{1-\alpha-\tau}\right)\xi^{\mathrm{T}}\tilde{Q}(z)^{-1}\xi \quad (\text{by def. of } \tau).\end{aligned}$$

Thus

$$\forall \xi \in \mathbb{R}^n, \quad \xi^{\mathrm{T}}Q(z)^{-1}\xi \leqslant \frac{1-\alpha}{1-\alpha-\tau}\xi^{\mathrm{T}}\tilde{Q}(z)^{-1}\xi. \tag{8.22}$$

We shall now show (8.17). Note that $\tilde{m} = m+1$, $\tilde{a}_i = a_i$, $1 \leqslant i \leqslant m$, $a_{\tilde{m}} = c$, $\tilde{b}_i = b_i$, $1 \leqslant i \leqslant m$, $b_{\tilde{m}} = \beta$, and $\tilde{\sigma}_{\tilde{m}}(x) = c^{\mathrm{T}}\tilde{H}(x)^{-1}c/(c^{\mathrm{T}}x-\beta)^2$. By Lemma 1,

$$\begin{aligned}
-\nabla F(z) &= \sum_{i=1}^{m} \sigma_i(z)\frac{a_i}{a_i^{\mathrm{T}}z-b_i} \\
&= \sum_{i=1}^{m} \tilde{\sigma}_i(z)\frac{a_i}{a_i^{\mathrm{T}}z-b_i} - \sum_{i=1}^{m}\left(\tilde{\sigma}_i(z)-\sigma_i(z)\right)\frac{a_i}{a_i^{\mathrm{T}}z-b_i} \\
&= \sum_{i=1}^{\tilde{m}} \tilde{\sigma}_i(z)\frac{\tilde{a}_i}{\tilde{a}_i^{\mathrm{T}}z-\tilde{b}_i} - \sum_{i=1}^{m}\left(\tilde{\sigma}_i(z)-\sigma_i(z)\right)\frac{a_i}{a_i^{\mathrm{T}}z-b_i} \\
&\quad - \frac{c^{\mathrm{T}}\tilde{H}(z)^{-1}c}{(c^{\mathrm{T}}z-\beta)^2}\frac{c}{c^{\mathrm{T}}z-\beta} \\
&= -\left(\nabla\tilde{F}(z)+u+\nu\right),
\end{aligned}$$

where

$$u = \sum_{i=1}^{m}\left(\tilde{\sigma}_i(z)-\sigma_i(z)\right)\frac{a_i}{a_i^{\mathrm{T}}z-b_i} \quad \text{and} \quad \nu = \frac{c^{\mathrm{T}}\tilde{H}(z)^{-1}c}{(c^{\mathrm{T}}z-\beta)^2}\frac{c}{c^{\mathrm{T}}z-\beta}.$$

To obtain a bound on $\nabla F(z)^{\mathrm{T}}Q(z)^{-1}\nabla F(z)$ we shall upper bound each of $\nabla\tilde{F}(z)^{\mathrm{T}}Q(z)^{-1}\nabla\tilde{F}(z)$, $u^{\mathrm{T}}Q(z)^{-1}u$ and $\nu^{\mathrm{T}}Q(z)^{-1}\nu$.

Since $\alpha \leqslant \delta$ and $\tilde{\mu}(\tilde{\omega}) \leqslant 1$ we have that

$$\tilde{F}(z) - \tilde{F}(\tilde{\omega}) \leqslant \alpha^2\tilde{\mu}(\tilde{\omega}) \leqslant \delta\sqrt{\tilde{\mu}(\tilde{\omega})}\,.$$

So from Lemma 10 it follows that

$$\nabla\tilde{F}(z)^{\mathrm{T}}\tilde{Q}(z)^{-1}\nabla\tilde{F}(z) \leqslant \frac{1}{0.14}\left(\tilde{F}(z)-\tilde{F}(\tilde{\omega})\right) \leqslant 7.2\,\alpha^2\tilde{\mu}(\tilde{\omega}),$$

and that

$$\tilde{\mu}(\tilde{\omega}) \leqslant 1.5\tilde{\mu}(z).$$

So

$$\begin{aligned}
\nabla\tilde{F}(z)^{\mathrm{T}}\tilde{Q}(z)^{-1}\nabla\tilde{F}(z) &\leqslant 7.2\,\alpha^2\tilde{\mu}(\tilde{\omega}) \\
&\leqslant 11\,\alpha^2\tilde{\mu}(z) \\
&\leqslant \frac{11\,\alpha^2(1-\alpha)\mu(z)}{1-\alpha-\tau} \quad \text{(by 8.21)}.
\end{aligned}$$

Thus

$$\begin{aligned}
\nabla\tilde{F}(z)^{\mathrm{T}}Q(z)^{-1}\nabla\tilde{F}(z) &\leqslant \frac{1-\alpha}{1-\alpha-\tau}\nabla\tilde{F}(z)^{\mathrm{T}}\tilde{Q}(z)^{-1}\nabla\tilde{F}(z) \quad \text{(by 8.22)} \\
&\leqslant \frac{11\,\alpha^2(1-\alpha)^2\mu(z)}{(1-\alpha-\tau)^2}.
\end{aligned}$$

Then noting that $\alpha \leqslant \delta \leqslant 10^{-4}$, $\tau \leqslant \alpha/(1-\alpha)$, and $\mu(z) \leqslant 1$, we get that

$$\nabla \tilde{F}(z)^{\mathrm{T}} Q(z)^{-1} \nabla \tilde{F}(z) \leqslant 12 \min\left\{\alpha\delta\mu(z),\ \delta^2\sqrt{\mu(z)}\right\}. \tag{8.23}$$

From Lemma 12 we have that

$$\begin{aligned} u^{\mathrm{T}} Q(z)^{-1} u &\leqslant \frac{\left(c^{\mathrm{T}} H(z)^{-1} c\right)^2}{\left(c^{\mathrm{T}} z - \beta\right)^4} \\ &\leqslant \tau^2 \quad \text{(by 8.20)} \\ &\leqslant \frac{1}{(1-\alpha)^2} \min\left\{\alpha\tilde{\mu}(z),\ \alpha^{3/2}\sqrt{\tilde{\mu}(z)}\right\} \quad \text{(by def. of } \tau) \\ &\leqslant \frac{1}{(1-\alpha-\tau)(1-\alpha)} \min\left\{\alpha\mu(z),\ \alpha^{3/2}\sqrt{\mu(z)}\right\} \quad \text{(by 8.21)}. \end{aligned}$$

Then noting that $\alpha \leqslant \delta \leqslant 10^{-4}$, $\tau \leqslant \alpha/(1-\alpha)$, we get that

$$u^{\mathrm{T}} Q(z)^{-1} u \leqslant 1.01 \min\left\{\alpha\mu(z),\ \delta^{3/2}\sqrt{\mu(z)}\right\}. \tag{8.24}$$

Next, we have that

$$\nu^{\mathrm{T}} Q(z)^{-1} \nu = \frac{\left(c^{\mathrm{T}} \tilde{H}(z)^{-1} c\right)^2}{\left(c^{\mathrm{T}} z - \beta\right)^4} \frac{c^{\mathrm{T}} Q(z)^{-1} c}{\left(c^{\mathrm{T}} z - \beta\right)^2}.$$

From the definition of $\mu(z)$ and Claim 2 it follows that

$$\frac{c^{\mathrm{T}} Q(z)^{-1} c}{\left(c^{\mathrm{T}} z - \beta\right)^2} \leqslant \frac{1}{\mu(z)} \frac{c^{\mathrm{T}} H(z)^{-1} c}{\left(c^{\mathrm{T}} z - \beta\right)^2}$$

and by Lemma 11(b),

$$c^{\mathrm{T}} \tilde{H}(z)^{-1} c \leqslant c^{\mathrm{T}} H(z)^{-1} c.$$

Thus

$$\begin{aligned} \nu^{\mathrm{T}} Q(z)^{-1} \nu &\leqslant \frac{\left(c^{\mathrm{T}} H(z)^{-1} c\right)^3}{\mu(z)\left(c^{\mathrm{T}} z - \beta\right)^6} \\ &\leqslant \frac{\tau^3}{\mu(z)} \quad \text{(by 8.20)} \\ &\leqslant \frac{\tilde{\mu}(z)}{(1-\alpha)^3 \mu(z)} \min\left\{\alpha\tilde{\mu}(z),\ \alpha^{3/2}\sqrt{\tilde{\mu}(z)}\right\} \quad \text{(by def. of } \tau) \\ &\leqslant \frac{1}{(1-\alpha)(1-\alpha-\tau)^2} \min\left\{\alpha\mu(z),\ \alpha^{3/2}\sqrt{\mu(z)}\right\} \quad \text{(by 8.21)}. \end{aligned}$$

Then noting that $\alpha \leqslant \delta \leqslant 10^{-4}$, $\tau \leqslant \alpha/(1-\alpha)$, we get that

$$\nu^{\mathrm{T}} Q(z)^{-1} \nu \leqslant 1.01 \min\left\{\alpha\mu(z),\ \delta^{3/2}\sqrt{\mu(z)}\right\}. \tag{8.25}$$

Using the relation $x^{\mathrm{T}}y \leqslant \|x\|_2 \|y\|_2$ we get that

$$\begin{aligned}&\nabla F(z)^{\mathrm{T}} Q(z)^{-1} \nabla F(z)\\&\quad= \left(\nabla \tilde{F}(z) + u + \nu\right)^{\mathrm{T}} Q(z)^{-1}\left(\nabla \tilde{F}(z) + u + \nu\right)\\&\quad\leqslant \left(\left(\nabla \tilde{F}(z)^{\mathrm{T}} Q(z)^{-1} \nabla \tilde{F}(z)\right)^{1/2} + \left(u^{\mathrm{T}} Q(z)^{-1} u\right)^{1/2} + \left(\nu^{\mathrm{T}} Q(z)^{-1} \nu\right)^{1/2}\right)^2.\end{aligned} \tag{8.26}$$

Then from (8.23)–(8.26) and the observation that $\delta \leqslant 10^{-4}$, it follows that

$$\nabla F(z)^{\mathrm{T}} Q(z)^{-1} \nabla F(z) \leqslant \min\left\{5\alpha\mu(z),\ \delta\sqrt{\mu(z)}\right\}$$

(8.17) then follows. □

8.1. Proofs of Lemmas 11 to 14

In this section we shall prove Lemmas 11 to 14.

Proof of Lemma 11. $H(z)$ may be expressed as

$$\begin{aligned}H(z) &= \tilde{H}(z) - \frac{cc^{\mathrm{T}}}{(c^{\mathrm{T}}z - \beta)^2}\\&= \tilde{H}(z)^{1/2}\left(I - \frac{\tilde{H}(z)^{-1} cc^{\mathrm{T}} \tilde{H}(z)^{-1/2}}{(c^{\mathrm{T}}z - \beta)^2}\right)\tilde{H}(z)^{1/2}.\end{aligned} \tag{8.27}$$

Thus

$$\forall \xi \in \mathbb{R}^n, \quad \xi^{\mathrm{T}}\tilde{H}(z)\xi \geqslant \xi^{\mathrm{T}}H(z)\xi \geqslant \xi^{\mathrm{T}}\tilde{H}(z)\xi\left(1 - \frac{c^{\mathrm{T}}\tilde{H}(z)^{-1}c}{(c^{\mathrm{T}}z - \beta)^2}\right). \tag{8.28}$$

(a) follows from (8.28).

Next, note that $\tilde{H}(z)^{-1}$ may be expressed as

$$\tilde{H}(z)^{-1} = H(z)^{-1} - \frac{H(z)^{-1}cc^{\mathrm{T}}H(z)^{-1}}{(c^{\mathrm{T}}z - \beta)^2 + c^{\mathrm{T}}H(z)^{-1}c} \tag{8.29}$$

(b) follows directly from (8.29).

We have that for $1 \leqslant i \leqslant m$,

$$\begin{aligned}\sigma_i(z) - \tilde{\sigma}_i(z) &= \frac{a_i^{\mathrm{T}}H(z)^{-1}a_i}{(a_i^{\mathrm{T}}z - b_i)^2} - \frac{\tilde{a}_i^{\mathrm{T}}\tilde{H}(z)^{-1}\tilde{a}_i}{(\tilde{a}_i^{\mathrm{T}}z - \tilde{b}_i)^2}\\&= \frac{a_i^{\mathrm{T}}H(z)^{-1}a_i - a_i^{\mathrm{T}}\tilde{H}(z)^{-1}a_i}{(a_i^{\mathrm{T}}z - b_i)^2}.\end{aligned}$$

From (8.29) it follows that

$$a_i^{\mathrm{T}}H(z)^{-1}a_i - a_i^{\mathrm{T}}\tilde{H}(z)^{-1}a_i = \frac{(a_i^{\mathrm{T}}H(z)^{-1}c)^2}{(c^{\mathrm{T}}z - \beta)^2 + c^{\mathrm{T}}H(z)^{-1}c}.$$

We can then conclude that

$$\sigma_i(z) - \tilde{\sigma}_i(z) = \frac{\left(a_i^{\mathrm{T}} H(z)^{-1} c\right)^2}{\left((c^{\mathrm{T}} z - \beta)^2 + c^{\mathrm{T}} H(z)^{-1} c\right)\left(a_i^{\mathrm{T}} z - b_i\right)^2}. \tag{8.30}$$

(c) follows from (8.30).

Next, since $x^{\mathrm{T}} y \leqslant \| x \|_2 \| y \|_2$, we have that

$$\begin{aligned} \frac{\left(a_i^{\mathrm{T}} H(z)^{-1} c\right)^2}{\left(a_i^{\mathrm{T}} z - b_i\right)^2} &\leqslant \frac{\left(a_i^{\mathrm{T}} H(z)^{-1} a_i\right)\left(c^{\mathrm{T}} H(z)^{-1} c\right)}{\left(a_i^{\mathrm{T}} z - b_i\right)^2} \\ &= \sigma_i(z) c^{\mathrm{T}} H(z)^{-1} c. \end{aligned} \tag{8.31}$$

(d) follows from (8.30) and (8.31).

Finally, from (8.27) we get that

$$\begin{aligned} \det(H(z)) &= \det\left(\tilde{H}(z)\right) \det\left(I - \frac{\tilde{H}(z)^{-1/2} c c^{\mathrm{T}} \tilde{H}(z)^{-1/2}}{\left(c^{\mathrm{T}} z - \beta\right)^2} \right) \\ &= \det\left(\tilde{H}(z)\right)\left(1 - \frac{c^{\mathrm{T}} \tilde{H}(z)^{-1} c}{\left(c^{\mathrm{T}} z - \beta\right)^2} \right) \end{aligned}$$

and thus

$$\tilde{F}(z) - F(z) = -\tfrac{1}{2} \ln\left(1 - \frac{c^{\mathrm{T}} \tilde{H}(z)^{-1} c}{\left(c^{\mathrm{T}} z - \beta\right)^2} \right).$$

(e) then follows from the relation between $c^{\mathrm{T}}\tilde{H}(z)^{-1}c$ and $c^{\mathrm{T}}H(z)^{-1}c$ given by (b). □

Proof of Lemma 12. By Lemma 11(c), we have that for $1 \leqslant i \leqslant m$,

$$\tilde{\sigma}_i(z) - \sigma_i(z) = \frac{-\left(a_i^{\mathrm{T}} H(z)^{-1} c\right)^2}{\left((c^{\mathrm{T}} z - \beta)^2 + c^{\mathrm{T}} H(z)^{-1} c\right)\left(a_i^{\mathrm{T}} z - b_i\right)^2}. \tag{8.32}$$

Next, since $x^{\mathrm{T}} y \leqslant \| x_2 \| y \|_2$, we have that

$$\begin{aligned} \frac{\left(a_i^{\mathrm{T}} H(z)^{-1} c\right)^2}{\left(a_i^{\mathrm{T}} z - b_i\right)^2} &\leqslant \frac{\left(a_i^{\mathrm{T}} H(z)^{-1} a_i\right)\left(c^{\mathrm{T}} H(z)^{-1} c\right)}{\left(a_i^{\mathrm{T}} z - b_i\right)^2} \\ &= \sigma_i(z) c^{\mathrm{T}} H(z)^{-1} c. \end{aligned} \tag{8.33}$$

Let D be an $m \times m$ diagonal matrix such that the ith diagonal entry D_{ii} is given by $D_{ii} = \sqrt{\sigma_i(z)} / (a_i^{\mathrm{T}} z - b_i)$. Then $Q(z)$ may be written as

$$Q(z) = A^{\mathrm{T}} D^2 A,$$

and u may be written as

$$u = A^{\mathrm{T}} D \bar{u},$$

where $\bar{u}^{\mathrm{T}} = (\bar{u}_1, \ldots, \bar{u}_m)$ and

$$\bar{u}_i = \frac{\tilde{\sigma}_i(z) - \sigma_i(z)}{\sqrt{\sigma_i(z)}}, \quad 1 \leqslant i \leqslant m.$$

Thus we may express $u^{\mathrm{T}}Q(z)^{-1}$ as

$$u^{\mathrm{T}}Q(z)^{-1}u = \bar{u}^{\mathrm{T}}DA(A^{\mathrm{T}}D^2A)^{-1}A^{\mathrm{T}}D\bar{u}.$$

Moreover, since $\| DA(A^{\mathrm{T}}D_2A)^{-1}A^{\mathrm{T}}D\|_2 \leqslant 1$ we get that

$$u^{\mathrm{T}}Q(z)^{-1}u \leqslant \bar{u}^{\mathrm{T}}\bar{u}.$$

Thus

$$\begin{aligned} u^{\mathrm{T}}Q(z)^{-1}u &\leqslant \bar{u}^{\mathrm{T}}\bar{u} \\ &= \sum_{i=1}^{m}\left(\frac{\tilde{\sigma}_i(z) - \sigma_i(z)}{\sqrt{\sigma_i(z)}}\right)^2 \\ &\leqslant \sum_{i=1}^{m} \frac{\left(a_i^{\mathrm{T}}H(z)^{-1}c\right)^2}{\left(c^{\mathrm{T}}z-\beta\right)^4\left(a_i^{\mathrm{T}}z-b_i\right)^2\sigma_i(z)} \frac{\left(a_i^{\mathrm{T}}H(z)^{-1}c\right)^2}{\left(a_i^{\mathrm{T}}z-b_i\right)^2} \quad \text{(by 8.32)} \\ &\leqslant \sum_{i=1}^{m} \frac{\left(a_i^{\mathrm{T}}H(z)^{-1}c\right)^2}{\left(c^{\mathrm{T}}z-\beta\right)^4\left(a_i^{\mathrm{T}}z-b_i\right)^2\sigma_i(z)}\sigma_i(z)c^{\mathrm{T}}H(z)^{-1}c \quad \text{(by 8.33)} \\ &\leqslant \frac{c^{\mathrm{T}}H(z)^{-1}c}{\left(c^{\mathrm{T}}z-\beta\right)^4}\sum_{i=1}^{m}\frac{\left(a_i^{\mathrm{T}}H(z)^{-1}c\right)^2}{\left(a_i^{\mathrm{T}}z-b_i\right)^2}. \end{aligned}$$

Note that

$$c^{\mathrm{T}}H(z)^{-1}c = c^{\mathrm{T}}H(z)^{-1}H(z)H(z)^{-1}c = \sum_{i=1}^{m}\frac{\left(a_i^{\mathrm{T}}H(z)^{-1}c\right)^2}{\left(a_i^{\mathrm{T}}z-b_i\right)^2}.$$

So we can conclude that

$$u^{\mathrm{T}}Q(z)^{-1}u \leqslant \frac{\left(c^{\mathrm{T}}H(z)^{-1}c\right)^2}{\left(c^{\mathrm{T}}z-\beta\right)^4}. \quad \square$$

Proof of Lemma 13. From (a) and (b) in Lemma 11 it follows that for all $\xi \in \mathbb{R}^n$,

$$\xi^{\mathrm{T}}\tilde{H}(z)\xi \leqslant \left(1 + \frac{c^{\mathrm{T}}H(z)^{-1}c}{\left(c^{\mathrm{T}}z-\beta\right)^2}\right)\xi^{\mathrm{T}}H(z)\xi \leqslant (1+\tau)\xi^{\mathrm{T}}H(z)\xi. \tag{8.34}$$

Since $\tilde{A}_i = a_i$, $\tilde{b}_i = b_i$, $1 \leqslant i \leqslant m$, from the definition of $\tilde{Q}(z)$ we get that

$$\xi^{\mathrm{T}}\tilde{Q}(z)\xi \geqslant \sum_{i=1}^{m}\tilde{\sigma}_i(z)\frac{\left(a_i^{\mathrm{T}}\xi\right)^2}{\left(a_i^{\mathrm{T}}z-b_i\right)^2}. \tag{8.35}$$

By Lemma 11(e), we get that

$$\tilde{\sigma}_i(z) \geqslant \left(1 + \frac{c^{\mathrm{T}} H(z)^{-1} c}{(c^{\mathrm{T}} z - \beta)^2}\right) \sigma_i(z) \geqslant (1-\tau)\sigma_i(z).$$

So from (8.35) it follows that

$$\xi^{\mathrm{T}} \tilde{Q}(z) \xi \geqslant (1-\tau) \sum_{i=1}^{m} \sigma_i(z) \frac{(a_i^{\mathrm{T}} \xi)^2}{(a_i^{\mathrm{T}} z - b_i)^2} = (1-\tau) \xi^{\mathrm{T}} Q(z) \xi. \tag{8.36}$$

From (8.36) and Claim 2 it follows that for all $\xi \in \mathbb{R}^n$,

$$\xi^{\mathrm{T}} \tilde{Q}(z)^{-1} \xi \leqslant \frac{1}{1-\tau} \xi^{\mathrm{T}} Q(z)^{-1} \xi. \tag{8.37}$$

$\tilde{\mu}(z)$ is lower bounded as follows. We have that for all $\xi \in \mathbb{R}^n$,

$$\begin{aligned} \xi^{\mathrm{T}} \tilde{Q}(z) \xi &\geqslant (1-\tau) \xi^{\mathrm{T}} Q(z) \xi \quad \text{(by 8.36)} \\ &\geqslant \mu(z)(1-\tau) \xi^{\mathrm{T}} H(z) \xi \quad \text{(by def. of } \mu(z)) \\ &\geqslant \frac{\mu(z)(1-\tau)}{(1+\tau)} \xi^{\mathrm{T}} \tilde{H}(z) \xi \quad \text{(by 8.34)}. \end{aligned}$$

Hence

$$\tilde{\mu}(z) \geqslant \frac{\mu(z)(1-\tau)}{(1+\tau)}. \tag{8.38}$$

We have that

$$\xi^{\mathrm{T}} \tilde{Q}(z) \xi = \sum_{i=1}^{m} \tilde{\sigma}_i(z) \frac{(a_i^{\mathrm{T}} \xi)^2}{(a_i^{\mathrm{T}} z - b_i)^2} + \frac{c^{\mathrm{T}} \tilde{H}(z)^{-1} c}{(c^{\mathrm{T}} z - \beta)^2} \frac{(c^{\mathrm{T}} \xi)^2}{(c^{\mathrm{T}} z - \beta)^2}. \tag{8.39}$$

By Lemma 11(b), $c^{\mathrm{T}} \tilde{H}(z)^{-1} c \leqslant c^{\mathrm{T}} H(z)^{-1} c$, and by Lemma 11(d), $\tilde{\sigma}_i(z) \leqslant \sigma_i(z)$, $1 \leqslant i \leqslant m$. So from (8.39) it follows that

$$\begin{aligned} \xi^{\mathrm{T}} \tilde{Q}(z) \xi &\leqslant \sum_{i=1}^{m} \sigma_i(z) \frac{(a_i^{\mathrm{T}} \xi)^2}{(a_i^{\mathrm{T}} z - b_i)^2} + \frac{c^{\mathrm{T}} H(z)^{-1} c}{(c^{\mathrm{T}} z - \beta)^2} \frac{(c^{\mathrm{T}} \xi)^2}{(c^{\mathrm{T}} z - \beta)^2} \\ &= \xi^{\mathrm{T}} \left(Q(z) + \frac{c^{\mathrm{T}} H(z)^{-1} c}{(c^{\mathrm{T}} z - \beta)^2} \frac{c c^{\mathrm{T}}}{(c^{\mathrm{T}} z - \beta)^2} \right) \xi \\ &= \xi^{\mathrm{T}} Q(z)^{1/2} \left(I + \frac{c^{\mathrm{T}} H(z)^{-1} c}{(c^{\mathrm{T}} z - \beta)^2} \frac{Q(z)^{-1/2} c c^{\mathrm{T}} Q(z)^{-1/2}}{(c^{\mathrm{T}} z - \beta)^2} \right) Q(z)^{1/2} \xi \\ &\leqslant \xi^{\mathrm{T}} Q(z) \xi \left(1 + \frac{c^{\mathrm{T}} H(z)^{-1} c}{(c^{\mathrm{T}} z - \beta)^2} \frac{c^{\mathrm{T}} Q(z)^{-1} c}{(c^{\mathrm{T}} z - \beta)^2} \right). \end{aligned}$$

Since $\forall \xi \in \mathbb{R}^n$, $\xi^{\mathrm{T}} Q(z) \xi \geqslant \mu(z) \xi^{\mathrm{T}} H(z) \xi$, by Claim 2 we get that $c^{\mathrm{T}} Q(z)^{-1} c \leqslant (1/\mu(z)) c^{\mathrm{T}} H(z)^{-1} c$. It then follows that for all $\xi \in \mathbb{R}^n$,

$$\xi^{\mathrm{T}} \tilde{Q}(z) \xi \leqslant \xi^{\mathrm{T}} Q(z) \xi \left(1 + \frac{\left(c^{\mathrm{T}} H(z)^{-1} c\right)^2}{\mu(z)\left(c^{\mathrm{T}} z - \beta\right)^4} \right)$$

$$\leqslant \left(1 + \frac{\tau^2}{\mu(z)} \right) \xi^{\mathrm{T}} Q(z) \xi. \tag{8.40}$$

So from Claim 2 we can conclude that for all $\xi \in \mathbb{R}^n$,

$$\xi^{\mathrm{T}} \tilde{Q}(z)^{-1} \xi \geqslant \frac{1}{\left(1 + \tau^2/\mu(z)\right)} \xi^{\mathrm{T}} Q(z)^{-1} \xi. \tag{8.41}$$

Furthermore,

$$\xi^{\mathrm{T}} Q(z) \xi \geqslant \frac{1}{\left(1 + \tau^2/\mu(z)\right)} \xi^{\mathrm{T}} \tilde{Q}(z) \xi \quad \text{(by 8.40)}$$

$$\geqslant \frac{\tilde{\mu}(z)}{\left(1 + \tau^2/\mu(z)\right)} \xi^{\mathrm{T}} \tilde{H}(z) \xi \quad \text{(by def. of } \tilde{\mu}(z))$$

$$\geqslant \frac{\tilde{\mu}(z)}{\left(1 + \tau^2/\mu(z)\right)} \xi^{\mathrm{T}} H(z) \xi \quad \text{(by Lemma 11(a))}.$$

Thus

$$\mu(z) \geqslant \frac{\tilde{\mu}(z)}{\left(1 + \tau^2/\mu(z)\right)}. \tag{8.42}$$

Lemma 13 follows from (8.37), (8.38), (8.41) and (8.42). □

Proof of Lemma 14. We shall first prove a useful property about positive linear combinations, namely Property PLC.

Property PLC. Let $G = [g_1, g_2, \ldots, g_p]$, $g_j \in \mathbb{R}^n$, $1 \leqslant j \leqslant p$, let the columns of G span $\mathbb{R}^n$, and let $\lambda > 0$ be a lower bound on the smallest eigenvalue of GG^{T}. Suppose that $\| \sum_{j=1}^{p} g_j \|_2 \leqslant \sqrt{\lambda}/(2pn)$. Then

$$\forall \xi \in \mathbb{R}^n \text{ s.t. } \| \xi \|_2 = 1, \quad \min_{1 \leqslant j \leqslant p} \left\{ g_j^{\mathrm{T}} \xi \right\} \leqslant -\frac{\sqrt{\lambda}}{4p}.$$

Let $u = G^{\mathrm{T}} \xi$ where ξ is some unit vector in $\mathbb{R}^n$. Then $\| u \|_2 \geqslant \sqrt{\lambda}$ and

$$\sum_{j=1}^{p} | u_j | \geqslant \sqrt{\lambda}. \tag{8.43}$$

We have that

$$\sum_{j=1}^{p} u_j = \xi^{\mathrm{T}} \left(\sum_{j=1}^{p} g_j \right) \leqslant \left\| \sum_{j=1}^{p} g_j \right\|_2 \leqslant \frac{\sqrt{\lambda}}{2pn}.$$

Thus

$$\sum_{j=1}^{p} u_j = \sum_{u_j>0} |u_j| - \sum_{u_j<0} |u_j| \leqslant \frac{\sqrt{\lambda}}{2pn}.$$

So we get that

$$\sum_{j=1}^{p} |u_j| \leqslant 2 \sum_{u_j<0} |u_j| + \frac{\sqrt{\lambda}}{2pn}.$$

From (8.43) it then follows that

$$\sum_{u_j<0} |u_j| \geqslant \frac{\sqrt{\lambda}}{2}\left(1 - \frac{1}{2pn}\right) \geqslant \frac{\sqrt{\lambda}}{4}.$$

Then noting that $u_j = g_j^{\mathrm{T}}\xi$ we get that

$$\min_{1\leqslant j\leqslant p} \{g_j^{\mathrm{T}}\xi\} = -\max_{u_j<0}\{|u_j|\} \leqslant -\frac{\sqrt{\lambda}}{4p}.$$

That concludes the proof of Property PLC.

Let $w = \nabla F(z)$ and let $r = 1.1\sqrt{\delta}$. We shall next show that the equation $\nabla F(x) = tw$ has a unique solution for $t > 0$ and that all points x satisfying $\nabla F(x) = tw$, $0 < t \leqslant 1$, lie in the region $\Sigma(z, r)$. (Note that $\Sigma(z, r) = \{x: |a_i^{\mathrm{T}}(x-z)/(a_i^{\mathrm{T}}z - b_i)| \leqslant r,\ 1 \leqslant i \leqslant m\}$.) Then the fact that for arbitrarily small positive t, the solution to $\nabla F(x) = tw$ lies in $\Sigma(z, r)$ will be used to show that the polytope is bounded.

Let $\psi_t(x) = F(x) - tw$, and let Interior(P) denote the interior of the polytope P. $\psi_t(x)$ has a minimum in Interior(P) for all $t > 0$ which is seen as follows. Since $\nabla F(z) - w = 0$, by Lemma 1 we have that

$$\sum_{i=1}^{m} \frac{\sigma_i(z)}{a_i^{\mathrm{T}}z - b_i} a_i + w = 0.$$

Note that A has linearly independent columns. So from Property PLC above (with $g_i = (\sigma_i(z)/(a_i^{\mathrm{T}}z - b_i))a_i$, $1 \leqslant i \leqslant m$, $g_{m+1} = w$) it follows that there exists a $\hat{\lambda} > 0$ such that

$$\forall \xi \in \mathbb{R}^n,\quad \|\xi\|_2 = 1,\quad \min\left\{\min_{1\leqslant i\leqslant m}\left\{\frac{\sigma_i(z)}{a_i^{\mathrm{T}}z - b_i} a_i^{\mathrm{T}}\xi\right\},\ w^{\mathrm{T}}\xi\right\} \leqslant -\hat{\lambda}.$$

Thus if we move along any direction from z either the distance to some boundary of P must decrease at some minimum rate or the function $tw^{\mathrm{T}}x$, $t > 0$, must decrease at some minimum rate (depending on t). So for each point x in Interior(P) that is either sufficiently close to a boundary of P or outside a sufficiently large sphere centered at z, $\psi_t(x) > \psi_t(z)$. Hence the problem of minimizing $\psi_t(x)$ over Interior(P) is equivalent to the problem of minimizing $\psi_t(x)$ over a closed set which is the intersection of a large sphere and a slightly shrunken version of P. Thus by the theorem of Weierstrass [5] $\psi_t(x)$ has a minimum in Interior(P) for $t > 0$, and the equation $\nabla F(x) = tw$ has a solution in Interior(P) for $t>0$. By the strict convexity of $F(x)$ the solution for each such t is unique. Furthermore from the Implicit Function Theorem [1,2] it follows that the function $x=x(t)$ defined by the equation $\nabla F(x)=tw$, $t>0$, is analytic, since the coordinate functions of $\nabla F(x)$ are analytic functions of x over the interior of P.

Note that Lemmas 7 and 8 hold for the trajectory defined by $\nabla F(x)=tw$, $t>0$. Suppose there exists a $\bar{t}$, $0<\bar{t}\leqslant 1$ such that $x(\bar{t})\notin \Sigma(z,r)$. Then by Lemma 8 (with $\hat{x}=z$, $\hat{t}=1$) we get that

$$1-\hat{t}\geqslant \frac{\left(r-\frac{1}{2}r^2\right)(1-r)^2(\mu(z))^{1/4}}{(1+r)^3\sqrt{w^{\mathrm{T}}Q(z)^{-1}w}}$$

and noting that $r=1.1\sqrt{\delta}\leqslant 0.011$ we have that

$$w^{\mathrm{T}}Q(z)^{-1}w\geqslant 1.02\,\delta\sqrt{\mu(z)}\,.$$

However, this cannot happen since by assumption $w^{\mathrm{T}}Q(z)^{-1}w\leqslant \delta\sqrt{\mu(z)}$. Thus for $0<t\leqslant 1$, $x(t)\in\Sigma(z,r)$.

We shall now show that P is bounded. Let $G(t)=[g_1(t),\dots,g_m(t)]$ where $g_i(t)=\sigma_i(x(t))/(a_i^{\mathrm{T}}x(t)-b_i)a_i$, $1\leqslant i\leqslant m$. Since $x(t)\in\Sigma(z,r)$, $0<t\leqslant 1$, from equation (7.16) in Lemma 5 (with $\hat{x}=z$) it follows that

$$\forall t,\ 0<t\leqslant 1,\quad \frac{(1-r)^2}{(1+r)^3}\frac{\sigma_i(z)}{a_i^{\mathrm{T}}z-b_i}\leqslant\frac{\sigma_i(x(t))}{a_i^{\mathrm{T}}x(t)-b_i}\leqslant\frac{(1+r)^2}{(1-r)^3}\frac{\sigma_i(z)}{a_i^{\mathrm{T}}z-b_i}.$$

Then noting that $z=x(1)$, for $0<t\leqslant 1$ we can express $G(t)G(t)^{\mathrm{T}}$ as

$$G(t)G(t)^{\mathrm{T}}=G(1)D(t)G(1)^{\mathrm{T}}$$

where $D(t)$ is an $m\times m$ diagonal matrix and each diagonal entry in $D(t)$ lies in the interval $[(1-r)^4/(1+r)^6,\ (1+r)^4/(1-r)^6]$. Thus

$$\min_{\xi^{\mathrm{T}}\xi=1}\left\{\xi^{\mathrm{T}}G(t)G(t)^{\mathrm{T}}\xi\right\}\geqslant\frac{(1-r)^4}{(1+r)^6}\min_{\xi^{\mathrm{T}}\xi=1}\left\{\xi^{\mathrm{T}}G(1)G(1)^{\mathrm{T}}\xi\right\}.$$

Since A has linearly independent columns, $G(1)G(1)^{\mathrm{T}}$ is non-singular, and so there exists a $\lambda^*>0$ such that

$$\forall t,\ 0<t\leqslant 1,\quad \text{the smallest eigenvalue of } G(t)G(t)^{\mathrm{T}} \text{ is at least } \lambda^*.$$

We can then find a sufficiently small positive t^* such that

$$\left\|\sum_{i=1}^{m}g_i(t^*)\right\|_2=\|\nabla F(x(t^*))\|_2=\|t^*w\|_2\leqslant\frac{(\lambda^*)^{1/2}}{2mn}.$$

So from Property PLC above we may conclude that for all unit vectors ξ

$$\min_{1\leqslant i\leqslant m}\left\{\frac{\sigma_i(x(t^*))}{a_i^{\mathrm{T}}x(t^*)-b_i}a_i^{\mathrm{T}}\xi\right\}\leqslant-\frac{(\lambda^*)^{1/2}}{4m}.$$

As a result if we move along any direction from z the distance to some boundary of P must decrease at a certain minimum rate and it then follows that P is bounded. $\square$

References

[1] G.A. Bliss, *Lectures on the Calculus of Variations*, Phoenix Science Series (The University of Chicago Press, Chicago, IL, 1946).

[2] S. Bochner and W.T. Martin, *Several Complex Variables* (Princeton University Press, Princeton, NJ, 1948).

[3] D. Coppersmith and S. Winograd, "Matrix multiplication via arithmetic progressions," *Proceedings 19th Annual ACM Symposium Theory of Computing* (1987) pp. 1–6.

[4] M. Grotschel, L. Lovasz and A. Schrijver, *Geometric Algorithms and Combinatorial Optimization* (Springer, Berlin, 1988).

[5] M. Minoux, *Mathematical Programming: Theory and Algorithms* (Wiley, New York, 1986).

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

[7] P.M. Vaidya, "An algorithm for linear programming that requires $O(((m+n)n^2+(m+n)^{1.5}n)L)$ arithmetic operations," *Proceedings 19th Annual ACM Symposium Theory of Computing* (1987) pp. 29–38; also in *Mathematical Programming* 47 (1990) 175–201.

[8] P.M. Vaidya, "Speeding-up linear programming using fast matrix multiplication," *Proceedings 30th Annual IEEE Symposium Foundations of Computer Science* (1989) pp. 338–343.

[9] P.M. Vaidya, "A new algotrithm for minimizing convex functions over convex sets," *Proceedings 30th Annual IEEE Symposium Foundations of Computer Science* (1989) pp. 332–337.

[10] Gy. Sonevand, "An analytical center for polyhedrons and new classes of global algorithms for linear (smooth convex) programming," Preprint, Department of Numerical Analysis, Institute of Mathematics, Eotvos University (Budapest, 1989) pp. 6–8.

[11] J.L. Goffin and J.P. Vial, "Cutting planes and column generation techniques with the projective algorithm," CORE Discussion Paper 8829, CORE, Universite Catholique de Louvain (Louvain, Belgium, 1988).

[12] Y. Ye, "A potential reduction algorithm allowing column generation," Working paper, Department of Management Sciences, University of Iowa (Iowa City, IA, 1989).

[13] S.P.Tarasov, L.G. Khachiyan and I.I. Erlich, "The method of inscribed ellipsoids," *Soviet Mathematics Doklady* 37(1) (1988).