Mathematical Programming 47 (1990) 53-63
North-Holland

ON A SUBPROBLEM OF TRUST REGION ALGORITHMS FOR CONSTRAINED OPTIMIZATION

Y. YUAN

Department of Applied Mathematics and Theoretical Physics, University of Cambridge, England CB3 9EW

Received 6 November 1987
Revised manuscript received August 1988

We study a subproblem that arises in some trust region algorithms for equality constrained optimization. It is the minimization of a general quadratic function with two special quadratic constraints. Properties of such subproblems are given. It is proved that the Hessian of the Lagrangian has at most one negative eigenvalue, and an example is presented to show that the Hessian may have a negative eigenvalue when one constraint is inactive at the solution.

Key words: Subproblem, constrained optimization, quadratic constraints, trust region algorithms.

1. Introduction

In this paper, we study a subproblem that arises in some trust region algorithms for equality constrained optimization. It has the form

$$\min_{d\in\mathbb{R}^n} g^{\mathrm{T}}d+\tfrac{1}{2}d^{\mathrm{T}}Bd \tag{1.1}$$

subject to

$$\|d\|_2\leq \Delta, \tag{1.2}$$

$$\|A^{\mathrm{T}}d+c\|_2\leq \xi, \tag{1.3}$$

where $g\in\mathbb{R}^n$, $B\in\mathbb{R}^{n\times n}$, $A\in\mathbb{R}^{n\times m}$, $c\in\mathbb{R}^m$, $\Delta>0$, $\xi\geq 0$ and B is a symmetric matrix. Problem (1.1)-(1.3) appears in some trust region algorithms for constrained optimization where the original problem is to minimize a general nonlinear function $F(x)$ subject to $c(x)=(c_1(x),\ldots,c_m(x))^{\mathrm{T}}=0$. At the kth iteration, an estimate x_k is known, a subproblem is constructed by using a quadratic approximation of $F(x_k+d)$ and a linear approximation of $c(x_k+d)$. That is,

$$\min_{d\in\mathbb{R}^n} d^{\mathrm{T}}\nabla F(x_k)+\tfrac{1}{2}d^{\mathrm{T}}B_kd\approx F(x_k+d)-F(x_k) \tag{1.4}$$

subject to

$$\|c(x_k)+d^{\mathrm{T}}\nabla c(x_k)\|_2\leq \xi_k, \tag{1.5}$$

and a trust region condition

$$\|d\|_2\leq \Delta_k. \tag{1.6}$$

Research supported by a Research Fellowship of Fitzwilliam College, Cambridge, and by a research grant from the Chinese Academy of Sciences.

$\Delta_k > 0$, $\xi_k \geq 0$ and B_k symmetric are updated every iteration. It can be seen that the subproblem (1.4)–(1.6) has the form of (1.1)–(1.3). More details about trust region algorithms for constrained optimization can be found in Celis, Dennis and Tapia (1985) and Powell and Yuan (1986). Usually the parameter ξ in (1.3) satisfies

$$\|c\|_2 \geq \xi \geq \xi_{\min} = \min_{\|d\|_2 \leq \Delta} \|A^{\mathrm{T}}d + c\|_2 \tag{1.7}$$

and $\xi = \|c\|_2$ only if $\xi_{\min} = \|c\|_2$. One can easily see that (1.2)–(1.3) has no feasible solution if $\xi < \xi_{\min}$. The condition $\xi \leq \|c\|_2$ is not used in the paper, but this restriction is motivated from the fact that $\xi_k \leq \|c(x_k)\|_2$ in (1.5) as it is desirable to reduce the linearized constraint violation.

In the case when $\xi = \xi_{\min}$, it is easy to deduce from the strict convexity of (1.2) that either there is only one feasible solution of (1.2)–(1.3) or that

$$\xi = \xi_{\min} = \min_{d \in \mathbf{R}^n} \|A^{\mathrm{T}}d + c\|_2. \tag{1.8}$$

The case when there is only one feasible solution of (1.2)–(1.3) requires no further consideration, since this feasible solution must be the solution of the problem (1.1)–(1.3). Now assume that (1.8) holds. By the definition of $\xi_{\min}$ in (1.7), we have that

$$\|(A^{\mathrm{T}})^{+}c\|_2 \leq \Delta, \tag{1.9}$$

where $(\ )^{+}$ represents the minimum norm pseudo-inverse, and that condition (1.3) defines $A^{\mathrm{T}}d$ uniquely. By changing variables, we can see that (1.1)–(1.3) is equivalent to the following problem:

$$\min_{d \in \mathbf{R}^n} g^{\mathrm{T}}[-(A^{\mathrm{T}})^{+}c + d] + \tfrac{1}{2}[-(A^{\mathrm{T}})^{+}c + d]^{\mathrm{T}}B[-(A^{\mathrm{T}})^{+}c + d] \tag{1.10}$$

subject to

$$A^{\mathrm{T}}d = 0, \tag{1.11}$$

$$\|d\|_2 \leq \sqrt{\Delta^2 - \|(A^{\mathrm{T}})^{+}c\|_2^2}. \tag{1.12}$$

Now solutions to (1.11) have the form

$$d = (I - (A^{\mathrm{T}})^{+}A^{\mathrm{T}})\bar{d}, \tag{1.13}$$

where $\bar{d}$ is unconstrained, and (1.13) implies not only that $\|d\|_2 \leq \|\bar{d}\|_2$ but also that any d of this form can be generated from a $\bar{d}$ satisfying $\|d\|_2 = \|\bar{d}\|_2$. Therefore problem (1.10)–(1.12) is equivalent to

$$\min_{\bar{d} \in \mathbf{R}^n} (g - B(A^{\mathrm{T}})^{+}c)^{\mathrm{T}}(I - (A^{\mathrm{T}})^{+}A^{\mathrm{T}})\bar{d} + \tfrac{1}{2}\bar{d}^{\mathrm{T}}(I - (A^{\mathrm{T}})^{+}A^{\mathrm{T}})B(I - (A^{\mathrm{T}})^{+}A^{\mathrm{T}})\bar{d} \tag{1.14}$$

subject to

$$\|\bar{d}\|_2 \leq \sqrt{\Delta^2 - \|(A^{\mathrm{T}})^{+}c\|_2^2}. \tag{1.15}$$

This problem is to minimize a quadratic function within a trust region ball, which can be solved by known methods; for example, see Gay (1981) and Moré and Sorensen (1983).

Therefore in this paper, we concentrate our attention on the case when

$$\xi > \xi_{\min}. \tag{1.16}$$

In the next section we give some properties of the problem, mainly considering the signs of eigenvalues of the Hessian of the Lagrangian function at a solution. Our main result is that the Hessian of the Lagrangian has at most one negative eigenvalue if the Lagrangian multipliers are unique. We also show by examples that the Hessian may have one negative eigenvalue when only one constraint is active at the solution and it may have more than one negative eigenvalue if the Lagrangian multipliers are not unique. A short discussion is given in Section 3.

2. Properties of the problem

In this section we consider the optimality conditions of the problem (1.1)–(1.3). Our main result is as follows:

Theorem 2.1. *Let d^* be a global solution of the problem* (1.1)–(1.3). *Assume that* (1.16) *holds. Then there exist non-negative constants λ, μ such that*

$$(B+\lambda I+\mu AA^{\mathrm{T}})d^* = -(g+\mu Ac), \tag{2.1}$$

where λ and μ satisfy the complementarity conditions

$$\lambda(\Delta - \|d^*\|_2) = 0, \tag{2.2}$$

$$\mu(\xi - \|A^{\mathrm{T}}d^* + c\|_2) = 0. \tag{2.3}$$

Furthermore the matrix

$$H(\lambda, \mu) = B+\lambda I+\mu AA^{\mathrm{T}} \tag{2.4}$$

has at most one negative eigenvalue if the multipliers λ and μ are unique.

Proof. Due to (1.16), the feasible region of (1.2)–(1.3) is convex and has a nonempty interior, thus the Slater constraint qualification is satisfied. Hence the existence of non-negative numbers λ and μ such that (2.1)–(2.3) hold is a straightforward application of the Kuhn–Tucker theory of constrained optimization. To complete the proof, we need to prove that the matrix $H(\lambda, \mu)$ has no more than one negative eigenvalue if the multipliers λ, μ are unique.

First we consider the case when both constraints (1.2) and (1.3) are inactive at the solution, that is, $\|d^*\|_2 < \Delta$ and $\|A^{\mathrm{T}}d^* + c\|_2 < \xi$. Then d^* is a local minimum of the objective function in (1.1), so B is positive semi-definite, and the theorem is true.

In the case when just one of the constraints is active, the multiplier of the inactive constraint is zero, and the other multiplier is defined uniquely by (2.1). One can also show that the matrix $H(\lambda, \mu)$ has at most one negative eigenvalue, by applying a second order necessary condition (for example, see Fletcher, 1981, Theorem 9.3.1). We give a constructive proof below, however, since this approach provides a technique for proving the theorem when both constraints are active. Assuming the first constraint is active, that is $\|d^*\|_2 = \Delta$ and $\|A^T d^* + c\|_2 < \xi$, we define the set

$$G(d^*) = \{d \mid \|d\|_2 = 1;\ \|d^* + d\|_2 = \Delta;\ \|A^T(d^* + d) + c\|_2 \leq \xi\}. \tag{2.5}$$

The function

$$f(d) = g^T d + \tfrac{1}{2} d^T B d + \tfrac{1}{2}\lambda d^T d, \quad d \in \mathbb{R}^n, \tag{2.6}$$

is stationary at $d = d^*$, where λ is the Lagrange multiplier of the active constraint at d^*, and the definition of the problem implies $f(d^* + d) \geq f(d^*)$ for all d in (2.5). It follows that

$$d^T(B + \lambda I)d \geq 0 \tag{2.7}$$

for all $d \in G(d^*)$. Consider any non-zero vector $\bar{d}$ such that $\bar{d}^T d^* = 0$ and $\|\bar{d}\|_2 = 1$. Since $\|A^T d^* + c\|_2 < \xi$ and $\|d^*\|_2 = \Delta$, there exists $\delta > 0$ such that

$$\|A^T((d^* + \theta\bar{d})/\|d^* + \theta\bar{d}\|_2)\Delta + c\|_2 \leq \xi \tag{2.8}$$

for all $0 \leq \theta \leq \delta$. Thus the last condition of (2.5) holds if

$$d = ((d^* + \theta\bar{d})/\|d^* + \theta\bar{d}\|_2)\Delta - d^* = v(d^*, \theta), \tag{2.9}$$

say, and we see that $\|d^* + d\|_2 = \Delta$ too. Hence we have that

$$v(d^*, \theta)/\|v(d^*, \theta)\|_2 \in G(d^*). \tag{2.10}$$

Further, $\|d^* + \theta\bar{d}\|_2 = \Delta + O(\theta^2)$ because $\bar{d}^T d^* = 0$ and $\|d^*\|_2 = \Delta$. Therefore the definition (2.9) implies that $v(d^*, \theta) = \theta\bar{d} + O(\theta^2)$, which shows the limit

$$\lim_{\theta \to 0}(v(d^*, \theta)/\|v(d^*, \theta)\|_2) = \bar{d}. \tag{2.11}$$

Relations (2.7), (2.10) and (2.11) imply the inequality

$$\bar{d}^T H(\lambda, 0)\bar{d} \geq 0 \tag{2.12}$$

for all $\bar{d}$ such that $\bar{d}^T d^* = 0$. Thus the Hessian has at most one negative eigenvalue. Similarly, one can show that $H(0, \mu)$ has at most one negative eigenvalue if $\|d^*\|_2 < \Delta$ and $\|A^T d^* + c\|_2 = \xi$.

To complete our proof, we consider now the case when both constraints are active, that is

$$\|d^*\|_2 = \Delta \tag{2.13}$$

and

$$\|A^T d^* + c\|_2 = \xi. \tag{2.14}$$

Define the set

$$\Lambda(d^*) = \{d \mid \|d^* + d\|_2 = \Delta,\ \|A^T(d^* + d) + c\|_2 = \xi\}. \tag{2.15}$$

Since d^* is a global solution to the problem (1.1)–(1.3), the function

$$\Phi_{\lambda,\mu}(d) = g^{\mathrm{T}}d + \tfrac{1}{2}d^{\mathrm{T}}Bd + \tfrac{1}{2}\lambda d^{\mathrm{T}}d + \tfrac{1}{2}\mu\|A^{\mathrm{T}}d + c\|_2^2 \tag{2.16}$$

is stationary at $d = d^*$, where λ and μ are the Lagrange multipliers at d^*. Further, because $\Phi_{\lambda,\mu}(d^*+d) \geq \Phi_{\lambda,\mu}(d^*)$ for all $d \in \Lambda(d^*)$, it follows that

$$d^{\mathrm{T}}H(\lambda,\mu)d \geq 0, \quad d \in \Lambda(d^*), \tag{2.17}$$

where $H(\lambda,\mu)$ is the matrix (2.4). Define the vector

$$y^* = A(A^{\mathrm{T}}d^* + c). \tag{2.18}$$

Our assumption that λ and μ are unique in (2.1) ensures that d^* and y^* are linearly independent. For any vector $d_1 \in \mathbb{R}^n$ satisfying $\|d_1\|_2 = 1$,

$$d_1^{\mathrm{T}}d^* = 0 \quad \text{and} \quad d_1^{\mathrm{T}}y^* = 0, \tag{2.19}$$

there exist vectors d_i $(i = 2, 3, \ldots, n-2)$ such that

$$\begin{aligned} &d_i^{\mathrm{T}}d_j = \delta_{ij}, \quad i, j = 1, 2, \ldots, n-2, \\ &d_i^{\mathrm{T}}d^* = 0 \quad \text{and} \quad d_i^{\mathrm{T}}y^* = 0, \qquad i = 1, 2, \ldots, n-2. \end{aligned} \tag{2.20}$$

Now consider the system

$$\begin{aligned} &\|d^* + d\|_2 = \Delta, \qquad \|A^{\mathrm{T}}(d^*+d) + c\|_2 = \xi, \\ &d_1^{\mathrm{T}}d = t, \qquad d_i^{\mathrm{T}}d = 0, \quad i = 2, 3, \ldots, n-2, \end{aligned} \tag{2.21}$$

for sufficiently small $t > 0$. By the implicit function theorem, (2.21) has a unique 9solution

$$d(t) = (J^*)^{-1}\begin{pmatrix} 0 \\ 0 \\ t \\ 0 \\ \vdots \\ 0 \end{pmatrix} + o(t) \tag{2.22}$$

for sufficiently small $t > 0$, where J^* is the Jacobian matrix of (2.21) at $d = 0$. We see that

$$(J^*)^{-1} = [u^*, v^*, d_1, \ldots, d_{n-2}] \tag{2.23}$$

where $u^*, v^* \in \mathrm{span}(d^*, y^*)$, and it follows from (2.22) and (2.23) that

$$\lim_{t\to 0_+}(d(t)/\|d(t)\|_2) = d_1. \tag{2.24}$$

Hence, since $d(t) \in \Lambda(d^*)$, we deduce from (2.17) that

$$d_1^{\mathrm{T}}H(\lambda,\mu)d_1 \geq 0. \tag{2.25}$$

The freedom in the choice of d_1 shows that (2.17) holds for all d satisfying

$$d^{\mathrm{T}}d^* = 0 \quad \text{and} \quad d^{\mathrm{T}}y^* = 0. \tag{2.26}$$

If the matrix $H(\lambda, \mu)$ had two negative eigenvalues, there would exist two orthogonal unit vectors $e_1, e_2 \in \mathbb{R}^n$ such that

$$d^T H(\lambda, \mu) d < 0 \tag{2.27}$$

for all non-zero vectors d in the subspace span(e_1, e_2). Therefore, because (2.26) implies (2.17), the system

$$(\alpha e_1 + \beta e_2)^T d^* = 0, \qquad (\sigma e_1 + \beta e_2)^T y^* = 0, \qquad \alpha^2 + \beta^2 \neq 0, \tag{2.28}$$

in α and β has no solution. Thus the matrix

$$\begin{pmatrix} e_1^T d^* & e_2^T d^* \\ e_1^T y^* & e_2^T y^* \end{pmatrix} \tag{2.29}$$

is non-singular. Now the system

$$\|d^* + d\|_2 = \Delta, \qquad \|A^T(d^* + d) + c\|_2 = \xi, \qquad d \in \text{span}(e_1, e_2), \tag{2.30}$$

represents the intersection of a circle with an ellipse in $\mathbb{R}^2$. If (2.30) has no non-zero solutions, it can be easily seen that $(\alpha, \beta) = (0, 0)$ is a minimizer or maximizer of the function $\|A^T[d^* + (\alpha e_1 + \beta e_2)] + c\|_2 - \xi$ in α and β subject to $\|d^* + (\alpha e_1 + \beta e_2)\|_2 = \Delta$, which implies $(e_1^T y^* \ e_2^T y^*) = \tau(e_1^T d^* \ e_2^T d^*)$ for some $\tau \in \mathbb{R}$. This contradicts the nonsingularity of (2.29). Hence (2.30) has a nonzero solution, say $\tilde{d}$. It follows from (2.27) that

$$\Phi_{\lambda,\mu}(d^* + \tilde{d}) = \Phi_{\lambda,\mu}(d^*) + \tfrac{1}{2}\tilde{d}^T H(\lambda, \mu)\tilde{d} < \Phi_{\lambda,\mu}(d^*), \tag{2.31}$$

which contradicts the fact that d^* is the minimum of $\Phi_{\lambda,\mu}(d)$ subject to (2.13) and (2.14). This contradiction proves that $H(\lambda, \mu)$ has at most one negative eigenvalue. □

We have noted that the matrix $H(\lambda, \mu)$ is positive semidefinite if both constraints are inactive. However the result for the case when only one constraint is inactive seems too pessimistic, because one might guess that the matrix $H(\lambda, \mu)$ is also positive semidefinite if either $\|d\|_2 < \Delta$ or $\|A^T d^* + c\|_2 < \xi$, because it is well known that the matrix $B + \lambda I$ is positive semi-definite if d^* is a global solution to problem (1.1)-(1.2) (for example, see Gay, 1981). However, Lemma 2.2 shows that this view is incorrect.

One can easily see that λ and μ are unique if at most one constraint is active at the solution or d^* and $A(A^T d^* + c)$ are linearly independent when two constraints are active. Hence, λ and μ are not unique if and only if both constraints are active and d^* and $A(A^T d^* + c)$ are linearly dependent. The following lemma shows that the matrix $H(\lambda, \mu)$ defined in (2.4) may have more than one eigenvalue if λ and μ are not unique.

Lemma 2.2. *The matrix $H(\lambda, \mu)$ may have a negative eigenvalue when one of the constraints is inactive, and it may have more than one negative eigenvalue if d^* and $A(A^T d^* + c)$ are linearly dependent.*

Proof. First we consider the following example: $n = m = 2$, $\Delta = 2$, $\xi = 1$ and

$$g = \begin{pmatrix} 2 \\ 0 \end{pmatrix}, \quad B = \begin{pmatrix} -2 & 0 \\ 0 & 2 \end{pmatrix}, \quad A = I, \quad c = \begin{pmatrix} -2 \\ 0 \end{pmatrix}. \tag{2.32}$$

Then, because the second component of d^* must be zero, it is easy to show that

$$d^* = \begin{pmatrix} 2 \\ 0 \end{pmatrix} \tag{2.33}$$

is the unique global solution to (1.1)-(1.3). One can verify that $\lambda^* = 1$, $\mu^* = 0$ are the multipliers, and that $H(\lambda^*, \mu^*)$ is the matrix

$$\begin{pmatrix} -1 & 0 \\ 0 & 3 \end{pmatrix}. \tag{2.34}$$

It has a negative eigenvalue, which proves the first part of the lemma.

Now we study another example: $n = m = 2$, $\Delta = 2$, $\xi = 1$ and

$$g = \begin{pmatrix} 0.5 \\ 0 \end{pmatrix}, \quad B = -I = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}, \quad A = I, \quad c = \begin{pmatrix} -1 \\ 0 \end{pmatrix}. \tag{2.35}$$

One can show that the unique global solution to (1.1)-(1.3) is

$$d^* = \begin{pmatrix} 2 \\ 0 \end{pmatrix}, \tag{2.36}$$

and that the Lagrange multipliers are $\lambda, \mu \geqslant 0$ satisfying $2\lambda + \mu = 1.5$. Hence the choice $\lambda = \frac{3}{4}$ and $\mu = 0$ gives $H(\lambda, \mu) = -0.25I$, which has two negative eigenvalues. This completes our proof. □

Next we consider the case when the multipliers λ and μ are not unique. We need the following lemma.

Lemma 2.3. *Let $C, D \in \mathbb{R}^{n \times n}$ be two symmetric matrices and let A and B be two closed sets in $\mathbb{R}^n$ such that*

$$A \cup B = \mathbb{R}^n. \tag{2.37}$$

If we have

$$x^T C x \geqslant 0, \quad x \in A, \qquad x^T D x \geqslant 0, \quad x \in B, \tag{2.38}$$

then there exists a $t \in [0, 1]$ such that the matrix

$$tC + (1 - t)D \tag{2.39}$$

is positive semidefinite.

Proof. The lemma is trivial if either A or B is empty. Hence we assume that A and B are nonempty sets. Without loss of generality, due to (2.38), we can also assume that $-A = A$ and $-B = B$.

Let $v(t)$ be the least eigenvalue of the matrix (2.39). If $v(t)\geqslant 0$ for some $t\in[0,1]$, the lemma is true. Therefore we assume that

$$v(t)<0 \tag{2.40}$$

for all $t\in[0,1]$. Define the set

$$S(t)=\{x\,|\,x^T(tC+(1-t)D)x=v(t)\}\cap\Gamma \tag{2.41}$$

where $\Gamma=\{x\,|\,\|x\|_2\leqslant 1\}$. By the definition of $v(t)$, $S(t)$ is an intersection of a subspace and the unit ball Γ. Further, $S(t)$ is non-empty for all $t\in[0,1]$ and because $S(t)$ is closed one can show that

$$S(t)\supseteq\lim_{t'\to t}S(t')\equiv\left\{x\,|\,x=\lim_{k\to\infty}x_k,\ x_k\in S(t_k),\ t=\lim_{k\to\infty}t_k\right\}. \tag{2.42}$$

It should bê pointed out that this definition of the limit of sets is not the standard one. Inequalities (2.38) and (2.40) imply

$$S(0)\cap A\neq\emptyset, \tag{2.43}$$

where $\emptyset$ is the empty set. Let t_1 be the largest number in $[0,1]$ such that

$$S(t_1)\cap A\neq\emptyset, \tag{2.44}$$

which exists due to (2.42). If $t_1<1$, the definition of t_1, relation (2.37) and the fact that $S(t)$ is non-empty imply

$$S(t)\cap B=(S(t)\cap B)\cup(S(t)\cap A)=S(t)\neq\emptyset \tag{2.45}$$

for all $t\in(t_1,1]$. Hence, due to (2.42), it follows that

$$S(t_1)\cap B\neq\emptyset. \tag{2.46}$$

On the other hand, similar to (2.43), we can show that (2.46) is also true if $t_1=1$. One can easily see that either $S(t_1)$ is a connected set or it consists of two points, that is $S(t_1)=\{x_t,-x_t\}$. If $S(t_1)$ is connect, the relations (2.44) and (2.46) give

$$S(t_1)\cap A\cap B\neq\emptyset. \tag{2.47}$$

Otherwise, because $S(t_1)=\{x_t,-x_t\}$, relation (2.44) and the assumption $A=-A$ imply that $S(t_1)\subseteq A$. Similarly, $S(t_1)\subseteq B$. Hence (2.47) holds. Now (2.47) contradicts (2.38) because of the inequality (2.40). Therefore the lemma is true. □

This lemma is used in the proof of the following result.

Theorem 2.4. *Assume that the conditions of Theorem 2.1 hold, then there exists $(\lambda,\mu)\in\Omega$ such that the matrix (2.4) has at most one negative eigenvalue, where Ω is the set of Lagrangian multipliers.*

Proof. Assume that d^* is a global solution of (1.1)–(1.3). If the Lagrangian multipliers are unique, Ω has only one element and the theorem is true by Theorem 2.1.

In the case when the multipliers are not unique, it can be easily seen that both constraints are active at the solution and d^* and $y^*=A(A^{\mathrm{T}}d^*+c)$ are linearly dependent. Due to inequality (1.16), the feasible region (1.2)–(1.3) contains some interior points, so there exists a positive number ω such that

$$y^*=\omega d^*. \tag{2.48}$$

It follows from the Kuhn–Tucker condition (2.1) and the non-negativity of λ and μ that there exists a nonnegative number ψ such that

$$-(g+Bd^*)=\psi d^*. \tag{2.49}$$

For $t>0$, define the sets

$$\begin{aligned} A(t)&=\{d\,|\,\|d^*+td\|_2=\Delta;\ \|a^{\mathrm{T}}(d^*+td)+c\|_2\leq\xi;\ \|d\|_2=1\},\\ B(t)&=\{d\,|\,\|d^*+td\|_2\leq\Delta;\ \|A^{\mathrm{T}}(d^*+td)+c\|_2=\xi;\ \|d\|_2=1\},\end{aligned} \tag{2.50}$$

and let

$$\lim_{t\to0_+}A(t)=A,\qquad \lim_{t\to0_+}B(t)=B, \tag{2.51}$$

where the limits are defined by (2.42). It can be seen that A and B are closed sets. Using (2.48), one can show that

$$A\cup B=\{d\,|\,d^{\mathrm{T}}d^*=0;\ \|d\|_2=1\}. \tag{2.52}$$

Because d^* is a solution to (1.1)–(1.3), $d=0$ is a minimum of the function

$$g^{\mathrm{T}}(d^*+td)+\tfrac{1}{2}(d^*+td)^{\mathrm{T}}(B+\psi I)(d^*+td),\quad d\in A(t). \tag{2.53}$$

It follows from (2.49) that $d^{\mathrm{T}}(B+\psi I)d$ is nonnegative for all $d\in A(t)$, which implies

$$d^{\mathrm{T}}(B+\psi I)d\geq0,\quad d\in A. \tag{2.54}$$

Similarly, one can show that

$$d^{\mathrm{T}}(B+(\psi/\omega)AA^{\mathrm{T}})d\geq0,\quad d\in B. \tag{2.55}$$

Applying Lemma 2.3 in the linear subspace $\{d\,|\,d^{\mathrm{T}}d^*=0\}$ we deduce that there exists a number $t\in[0,1]$ such that

$$d^{\mathrm{T}}(B+t\psi I+(1-t)(\psi/\omega)AA^{\mathrm{T}})d\geq0, \tag{2.56}$$

for all d such that $d^{\mathrm{T}}d^*=0$. Thus the matrix $H(t\psi,(1-t)\psi/\omega)$ has at most one negative eigenvalue, $t\psi$ and $(1-t)\psi/\omega$ being Lagrange multipliers. This completes our proof. □

We also have the following sufficient condition for problem (1.1)–(1.3).

Theorem 2.5. *If d^* is a feasible point of (1.2)–(1.3), if there are two multipliers λ, $\mu\geq0$ such that (2.1)–(2.3) hold, and if the matrix (2.4) is positive semi-definite, then d^* is a global solution of the problem (1.1)–(1.3).*

Proof. If d^* is not a global solution, there exists a feasible vector $\bar{d}$ such that

$$g^T\bar{d}+\tfrac{1}{2}\bar{d}^TB\bar{d}<g^Td^*+\tfrac{1}{2}d^{*T}Bd^*. \tag{2.57}$$

Moreover, because $H(\lambda, \mu)$ is positive semi-definite, d^* is a global minimizer of the function (2.16). Therefore we have

$$\lambda\|\bar{d}\|_2^2+\mu\|A^T\bar{d}+c\|_2^2>\lambda\|d^*\|_2^2+\mu\|A^Td^*+c\|_2^2, \tag{2.58}$$

which, due to (2.2) and (2.3), implies

$$\lambda(\|\bar{d}\|_2^2-\Delta^2)+\mu(\|A^T\bar{d}+c\|_2^2-\xi^2)>0. \tag{2.59}$$

However this contradicts the feasibility of $\bar{d}$ and the nonnegativity of λ and μ. Therefore the theorem is true. □

3. Discussion

It is shown that the Hessian of the Lagrangian at the solution has at most one negative eigenvalue. Unfortunately, the Hessian may have a negative eigenvalue even if only one constraint is active at the solution and it may even have two negative eigenvalues if the gradients of both active constraints are linearly dependent.

We are now studying the problem of constructing an algorithm for solving (1.1)-(1.3). If the matrix B is positive definite, for any $\lambda, \mu \geq 0$ one can always define $d(\lambda, \mu)$ such that (2.1) holds, that is

$$d(\lambda, \mu)=-H(\lambda, \mu)^{-1}(g+\mu Ac). \tag{3.1}$$

Then it is sufficient to solve the system

$$\lambda(\Delta^2-\|d(\lambda, \mu)\|_2^2)=0, \qquad \mu(\xi^2-\|A^Td(\lambda, \mu)+c\|_2^2)=0, \tag{3.2}$$

subject to $d(\lambda, \mu)$ being a feasible point of (1.2)-(1.3), that is

$$\Delta^2-\|d(\lambda, \mu)\|_2^2\geq 0, \qquad \xi^2-\|A^Td(\lambda, \mu)+c\|_2^2\geq 0. \tag{3.3}$$

An initial guess $\lambda=\mu=0$ satisfies (3.2) but not (3.3) if $d(0,0)=-B^{-1}g$ is not a solution to (1.1)-(1.3). We have tried an algorithm that is based on Newton's method for solving (3.2) subject to the condition (3.3). Numerical results show that the algorithm performs reasonably well. However more work is needed to analyze the convergence properties of the algorithm. In the case when B has negative eigenvalues, trying to solve problem (2.1)-(2.3) by considering problem (3.1)-(3.3) may have difficulties, since the matrix $H(\lambda, \mu)$ may be singular. Another possible way for solving (2.1)-(2.3) is that for any feasible point d of (1.2)-(1.3), we let $(\lambda(d), \mu(d))$ be a least squares solution of (2.1)-(2.3) subject to $\lambda\geq 0$, $\mu\geq 0$. Then at each iteration, we update d such that the residual of (2.1)-(2.3) (with $\lambda=\lambda(d)$ and $\mu=\mu(d)$) decreases (for example, Newton's step may be suitable) and such that the new d also satisfies (1.2)-(1.3).

Acknowledgement

The author would like to thank Professor M.J.D. Powell for his reading an early version of the paper, for his suggestions and corrections, and also for his constant help and encouragement. I am also indebted to the referees for their valuable suggestions and comments.

References

M.R. Celis, J.E. Dennis and R.A. Tapia, "A trust region strategy for nonlinear equality constrained optimization," in: P.T. Boggs, R.H. Byrd and R.B. Schnabel, eds., *Numerical Optimization* (SIAM, Philadelphia, PA 1985) pp. 71–82.

R. Fletcher, *Practical Methods for Optimization, Vol. 2: Constrained Optimization* (Wiley, Chichester, 1981).

D.M. Gay, "Computing optimal locally constrained steps," *SIAM Journal on Scientific and Statistical Computing* 2 (1981) 186–197.

J.J. Moré and D.C. Sorensen, "Computing a trust region step," *SIAM Journal on Scientific and Statistical Computing* 4 (1983) 553–572

M.J.D. Powell and Y. Yuan, "A trust region algorithm for equality constrained optimization," Report DAMTP 1986/NA2, University of Cambridge (Cambridge, UK).