Journal of Computational Mathematics, Vol.9, No.4, 1991, 348–359.

A DUAL ALGORITHM FOR MINIMIZING A QUADRATIC FUNCTION WITH TWO QUADRATIC CONSTRAINTS*[1)]

Yuan Ya-xiang
(Department of Applied Mathematics and Theoretical Physics, University of Cambridge, England)

Abstract

In this paper, we present a dual algorithm for minimizing a convex quadratic function with two quadratic constraints. Such a minimization problem is a subproblem that appears in some trust region algorithms for general nonlinear programming. Some theoretical properties of the dual problem are given. Global convergence of the algorithm is proved and a local superlinear convergence result is presented. Numerical examples are also provided.

§1. The Problem

In this paper, we present a dual algorithm for minimizing a convex quadratic function with two special quadratic constraints. The problem has the form:

$$\min_{d\in\Re^n} \Phi(d) \equiv g^T d + \frac{1}{2}d^T Bd, \tag{1.1}$$

subject to

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

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

where $g \in \Re^n, B \in \Re^{n\times n}, A \in \Re^{n\times m}, c \in \Re^m, \Delta > 0, \xi \ge 0$ and B is a symmetric matrix. Problem (1.1)–(1.3) is a subproblem of some trust region algorithms for constrained optimization (for example, see Celis, Dennis and Tapia, 1985; and Powell and Yuan, 1986). Some theoretical properties of the problem are presented in Yuan (1987) for general B, but now we restrict attention to the case when B is positive definite, because we have not yet found a reliable method for computing the global solution in the general case.

The algorithm, given in Section 3, is based on Newton's method for the dual program of the following problem:

$$\min_{d\in\Re^n} \Phi(d) \equiv g^T d + \frac{1}{2}d^T Bd, \tag{1.4}$$

subject to

$$\|d\|_2^2 \le \Delta^2, \tag{1.5}$$

$$\|A^T d + c\|_2^2 \le \xi^2, \tag{1.6}$$

---

* Received August 16, 1988.
[1)] Research Supported by a Research Fellowship of Fitzwilliam College, Cambridge and by a research grant from the Chinese Academy of Sciences.

which is equivalent to (1.1)–(1.3).

In the next section, we give some theoretical properties of the dual problem. Then an algorithm is presented in Section 3, and convergence properties of the algorithm are given in Section 4. Numerical results are reported in Section 5, and finally a short discussion is given in Section 6.

§2. Dual Theory

For the dual variables $\lambda \geq 0, \mu \geq 0$ we define the matrix

$$H(\lambda,\mu) = B + \lambda I + \mu AA^T, \tag{2.1}$$

and the vector

$$d(\lambda,\mu) = -H(\lambda,\mu)^{-1}(g + \mu Ac). \tag{2.2}$$

We also define the function

$$\Psi(\lambda,\mu) = \Phi(d(\lambda,\mu)) + \frac{1}{2}\lambda(\|d(\lambda,\mu)\|_2^2 - \Delta^2) + \frac{1}{2}\mu(\|A^T d(\lambda,\mu) + c\|_2^2 - \xi^2). \tag{2.3}$$

The dual problem for (1.4)–(1.6) is

$$\max_{(\lambda,\mu)\in\Re_+^2} \Psi(\lambda,\mu), \tag{2.4}$$

where we use the notation $\Re_+^2 = \{\lambda \geq 0, \mu \geq 0\}$. The relation of the dual problem to the primal problem is given in Lemma 2.2 below. One advantage of working with the dual problem (2.4) is that it has only two variables. Moreover, because gradients and second-order derivatives of $\Psi(\lambda,\mu)$ can be easily computed, (2.4) can be solved by applying Newton's method. Because the vector (2.2) is the value of $d(\lambda,\mu)$ that minimises the righthand side of expression (2.3), direct calculations show that

$$\nabla\Psi(\lambda,\mu) = \frac{1}{2}\begin{pmatrix} \|d(\lambda,\mu)\|_2^2 - \Delta^2 \\ \|A^T d(\lambda,\mu) + c\|_2^2 - \xi^2 \end{pmatrix}, \tag{2.5}$$

$$\nabla^2\Psi(\lambda,\mu) = -\begin{pmatrix} d(\lambda,\mu)^T H(\lambda,\mu)^{-1} d(\lambda,\mu) & d(\lambda,\mu)^T H(\lambda,\mu)^{-1} y(\lambda,\mu) \\ d(\lambda,\mu)^T H(\lambda,\mu)^{-1} y(\lambda,\mu) & y(\lambda,\mu)^T H(\lambda,\mu)^{-1} y(\lambda,\mu) \end{pmatrix}, \tag{2.6}$$

where $y(\lambda,\mu)$ is the vector

$$y(\lambda,\mu) = A(A^T d(\lambda,\mu) + c). \tag{2.7}$$

It is easy to see that $\Psi(\lambda,\mu)$ is a concave function. Another advantage of working with the dual problem (2.4) is that, as shown in the following lemma, the gradient and the Jacobian of $\Psi(\lambda,\mu)$ are both bounded above, even if the constraints (1.2) and (1.3) are inconsistent.

Lemma 2.1. *Let $d(\lambda,\mu)$ be defined by (2.2). Then*

$$\max_{(\lambda,\mu)\in\Re_+^2} \|d(\lambda,\mu)\|_2 \tag{2.8}$$

is finite. Consequently, $\nabla\Psi(\lambda,\mu)$ and $\nabla^2\Psi(\lambda,\mu)$ are bounded above in $\Re_+^2$.

Proof. The definition (2.2) shows that

$$\|d(\lambda,\mu)\|_2 \leq \|H(\lambda,\mu)^{-1}g\|_2 + \|H(\lambda,\mu)^{-1}\mu Ac\|_2. \tag{2.9}$$

The first term of the right-hand side of (2.9) is bounded above by

$$\|B^{-1}\|_2\|g\|_2. \tag{2.10}$$

Using the relation

$$Ac = AA^+Ac = AA^T(A^+)^Tc, \tag{2.11}$$

it can be shown that

$$\begin{aligned}\|H(\lambda,\mu)^{-1}\mu Ac\|_2 &\le \|(B+\mu AA^T)^{-1}\mu Ac\|_2 = \|(B+\mu AA^T)^{-1}(B+\mu AA^T)(A^+)^Tc \\ &-(B+\mu AA^T)^{-1}B(A^+)^Tc\|_2 \le \|(A^+)^Tc\|_2 + \|B^{-1}\|_2\|B\|_2\|(A^+)^Tc\|_2.\end{aligned} \tag{2.12}$$

Therefore we have

$$\|d(\lambda,\mu)\|_2 \le \|B^{-1}\|_2\|g\|_2 + \|(A^+)^Tc\|_2 + \|B^{-1}\|_2\|B\|_2\|(A^+)^Tc\|_2 \tag{2.13}$$

for all $(\lambda,\mu) \in R^2_+$. This completes our proof.

Standard properties of dual problems give us the following results.

Lemma 2.2. *Using the notation*

$$\xi_{\min} = \min_{\|d\|_2\le\Delta} \|A^Td + c\|_2, \tag{2.14}$$

we have

1) *if* $\xi = \xi_{\min}$, *then either there is only one feasible solution of* (1.2)–(1.3), *or* (1.1)–(1.3) *can be reduced to a simpler problem which has the form of* (1.1)–(1.2);

2) *if* $\xi < \xi_{\min}$, *that is the original problem has no feasible points, the function* $\Psi(\lambda,\mu)$ *is not bounded above;*

3) *if* $\xi > \xi_{\min}$, *the dual problem* (2.4) *has a finite solution* $(\lambda^*,\mu^*) \in \Re^2_+$, *and for any solution* (λ,μ) *of* (2.4), $d(\lambda,\mu)$ *is the unique solution to the original problem* (1.1)–(1.3);

4) *if* $\xi > \xi_{\min}$ *and there is more than one solution to problem* (2.4), *for any solution* (λ^*,μ^*) *the equation*

$$\nabla\Psi(\lambda^*,\mu^*) = 0 \tag{2.15}$$

holds, and the set of solutions of (2.4) *is a line segment of the form*

$$\begin{pmatrix}\lambda\\ \mu\end{pmatrix} = \begin{pmatrix}\alpha\bar\lambda\\ (1-\alpha)\bar\mu\end{pmatrix}, \quad 0\le\alpha\le 1, \tag{2.16}$$

for some $\bar\lambda > 0$ *and* $\bar\mu > 0$.

Proof. 1) is proved in Yuan (1987). 2) and 3) can be shown by applying the standard dual theory on convex programming (for example, Rockafallar, 1970). So we only need to prove 4).

Assume that $\xi > \xi_{\min}$ and that $(\lambda^*,\mu^*)^T$ is a non-zero solution to the maximization problem (2.4). Due to the uniqueness in 1), it follows from (2.1), (2.2) and (2.7) that

$$(\lambda-\lambda^*)d^* + (\mu-\mu^*)y^* = 0, \tag{2.17}$$

where $d^* = d(\lambda^*,\mu^*), y^* = y(\lambda^*,\mu^*)$ and (λ,μ) is any solution of problem (2.4). Because (λ,μ) maximizes (2.4), we have

$$\begin{cases}\lambda(\|d^*\|_2^2-\Delta^2)=0,\\ \mu(\|A^Td^*+c\|_2^2-\xi^2)=0.\end{cases} \tag{2.18}$$

From (2.17), we see that $d^* = y^* = 0$ or all solutions (λ, μ) of (2.4) lie on a straight line. If $d^* = y^* = 0$ and $\Delta > 0$, our assumption $\xi > \xi_{\min}$ and the relation (2.18) show that $\lambda = \mu = 0$, which contradicts the assumption that problem (2.4) has more than one solution. Thus all the solutions of (2.4) lie on a straight line. Due to the fact $\xi > \xi_{\min}$, it can be shown that $(d^*)^T y^* > 0$. Further, $(\lambda, \mu) \in \Re^2_+$ is a solution to (2.4) if and only if it satisfies (2.17). Now all solution $(\lambda, \mu) \in \Re^2_+$ of (2.17) are vectors in (2.16) if we let $\bar\lambda = \lambda^* + \mu^*(d^*)^T y^*/\|d^*\|_2^2$ and $\bar\mu = \mu^* + \lambda^*\|d^*\|_2^2/(d^*)^T y^*$. This completes our proof.

When problem (1.1)–(1.3) is a subproblem derived from an algorithm for nonlinear constrained optimisation, it is normally true that

$$\xi \geq \xi_{\min}, \tag{2.19}$$

and it is normally known when $\xi = \xi_{\min}$. In the latter case, as stated in Lemma 2.2, either there is only one feasible point of (1.2)–(1.3) or problem (1.1)–(1.3) can be reduced to a simpler problem. The simpler calculation is to minimize a convex quadratic function in a ball, which can be solved by algorithms in Gay (1981) and Moré and Sorensen (1983). Therefore, because our work is motivated by the needs of trust region algorithms for nonlinear constraints, from now on we assume that

$$\xi > \xi_{\min} \tag{2.20}$$

Lemma 2.3. *Condition* (2.20) *implies that the set*

$$\{(\lambda, \mu) | \Psi(\lambda, \mu) \geq \Psi(0,0), \quad (\lambda, \mu) \in \Re^2_+\} \tag{2.21}$$

is bounded.

Proof. When (2.20) is satisfied, there exists $\hat d \in R^n$ such that $\|\hat d\|_2^2 < \Delta^2$ and $\|A^T \hat d + c\|_2^2 < \xi^2$. Since $d(\lambda, \mu)$ is calculated to minimize the right-hand side of expression (2.3), we have the bound

$$\Psi(\lambda, \mu) \leq \Phi(\hat d) + \frac{\lambda}{2}(\|\hat d\|_2^2 - \Delta^2) + \frac{\mu}{2}(\|A^T \hat d + c\|_2^2 - \xi^2). \tag{2.22}$$

Hence $\Psi(\lambda, \mu) \to -\infty$ if $\max[\lambda, \mu] \to \infty$. Therefore the lemma is true.

This result tells us that, if an algorithm for solving (2.4) has the property that $\Psi(\lambda_{k+1}, \mu_{k+1}) \geq \Psi(\lambda_k, \mu_k)$, then the sequence $\{(\lambda_k, \mu_k)\ (k = 1, 2, 3, \cdots)\}$ remains bounded if $\lambda_1 = \mu_1 = 0$.

§3. A Dual Algorithm

The algorithm presented below is iterative. At each iteration, an estimate of the solution (λ, μ) is known. Then, an acceptable step $(\delta\lambda, \delta\mu)$ is calculated and we let the next iterate be $(\lambda, \mu) + (\delta\lambda, \delta\mu)$.

Our algorithm is based on Newton's method for the dual problem (2.4). Newton's step for (2.4) is the solution $(\delta\lambda, \delta\mu)^T$ of the equation

$$\nabla\Psi(\lambda, \mu) + \nabla^2\Psi(\lambda, \mu)\begin{pmatrix}\delta\lambda \\ \delta\mu\end{pmatrix} = \begin{pmatrix}0 \\ 0\end{pmatrix}. \tag{3.1}$$

When $\nabla^2\Psi(\lambda, \mu)$ is nonsingular, this step is

$$\rho = -(\nabla^2\Psi(\lambda, \mu))^{-1}\nabla\Psi(\lambda, \mu), \tag{3.2}$$

but, because the matrix $\nabla^2\Psi(\lambda,\mu)$ may be singular, we use the generalised Newton step

$$\bar{p} = -(\nabla^2\Psi(\lambda,\mu))^{+}\nabla\Psi(\lambda,\mu) \tag{3.3}$$

as a trial step. Since $\nabla^2\Psi(\lambda,\mu)$ is a two by two negative semi-definite matrix, its generalised inverse is easy to calculate. It can be seen that ρ is the maximiser of the quadratic function

$$Q(s) = s^T\nabla\Psi(\lambda,\mu) + \frac{1}{2}s^T\nabla^2\Psi(\lambda,\mu)s, \quad s \in \Re^2, \tag{3.4}$$

when $\nabla^2\Psi(\lambda,\mu)$ is nonsingular. But in the case when $\nabla^2\Psi(\lambda,\mu)$ is singular, the generalised Newton step $\bar{p}$ is only a maximizer of (3.4) in the range space of $\nabla^2\Psi(\lambda,\mu)$. Further, for any positive number M satisfying

$$M \geq -\text{ Trace }[\nabla^2\Psi(\lambda,\mu)] = d(\lambda,\mu)^T H(\lambda,\mu)^{-1}d(\lambda,\mu) + y(\lambda,\mu)^T H(\lambda,\mu)^{-1}y(\lambda,\mu), \tag{3.5}$$

it can be shown that

$$Q\Big(\frac{1}{M}\nabla\Psi(\lambda,\mu)\Big) > Q(\bar{p}) \tag{3.6}$$

if

$$\bar{p}^T\nabla\Psi(\lambda,\mu) < \frac{1}{M}\|\nabla\Psi(\lambda,\mu)\|_2^2. \tag{3.7}$$

Therefore, at every iteration, a number M satisfying (3.5) is available and we use a steepest ascent step

$$\hat{p} = \frac{1}{M}\nabla\Psi(\lambda,\mu) \tag{3.8}$$

whenever inequality (3.7) holds. It is noticed that (3.7) is satisfied only when the matrix $\nabla^2\Psi(\lambda,\mu)$ is singular. Lemma 2.3 tells us that all solutions of the dual problem (2.4) are in a bounded set. Thus, we also use the steepest ascent step $\hat{p}$ as a trial step if the generalised Newton step too large, that is

$$\|\bar{p}\|_2 > s, \tag{3.9}$$

where s is a parameter updated at each iteration.

At the boundary, we search along the boundary if necessary. At a point $(0,\mu)^T$, if $\|d(0,\mu)\|_2^2 - \Delta^2 \leq 0$ (this may indicate the case when (1.2) of the original problem is inactive) or the first component of the calculated trial step (either $\bar{p}$ or $\hat{p}$) is negative (the trial step is infeasible), we then use the 'projected steepest ascent direction':

$$\begin{pmatrix} \delta\lambda \\ \delta\mu \end{pmatrix} = \begin{pmatrix} 0 \\ (\|A^Td + c\|_2^2 - \xi^2)/2y^TH^{-1}y \end{pmatrix}, \tag{3.10}$$

where $d = d(\lambda,\mu), y = y(\lambda,\mu)$ and $H = H(\lambda,\mu)$. Similarly, at a boundary point $(\lambda,0)^T$ if $\|A^Td+c\|_2^2 - \xi^2 \leq 0$ or if the second component of the trial step is negative, we use the step

$$\begin{pmatrix} \delta\lambda \\ \delta\mu \end{pmatrix} = \begin{pmatrix} (\|d\|_2^2 - \Delta_2^2)/2d^TH^{-1}d \\ 0 \end{pmatrix}. \tag{3.11}$$

A step is truncated if it makes the new point $(\lambda + \delta\lambda \;\; \mu + \delta\mu)^T$ infeasible. That is, we choose the largest $t \in (0,1]$ that satisfies

$$\begin{pmatrix} \lambda \\ \mu \end{pmatrix} + t\begin{pmatrix} \delta\lambda \\ \delta\mu \end{pmatrix} \in \Re_+^2. \tag{3.12}$$

Our condition for accepting a trial step is that either

$$(\nabla\Psi(\lambda+t\delta\lambda,\mu+t\delta\mu))^T\begin{pmatrix}\delta\lambda\\ \delta\mu\end{pmatrix}\geq 0 \tag{3.13}$$

or

$$\Psi(\lambda+t\delta\lambda,\mu+t\delta\mu)\geq\Psi(\lambda,\mu)+vt(\nabla\Psi(\lambda,\mu))^T\begin{pmatrix}\delta\lambda\\ \delta\mu\end{pmatrix} \tag{3.14}$$

holds, where $v\in(0,0.5)$ is a pre-set constant. If a trial step $\bar{\rho}$ is unacceptable, we replace it by $\hat{\rho}$. Then we keep increasing M by twice until the step $\hat{\rho}$ is acceptable. It can be seen that (3.13) holds if

$$M\geq\max_{(\lambda,\mu)^T\in\Re^2_+}\|\nabla^2\Psi(\lambda,\mu)\|_2. \tag{3.15}$$

Due to Lemma 2.1, the right-hand side of (3.15) is finite. Because $\Psi(\lambda,\mu)$ is concave, if (3.13) holds but (3.14) fails, for any θ such that

$$\Psi(\lambda+t\theta\delta\lambda,\ \mu+t\theta\delta\mu)\geq\Psi(\lambda,\mu)+vt\theta(\nabla\Psi(\lambda,\mu))^T\begin{pmatrix}\delta\lambda\\ \delta\mu\end{pmatrix} \tag{3.16}$$

we have

$$\Psi(\lambda+t\delta\lambda,\ \mu+t\delta\mu)\geq\Psi(\lambda+t\theta\delta\lambda,\ \mu+t\theta\delta\mu). \tag{3.17}$$

We can easily show (3.16) holds for

$$\theta=\frac{2(1-v)}{Mt}\left\|\begin{pmatrix}\delta\lambda\\ \delta\mu\end{pmatrix}\right\|_2^{-2}(\nabla\Psi(\lambda,\mu))^T\begin{pmatrix}\delta\lambda\\ \delta\mu\end{pmatrix}, \tag{3.18}$$

where M satisfies (3.15). Therefore

$$\Psi(\lambda+t\delta\lambda,\mu+t\delta\mu)\geq\Psi(\lambda,\mu)+\frac{2v(1-v)}{M}\left\|\begin{pmatrix}\delta\lambda\\ \delta\mu\end{pmatrix}\right\|_2^{-2}\left[(\nabla\Psi(\lambda,\mu))^T\begin{pmatrix}\delta\lambda\\ \delta\mu\end{pmatrix}\right]^2. \tag{3.19}$$

Thus, either (3.14) or (3.19) is satisfied for an acceptable step.

If a trial step (3.10) is unacceptable, it can be shown that

$$\|A^Td(0,\mu)+c\|_2^2-\xi^2<0 \tag{3.20}$$

and

$$\|A^Td(0,\mu+\delta\mu)+c\|_2^2-\xi^2>0, \tag{3.21}$$

where $\delta\mu$ is defined in (3.10). Then we set

$$\delta\mu:=\delta\mu\frac{\xi^2-\|A^Td(0,\mu)+c\|_2^2}{\|A^Td(0,\mu+\delta\mu)+c\|_2^2-\|A^Td(0,\mu)+c\|_2^2}. \tag{3.22}$$

We can verify that the new trial step $(0,\delta\mu)^T$ satisfies inequality (3.13), because the function $\|A^Td(0,\mu)+c\|_2^2-\xi^2$ (in μ) is convex for all $\mu\geq 0$. Similarly, if a trial step (3.11) is unacceptable, we set

$$\delta\lambda:=\delta\lambda\frac{\Delta^2-\|d(\lambda,0)\|_2^2}{\|d(\lambda+\delta\lambda,0)\|_2^2-\|d(\lambda,0)\|_2^2}, \tag{3.23}$$

and then the new step $(\delta\lambda,0)^T$ is acceptable.

Now the details of the algorithm can be given as follows:

Algorithm 3.1.

Step 0. Set $\lambda_1 = 0, \mu_1 = 0, M_0 > 0, s_0 > 0$ and $k = 1$.

Step 1.Factorize $H(\lambda_k, \mu_k) = LL^T$; Calculate $d(\lambda_k, \mu_k)$.

Step 2. Calculate $w_1 = (\|d\|_2^2 - \Delta^2)/2, w_2 = (\|A^T d + c\|_2^2 - \xi^2)/2$; Convergence test; if convergence then stop; $M_k = \max\{M_{k-1}, d^T H^{-1} d + y^T H^{-1} y\}$; $s_k = \max\{s_{k-1}, \|\nabla\Psi\|_2/M_k\}$; If $\lambda_k = 0$ and $w_1 \le 0$ go to Step 4; If $\mu_k = 0$ and $w_2 \le 0$ go to Step 5.

Step 3. Calculate the generalized Newton step and set $(\delta\lambda_k\ \delta\mu_k)^T = \bar{p}$; Set $(\delta\lambda_k\ \delta\mu_k)^T = \hat{p}$ if either (3.7) or (3.9) holds; If $\lambda_k = 0$ and $\delta\lambda_k < 0$ go to Step 4; If $\mu_k = 0$ and $\delta\mu_k < 0$ go to Step 5; Go to Step 6.

Step 4. Calculate the step by (3.10) and go to Step 6.

Step 5. Calculate the step by (3.11).

Step 6. Truncate the step if necessary (calculate $t_k = 1/\max\{1, -\delta\lambda_k/\lambda_k, -\delta\mu_k/\mu_k\}$); If either (3.13) or (3.14) holds then go to Step 8.

Step 7. If $(\delta\lambda_k\ \delta\mu_k)^T$ is calculated by (3.10) (or (3.11)) then adjust $\delta\mu_k$ (or $\delta\lambda_k$) by (3.22) (or 3.23)) and go to Step 8; Set $(\delta\lambda_k\ \delta\mu_k)^T = \hat{p}$; Calculate the smallest nonnegative integer $I(k)$ such that either (3.13) or (3.14) holds for $2^{-I(k)}(\delta\lambda_k\ \delta\mu_k)^T$; Set $(\delta\lambda_k\ \delta\mu_k)^T = 2^{-I(k)}(\delta\lambda_k\ \delta\mu_k)^T$; Set $M_k = 2^{I(k)}M_k/t_k$ if $I(k) > 0$.

Step 8. Set $\lambda_{k+1} = \lambda_k + t_k\delta\lambda_k; \mu_{k+1} = \mu_k + t_k\delta\mu_k$; Set $k = k+1$ and go to Step 1.

The following lemma shows that $I(k) = 0$ for all large k; therefore at most two trial steps are calculated in each iteration when k is sufficiently large.

Lemma 3.2. *There are only finitely many k such that $I(k) > 0$. Consequently, $M_k (k = 1, 2, \cdots)$ is bounded above.*

Proof. If there are infinitely many k such that $I(k) > 0$, we can show

$$\lim_{k \to \infty} M_k = \infty. \tag{3.24}$$

Hence there exists an integer K_0 such that M_k satisfies (3.15) for all $k \ge K_0$. Therefore $\hat{p}$ is an acceptable step, which implies $M_k = M_{K_0}$ for all $k \ge K_0$. This contradicts the limit (3.24). Thus the lemma is true.

It is easy to see that $(\lambda\ \mu)^T$ solves problem (2.4) if and only if

$$w_1 \le 0, \quad w_2 \le 0, \quad w_1\lambda = 0, \quad w_2\mu = 0, \tag{3.25}$$

where w_1 and w_2 are defined in Step 2 of Algorithm 3.1. Hence a practical condition for terminating the algorithm is requiring that (3.25) be satisfied within a prescribed tolerance error $\varepsilon > 0$.

The reason for us to study problem (1.1)–(1.3) is that it is a subproblem of some trust region algorithms for nonlinear constrained optimization. To the original constrained optimization problem, (1.1)–(1.3) is only a subproblem appearing in every iteration, so it seems unwise to solve the subproblem to a very high accuracy. Hence it is desirable to have a relaxation stopping criterion that is suitable for application to trust region methods. We suggest that the following conditions be used:

$$\begin{cases} \|d(\lambda, \mu)\|_2^2 - \Delta^2 \le \eta\Delta^2, \\ \|A^T d(\lambda, \mu) + c\|_2^2 - \xi^2 \le \eta[\|c\|_2^2 - \xi^2] \end{cases} \tag{3.26}$$

and

$$\begin{cases} \|d(\lambda,\mu)\|_2^2 - \Delta^2 \geq -\eta\Delta^2, & \text{if } \lambda > 0, \\ \|A^T d(\lambda,\mu) + c\|_2^2 - \xi^2 \geq -\eta[\xi^2 - \xi_{\min}^2] & \text{if } \mu > 0 \end{cases} \tag{3.27}$$

where η is a given number in $(0, 1/2)$.

§4. Convergence Properties

For the convenience of our analysis, we define the vector

$$r(\lambda,\mu) = \nabla\Psi(\lambda,\mu) \tag{4.1}$$

for $(\lambda,\mu) \in \Re_+^2$, except that

$$r(\lambda,\mu) = \begin{cases} e_2 e_2^T \nabla\Psi(\lambda,\mu) & \text{if } \lambda = 0 \text{ and } \|d\|_2^2 - \Delta^2 \leq 0, \\ e_1 e_1^T \nabla\Psi(\lambda,\mu) & \text{if } \mu = 0 \text{ and } \|A^T d + c\|_2^2 - \xi^2 \leq 0, \\ (0\ 0)^T & \text{if } \lambda = \mu = 0, \|d\|_2^2 \leq \Delta^2 \text{ and } \|A^T d + c\|_2^2 \leq \xi^2, \end{cases} \tag{4.2}$$

where $e_1 = (1, 0)^T$ and $e_2(0\ 1)^T$. And we use the notation

$$z_k = \|r(\lambda_k,\mu_k)\|_2 \tag{4.3}$$

for all k. It can be shown that $(\lambda\ \mu)^T$ solves problem (2.4) if and only if $r(\lambda,\mu) = 0$. Hence we need to prove $z_k \to 0$. First we have the following lemma.

Lemma 4.1. *Let $(\delta\lambda_k\ \delta\mu_k)^T$ be the step at the k-th iteration. We have*

$$\alpha_k = w_k^T \nabla\Psi_k / \|w_k\|_2 \|\nabla\Psi_k\|_2 \geq \min\{1, z_k/M_k s_k\}, \tag{4.4}$$

where $w_k = (\delta\lambda_k\ \delta\mu_k)^T$ and $\nabla\Psi_k = \nabla\Psi(\lambda_k,\mu_k)$.

Proof. If the step $w_k = \hat{p}$, (4.4) is trivial because $\alpha_k = 1$. When the step w_k is the generalized Newton step $\bar{p}$, we have

$$\alpha_k \geq \frac{1}{M_k}\|\nabla\Psi_k\|_2^2/\|\bar{p}\|_2\|\nabla\Psi_k\|_2 \geq z_k/M_k s_k, \tag{4.5}$$

which shows (4.4) is true. Now we assume the step w_k is defined by (3.10). It can be seen that either $\|d(\lambda_k,\mu_k)\|_2^2 - \Delta^2 \leq 0$ or $e_1^T\bar{p} < 0$. In the first case, we have

$$\alpha_k = |e_2^T\nabla\Psi_k|/\|\nabla\Psi_k\|_2 = z_k/\|\nabla\Psi_k\|_2 \geq z_k/M_k s_k. \tag{4.6}$$

If $e_1^T\bar{p} < 0$ and $\|d(\lambda_k,\mu_k)\|_2^2 - \Delta^2 \geq 0$, it follows that

$$\begin{aligned}(e_2^T\nabla\Psi_k)e_2^T\bar{p} &= \bar{p}^T\nabla\Psi_k - (e_1^T\nabla\Psi_k)e_1^T\bar{p} = \bar{p}^T\nabla\Psi_k - \frac{1}{2}(\|d(\lambda_k,\mu_k)\|_2^2 - \Delta^2)e_1^T\bar{p} \\ &\geq \bar{p}^T\nabla\Psi_k \geq \|\nabla\Psi_k\|_2^2/M_k.\end{aligned} \tag{4.7}$$

Thus, we have the inequality

$$\alpha_k = |e_2^T\nabla\Psi_k|/\|\nabla\Psi_k\|_2 \geq \|\nabla\Psi_k\|_2/\|\bar{p}\|_2 M_k \geq z_k/M_k s_k, \tag{4.8}$$

which implies (4.6). Similarly, it can be proved that (4.6) is also true when $(\delta\lambda_k\ \delta\mu_k)^T$ is defined by (3.11).

Using the above lemma, we can prove the global convergence of Algorithm 3.1.

Theorem 4.2. *The sequence $\{(\lambda_k\ \mu_k)^T; k = 1,2,3,\cdots\}$ generated by Algorithm 3.1 converges in the sense that*

$$\lim_{k\to\infty} z_k = 0 \tag{4.9}$$

and

$$\lim_{k\to\infty} d(\lambda_k, \mu_k) = d^*, \tag{4.10}$$

where d^* is the unique solution of problem (1.1)–(1.3).

Proof. Because $\Psi(\lambda,\mu)$ is a concave function and $\Psi(\lambda_{k+1},\mu_{k+1}) > \Psi(\lambda_k,\mu_k)$, it can be shown that (4.9) is equivalent to

$$\lim_{k\to\infty} \inf z_k = 0. \tag{4.11}$$

First we consider lower bounds for the length of step $w_k = (\delta\lambda_k\ \delta\mu_k)^T$. It can be seen that

$$\|w_k\|_2 = \|\nabla\Psi_k\|_2/M_k \ge z_k/M_k \tag{4.12}$$

if $w_k = \hat{p}$, and

$$\|w_k\|_2 = \|(\nabla^2\Psi_k)^+\nabla\Psi_k\|_2/M_k \ge \|\nabla\Psi_k\|_2/\|\nabla^2\Psi_k\|_2 \ge z_k/\|\nabla^2\Psi_k\|_2 \tag{4.13}$$

if $w_k = \bar{p}$. For w_k defined by (3.10), we have

$$\|w_k\|_2 = z_k/y(\lambda_k,\mu_k)^T H(\lambda_k,\mu_k)^{-1} y(\lambda_k,\mu_k) \ge z_k/M_k \tag{4.14}$$

if $\|d(\lambda_k,\mu_k)\|_2^2 - \Delta^2 \le 0$. Otherwise, applying (4.7) and inequality $M_k \ge y_k^T H_k^{-1} y_k$, we can show

$$\|w_k\|_2 = |e_2^T\nabla\Psi_k|/y_k^T H_k^{-1} y_k \ge z_k^2/M_k^2 s_k, \tag{4.15}$$

where we use $y_k = y(\lambda_k,\mu_k)$ and $H_k^{-1} = H(\lambda_k,\mu_k)^{-1}$. Therefore

$$\|w_k\|_2 \ge \min\{1, z_k/M_k s_k\} z_k/M_k, \tag{4.16}$$

if w_k is defined by (3.10). If the step (3.10) is modified by (3.22), we can show that

$$\|w_k\|_2 \ge \min\{1, z_k/M_k s_k\} 2z_k^2/\bar{M} M_k \tag{4.17}$$

where $\bar{M} = \max_{\mu\ge 0} \|A^T d(0,\mu) + c\|_2^2$. Similarly, when $(\delta\lambda_k\ \delta\mu_k)^T$ is defined by either (3.11) or (3.23), we have

$$\|w_k\|_2 \ge \min\{1, 2z_k/\hat{M}\} \min\{1, z_k/M_k s_k\} z_k/M_k, \tag{4.18}$$

where $\hat{M} = \max_{\lambda\ge 0} \|d(\lambda,0)\|_2^2$.

If (4.11) is false, there exists a constant $\tau > 0$ such that

$$z_k \ge \tau \tag{4.19}$$

for all k. Now (4.12)–(4.19) show for some constant $\beta > 0$

$$\|w_k\|_2 \ge \beta \tag{4.20}$$

for all k. Since the condition for accepting a step ensures that either (3.14) or (3.19) holds, we have

$$\Psi(\lambda_{k+1},\mu_{k+1}) \ge \Psi(\lambda_k,\mu_k) + \min\left\{ v t_k z_k \beta \min\{1, z_k/M_k s_k\}, \frac{2v(1-v)z_k^2}{M^*}[\min\{1, z_k/M_k s_k\}]^2 \right\} \tag{4.21}$$

where $M^* = \max_{\lambda \geq 0, \mu \geq 0} \|\nabla^2 \Psi(\lambda, \mu)\|_2$. Due to (4.19) and (4.21), there exist positive constants β_1 and β_2 such that

$$\Psi(\lambda_{k+1}, \mu_{k+1}) \geq \Psi(\lambda_k, \mu_k) + \min\{\beta_1 t_k, \beta_2\}. \tag{4.22}$$

Because $\Psi(\lambda_k, \mu_k)$ $(k = 1, 2, \cdots)$ is bounded above,

$$\sum_{k=1}^{\infty} \min\{t_k \beta_1, \beta_2\} < \infty, \tag{4.23}$$

which implies

$$\sum_{k=1}^{\infty} t_k < \infty. \tag{4.24}$$

Therefore, since $(\delta\lambda_k \ \delta\mu_k)^T$ is bounded, the sequence $(\lambda_k \ \mu_k)^T$ converges, say to $(\lambda^* \ \mu^*)^T$. (4.24) also shows that $\min\{\lambda_k, \mu_k\} = 0$ for all large k. However $\lambda_k = O(\mu_k = 0)$ and $t_k < 1$ imply $\mu_{k+1} = 0(\lambda_{k+1} = 0)$ for sufficiently large k. Thus $(\lambda^* \ \mu^*)^T = (0 \ 0)^T$, which contradicts the inequality

$$\Psi(\lambda^*, \mu^*) \geq \lim_{k \to \infty} \Psi(\lambda_k, \mu_k) \geq \Psi(\lambda_2, \mu_2) > \Psi(0, 0). \tag{4.25}$$

This shows that (4.11) holds. Hence (4.9) is true and

$$\lim_{k \to \infty} \text{dist} \left(\begin{pmatrix} \lambda_k \\ \mu_k \end{pmatrix}, \Pi \right) = 0, \tag{4.26}$$

where Π is the set of solutions of the dual problem (2.4). Now (4.10) follows from (4.26), the boundedness of Π and $d(\lambda, \mu) = d^*$ for all $(\lambda \ \mu)^T \in \Pi$.

The local convergence result is as follows:

Theorem 4.3. *Assume* $\{(\lambda_k \ \mu_k)^T; k = 1, 2, 3, \cdots\}$ *generated by Algorithm* 3.1 *converges to* $(\lambda^* \ \mu^*)^T$*, furthermore,* $\nabla^2 \Psi(\lambda^*, \mu^*)$ *is positive definite if*

1) $\lambda^* > 0, \mu^* > 0$; or

2) $\lambda^* = 0, \|d^*\|_2^2 - \Delta^2 = 0$; or

3) $\mu^* = 0, \|A^T d^* + c\|_2^2 - \xi^2 = 0$.

Then $(\lambda_k \ \mu_k)^T$ converges to $(\lambda^* \ \mu^*)^T$ Q-superlinearly.

Proof. In case 1) Newton's step will be taken for all large k; hence Q-superlinear convergence follows.

In Cases 2) and 3), either Newton's step or a step along the boundary ((3.10) or (3.11)) will give superlinear convergence.

To complete the proof, we now consider the case when $\lambda^* = 0$ and $\|d^*\|_2^2 - \Delta^2 < 0$ and the case when $\mu^* = 0$ and $\|A^T d^* + c\|_2^2 - \xi^2 < 0$. In both cases, it can be easily shown that the iterates shall converge to the solution along the boundary and the step (3.10) or (3.11) is used. In the first case, $\lambda^* = 0$ and $\|d^*\|_2^2 - \Delta^2 < 0$, and the algorithm is exactly the same as the Newton-Raphson method for solving the equation $\|A^T d(0, \mu) + c\|_2^2 - \xi^2 = 0$. Since the function $\|A^T d(0, \mu) + c\|_2^2 - \xi^2 = 0$ is convex for all $\mu \geq 0$, Newton's step gives superlinear convergence and the iterates converge to the solution monotonically. It can be easily shown that a superlinear convergent step shall satisfy condition (3.14). Thus superlinear convergence of the iterates is proved.

§5. Numerical Results

Our algorithm is implemented in Fortran, and the following problems are solved by the algorithm. $n = 4$ and $\Delta = 1$ in all the problems.

Problem 1. $m = 1, \xi = 0.5, g = (0.5, 1, 1, 1)^T, B = I, c = -1$ and $A = (1, 0, 0, 0)^T$.

Problem 2. $m = 2, \xi = \sqrt{2}, g = (0, -0.5, 0, 0)^T, B = \text{diag}\,[1, 2, 3, 4], c = (-2, 0)^T$ and $A = (I_{2\times 2} O_{2\times 2})^T$.

Problem 3. $m = 3, \xi = 0.6, g = (-3, -4, -5, 0)^T, B = I, c = (-0.3, -0.4, -0.5)^T$ and $A = (I_{3\times 3} O_{3\times 1})^T$.

Problem 4. $m = 4, \xi = \sqrt{3}, g = (-5, -1, -1, -1)^T, B = \text{diag}\,[1, 1/2, 1/3, 1/4], c = (1, 1, 1, -0.5)^T$ and $A = (A_{i,j})_{4\times 4}$ where $A_{i,j} = 1$ for all $i, j = 1, 2, 3, 4$.

Problem 5. As Problem 4, except that $A_{i,j} = 0.1$ for all $i \neq j$.

The calculations were done by an IBM 3081 computer. The convergence test is that the conditions in (3.25) hold within the error ε. We set $v = 0.01$ (see (3.14)) and $\varepsilon = 10^{-12}$ in all problems. Step length of one is accepted at every iterations of all the 5 problems. The results are as follows:

Problem 1 is solved after 8 iterations, the solution found is $(0.5000000, -0.5000000, -0.5000000, -0.5000000)^T$, and the multipliers are $\lambda = 1.0000000, \mu = 3.0000000$.

Problem 2 is solved after 6 iterations, the solution found is $(0.6008422, 0.2058095, 0.0000000, 0.0000000)^T$, and the multipliers are $\lambda = 0.0000000, \mu = 0.4294314$.

Problem 3 is solved after 11 iterations, the solution found is $(0.4242641, 0.5656854, 0.7071068, 0.0000000)^T$, and the multipliers are $\lambda = 6.0710678, \mu = 0.0000000$.

Problem 4 is solved after 14 iterations, the solution found is $(0.8525473, -0.2878663, -0.3040365, -0.3128225)^T$, and the multipliers are $\lambda = 2.6337108, \mu = 0.8301400$.

Problem 5 is solved after 11 iterations, the solution found is $(0.5827114, -0.4720780, -0.4955691, -0.4381793)^T$, and the multipliers are $\lambda = 1.1614503, \mu = 2.2895731$.

§6. Discussion

The advantages of Algorithm 3.1 are that it works on a dual program which has only two variables, and that Newton's method can be used because the gradients and the Jacobian of the objective function in the dual program can be easily computed. Since Ψ is concave, the global convergence and local superlinear convergence results can be also easily proved.

One point that should be mentioned is that eventually there is no line search in the algorithm. For large k, either Newton's step or a steepest acsent step will be taken except that at the boundary maybe a projected newton's step is used.

The convergence results in Section 4 are based on the assumption $\xi > \xi_{\min}$. If $\xi = \xi_{\min}$, it is possible for the set of solutions of (2.4) to be unbounded. Fortunately, when (1.1)–(1.3) is a subproblem of a trust region algorithm for nonlinear constrained optimization, it is normally known if $\xi = \xi_{\min}$.

One may also be interested in the upper and lower bounds of the final value $\Psi(\lambda^*, \mu^*) = \Phi(d^*)$. It is easy to see that $\Psi(\lambda_k, \mu_k) \le \Psi(\lambda^*, \mu^*)$ for all k. We can show that $\Psi(\lambda^*, \mu^*) \le \Phi(d(\lambda, \mu))$ for any $d(\lambda, \mu)$ satisfies the constrained condition (1.2)–(1.3). Hence for any k, if $d(\lambda_k, \mu_k)$ is a feasible point of (1.2)–(1.3), we have $\Psi(\lambda^*, \mu^*) \le \Phi(d(\lambda_k, \mu_k))$. Another upper bound for $\Psi(\lambda^*, \mu^*)$ is $\|g\|_2\Delta + \|B\|_2\Delta^2/2$ since $\Phi(d^*) \le \|g\|_2\Delta + \|B\|_2\Delta^2/2$.

Acknowledgement. The author would like to thank Professor M.J.D. Powell for his constant help and encouragement and for his reading an early version of the paper, and is indebted to him for his suggestions and comments.

References

[1] 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,* 1985), 71–82.

[2] R. Fletcher, Practical Methods for Optimization, Vol. 2, Constrained Optimization, John Wiley and Sons, Chichester, 1981.

[3] D.M. Gay, Computing optimal locally constrained steps, *SIAM J. Sci. Stat. Comput.,* **2** (1981), 186–197.

[4] J.J. Moré and D.C. Sorensen, Computing a trust region step, *SIAM J. Sci. Stat. Comput.,* **4** (1983), 553–572.

[5] M.J.D. Powell and Yuan Ya-xiang, A trust region algorithm for equality constrained optimization, *Mathematical Programming,* **49** (1991), 189–211.

[6] R.T. Rockafellar, Convex Analysis, Princeton University Press, Princeton, New Jersey, 1970.

[7] Yuan Ya-xiang, On a subproblem of trust region algorithms for constrained optimization, *Mathematical Programming,* **47** (1990), 53–63.