International Institute for Applied Systems Analysis
IIASA
www.iiasa.ac.at

A Primal-Dual Quasi-Newton Method for Constrained Optimization

Nakayama, H. and Orimo, M.

IIASA Collaborative Paper
October 1984

NOT FOR QUOTATION
WITHOUT PERMISSION
OF THE AUTHOR

A PRIMAL-DUAL QUASI-NEWTON METHOD FOR CONSTRAINED OPTIMIZATION

H. Nakayama
M. Orimo

October 1984
CP-84-44

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
A-2361 Laxenburg, Austria

PREFACE

In this paper, the authors propose an efficient new method for constrained optimization which they call the *primal-dual quasi-Newton method*. The main feature of this method is that it improves both the Hessian of the Lagrangian and that of the dual objective function using quasi-Newton methods. Several variants of the method are possible: the properties of these methods are described and the computational results obtained for some test problems are given.

This research was carried out in collaboration with the Interactive Decision Analysis Project in the System and Decision Sciences Program.

ANDRZEJ WIERZBICKI

Chairman
System and Decision Sciences
Program

ABSTRACT

One of the most important developments in nonlinear constrained optimization in recent years has been the recursive quadratic programming (RQP) method suggested by Wilson, Han, Powell and many other researchers. It is clear that the role of the auxiliary quadratic programming problem is to calculate (implicitly) the inverse Hessian of the dual objective function. We describe the Hessian of the Lagrangian and that of the dual objective function as the primal Hessian and the dual Hessian, respectively. In this paper, a new method for constrained optimization, called the primal-dual quasi-Newton method, is proposed. The main feature of this method is that it improves (explicitly) both the primal Hessian and the dual Hessian using quasi-Newton methods. Several variants of the primal-dual quasi-Newton method are possible: the properties of these methods are described and the computational results obtained for some test problems are given.

A PRIMAL-DUAL QUASI-NEWTON METHOD FOR CONSTRAINED OPTIMIZATION

H. Nakayama* and M. Orimo**

*Department of Applied Mathematics,
Konan University, Okamoto, Higashinada,
Kobe 658, Japan

**Hitachi System Development,
Kawasaki 215, Japan

1. INTRODUCTION

Consider the following problem:

$$\begin{aligned} &\text{Minimize} \quad f(x) \\ &\text{subject to } h_i(x) = 0, \ i = 1,\dots,m \\ &\qquad\qquad x \in E^n , \end{aligned} \tag{A}$$

where E^n is an n-dimensional Euclidean space.

The recursive quadratic programming (RQP) method has been recognized as an effective means of solving general nonlinear problems of this type. It does not make any assumptions about the functions f and h_i, except that they should be smooth (in some appropriate sense) [1-3]. The Lagrangian associated with problem A is defined by

$$L(x,u) = f(x) + u^T h(x),$$

where $u^T = (u_1,\dots,u_m)$ and $h = (h_1,\dots,h_m)^T$. The RQP algorithm can then be summarized as follows:

Presented at the XI International Symposium on Mathematical Programming in Bonn (August 1982).

(i) *Choice of search direction* Δx^k
Determine Δx^k by solving the following auxiliary quadratic programming problem:

$$\text{Minimize} \quad f_x(x^k)\Delta x + \tfrac{1}{2}\Delta x^T B^k \Delta x$$

$$\text{subject to} \quad h_x(x^k)\Delta x + h(x^k) = 0 ,$$

where the i-th row vector of the matrix h_x is the gradient of h_i with respect to x.

(ii) *Line search;* $x^{k+1} = x^k + \alpha^k \Delta x^k$
Assuming the penalty function

$$P(x;c) = f(x) + c\sum_{i=1}^{m} |h_i(x)|,$$

where c is sufficiently large, the step-size parameter α^k is given by

$$\alpha^k = \arg\min_{\alpha} P(x^k + \alpha\Delta x^k; c) \quad .$$

(iii) *Improvement of* B^k
Improve B^k using some quasi-Newton method, such as the BFGS method:

$$B^{k+1} = B^k - \frac{B^k s s^T B^k}{s^T B^k s} + \frac{y y^T}{s^T y} ,$$

where

$$s = x^{k+1} - x^k$$

$$y = L_x(x^{k+1}, u^{k+1}) - L_x(x^k, u^{k+1})$$

and u^{k+1} is the Lagrange multiplier obtained by solving the auxiliary quadratic programming problem described in (i).

The RQP method has several good features and is of considerable importance [2]. On the other hand, however, the method also has some drawbacks in that it is necessary to solve a succession of auxiliary quadratic programming problems and the line search has to be made along the non-smooth function P(x;c) in order to ensure global convergence. Before considering how to overcome these difficulties, we shall first look at the role of auxiliary quadratic programming in RQP.

From the Kuhn-Tucker condition for the auxiliary quadratic programming problem, we have

$$\begin{pmatrix} B^k & h_x^T(x^k) \\ h_x(x^k) & 0 \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta u \end{pmatrix} = - \begin{pmatrix} L_x(x^k, u^k) \\ h(x^k) \end{pmatrix} , \tag{1}$$

where $\Delta u = u^{k+1} - u^k$. The Kuhn-Tucker condition for problem A is given by

$$L_x(x,u) = 0 \tag{2}$$

$$h(x) = 0 \quad . \tag{3}$$

Applying the Newton-Raphson method to equations (2) and (3), we have

$$\begin{pmatrix} L_{xx}(x^k, u^k) & h_x^T(x^k) \\ h_x(x^k) & 0 \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta u \end{pmatrix} = - \begin{pmatrix} L_x(x^k, u^k) \\ h(x^k) \end{pmatrix} . \tag{4}$$

Comparing (4) with (1), it is clear that RQP is essentially equivalent to the Newton-Raphson method for equations (2) and (3) with $L_{xx}(x^k, u^k)$ approximated by B^k. Tanabe has recently reported an attractive unified approach to a class of (quasi) Newton methods for constrained optimization which includes RQP, the generalized reduced gradient method and the gradient projection method as special cases differing only in the approximation of L_{xx} [4].

Now suppose that L_{xx} is non-singular and h_x has full rank. Then, assuming that

$$\det\begin{pmatrix} L_{xx} & h_x^T \\ h_x & 0 \end{pmatrix} \neq 0 \quad ,$$

we have

$$\begin{pmatrix} L_{xx} & h_x^T \\ h_x & 0 \end{pmatrix}^{-1} = \begin{pmatrix} L_{xx}^{-1}+L_{xx}^{-1}h_x^T(-h_xL_{xx}^{-1}h_x^T)^{-1}h_xL_{xx}^{-1} & -L_{xx}^{-1}h_x^T(-h_xL_{xx}^{-1}h_x^T)^{-1} \\ -(-h_xL_{xx}^{-1}h_x^T)^{-1}h_xL_{xx}^{-1} & (-h_xL_{xx}^{-1}h_x^T)^{-1} \end{pmatrix}.$$

Therefore, (4) yields

$$\Delta u = -(h_xL_{xx}^{-1}h_x^T)^{-1}(h-h_xL_{xx}^{-1}L_x) \tag{5}$$

$$\Delta x = -L_{xx}^{-1}L_x - L_{xx}^{-1}h_x^T\Delta u \quad . \tag{6}$$

Here all the functions are evaluated at x^k and u^k. It has already been shown [5,6] that the update scheme (5)-(6) is equivalent to certain existing methods, for example, the Bard-Greenstadt method [7], the multiplier method for inexact unconstrained minimization [6], and the diagonalized multiplier method [5]. Note that two kinds of inverse matrices L_{xx}^{-1} and $(-h_xL_{xx}^{-1}h_x^T)^{-1}$ appear in (5) and (6). We refer to L_{xx} as the *primal Hessian* and to $-h_xL_{xx}^{-1}h_x^T$ as the *dual Hessian*. (The name of the latter originates from the fact that it is the Hessian of the dual objective function $\Phi(u) = \min_x L(x,u)$ associated with problem A.) One interpretation of RQP is therefore that the approximation of the primal Hessian is improved by some quasi-Newton method and the inverse of the dual Hessian is calculated implicitly by solving the auxiliary quadratic programming problem. Based on this consideration, we shall suggest a method, called the *primal-dual quasi-Newton method*, which approximates both the primal Hessian and the dual Hessian using some quasi-Newton method.

2. THE PRIMAL-DUAL QUASI-NEWTON METHOD

Let H_1 and H_2 approximate the inverses of the primal and dual Hessians, respectively. The Newton-Raphson update (5)-(6) is then reduced to

$$\Delta u = -H_2(h-h_x H_1 L_x) \tag{7}$$

$$\Delta x = -H_1 L_x - H_1 h_x^T \Delta u \quad . \tag{8}$$

In primal-dual quasi-Newton methods, the matrices H_1 and H_2 are improved by an appropriate quasi-Newton method, for example, using the BFGS update

$$H^{k+1} = \left(I - \frac{sy^T}{y^Ts}\right) H^k \left(I - \frac{ys^T}{y^Ts}\right) + \frac{ss^T}{y^Ts} \quad , \tag{9}$$

where we take

$$s = x^k - x^{k-1} \text{ and } y = L_x(x^k,u^k) - L_x(x^{k-1},u^{k-1}) \tag{10}$$

for H_1 and

$$s = u^k - u^{k-1} \text{ and } y = h(x^k) - h(x^{k-1}) \tag{11}$$

for H_2. It should be noted here that the gradient of the dual objective function $\Phi(u)$ is given by $h(x(u))$, where $x(u) = \arg\min_x L(x,u)$. If x^k is not a minimizer of $L(x,u^k)$ and is determined merely from $x^{k+1} = x^k + \Delta x^k$, with Δx given by (8), then the algorithm based on (7)-(11) does not necessarily perform very well because the estimate of the gradient of the dual objective function is generally not good enough. We therefore suggest the following method:

(i) First, for a given multiplier u^k, determine the $\hat{x}^k$ that minimizes $L(x,u^k)$.

(ii) From (7), (8) and $L_x(\hat{x}^k,u^k) = 0$, the next search direction from the point $(\hat{x}^k,u^k)$ is

$$\Delta u^k = -H_2^k h(\hat{x}^k) \tag{12}$$

$$\Delta x^k = -H_1^k h_x^T(\hat{x}^k)\Delta u^k \quad , \tag{13}$$

where H_1^k and H_2^k are approximations of $L_{xx}^{-1}(\hat{x}^k,u^k)$ and $(-h_x(\hat{x}^k)H_1^k h_x^T(\hat{x}^k))^{-1}$, respectively, and are improved using (9)-(11).

Note 2.1 Henceforth, we shall assume that L_{xx} is positive definite for all (x,u). The matrix

$$\begin{pmatrix} L_{xx}(\hat{x}^k,u^k) & h_x^T(\hat{x}^k) \\ h_x(\hat{x}^k) & 0 \end{pmatrix}$$

is then nonsingular, where $\hat{x}^k$ minimizes $L(x,u^k)$.

We can now interpret this procedure geometrically as follows: For $x(u) = \arg\min_x L(x,u)$, the dual objective function is given by $\Phi(u) = L(x(u),u)$. Assuming that L_{xx} is positive definite and that the functions f and h_i are smooth (in some appropriate sense), we obtain $x(u)$ by solving $L_x(x,u) = 0$; $x(u)$ also has some appropriate smoothness. Then, taking Δx^k and Δu^k such that $L_x(\hat{x}^k+\Delta x^k,\ u^k+\Delta u^k) = 0$, the solution to

$$L_x(\hat{x}^k+\Delta x^k,\ u^k+\Delta u^k) \cong L_{xx}(\hat{x}^k,u^k)\Delta x^k + h_x^T(\hat{x}^k)\Delta u^k = 0 \tag{14}$$

yields

$$\Delta x^k = -L_{xx}^{-1}(\hat{x}^k,u^k)h_x^T(\hat{x}^k)\Delta u^k \ . \tag{15}$$

Here, Δu^k is given by

$$\Delta u^k = -(-h_x(\hat{x}^k)L_{xx}^{-1}(\hat{x}^k,u^k)h_x^T(\hat{x}^k))^{-1}h(\hat{x}^k), \tag{16}$$

using the Newton method for maximizing the dual objective function. Observe that the Δx^k defined by (13) and the Δu^k defined by (12) are identical to the corresponding definitions in (15)-(16) with $L_{xx}^{-1}(\hat{x}^k,u^k)$ and $(-h_x(\hat{x}^k)L_{xx}^{-1}(\hat{x}^k,u^k)h_x^T(\hat{x}^k))^{-1}$ replaced by H_1 and H_2, respectively. We can therefore say that the search based on (12)-(13) is carried out on the tangent space of the solution surface $\{(x,u) \mid L_x(x,u) = 0\}$ by considering $L_x(x,u) = 0$ as a new constraint. Minimization of $L(x,u)$ over x corresponds to projection onto the constraint surface $L_x(x,u) = 0$. This is illustrated in Fig. 1.

An algorithm based on the suggested primal-dual quasi-Newton method can be summarized as follows:

Step 1. Take initial values (x^0,u^0) and convergence parameters ε_1 and ε_2. Set $H_1 = I$, $H_2 = -I$ and $k = 0$.

Step 2. Solve the unconstrained problem $\min L(x,u^k)$ using an appropriate quasi-Newton method, for example, the BFGS method:

(2-i) $x^{k,0} = x^k$, $H_1^{k,0} = H_1^k$; $i = 0$.

(2-ii) If $\|L_x(x^{k,i},u^k)\| < \varepsilon_1$, then $\hat{x}^k = x^{k,i}$, $H_1^k = H_1^{k,i}$ and go to Step 3. Otherwise, go to (2-iii).

(2-iii) Calculate $x^{k,i+1} = x^{k,i} + \alpha_i \Delta x_1^i$, where

$$\Delta x_1^i = -H_1^{k,i} L_x(x^{k,i},u^k)$$

$$\alpha_i = \arg\min_{\alpha} L(x^{k,i}+\alpha\Delta x_1^i,u^k).$$

(2-iv) Set

$$s = x^{k,i+1}-x^{k,i}$$

$$y = L_x(x^{k,i+1},u^k) - L_x(x^{k,i},u^k)$$

and improve $H_1^{k,i+1}$ using the BFGS update (9).

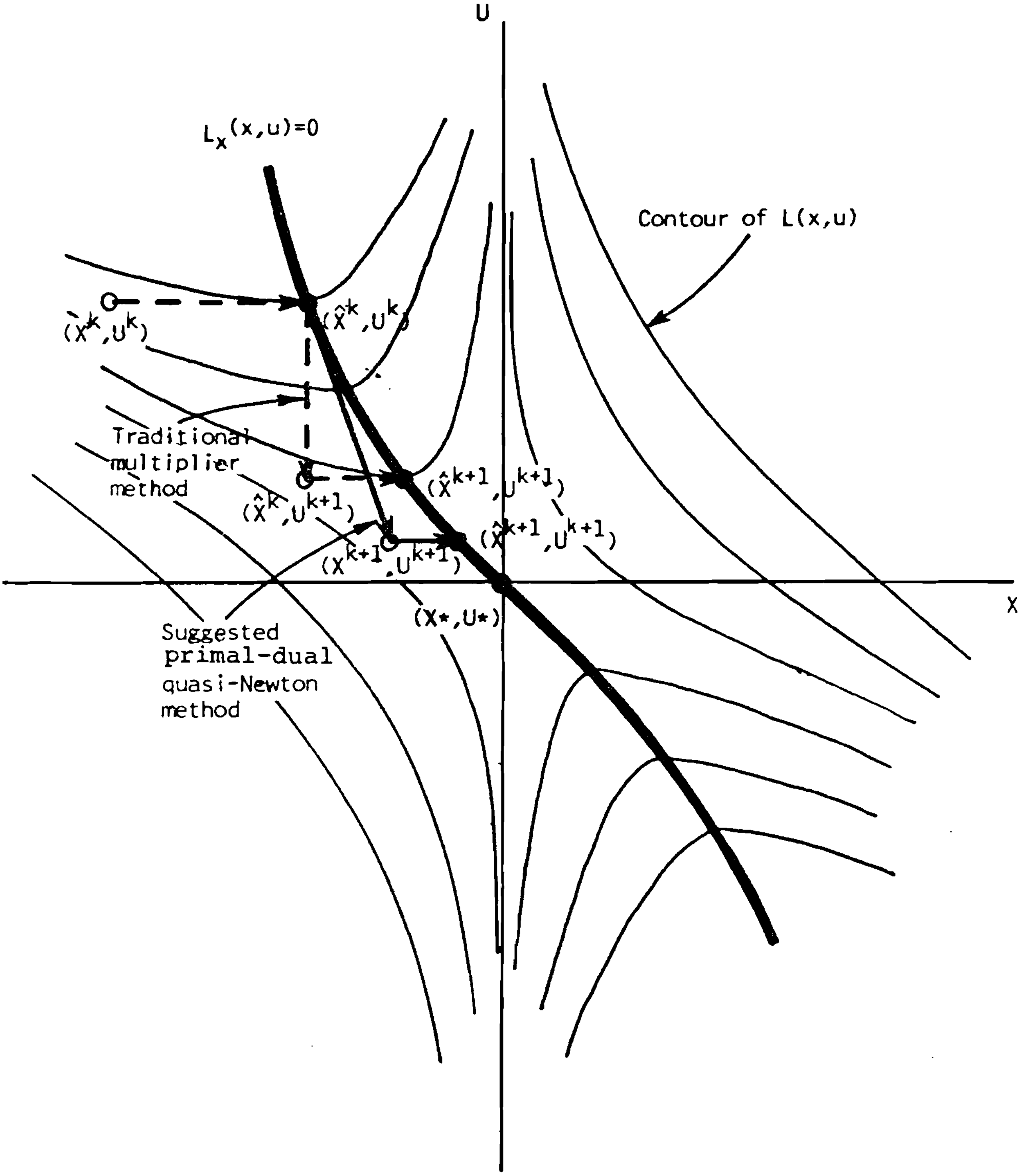

Fig. 1. Geometric interpretation of the primal-dual quasi-Newton method.

(2-v) Set $i = i + 1$ and go to (2-ii).

Step 3. If $k = 0$, then go to Step 4. Otherwise set

$$s = u^k - \tilde{u}$$

$$y = h(\hat{x}^k) - h(\tilde{x})$$

and improve H_2^k using the BFGS update (9).

Step 4. Set $\tilde{x} = \hat{x}^k$ and $\tilde{u} = u^k$.

Step 5. Calculate

$$x^{k+1} = \hat{x}^k + \beta_k \Delta x^k$$

$$u^{k+1} = u^k + \beta_k \Delta u^k$$

where Δx^k and Δu^k are given by (12)-(13) and the step-size parameter β_k is determined as described in the following section.

Step 6. If $\|h(x^{k+1})\| < \varepsilon_2$ and $\|L_x(x^{k+1}, u^{k+1})\| < \varepsilon_1$, then stop. Otherwise, set $k = k + 1$ and go to Step 2.

3. A METHOD OF LINE SEARCH

As stated in the previous section, the update

$$u^{k+1} = u^k - \beta_k H_2^k h(x^k) \tag{17}$$

follows from the quasi-Newton method for maximizing the dual objective function $\Phi(u) = \min_x L(x,u)$ or, equivalently, for solving $\Phi_u(u) = h(x(u)) = 0$. We therefore determine the step-size parameter β_k in such a way that some norm of $h(x)$ is minimized. Here we suggest

$$\|h(x)\|^2_{-H_2} = h^T(x)(-H_2)h(x) \tag{18}$$

as line-search objective function.

Letting

$$\psi(\beta) = \| h(x^k + \beta\Delta^k) \|^2_{-H_2^k} ,$$

we have

$$\psi'(0) = -2h^T(\hat{x}^k) H_2^k h_x(\hat{x}^k) \Delta x^k$$
$$= -2h^T(\hat{x}^k) H_2^k h_x(\hat{x}^k) H_1^k h_x^T(\hat{x}^k) H_2^k h(\hat{x}^k) . \qquad (19)$$

Since H_1^k and $-H_2^k$ are positive definite under the BFGS update, relation (19) yields $\psi'(0) < 0$, assuming that $h_x(\hat{x}^k)$ has maximum rank. This means that the Δx^k given by (13) ensures a search direction in which $\psi(\beta)$ is decreasing.

The reason why $\beta_k = \arg\min \psi(\beta)$ is also used as the step-size parameter when updating u may be understood by taking into account the following relationship between the minimization of $\| h(x) \|^2_{-H_2}$ and the maximization of the dual objective function $\Phi(u)$. Define

$$B_2^k = -h_x H_1^k h_x^T .$$

Then

$$\Phi(u^k + \Delta u) \cong \tfrac{1}{2}\Delta u^T B_2^k \Delta u + h^T \Delta u + L , \qquad (20)$$

where the right-hand side is evaluated at $(\hat{x}^k, u^k)$. On the other hand, since $h(\hat{x}^k + \Delta x) \cong h - h_x H_1^k h_x^T \Delta u$, we have

$$\| h(\hat{x}^k + \Delta x) \|^2_{-H_2^k} \cong -2(\tfrac{1}{2}\Delta u^T B_2^k H_2^k B_2^k \Delta u + h^T H_2^k B_2^k \Delta u) - h^T H_2^k h .$$

Suppose that H_2^k is a sufficiently good approximation of $(B_2^k)^{-1}$, i.e., $H_2^k B_2^k \cong I$. Then from (20) we have

$$\|h(\hat{x}^k+\Delta x)\|^2_{-H_2^k} \cong -2(\tfrac{1}{2}\Delta u^T B_2^k \Delta u + h^T \Delta u) - h^T H_2^k h$$

$$\cong -2\Phi(u^k+\Delta u) - h^T H_2^k h + 2L \quad .$$

Hence, if H^k is a sufficiently good approximation of $(B^k)^{-1}$, then for the β^k that minimizes $\psi(\beta) = \|h(\hat{x}^k+\beta\Delta x^k)\|^2_{-H_2^k}$, $u^k+\beta^k\Delta u^k$ maximizes the dual objective function $\Phi(u^k+\beta\Delta u^k)$.

4. AN EXTENSION TO NONCONVEX AND/OR INEQUALITY CONSTRAINED CASES

We have so far assumed that L_{xx} is positive definite. However, in cases where L_{xx} is not always positive definite, we can develop a similar argument by using some appropriate augmented Lagrangian instead of the conventional Lagrangian. It is shown in [8] that

$$\tilde{L}(x,u,v;c) = f(x) + u^T h(x) + ch^T(x)h(x) \tag{21}$$

$$+ \sum_{i\in I} (v_i g_i(x) + cg_i^2(x)) + \sum_{i\notin I} \frac{v_i^2 g_i(x)}{v_i - cg_i(x)} \quad ,$$

where $I = \{i \mid g_i(x) \geq 0,\ 1 \leq i \leq r\}$, is an augmented Lagrangian for problem A with additional inequality constraints $g_i(x) \leq 0$, $i = 1,\ldots,r$.

The optimal solutions x^*, u^* and v^* are clearly obtained as the solutions to

$$\tilde{L}_x(x,u,v;c) = 0$$

$$\tilde{L}_u(x,u,v;c) = 0$$

$$\tilde{L}_v(x,u,v;c) = 0$$

$$v \geqq 0 \quad ,$$

where

$$\tilde{L}_x(x,u,v;c) = f_x(x) + h_x(x)(u+2ch(x)) + \sum_{i\in I}(v_i+2cg_i(x))g_{i_x}(x) + \sum_{i\notin I}\frac{v_i^3}{(v_i-cg_i(x))^2}g_i(x)$$

$$\tilde{L}_u(x,u,v;c) = h(x)$$

$$\tilde{L}_v(x,u,v;c) = \begin{cases} g_i(x), & i\in I \\ \dfrac{v_i g_i(x)(v_i-2cg_i(x))}{(v_i-cg_i(x))^2} & i\notin I\ . \end{cases}$$

Note that the condition of complementary slackness is embedded in the equation $\tilde{L}_x = 0$. Moreover, it is known that $\tilde{L}_{xx}$ is positive definite even in nonconvex cases as long as c is sufficiently large [8]. Therefore, the suggested primal-dual quasi-Newton method can be modified for use in this case simply by taking the additional condition $v \geqq 0$ into account. (This constraint is easily handled by the gradient projection method: set $v_i = 0$, if $v_i < 0$.)

5. NUMERICAL EXAMPLES

It is clear from the previous sections that the suggested primal-dual quasi-Newton method can be regarded as an accelerated multiplier method. Hence, its convergence can be verified in the same way as that of the multiplier method [9] or the diagonalized multiplier method [10]. Another extension of multiplier methods has been made by Kameyama and others [11], who modified the traditional multiplier method in such a way that the Lagrange multipliers are updated by some quasi-Newton method for maximizing the dual objective function. This method was named the quasi-Newton multiplier method, and may be regarded as another type of primal-dual quasi-Newton method. Unlike the primal-dual quasi-Newton method suggested in this paper, however, the quasi-Newton multiplier method only updates the Lagrange multipliers in the maximization of the dual objective function. Note that in the quasi-Newton multiplier method it becomes virtually impossible to carry out the line search required to update the Lagrange multipliers because each estimation of the step-size parameter requires an infinite number of steps in the unconstrained minimization of the (augmented) Lagrangian. We shall now compare these methods using a few test problems.

Example 1 (Rosen-Suzuki problem)

Minimize

$$f(x) = x_1^2 + x_2^2 + 2x_3^2 + x_4^2 - 5x_1 - 5x_2 - 21x_3 + 7x_4$$

subject to

$$g_1(x) = x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_1 - x_2 + x_3 - x_4 - 8 \leqq 0$$

$$g_2(x) = x_1^2 + 2x_2^2 + x_3^2 + 2x_4^2 - x_1 - x_4 - 10 \leqq 0$$

$$g_3(x) = 2x_1^2 + x_2^2 + x_3^2 + 2x_1 - x_2 - x_4 - 5 \leqq 0 \, .$$

The optimal solution is $x^* = (0, 1, 2, -1)$, $v^* = (1, 0, 2)$ and $f(x^*) = -44$. The results obtained on applying the multiplier method, the quasi-Newton multiplier method and the proposed primal-dual quasi-Newton method to this problem are given in Table 1. The following values were taken: $x^0 = (0, 0, 0, 0)$, $v^0 = (0, 0, 0)$, penalty parameter $c = 1$, and convergence parameters $\varepsilon_1 = 10^{-6}$ and $\varepsilon_2 = 10^{-3}$.

Example 2 (Powell's problem)

Minimize

$$f(x) = x_1 x_2 x_3 x_4 x_5$$

subject to

$$h_1(x) = x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2 - 10 = 0$$

Table 1. Results obtained on applying various methods to the Rosen-Suzuki problem.

Method	f(x)	g(x)	Number of iterations for dual optimization problem	Number of iterations for unconstrained minimization problem	CPU time (ms)
Multiplier method	-44.000	g_1=-0.21×10^{-10} g_2=-1.0000 g_3=0.32×10^{-10}	9	32	45
Quasi-Newton multiplier method	-44.000	g_1=-0.19×10^{-8} g_2=-1.0000 g_3=0.86×10^{-8}	9	24	46
Proposed primal-dual quasi-Newton method	-44.000	g_1=-0.76×10^{-9} g_2=-1.0000 g_3=-0.60×10^{-9}	5	16	41

$$h_2(x) = x_2x_3 - 5x_4x_5 = 0$$

$$h_3(x) = x_1^3 + x_2^3 + 1 = 0 \quad .$$

The optimal solution is $x^* = (-1.71714, 1.59571, 1.82725, -0.763643, -0.763643)$, $u^* = (0.74446, -0.703575, 0.096806)$ and $f(x^*) = -2.91970$. The initial values were taken as $x^0 = (-2, 2, 2, -1, -1)$ and $u^0 = (0, 0, 0)$, while the penalty and convergence parameters were respectively $c = 0.5$, $\varepsilon_1 = 10^{-6}$ and $\varepsilon_2 = 10^{-3}$. The results are given in Table 2.

The augmented Lagrangian (21) was used in each case. When using the traditional multiplier method [8], we increased the penalty parameter in such a way that $c_{k+1} = 2c_k$ at each update of the Lagrange multipliers. In general, as the penalty parameter

Table 2. Results obtained on applying various methods to Powell's problem.

Method	f(x)	g(x)	Number of iterations for dual optimization problem	Number of iterations for unconstrained minimization problem	CPU time (ms)
Multiplier method	-2.9197	$h_1=0.15\times10^{-9}$ $h_2=0.45\times10^{-9}$ $h_3=-0.12\times10^{-9}$	7	30	51
Quasi-Newton multiplier method	-2.9197	$h_1=0.41\times10^{-8}$ $h_2=-0.14\times10^{-8}$ $h_3=-0.20\times10^{-8}$	8	21	46
Proposed primal-dual quasi-Newton method	-2.9197	$h_1=-0.11\times10^{-8}$ $h_2=-0.39\times10^{-8}$ $h_3=0.26\times10^{-7}$	5	15	41

c increases, the contour of the dual objective function approaches a circle and hence the dual objective function can be maximized more easily. This explains why the multiplier method with monotonically-increasing c and the quasi-Newton multiplier method have a similar rate of convergence for the dual maximization problem in our experiments. However, as the penalty parameter c increases, the unconstrained minimization problem becomes ill-conditioned and hence more difficult to solve. In fact, our experiments show that the multiplier method requires more iterations than the quasi-Newton multiplier method for the unconstrained minimization of $\tilde{L}(x,u,v;c)$. Our experiments also show that the proposed primal-dual quasi-Newton method has better convergence properties than the other two methods considered.

Unfortunately, we do not have any QP program as yet, and so we could not compare our method with the RQP method directly.

However, Fukushima [12] describes the results of two experiments in which the RQP method was applied to Example 1:

(i) when the line search was made along the function P(x;c) given in Section 1, the RQP method converges after 8 iterations

(ii) when no line search was made and the step-size parameter α_k was assumed to be constant and equal to 1, the RQP method converges after 12 iterations.

It was also pointed out that (i) consumed more CPU time than (ii), because the line search is made more difficult by the non-smoothness of the objective function. Recall that the RQP method requires both the updating of B^k and the solution of an auxiliary quadratic programming problem (which is equivalent to calculating the inverse of the dual Hessian) at each iteration. Therefore, there were a total of 24 updates of H_1 and H_2 in the case of (ii) above. The number of updates in our proposed method is 21. Thus, the suggested primal-dual quasi-Newton method seems to have the advantage that it does not require the solution of successive auxiliary quadratic programming problems and, moreover, the line search is very easy.

6. CONCLUDING REMARKS

In this paper, we have proposed an effective method for constrained optimization which we call the *primal-dual quasi-Newton method*. The main feature of this method is that it approximates the inverses of both the primal Hessian and the dual Hessian by H_1 and H_2, respectively, and then improves these approximations by some quasi-Newton method. Note that the RQP method implicitly calculates the inverse of the dual Hessian by solving an auxiliary quadratic programming problem, while Tapia's diagonalized multiplier method requires explicit calculation of the inverse of the dual Hessian. The suggested primal-dual quasi-Newton method can also be regarded as an extension of multiplier methods. Although the method requires infinite steps in the unconstrained minimization problem, the number of iterations expected in practice is quite small (except for the first step) because the initial point for each unconstrained minimization is forced by the method to be near the true minimum of the unconstrained optimization problem (see Fig. 1). In the neighborhood of the optimal solution (x^*,u^*,v^*), the unconstrained minimization problem is considered to be solved with sufficient accuracy in one step, and hence the search direction of the primal-dual quasi-Newton method becomes equivalent to that of the RQP and diagonalized multiplier methods. However, taking the ease of line search into account, the primal-dual quasi-Newton method seems to be the most efficient. In addition, this method has the advantage that it is not necessary to solve auxiliary quadratic programming problems nor to calculate the inverse of the dual Hessian explicitly. On the other hand, the primal-dual quasi-Newton method uses an augmented Lagrangian including a penalty parameter to ensure that the primal Hessian is positive definite. The arbitrary value assigned to the penalty parameter is a drawback of the primal-dual quasi-Newton method. However, this problem also arises to some extent in the RQP and diagonalized multiplier methods.

The idea of considering $L_x = 0$ as an additional constraint is very interesting. One obvious possibility is to include the constraint $L_x = 0$ in the augmented Lagrangian as a penalty term. In fact, Pillo and Grippo [13] and Boggs and Tolle [14] did just this, but for a completely different reason. (They wished to make the augmented Lagrangian convex with respect to both x and u.) The suggested primal-dual quasi-Newton method can also be regarded as a method which projects (x,u) onto the constraint $L_x = 0$ by solving the unconstrained problem min L(x,u) while finding the solution to h(x) = 0 (in other words, finding the saddle point of L(x,u)). Other methods for handling the constraint $L_x = 0$ (e.g., a GRG-like method) are of course possible. This will be discussed in a forthcoming paper.

ACKNOWLEDGEMENTS

This research was initiated by a problem suggested by Dr. J. L. Greenstadt during the stay of one of the authors, H. Nakayama, as a post-doctoral fellow at the IBM Palo Alto Scientific Center. This author would like to express his sincere gratitude to

Dr. Greenstadt for helpful advice and suggestions. He is also indebted to Dr. L. F. Escudero for providing much information and encouragement.

REFERENCES

1. S.P. Han. A globally convergent method for nonlinear programming. *Journal of Optimization Theory and Applications,* 22 (1977) 297-309.

2. M.J.D. Powell. Optimization algorithms in 1979. *Algorithms Newsletter of Mathematical Programming Society,* 5 (1981) 2-17.

3. L.F. Escudero. A projected Lagrangian method for nonlinear programming. Report No. G320-3401, IBM Palo Alto Scientific Center, 1980.

4. K. Tanabe. Feasibility-improving gradient-acute-projection method: a unified approach to nonlinear programming. In T. Yamamoto and K. Tanabe (Eds.), *The Newton Method and Related Topics.* Kino-kuniya Book-store Co., 1981.

5. R.A. Tapia. Diagonalized multiplier methods and quasi-Newton methods for constrained optimization. *Journal of Optimization Theory and Applications,* 22 (1977) 135-194.

6. H. Nakayama. Some remarks on recent developments in constrained optimization. Memorandum, IBM Palo Alto Scientific Center, 1979.

7. T. Bard and J.L. Greenstadt. A modified Newton method for optimization with equality constraints. In R. Fletcher (Ed.), *Optimization.* Academic Press, 1969.

8. H. Nakayama, H. Sayama and Y. Sawaragi. A generalized Lagrangean and multiplier method. *Journal of Optimization Theory and Applications,* 17 (1976) 133-145.

9. D.P. Bertsekas. Multiplier methods: a survey. *Automatica*, 12 (1976) 133-145.

10. R.H. Byrd. Local convergence of the diagonalized methods of multipliers. *Journal of Optimization Theory and Applications*, 26 (1978) 485-500.

11. Y. Kameyama, H. Sayama and K. Inaba. Quasi-Newton multiplier method in nonlinear programming. *Transactions of the Society of Instrument and Control Engineers*, 12 (1976) 34-40 (in Japanese).

12. M. Fukushima. Quasi-Newton methods. *Systems and Control*, 24 (1980) 668-679 (in Japanese).

13. G.D. Pillo and L. Grippo. A new class of augmented Lagrangeans in nonlinear programming. *SIAM Journal on Control and Optimization*, 17 (1979) 618-628.

14. P.T. Boggs and J.W. Tolle. Augmented Lagrangians which are quadratic in the multiplier. *Journal of Optimization Theory and Applications*, 31 (1980) 17-26.