Journal of Global Optimization **24:** 187–203, 2002.

Augmented Lagrangian Duality and Nondifferentiable Optimization Methods in Nonconvex Programming

RAFAIL N. GASIMOV
Department of Industrial Engineering, Osmangazi University, Bademlik 26030, Eskişehir, Turkey; E-mail: gasimovr@ogu.edu.tr

Abstract. In this paper we present augmented Lagrangians for nonconvex minimization problems with equality constraints. We construct a dual problem with respect to the presented here Lagrangian, give the saddle point optimality conditions and obtain strong duality results. We use these results and modify the subgradient and cutting plane methods for solving the dual problem constructed. Algorithms proposed in this paper have some advantages. We do not use any convexity and differentiability conditions, and show that the dual problem is always concave regardless of properties the primal problem satisfies. The subgradient of the dual function along which its value increases is calculated without solving any additional problem. In contrast with the penalty or multiplier methods, for improving the value of the dual function, one need not to take the 'penalty like parameter' to infinity in the new methods. In both methods the value of the dual function strongly increases at each iteration. In the contrast, by using the primal-dual gap, the proposed algorithms possess a natural stopping criteria. The convergence theorem for the subgradient method is also presented.

Key words: Nonconvex programming; Augmented Lagrangian; Duality with zero gap; Subgradient method; Cutting plane method

Mathematical Subject Classifications: 90C26; 90C48; 49M30

1. Introduction

Let X be any topological linear space, let $S \subset X$ be a nonempty subset of X, let Y be a real normed space and let Y^* be its dual.

We will be concerned with the nonlinear programming problem:

$$(P) \quad \text{minimize } f_0(x) \text{ over all } x \text{ in } S \text{ satisfying } f(x) = 0$$

where f_0 is a real-valued function defined S and f is a mapping of S into Y.

We will denote by $\|\cdot\|$ the norm of Y, by $\|\cdot\|_*$ the norm of Y^*, and the value of the linear continuous functional $y^* \in Y^*$ at the point $y \in Y$ by $\langle y, y^* \rangle$.

For every $x \in X$ and $y \in Y$ let

$$\Phi(x, y) = \begin{cases} f_0(x) & \text{if } x \in S \text{ and } f(x) = y \\ +\infty & \text{otherwise} \end{cases} \tag{1}$$

and define the perturbation function associated with (P) as

$$h(y) = \inf_{x \in S} \Phi(x, y)\,. \tag{2}$$

Let us denote the convex case of the problem (P) by (P_0). It is well-known that the ordinary Lagrangian function is defined as

$$L_0(x, y^*) = -\sup_{y \in Y} \{\langle y, y^* \rangle - \Phi(x, y)\} . \quad (3)$$

This definition leads to the following expression which allows to construct a dual problem in the convex case:

$$L_0(x, y^*) = f_0(x) + \langle f(x), y^* \rangle \quad \text{for} \quad (x, y^*) \in X \times Y^* .$$

This definition corresponds to the dual problem

$$(P_0^*) \quad \text{maximize } g_0(y^*) \text{ over all } y^* \in Y^* ,$$

where

$$g_0(y^*) = \inf_{x \in S} L_0(x, y^*) .$$

The optimal values in problems (P_0) and (P_0^*) can be expressed as

$$\inf P_0 = \inf_{x \in S} \sup_{y^* \in Y^*} L_0(x, y^*) \text{ and } \sup P_0^* = \sup_{y^* \in Y^*} \inf_{x \in S} L_0(x, y^*) ,$$

respectively. It is well-known that the optimal values in these two problems satisfy

$$\inf P_0 \geqslant \sup P_0^* , \quad (4)$$

and a necessary and sufficient condition for the equality to hold is the existence of a saddle point of the Lagrangian L_0; cf. Rockafellar (1970), Ekeland and Temam (1976) and Bazaraa et al. (1993). It was also proved that a pair of vectors $x \in S$ and $u \in Y^*$ furnishes a saddle point of the Lagrangian L_0 on $S \times Y^*$ if and only if

$$\left.\begin{array}{l} x \text{ is a (globally) optimal solution to } (P), \\ h(y) \geqslant h(0) + \langle y, u \rangle \text{ for all } y, \end{array}\right\} \quad (5)$$

where h is the perturbation function defined by (2). Unfortunately, without convexity and assumptions dependent on it, one cannot very well ensure the existence of some u for which the inequality in (5) holds. A number of authors have addressed the question of whether the discrepancy in (4), which is called a ‘duality gap’ in nonconvex programming, could be eliminated by changing the Lagrangian function. Such a Lagrangian, called the augmented Lagrange function, may be obtained by adding linear or quadratic penalty expressions to L_0. In the finite dimensional case, linear penalty expressions was introduced by Pietrzykowski (1969) and Zangwill (1967). Quadratic penalty expressions was first proposed in the inequality case by Courant (1943). The addition of squared constrained terms to the ordinary Lagrangian was first proposed by Hestenes (1969) and Powell (1969) in association with a special numerical approach—called the method of multipliers—to problems having equality constriants only; cf. Fletcher (1979), Haarhoff and Buys (1970), Kort and Bertsekas (1972) and Poljak (1970). Penalty expressions with a possible mixture of linear and quadratic pieces have been suggested by Rockafellar

(1993). All these approaches was generalized by Rockafellar and Wets (1998). They introduced a general augmenting function σ: $R^m \to \bar{R}$, which is proper, lower semicontinuous and convex, with min $\sigma = 0$, arg min $\sigma = \{0\}$ and defined an augmented Lagrangian with penalty parameter $r > 0$ in the form $\bar{l}(x, y, r) = \inf_u \{\Phi(x, u) + r\sigma(u) - \langle y, u\rangle\}$, where $f_0 = \Phi(\cdot, 0)$. In particular, the augmented Lagrangian generated with the augmenting function $\sigma(u) = \|u\|$ was termed a 'sharp Lagrangian'.

In recent years the duality of nonconvex optimization problems, in more generalized form, was studied in the framework of the so-called abstract convexity (see Pallaschke and Rolewicz, 1997; Rubinov, 2000; Singer, 1997). Abstract convexity has found many applications in the study of problems of mathematical analysis and optimization. The books by D. Palaschke and Rolewicz (1997) and by I. Singer (1997) contain detailed presentations of many results of abstract convex analysis, which are concentrated around notions of subdifferentials, conjugations and dualities. Special kinds of nonlinear analogues of Lagrange and penalty functions are studied in the excellent book by A.M. Rubinov (2000).

Computational considerations in nonconvex problems have led to algorithms based on Lagrangian duality. Naturally, the differentiability properties of dual functions are a very important determinant of the type of dual method that is appropriate for a given problem.

The main purpose of this paper is to present solution algorithms for nonconvex mathematical programming problems with respect to an augmented Lagrangian of the special type. In this paper we modify the subgradient and cutting plane methods for maximizing the nondifferentiable dual function. We consider for simplicity, equality constraint case and obtain the duality relations with respect to the constructed here Lagrangian. In order to calculate augmented Lagrange function we use the formula (3)—a natural definition of the Lagrange functions, by taking $-c\|y\| + \langle y, y^*\rangle$ instead of $\langle y, y^*\rangle$. We present conditions under which a duality gap can be eliminated by using such a Lagrangian for nonconvex mathematical programming problems. As was mentioned above, such a Lagrangian may be considered as the sharp Lagrangian presented by Rockafellar and Wets (1998), on the one hand and as a special case of the extended Lagrangians defined in the book by Rubinov (2000), on the other hand. Therefore some theorems concerning the duality results are presented without proofs.

Note that subgradient methods were first introduced in the middle 60s; the works of Demyanov (1968), Poljak (1969a, 1969b, 1970) and Shor (1985, 1995) were particularly influental. The convergence rate of subgradient methods is discussed in Goffin (1977). Cutting plane methods were proposed by Cheney and Goldstein (1959) and Kelley (1960). For further study of this subject see Bertsekas (1995) and Bazaraa et al. (1993). These methods were used for solving dual problems obtained by using ordinary Lagrangians or problems satisfying convexity conditions. However, our main purpose is to find an optimal solution to the primal problem and unfortunately, without convexity assumptions we cannot very well ensure the equality of optimal values of primal and dual problems constructed by this manner. We show that a dual function constructed in this paper is always concave. Without solving any additional problem, we calculate the subgradient of the dual function along which its value increases. We use this result in the subgradient and cutting plane methods for improving the value of the dual function and for choosing the better values of the primal and dual variables.

2. Duality

Consider the primal mathematical programming problem defined as:

$$(P) \quad \inf P = \inf f_0(x) \text{ subject to } x \in S,\ f(x) = 0 .$$

Every element $x \in D$, where D is a feasible set defined as $D = \{x \in S \mid f(x) = 0\}$, such that $f_0(x) = \inf P$ will be termed a solution of (P). By a general definition of Lagrangians the augmented Lagrangian L associated with (P) will be defined as

$$L(x,\ u,\ c) = \inf_{y \in Y} \{\Phi(x,\ y) + c\|y\| - \langle y,\ u\rangle\} ,$$

for $x \in X,\ y \in Y,\ c \in [0,\ +\infty)$, where the function $\Phi(x, y)$ is defined in (1). By using the definition of Φ, we can concretize the augmented Lagrangian associated with (P):

$$L(x, u, c) = \inf_{y = g(x)} \{\Phi(x, y) + c\|y\| - \langle y, u\rangle\} = f_0(x) + c\|f(x)\| - \langle f(x),\ u\rangle ,$$

for $x \in S$, $u \in Y^*$, and $c \in [0,\ +\infty)$.

The dual function H is defined as:

$$H(u, c) = \inf_{x \in S} L(x, u, c), \text{ for } u \in Y^*, c \in [0,\ +\infty) \equiv R_+ . \tag{6}$$

Then, a dual problem of (P) is given by

$$(P^*) \quad \sup P^* = \sup_{(u,c) \in Y^* \times R_+} H(u,\ c) .$$

Any element $(u,\ c) \in Y^* \times R_+$ with $H(u,\ c) = \sup P^*$ is termed a solution of (P^*).

LEMMA 1. *For every $u \in Y^*$, $y \in Y$, $y \neq 0$ and for every $r \in R_+$ there exists $c \in R_+$ such that $c\|y\| - \langle y,\ u\rangle > r$.*

Proof. Let $u \in Y^*$, $y \in Y$, $y \neq 0$ and $r \in R_+$. We choose $c \in R_+$ with $c > \|u\| + r/\|y\|$. Then $c\|y\| - \|u\|_* \cdot \|y\| > r$. Since $\|u\|_* \cdot \|y\| \geqslant \langle y,\ u\rangle$ we have $c\|y\| - \langle y,\ u\rangle > r$. $\square$

It follows from this lemma that

$$\sup_{(u,c) \in Y^* \times R_+} L(x,\ u,\ c) = \begin{cases} f_0(x), & f(x) = 0 \\ +\infty, & f(x) \neq 0 . \end{cases}$$

Hence,

$$\inf_{x\in S}\ \sup_{(u,c)\in Y^*\times R_+} L(x, u, c) = \inf\{f_0(x)|x\in S,\ f(x)=0\} = \inf P\ . \tag{7}$$

This means that the value of a mathematical programming problem with equality constraints can be represented as (7), regardless of properties the original problem satisifes.

Proofs of the following four theorems are analogous to the proofs of similar theorems presented for augmented Lagrangian functions with quadratic penalty expressions or with more general augmenting functions. See, for example, Rockafellar (1993) and Rockafellar and Wets (1998).

THEOREM 1. *inf* $P \geqslant$ *sup* P^*.

THEOREM 2. *Suppose that inf P is finite. Then a pair of elements* $\bar{x}\in S$ *and* $(\bar{u}, \bar{c})\in Y^*\times R_+$ *furnishes a saddle point of the augmented Lagrangian L on* $S\times (Y^*\times R_+)$ *if and only if* $\bar{x}$ *is a solution to* (P), $(\bar{u}, \bar{c})$ *is a solution to* (P^*) *and inf* $P =$ *sup* P^*.

THEOREM 3. *A pair of vectors* $x\in S$ *and* $(u, c)\in Y^*\times R_+$ *furnishes a saddle pont of the augmented Lagrangian L on* $S\times(Y^*\times R_+)$ *if and only if*

$$\left.\begin{array}{l} x \text{ is a solution to } (P)\,,\\ h(y)\geqslant h(0)+\langle y,\ u\rangle - c\|y\| \text{ for all } y\,, \end{array}\right\}$$

where h is a perturbation function defined by (2). *When this holds, any* $a > c$ *will have the propety that*

$$[x \text{ solves } (P)] \Leftrightarrow [x \text{ minimizes } L(z,\ u,\ a) \text{ over } z\in S]\ .$$

THEOREM 4. *Suppose in* (P) *that* f_0 *and f are continuous, S is compact, and a feasible solution exists. Then inf* $P =$ *sup* P^* *and there exists a solution to* (P). *Furthermore, in this case, the dual function H in* (P^*) *is concave and finite everywhere on* $Y^*\times R_+$, *so this maximization problem is effectively unconstrained.*

The following theorem will also be used as a stopping criteria in solution algorithms for dual problem in the next section.

THEOREM 5. *Let inf* $P =$ *sup* P^* *and suppose that for some* $(\bar{u}, \bar{c})\in Y^*\times R_+$, *and* $\bar{x}\in S$,

$$\min_{x\in S} L(x,\ \bar{u},\ \bar{c}) = f_0(\bar{x}) + \bar{c}\|f(\bar{x})\| - \langle f(\bar{x}),\ \bar{u}\rangle\ . \tag{8}$$

Then $\bar{x}$ *is a solution to* (P) *and* $(\bar{u}, \bar{c})$ *is a solution to* (P^*) *if and only if*

$$f(\bar{x}) = 0\ . \tag{9}$$

Proof. Necessity. If (8) is satisfied and $\bar{x}$ is a solution to (P) then $\bar{x}$ is feasible and therefore $f(\bar{x}) = 0$.

Sufficiency. Suppose to the contrary that (8) and (9) are satisfied but $\bar{x}$ and $(\bar{u}, \bar{c})$ are not solutions. Then, there exists $\tilde{x} \in D$ such that $f_0(\tilde{x}) < f_0(\bar{x})$. Hence

$$\begin{aligned} f_0(\tilde{x}) &< f_0(\bar{x}) = f_0(\bar{x}) + \bar{c}\|f(\bar{x})\| - \langle f(\bar{x}),\ \bar{u}\rangle = H(\bar{u},\ \bar{c}) \\ &= \min_{x\in S} L(x,\ \bar{u},\ \bar{c}) \leqslant \sup_{(u,c)\in Y^*\times R_+} \min_{x\in S} L(x,\ u,\ c) = \sup\ P^* \\ &= \inf\ P \leqslant f_0(\tilde{x})\ , \end{aligned}$$

which proves the theorem. □

3. Solving the dual problem

We have described several properties of the dual function in the previous section. In this section, we utilize these propeties to modify the subgradient and cutting plane methods for maximizing the dual function H. Theorems 2, 3 and 4 give necessary and sufficient conditions for an equality between inf P and sup P^*. Therefore, when the hypotheses of these theorems are satisfied, the maximization of the dual function H by using the subgradient method or the cutting plane method will give us the optimal value of the primal problem.

We assume throughout this section that $X = E^n$ and $Y = E^m$ are finite dimensional spaces, and the hypotheses of Theorem 4 are satisfied. We consider the dual problem

$$\begin{aligned} &\text{maximize } H(u, c) = \min_{x\in S} L(x, u, c) = \min_{x\in S} \{f_0(x) + c\|f(x)\| - u' f(x)\} \\ &\text{subject to } (u, c) \in F = E^m \times R_+\ , \end{aligned}$$

where u' is the transpose of the vector u, and $u' f(x)$ denotes the scalar product of vectors u and $f(x)$.

It will be convenient to introduce the following set:

$$S(u, c) = Arg \min_{x\in S}\ [f_0(x) + c\|f(x)\| - u' f(x)]\ .$$

The assertion of the following theorem can be obtained from the known theorems on the subdifferentials of the continuous maximum and minimum functions. See, for example, Polak (1997).

THEOREM 6. *Let S be a nonempty compact set in X and let f_0 and f be continuous, so that for any $(\bar{u}, \bar{c}) \in E^m \times R_+$, $S(\bar{u}, \bar{c})$ is not empty. If $\bar{x} \in S(\bar{u}, \bar{c})$, then $(-f(\bar{x}), \|f(\bar{x})\|)$ is a subgradient of H at $(\bar{u}, \bar{c})$.*

3.1. SUBGRADIENT METHOD

Initialization Step. Choose a vector (u_1, c_1) with $c_1 \geqslant 0$, let $k = 1$, and go to the main step.

Main Step 1. Given (u_k, c_k), solve the following subproblem:

$$\begin{aligned}&\text{Minimize } f_0(x) + c_k\|f(x)\| - u_k' f(x)\\&\text{subject to } x \in S\,.\end{aligned}$$

Let x_k be any solution. If $f(x_k) = 0$, the stop; by Theorem 5, (u_k, c_k) is a solution to (P^*), x_k is a solution to (P). Otherwise, go to step 2.

2. Let

$$u_{k+1} = u_k - s_k f(x_k),\ c_{k+1} = c_k + (s_k + \varepsilon_k)\|f(x_k)\|\,, \tag{10}$$

where s_k and ε_k are positive scalar stepsizes, replace k by $k + 1$, and repeat step 1.

The following theorem shows that in contrast with the subgradient methods developed for dual problems formulated by using ordinary Lagrangians, the new iterate strictly improves the cost for all values of the stepsizes s_k and ε_k.

THEOREM 7. *Suppose that the pair $(u_k, c_k) \in E^m \times R_+$ is not a solution to the dual problem and $x_k \in S\ (u_k, c_k)$. Then for a new iterate (u_{k+1}, c_{k+1}) calculated from* (10) *for all positive scalar stepsizes s_k and ε_k we have*:

$$0 < H(u_{k+1}, c_{k+1}) - H(u_k, c_k) \geqslant (2s_k + \varepsilon_k)\|f(x_k)\|^2\,.$$

Proof. Let $(u_k, c_k) \in E^m \times R_+$, $x_k \in S(u_k, c_k)$ and (u_{k+1}, c_{k+1}) is a new iterate calculated from (10) for arbitrary positive scalar stepsizes s_k and ε_k. Then by Theorem 6, the vector $(-f(x_k), \|f(x_k)\|) \in E^m \times R_+$ is a subgradient of a concave function H at (u_k, c_k) and by definition of subgradients we have:

$$\begin{aligned}&H(u_{k+1}, c_{k+1}) - H(u_k, c_k)\\&\leqslant (u_{k+1} - u_k)'(-f(x_k)) + (c_{k+1} - c_k)\|f(x_k)\|\\&= s_k\|f(x_k)\|^2 + (s_k + \varepsilon_k)\|f(x_k)\|^2 = (2s_k + \varepsilon_k)\|f(x_k)\|^2\end{aligned}$$

On the other hand

$$\begin{aligned}&H(u_{k+1}, c_{k+1})\\&= \min_{x\in S}\ \{f_0(x) + c_{k+1}\|f(x)\| - u_{k+1}' f(x)\}\\&= \min_{x\in S}\ \{f_0(x) + c_k\|f(x)\| - u_k' f(x) + (s_k + \varepsilon_k)\|f(x_k)\|\,\|f(x)\| + s_k f(x_k)' f(x)\}\\&\geqslant \min_{x\in S}\ \{f_0(x) + c_k\|f(x)\| - u_k' f(x) + (s_k + \varepsilon_k)\|f(x_k)\|\,\|f(x)\| - s_k\|f(x_k)\|\,\|f(x)\|\}\\&= \min_{x\in S}\ \{f_0(x) + c_k\|f(x)\| - u_k' f(x) + \varepsilon_k\|f(x_k)\|\,\|f(x)\|\}\\&= \min_{x\in S}\ \{f_0(x) + (c_k + \varepsilon_k\|f(x_k)\|)\|f(x)\| - u_k' f(x)\}\,.\end{aligned}$$

Now suppose that the last minimum attains for a some $\tilde{x} \in S$. If $f(\tilde{x})$ were zero, then by Theorem 5, the pair $(u_k, c_k + \varepsilon_k \| f(x_k) \|)$ would be a solution to the dual problem and theorefore

$$\min_{x \in S} \{ f_0(x) + (c_k + \varepsilon_k \| f(x_k) \|) \| f(x) \| - u_k' f(x) \}$$
$$> \min_{x \in S} \{ f_0(x) + c_k \| f(x) \| - u_k' f(x) \} = H(u_k, c_k) ,$$

because of (u_k, c_k) is not a solution. When $f(\tilde{x}) \neq 0$ then

$$\begin{aligned} &\min_{x \in S} \{ f_0(x) + (c_k + \varepsilon_k \| f(x_k) \|) \| f(x) \| - u_k' f(x) \} \\ &= f_0(\tilde{x}) + (c_k + \varepsilon_k \| f(x_k) \|) \| f(\tilde{x}) \| - u_k' f(\tilde{x}) \\ &> f_0(\tilde{x}) + c_k \| f(\tilde{x}) \| - u_k' f(\tilde{x}) \geqslant \min_{x \in S} \{ f_0(x) + c_k \| f(x) \| - u_k' f(x) \} = H(u_k, c_k) . \end{aligned}$$

Thus we have established that $H(u_{k+1}, c_{k+1}) > H(u_k, c_k)$. □

The following theorem demonstrates that for the certain values of stepsizes s_k and ε_k, the distance between the points generated by the algorithm and the solution to the dual problem decreases at each iteration (cf. Bertsekas (1995), Proposition 6.3.1).

THEOREM 8. *Let (u_k, c_k) be any iteation, which is not a solution to the dual problem, so $f(x_k) \neq 0$. Then for any dual solution $(\bar{u}, \bar{c})$, we have*

$$\| (\bar{u}, \bar{c}) - (u_{k+1}, c_{k+1}) \| < \| (\bar{u}, \bar{c}) - (u_k, c_k) \|$$

for all stepsizes s_k such that

$$0 < s_k < \frac{2(H(\bar{u}, \bar{c}) - H(u_k, c_k))}{5 \| f(x_k) \|^2} , \tag{11}$$

and $0 < \varepsilon_k < s_k$.

Proof. We have

$$\begin{aligned} &\| (\bar{u}, \bar{c}) - (u_{k+1}, c_{k+1}) \|^2 = \| \bar{u} - u_{k+1} \|^2 + | \bar{c} - c_{k+1} |^2 \\ &= \| \bar{u} - (u_k - s_k f(x_k)) \|^2 + | \bar{c} - (c_k + (s_k + \varepsilon_k) \| f(x_k) \|) |^2 \\ &= \| \bar{u} - u_k \|^2 + 2 s_k (\bar{u} - u_k)' f(x_k) + (s_k)^2 \| f(x_k) \|^2 \\ &\quad + (\bar{c} - c_k)^2 - 2(s_k + \varepsilon_k)(\bar{c} - c_k) \| f(x_k) \| + (s_k + \varepsilon_k)^2 \| f(x_k) \|^2 \\ &< \| \bar{u} - u_k \|^2 + 2 s_k (\bar{u} - u_k)' f(x_k) + (s_k)^2 \| f(x_k) \|^2 \\ &\quad + (\bar{c} - c_k)^2 - 2 s_k (\bar{c} - c_k) \| f(x_k) \| + (2 s_k)^2 \| f(x_k) \|^2 , \end{aligned}$$

where the last inequality is a result of inequalities $\bar{c} - c_k > 0$, $\| f(x_k) \| > 0$, and $0 < \varepsilon_k < s_k$. Now, by using the subgradient inequality

$$H(\bar{u}, \bar{c}) - H(u_k, c_k) \leqslant (\bar{u} - u_k)'(-f(x_k)) + (\bar{c} - c_k)\|f(x_k)\|,$$

we obtain

$$\|\bar{u} - u_{k+1}\|^2 + |\bar{c} - c_{k+1}|^2$$
$$< \|\bar{u} - u_k\|^2 + |\bar{c} - c_k|^2 - 2s_k(H(\bar{u}, \bar{c}) - H(u_k, c_k)) + 5(s_k)^2\|f(x_k)\|^2 . \tag{12}$$

It is straightforward to verify that for the range of stepsize of Eq. (11) the sum of the last two terms in the above relation is negative. Thus,

$$\|\bar{u} - u_{k+1}\|^2 + |\bar{c} - c_{k+1}|^2 < \|\bar{u} - u_k\|^2 + |\bar{c} - c_k|^2 . \qquad \square$$

The inequality (12) can also be used to establish the convergence result for the subgradient method.

THEOREM 9. *Assume that all conditions of Theorem* 4 *are satisfied. Let* (u_k, c_k) *be any iteration of the subgradient method. Suppose that each new iteration* (u_{k+1}, c_{k+1}) *is calculated from* (10) *for the stepwise*

$$s_k = \frac{\bar{H} - H(u_k, c_k)}{5\|f(x_k)\|^2} \text{ and } 0 < \varepsilon_k < s_k ,$$

where $\bar{H} = H(\bar{u}, \bar{c})$ *denotes the optimal dual value. Then* $H(u_k, c_k) \to \bar{H}$.

Proof. By taking $s_k = (\bar{H} - H(u_k, c_k))/5\|f(x_k)\|^2$ in (12) we obtain:

$$\|\bar{u} - u_{k+1}\|^2 + |\bar{c} - c_{k+1}|^2 < \|\bar{u} - u_k\|^2 + |\bar{c} - c_k|^2 - \frac{(\bar{H} - H(u_k, c_k))^2}{5\|f(x_k)\|^2}$$

which can be written in the form

$$(\bar{H} - H(u_k, c_k))^2 < 5\|f(x_k)\|^2[(\|\bar{u} - u_k\|^2 + |\bar{c} - c_k|^2)$$
$$- (\|\bar{u} - u_{k+1}\|^2 + |\bar{c} - c_{k+1}|^2)] . \tag{13}$$

It is obvious that, the sequence $\{\|\bar{u} - u_k\|^2 + |\bar{c} - c_k|^2\}$ is bounded from below (for example, by zero), and by Theorem 8, it is decreasing. Thus, $\{\|\bar{u} - u_k\|^2 + |\bar{c} - c_k|^2\}$ is a convergent sequence. Hence

$$\lim_{k\to\infty} [(\|\bar{u} - u_k\|^2 + |\bar{c} - c_k|^2) - (\|\bar{u} - u_{k+1}\|^2 + |\bar{c} - c_{k+1}|^2)] = 0 .$$

On the other hand, since S is a compact set and f is continuous, $\{5\|f(x_k)\|^2\}$ is a bounded sequence. Thus, (13) implies

$$H(u_k, c_k) \to \bar{H} . \qquad \square$$

Unfortunately, however, unless we know the dual optimal value $H(\bar{u}, \bar{c})$, which is rare, the range of stepsize is unknown. In practice, one can use the stepsize formula

$$s_k = \frac{\alpha_k(H_k - H(u_k, c_k))}{5\|f(x_k)\|^2}, \tag{14}$$

where H_k is an approximation to the optimal dual value and $0 < \alpha_k < 2$. By Theorem 7, the sequence $\{H(u_k, c_k)\}$ is increasing, therefore to estimate the optimal dual value from below, we can use the current dual value $H(u_k, c_k)$. As an upper bound, we can use any primal value $f_0(\bar{x})$ corresponding to a primal feasible solution $\bar{x}$.

Now we demonstrate the proposed algorithm on some problems. Note that, inequality constraint problems can be reduced to equality constraint one by adding slack variables to left hand sides of inequalities. The program package 'LINGO 6.0' has been applied for solving dual unconstrained problems. The stopping criteria is taken as $\|f(x^k)\| \leqslant 10^{-4}$. For description of results of the numerical examples we use the following notations:

- k is the number of iterations;
- (u^k, c^k) – is a vector of Lagrange multipliers at *kth* iteration;
- x^k is a minimizer of Lagrange function $L(x, u_k, c_k)$ over $x \in S$;
- H is the upper bound for the values of dual function;
- s^k is the stepsize parameter calculated at the *kth* iteration by the formula $s_k = (H - H(u_k, c_k))/5\|f(x_k)\|^2$;
- $\varepsilon^k = 0.95 s^k$;
- x^* is an optimal primal solution;
- $f_0^* = f_0(x^*)$.

EXAMPLE 1. (see Himmelblau, 1972) (Table 1)

$$f_0(x) + 1000 - x_1^2 - 2x_2^2 - x_3^2 - x_1x_2 - x_1x_3 \rightarrow \min$$

subject to

$$f_1(x) + x_1^2 + x_2^2 + x_3^2 - 25 = 0$$

$$f_2(x) = 8x_1 + 14x_2 + 7x_3 - 56 = 0$$

and

$$x_i \geqslant 0, \; i + 1, 2, 3$$

$$x^* = (3.512, 0.217, 3.552), \; f_0^* = 961.715$$

EXAMPLE 2. (see Khenkin, 1976) (Table 2)

Table 1. Summary of the computations for Example 1

k	u_1^k	u_2^k	c^k	x_1^k	x_2^k	x_3^k	H	$H(u^k, c^k)$	$\|f(x^k)\|$	s^k
1	0	0	0	5	5	5	2000	850	102.083	0.2207
2	−1.1035	−1.9643	4.3935	3.5113	0.2171	3.553	2000	961.7152	2.23$E-06$	

Table 2. Summary of the computations for Example 2

k	u^k	c^k	x_1^k	x_2^k	x_3^k	H	$H(u^k, c^k)$	$\|f(x^k)\|$	s^k
1	0	0	0	0	1.23457	2	0	2.734568	0.053491
2	−0.1463	0.2852	4.06E − 02	4.06E − 02	0	2	8.67E − 02	1.185017	0.272498
3	−0.4692	0.9149	2.29E − 01	2.29E − 01	0	2	0.3004191	5.89E − 08	

$$f_0(x) = 0.5(x_1 + x_2)^2 + 50(x_2 - x_1)^2 + \sin^2(\mathrm{x}_1 + \mathrm{x}_2) \to \min$$

subject to

$$f(x) = (x_1 - 1)^2 + (x_2 - 1)^2 + (x_3 - 1)^2 - 1.5 \leqslant 0\,.$$

$$x^* = (0.229014,\ 0.229014),\ f_0^* = 0.3004190265\,.$$

By adding a slack variable x_4 to the left-hand side of the single constraint we obtain an equality constrained problem.

EXAMPLE 3. (see Khenkin, 1976) (Table 3)

$$f_0(x) = 0.5(x_1 + x_2)^2 + 50(x_2 - x_1)^2 + x_3^2 + |x_3 - \sin(x_1 + x_2)| \to \min$$

subject to

$$f(x) = (x_1 - 1)^2 + (x_2 - 1)^2 + (x_3 - 1)^2 - 1.5 \leqslant 0\,.$$

$$x^* = (0.229014,\ 0.229014,\ 0.4421181),\ f_0^* = 0.3004190265\,.$$

By adding a slack variable x_4 to the left − hand side of the single constraint we obtain an equality constrained problem.

EXAMPLE 4. (see Floudas, 1999) (Table 4)

$$f(x,\ y) = [a,\ x] - 0.5[x,\ Qx] + by \to \min$$

subject to

Table 3. Summary of the computations for Example 3

k	u^k	c^k	x_1^k	x_2^k	x_3^k	x_4^k	H	$H(u^k, c^k)$	$\|f(x^k)\|$	s^k
1	0	0	0	0	0	0	2	0	1.5	0.177778
2	−0.2667	0.52	0.22904	0.22899	0.44218	0	2	0.3004193	1.18E − 07	

Table 4. Summary of the computations for Example 4

k	u_1^k	u_2^k	c_1^k	x_1^k	x_2^k	x_3^k	x_4^k	x_5^k	x_6^k	x_7^k	y	H	$H(u^k, c^k)$	$\|f(x^k)\|$	s^k
1	0	0	0	1	1	1	1	1	0	0	20	0	− 475.5	21.7313	0.201376
2	−1.7117	−4.0275	8.5335	0	1	0	1	1	0.5	0	20	0	− 361.5	0	

$$f_1(x) = 6x_1 + 3x_2 + 3x_3 + 2x_4 + x_5 - 6.5 \leqslant 0$$

$$f_2(x) = 10x_1 + 10x_3 + y - 20 \leqslant 0$$

and

$$0 \leqslant x_i \leqslant 1,\ i = 1,\ 2,\ 3,\ 4,\ 5,\ y \leqslant 0\ ,$$

where $a = (-10.5,\ -7.5,\ -3.5,\ -2.5,\ -1.5)$, $b = -10$, $Q = 100I$, and I is the identity matrix.

$$x^* = (0,\ 1,\ 0,\ 1,\ 1),\ y^* = 20,\ f^* = -361.5\ .$$

By adding the slack variables x_6 and x_7 to the left-hand sides of constraints we obtain an equality constrained problem.

EXAMPLE 5. Consider the following problem which is formed arbitrarily:

$$f_0(x) = x_1^3 + x_2^3 - x_3^2 \rightarrow \min$$

subject to

$$f_1(x) = x_1^2 + x_2^2 + x_3^2 - 9 = 0,$$

$$f_2(x) + x_1 - 2x_2{}^2 + x_3 - 2 = 0,$$

$$-3 \leqslant x_i \leqslant 3, \quad i = 1,\ 2,\ 3\ .$$

At iteration k for dual variables $(u^k,\ c^k)$, problem to be solved to give x^k and $H(u^k,\ c^k)$ is

$$\min\ [x_1{}^3 + x_2^3 - x_3^2 + c^k((x_1^2 + x_2^2 + x_3^2 - 9)^2 + (x_{\,1} - 2x_2^2 + x_3 - 2)^2)^{1/2}$$

$$- u_1^k(x_1^2 + x_2^2 + x_3^2 - 9) - u_2^k(x_1 - 2x_2^2 + x_3 - 2)]\ .$$

Because of the continuity of the functions f_0, f_1 and f_2 and the compactness of the set

$$S = \{x = (x_1,\ x_2,\ x_3) | -3 \leqslant x_i \leqslant 3, \quad i = 1,\ 2,\ 3\}$$

the optimal values of primal and dual problems must be equal. Table 5 below summarizes the computations obtained by using the modified subgradient method.

As seen from the Table 5, the values of primal and dual problems calculated at the second iteration are approximately equal. The approximative values of primal and dual solutions are $\bar{x} = (0.35,\ -0.78,\ 2.87)$ and $\bar{u} = (-8.4,\ 9.4)$, $\bar{c} = 18.9$ respectively.

Table 5. Summary of the computations for Example 5

k	$(u_1^k,\ u_1^k,\ c^k)$	$(x_1^k,\ x_2^k,\ x_3^k)$	$f_0(x^k)$	$H(u^k,\ c^k)$	$\Vert f(x^k)\Vert$	s^k
1	(0, 0, 0)	$(-3,\ -3,\ 3)$	-63	-63	26.9073	0.468
2	$(-8.4,\ 9.4,\ 18.9)$	$(0.35,\ -0.78,\ 2.87)$	-8.699210	-8.699213	1.335×10^{-5}	

3.2. COMBINATION WITH SUBGRADIENT AND CUTTING PLANE METHOD

We now discuss a strategy for solving the dual problem, in which at each iteration, a function that approximates the dual function is optimized. First, we present the main idea of the traditional cutting plane method (see, for example, Bazaraa et al., 1993; Bertsekas, 1995; and references therein).

Recall that the dual function H is defined by

$$H(u,\ c) = \inf\{f_0(x) + c\|f(x)\| - u'f(x) | x \in S\}\ .$$

Letting $z = H(u, c)$, the inequality $z \leqslant f_0(x) + c\|f(x)\| - u'f(x)$ must hold true for each $x \in S$. Hence the dual problem of maximizing $H(u,\ c)$ over $(u,\ c) \in E^m \times R_+$ is equivalent to the following problem:

$$\begin{aligned} &\text{maximize} \quad z \\ \text{s.t.} \quad & z \leqslant f_0(x) + c\|f(x)\| - u'f(x) \quad \text{for } x \in S, (u, c) \in E^m \times R_+\ . \end{aligned} \tag{15}$$

Note that the above problem is a linear program in the variables z, u, and c. Unfortunately, however, the constraints are infinite and are not known explicitly. Suppose that we have the points $x_1,...,x_{k-1}$ in S and consider the following problem:

$$\begin{aligned} &\text{maximize} \quad z \\ \text{s.t.} \quad & z \leqslant f_0(x_j) + c\|f(x_j)\| - u'f(x_j) \quad \text{for } j = 1,...,k-1\ , \\ & (u,\ c) \in E^m \times R_+ . \end{aligned} \tag{16}$$

The above problem is a linear program with a finite number of constraints and can be solved. Let (z_k, u_k, c_k) be an optimal solution. If this solution satisfies (15), then it is a solution to the dual problem. To check whether (15) is satisfied, consider the following subproblem:

$$\text{minimize} \quad f_0(x) + c_k\|f(x)\| - u_k'f(x), \quad \text{subject to } x \in S\ .$$

Let x_k be a solution to the above problem, so that

$$H(u_k,\ c_k) = f_0(x_k) + c_k\|f(x_k)\| - u_k'f(x_k)\ .$$

If

$$z_k \leqslant H(u_k, c_k)\ , \tag{17}$$

then $(u_k,\ c_k)$ is a solution to the dual problem. Otherwise, for $(u,\ c) = (u_k,\ c_k)$ the inequality (15) is not satisfied for $x = x_k$. Thus, we add the constraint $z \leqslant f_0(x_k) + c\|f(x_k)\| - u'f(x_k)$ to the constraints in (16), and resolve the linear program. Obviously the current optimal point $(z_k,\ u_k,\ c_k)$ contradicts this added constraint. Thus, this point is cut away and hence the name, cutting plane algorithm.

In this algorithm there is no guarantee that at each iteration the value of the dual function will increase. Therefore, modifying this method, we take at each iteration the best current value of the dual function and increase it by applying the iterates of the modified subgradient method proposed above. The second novelty that we present for this method is a stopping criteria. Since the condition (17) is only sufficient but not necessary for optimality, it may be happened that the procedure becomes to the optimal solution (u_k, c_k) but (17) is not still satisfied and therefore algorithm continues to make new iterates. To remove this situation we use the necessary and sufficient condition (9) for optimality presented in the Theorem 5.

4. Summary of the method

Initialization Step. Solve the problem: *minimize* $f_0(x)$, *subject to* $x \in S$.

Let $\tilde{x}_0$ be the solution. If $f(\tilde{x}_0) = 0$ then stop; $\tilde{x}_0$ is a solution to the primal problem. Otherwise, find a point $x_0 \in S$ such that $f(x_0) = 0$, let $u_0 = 0$, $c_0 = 0$, $k = 1$, $H(u_0, c_0) = f_0(\tilde{x}_0) + c_0\|f(\tilde{x}_0)\| - u_0' f(\tilde{x}_0)$, and go to the main step.

Main Step 1. Solve the following problem, which is usually referred to as the *master problem*.

$$\begin{aligned} &\text{maximize} \quad z \\ \text{subject to} \quad & z \leq f_0(x_j) + c\|f(x_j)\| - u' f(x_j) \quad \text{for } j+1, \ldots, k-1\,, \\ & (u, c) \in E^m \times R_+\,. \end{aligned}$$

Let $(z_k, \tilde{u}_k, \tilde{c}_k)$ be an optimal solution and go to the step 2.

2. Solve the first subproblem: $(SP1)$ minimize $f_0(x) + \tilde{c}_k\|f(x)\| - \tilde{u}_k' f(x)$,

subject to $x \in S$.

Let $\tilde{x}_k$ be an optimal point. If $f(\tilde{x}_k) = 0$ then stop; by Theorem 5, $\tilde{x}_k$ is a solution to the primal problem. Otherwise, let $H(\tilde{u}_k, \tilde{c}_k) = f_0(\tilde{x}_k) + \tilde{c}_k\|f(\tilde{x}_k)\| - \tilde{u}_k' f(\tilde{x}_k)$ go to the step 3.

3. Let

$$(\bar{u}_k, \bar{c}_k, \bar{x}_k) = \begin{cases} (\tilde{u}_k, \tilde{c}_k, \tilde{x}_k) & \text{if } H(\tilde{u}_k, \tilde{c}_k) \geq H(u_{k-1}, c_{k-1}), \\ (u_{k-1}, c_{k-1}, x_{k-1}) & \text{otherwise.} \end{cases}$$

and

$$u_k = \bar{u}_k - s_k f(\bar{x}_k),\; c_k = \bar{c}_k + (s_k + \varepsilon_k)\|f(\bar{x}_k)\|\,,$$

where $0 < \varepsilon_k < s_k$, and the stepsize parameter s_k may be calculated (cf. (14)) as:

$$s_k = \frac{\alpha_k(z_k - H(\bar{u}_k, \bar{c}_k))}{5\|f(x_k)\|^2}\,, \tag{18}$$

for $0 < \alpha_k < 2$ and solve the second *subproblem*

$(SP2)$ minimize $f_0(x) + c_k\|f(x)\| - u_k' f(x)$, subject to $x \in$.

Let x_k be an optimal point and let $H(u_k, c_k) = f_0(x_k) + c_k\|f(x_k)\| - u_k' f(x_k)$. If

$f(x_k) = 0$ then stop; x_k is a solution to the primal problem and (u_k, c_k) is a solution to the dual problem. Otherwise, replace k by $k+1$, and repeat step 1. □

REMARK 1. The presented above combined algorithm contains either iterations of the usual cutting plane method (calculated for the dual problem constructed with respect to the 'sharp' Lagrangian) and iterations of modified subgradient method. It was proved that every limit point of a sequence of dual solutions generated by the iterations of the cutting plane method is a dual optimal solution (see, for example, Proposition 6.3.2 in Bertsekas (1995). Since the dual function is continuous (under hypotheses of Theorem 4 the dual function is concave and finite everywhere on $R^m \times R_+$, which implies the continuity of this function on $R^m \times (0, +\infty)$), the corresponding sequence of dual values will converge to the optimal dual value which coincides with the optimal primal one (by Theorem 4). For the sequence, generated by the modified subgradient method, the convergence was proved above in Theorem 9. So, we can conclude that every limit point of the sequence of dual values generated by the combined method is an optimal value.

REMARK 2. Since $z_k \geqslant \sup P^* \geqslant H(\bar{u}_k, \bar{c}_k)$, for each k, we use z_k as an upper approximation for H_k in (14) and obtain (18).

REMARK 3. It seems that the use of 'Cutting Angle Method' (see Rubinov, 2000; and also Andramanov et al., 1997; Andramanov et al., 1999) for solving subproblems at the Main Step 1 of the Subgradient Method and at the Main Step 2 and 3 of the Combined Method may give better results for appropriate problems.

Acknowledgements

The author thanks Professor A.M. Rubinov fot helpful discussions on this subject. Moreover, the author gratefully acknowledges anonymous referees for their exceptionally careful reading and constructive criticism. Their useful comments improved the presentation of this paper and corrected some inaccuracies. The author also thanks O. Ustun for his help in solving test problems by applying the proposed method.

References

Andramanov, M.Yu., Rubinov, A.M. and Glover, B.M. (1997), Cutting Angle Method for Minimizing increasing Convex-Along-Rays Functions, Research Report 97/7, SITMS, University of Ballarat, Australia.

Andramanov, M.Yu., Rubinov, A.M. and Glover, B.M. (1999), Cutting Angle Methods in Global Optimization, *Applied Mathematics Letters* 12, 95–100.

Bazaraa, M.S., Sherali, H.D. and Shetty, C.M. (1993), *Nonlinear Programming. Theory and Algorithms*, John Wiley & Sons, Inc., New York.

Bertsekas, D.P. (1995), *Nonlinear Programming*, Athena Scientific, Belmont, MA.

Buys, J.D. (1972), Dual Algorithms for Constrained Optimization Problems, Doctoral Dissertation, University of Leiden, Leiden, the Netherlands.

Cheney, E.W. and Goldstein, A.A. (1959), Newton's Method for Convex Programming and Tchebycheff Approximation, *Numer. Math.* 1, 253–268.

Courant, R. (1943), Variational Methods for the Solution of Problems of Equilibrium and Vibrations, *Bull. Amer. Math. Soc.* 49, 1–23.

Demyanov, V.F. (1968), Algorithm for some Minimax Problems, *J. Computer and System Sciences* 2, 342–380.

Ekeland, I. and Temam, R. (1976) *Convex Analysis and Variational Problems*, Elsevier–North Holland, Amsterdam.

Fletcher, R. (1970), A Class of Methods for Nonlinear Programming with Termination and Convergence Properties, in Abadie, J. (ed.), *Integer and Nonlinear Programming*, North Holland, Amsterdam, pp. 157–173.

Floudas, C.A., et al. (1999), *Handbook of Test Problems in Local and Global Optimization*, Kluwer Academic Publishers, Dordrecht.

Goffin, J.L. (1977), On Convergent Rates of Subgradient Optimization Methods, *Math. Programming* 13, 329–347.

Haarhoff, P.C. and Buys, J.D. (1970), A New Method for the Optimization of a Nonlinear Function Subject to Nonlinear Constraints, *Comput. J.* 13, 178–184.

Hestenes, M.R. (1969), Multiplier and Gradient Methods, *J. Optim. Theory Appl.* 4, 303–320.

Himmelblau, D.M. (1972) *Applied Nonlinear Optimization*, McGraw-Hill Book Company, New York.

Kelley, J.E. (1960), The Cutting-Plane Method for Solving Convex programs, *J. Soc. Indust. Appl. Math.* 8, 703–712.

Khenkin, E.I. (1976), A Search Algorithm for General Problem of Mathematical Programming, *USSR Journal of Computational Mathematics and Mathematical Physics* 16, 61–71, (in Russian).

Kort, B.W. and Bertsekas, D.P. (1972), A New Penalty Function Method for Constrained Minimization, *Proc. IEEE Decision and Control Conference*, New Orleans, LA, pp. 162–166.

Krein, M.G. and Rutman, M.A. (1962), Linear Operators Leaving Invariant a Cone in a Banach Space, *Trans. Amer. Math. Soc., Providence, RI*, 10, 199–325

Pallaschke, D. and Rolewicz, S. (1997), *Foundations of Mathematical Optimization* (*Convex Analysis without Linearity*), Kluwer Academic Publishers, Dordrecht.

Pietrzykowski, T. (1969), An Exact Potential Method for Constrained Maxima, *SIAM J. Numer. Anal.* 6, 299–304.

Polak, E. (1997), *Optimization. Algorithms and Consistent Approximations*, Springer, Berlin.

Poljak, B.T. (1969a), Minimization of Unsmooth Functionals, *Z. Vychislitelnoy Matematiki i Matematicheskoy Fiziki* 9, 14–29.

Poljak, B.T. (1969b), The Conjugate Gradient Method in Extremal Problems, *Z. Vychislitelnoy Matematiki i Matematicheskoy Fiziki* 9, 94–112.

Poljak, B.T. (1970), Iterative Methods Using Lagrange Multipliers for Solving Extremal Problems with Constraints of the Equation Type, *Z. Vyçislitelnoy Mat. i Mat. Fiziki* 10, 1098–1106.

Powell, M.J.D. (1969), A Method for Nonlinear Constraints in Minimization Problems, in Fletcher, R. (ed.), *Optimization*, Academic Press, New York, pp. 283–298.

Rockafellar, R.T. (1970), *Convex Analysis*, Princeton University Press, Princeton, NJ.

Rockafellar, R.T. (1993), Lagrange Multipliers and Optimality, *SIAM Review* 35, 183–238.

Rockafellar, R.T. and Wets, R.J.-B. (1998), *Variational Analysis*, Springer, Berlin.

Rubinov, A.M. (2000), *Abstract Convexity and Global Optimization*, Kluwer Academic Publishers, Dordrecht.

Shor, N.Z. (1985), *Minimization Methods for Nondifferentiable Functions*, Springer, Berlin.

Shor, N.Z. (1995), Dual Estimates in Multiextremal Problems, *Journal of Global Optimzation* 7, 75–91.

Singer, I. (1997), *Abstract Convex Analysis*, John Wiley and Sons, Inc., New York.

Wierzbicki, A.P. (1971), A Penalty Function Shifting Method in Constrained Static Optimization and its Convergence Properties, *Arch. Automat. i Telemechaniki* 16, 395–415.

Zangwill, W.I. (1967), Nonlinear Programming via Penalty Functions, *Management Sci.* 13, 344–358.