Journal of Global Optimization **28:** 153–173, 2004.

On augmented Lagrangians for Optimization Problems with a Single Constraint

R. N. GASIMOV and A. M. RUBINOV
[1]*Department of Industrial Engineering, Osmangazi University, Bademlik 26030, Eskişehir, Turkey; E-mail: gasimovr@ogu.edu.tr;* [2]*SITMS, University of Ballarat Victoria 3353, Australia; E-mail: a.rubinov@ballarat edu.au*

(Received and accepted in revised form 21 January 2003)

Abstract. We examine augmented Lagrangians for optimization problems with a single (either inequality or equality) constraint. We establish some links between augmented Lagrangians and Lagrange-type functions and propose a new kind of Lagrange-type functions for a problem with a single inequality constraint. Finally, we discuss a supergradient algorithm for calculating optimal values of dual problems corresponding to some class of augmented Lagrangians.

Key words: Augmented Lagrangians, Lagrange-type functions, Supergradient method

1. Introduction

Classical Lagrange and penalty functions can be applied only for examination of some special classes of constrained optimization problems. Some generalizations of these functions have recently been studied. Currently there are two main types of such a generalization. One of them is the augmented Lagrangian, which is based on an augmentation of the classical Lagrange function by a certain augmenting function (see [6, 12] and references therein).

The fundamental of the other approach to generalization of Lagrangians is a nonlinear convolution of the objective and constraint functions (see [7, 9, 10, 12] and references therein). Such a convolution leads to nonlinear Lagrange-type functions. We establish some links between the two mentioned approaches.

It is well-known that each constrained optimization problem can be reformulated as a problem with a single inequality-constraint. Many complicated constructions become much simpler and more understandable for single-constrained problems. In this paper we examine the augmented Lagrangians with certain augmenting functions for problems with a single (either inequality or equality) constraint. In particular we study the so-called sharp Lagrangian [6] for such problems. The simple structure of the sharp augmenting function $\sigma(z)=|z|$ allows us to give an explicit description of the sharp augmented Lagrangian. By using this result we propose a new type of nonlinear Lagrangians for problems with an inequality constraint, for which the dual function can be easily expressed through the dual function of the problem with an equality constraint. This approach allows us to extend some results, obtained for problems with an equality constraint, to problems with an inequality constraint. We also examine a certain version of the supergradient method for solving the dual problem. First we consider problems with an equality constraint and generalize a version of this method proposed in [1] for sharp augmented Lagrangian, to a more general class of augmented Lagrangians. Then we show that a new type of nonlinear Lagrangians allows to use this method in solving the problems with an inequality constraint.

2. Preliminaries

2.1. CONSTRAINED OPTIMIZATION PROBLEM AND ITS REFORMULATION

Consider a metric space X and functions $f: X \to \mathbb{R}$ and $g: X \to \mathbb{R}^m$ where $\mathbb{R}^m$ is m-dimensional Euclidean space equipped with the coordinate-wise order relation. Consider the following constrained optimization problem $P(f,g)$:

$$\text{minimize } f(x) \text{ subject to } x \in X, g(x) \leqslant 0, \tag{2.1}$$

where $g(x) = (g_1(x), \ldots, g_m(x))$. Let

$$X_0 = \{x \in X : g(x) \leqslant 0\} \tag{2.2}$$

be the set of feasible solutions for $P(f,g)$ and let

$$M = \inf\{f(x) : x \in X_0\} \tag{2.3}$$

be the optimal value of $P(f,g)$. It is assumed that $M > -\infty$.

For some applications it is convenient to consider certain reformulations of the problem $P(f,g)$. In particular, $P(f,g)$ can be reformulated as the following problem $P(f,f_1)$ with the single constraint function f_1:

$$\text{minimize } f(x) \text{ subject to } x \in X, f_1(x) \leqslant 0, \tag{2.4}$$

where $f_1(x) = \max_{i=1,\ldots,m} g_i(x)$. Problems $P(f,g)$ and $P(f,f_1)$ have the same set of feasible solutions and the same objective function.

REMARK 2.1. A general mathematical programming problem:

$$\text{minimize} f(x) \text{ subject to } g_i(x) \leqslant 0 (i \in I), \quad h_j(x) = 0, (j \in J) \tag{2.5}$$

with finite I and J, also can be reformulated as (2.4) with

$$f_1(x) = \max(\max_{i \in I} g_i(x), \max_{j \in J} |h_j(x)|).$$

The approach proposed here is suitable for examining even more complicated problems, which lead to $f_1(x) = \min_{i \in I} \max_{j \in J} g_{ij}(x)$, where I and J are finite sets of indices and g_{ij} are certain functions.

An inequality constraint $f_1(x) \leqslant 0$ is called active if the problem $P(f, f_1)$ has a solution x_* such that $f_1(x_*) = 0$. In this case it is obvious that x_* is a solution to the problem $P^e(f, f_1)$ with the equality constraint:

$$\text{minimize } f(x) \text{ subject to } f_1(x) = 0. \tag{2.6}$$

Assume that the inequality constraint $f_1(x) \leqslant 0$ is obtained by the cutting of the negative part of a certain constraint function $\tilde{f}$: $f_1(x) = \tilde{f}^+(x)$, where $a^+ = \max(a, 0)$. (Such functions are used for penalization.) Then the constraint $f_1(x) \leqslant 0$ is active if and only if $P(f, f_1)$ has a solution.

2.2. AUGMENTED LAGRANGE FUNCTIONS

To define an augmented Lagrangian for a problem $P(f, g)$ we need to have two exogenous with respect to $P(f, g)$ functions. First of them is the so-called dualizing parameterization, that is a function $\bar{f}: X \times \mathbb{R}^m \to \bar{\mathbb{R}}$ such that $\bar{f}(x, 0) = f(x)$, where $\bar{\mathbb{R}} = [-\infty, +\infty]$. We need to have also an augmenting function σ for the augmentation of the classical Lagrangian. It is assumed that $\sigma: \mathbb{R}^m \to \mathbb{R}$ is a continuous function with the following properties:

$$\sigma(0) = 0, \qquad \sigma(z) > 0 \text{ if } z \neq 0. \tag{2.1}$$

Let $\Omega \subset \mathbb{R}^m \times \mathbb{R}_+$. The augmented Lagrange function $l: X \times \Omega \to \mathbb{R}$ corresponding to $\bar{f}$ and σ has the following form [3, 6]:

$$l(x, (y, r)) = \inf_{z \in \mathbb{R}^m} (\bar{f}(x, z) - [y, z] + r\sigma(z)), \qquad (y, r) \in \Omega. \tag{2.2}$$

(We denote the inner product of vectors y and z by $[y, z]$.) We shall consider here only the canonical dualizing parameterization function $\bar{f}$ defined on $X \times \mathbb{R}^m$ by

$$\bar{f}(x, z) = \begin{cases} f(x) & g(x) + z \leqslant 0; \\ +\infty & \text{otherwise}. \end{cases} \tag{2.3}$$

Then the augmented Lagrangian corresponding to an augmenting function σ has the form:

$$l(x, (y, r)) = \inf_{g(x) + z \leqslant 0} (f(x) - [y, z] + r\sigma(z)), \qquad (y, r) \in \Omega. \tag{2.4}$$

The dual function q is defined on Ω by

$$q(y, r) = \inf_{x \in X} l(x, (y, r)).$$

Under natural assumptions the weak duality property holds, that is, $M_* \leqslant M$, where M_* is the optimal value of the dual problem:

$$(D) \qquad \text{maximize } q(y, r) \text{ subject to } (y, r) \in \Omega.$$

It is well-known (see, for example [3, 6]) and easy to see that the dual function q is concave.

The augmented Lagrangian l^e for the problem $P^e(f,g)$ with the constraint function $g: X \to \mathbb{R}^m$ can be defined as

$$l^e(x,(y,r)) = f(x) + [y, g(x)] + r\sigma(g(x)).$$

This augmented Lagrangian corresponds to the dualizing parameterization

$$\bar{f}(x,z) = \begin{cases} f(x) & \text{if} \quad g(x) - z = 0, \\ +\infty, & \quad \text{otherwise}. \end{cases}$$

The function l^e generates the dual function $q^e(y,r) = \inf_{x\in X} l^e(x,(y,r))$ and the dual problem

$$(D^e) \qquad \text{maximize } q^e(y,r) \text{ subject to } (y,r) \in \Omega.$$

2.3. LAGRANGE-TYPE FUNCTIONS BASED ON CONVOLUTION FUNCTIONS

We consider one more type of nonlinear Lagrangians, which is based on convolution of the objective and constraint functions (see [3, 7, 9, 10, 12] and references therein). Let Ω be a set of parameters. Consider a function $h: \mathbb{R} \times \mathbb{R}^m \times \Omega \to \bar{\mathbb{R}}$, where $\bar{\mathbb{R}}$ is the extended real line. The function $L: X \times \Omega \to \bar{\mathbb{R}}$ defined by

$$L(x,\omega) = h(f(x), g(x), \omega)$$

is called the Lagrange-type function corresponding to the function h. The function h in this scheme, is called a convolution function. The dual function q is defined by $q(\omega) = \inf_{x\in X} L(x,\omega)$. The problem

$$(D) \qquad \text{maximize } q(\omega) \text{ subject to } \omega \in \Omega.$$

is called dual. Here we consider mainly convolution functions h of the form

$$h(u,v,\omega) = u + \chi(v,\omega), \qquad u \in \mathbb{R}, v \in \mathbb{R}^m, \omega \in \Omega. \tag{2.1}$$

The Lagrange-type function of the problem $P(f,g)$ corresponding to h has the form

$$L(x,\omega) = f(x) + \chi(g(x),\omega), \qquad x \in X, \omega \in \Omega. \tag{2.2}$$

Nonlinear Lagrange-type functions, which are based on more general convolution functions, have been studied in [7, 10, 12].

3. Links Between the Two Types of Generalized Lagrange Functions

Consider the problem $P(f,g)$ with $f: X \to \mathbb{R}$ and $g: X \to \mathbb{R}^m$. Let $l(x,(y,r))$ be the augmented Lagrangian of $P(f,g)$ corresponding to the canonical dualizing parameterization $\bar{f}$ defined by (2.3) and to an augmenting function σ. We have (see (2.4)):

$$l(x,(y,r)) = \inf_{g(x)+z\leqslant 0}(f(x)-[y,z]+r\sigma(z)). \tag{3.1}$$

Let $\Omega\subset\{(y,r):y\in\mathbb{R}^m, r\geqslant 0\}$. Assume that $(0,0)\in\Omega$ and $(y,r)\in\Omega$ implies $(y,r')\in\Omega$ for all $r'\geqslant 0$. We consider Ω as the set of parameters. Consider the convolution function $h:\mathbb{R}\times\mathbb{R}^m\times\Omega\to\mathbb{R}$ defined by

$$h(u,v,(y,r)) = \inf_{z+v\leqslant 0}(u-[y,z]+r\sigma(z)) = u+\inf_{z+v\leqslant 0}(-[y,z]+r\sigma(z)). \tag{3.2}$$

It is easy to see that the Lagrange-type function $L(x,(y,r))=h(f(x),g(x),(y,r))$ corresponding to (3.2) coincides with the augmented Lagrangian $l(x,y,r)$. Let

$$\chi(v,(y,r)) = \inf_{z+v\leqslant 0}(-[y,z]+r\sigma(z)). \tag{3.3}$$

Then $h(u,v,(y,r))=u+\chi(v,(y,r))$. We now give some properties of the function χ.

PROPOSITION 3.1. *Let χ be the function defined by* (3.3). *Then*

(1) $\chi(v,(0,0))=0$ *for all* $v\in\mathbb{R}^m$.
(2) *The function* $v\mapsto\chi(v,(y,r))$ *is increasing for each* $(y,r)\in\Omega$.
(3) $\chi(0,(y,r))\leqslant 0$ *for all* $(y,r)\in\Omega$.
(4) $\chi(v,(y,r))\leqslant 0$ *for* $v\leqslant 0$ *and* $(y,r)\in\Omega$.
(5) $\max_{(y,r)\in\Omega}\chi(v,(y,r))=0$ *for all* $v\leqslant 0$.

Proof. (1) This is clear.

(2) Since $v_1\leqslant v_2$ implies $\{z:z+v_2\leqslant 0\}\subset\{z:z+v_1\leqslant 0\}$ it follows that the function $v\mapsto\chi(v,(y,r))$ is increasing for each (y,r).
(3) $\chi(0,(y,r))=\inf_{z\leqslant 0}(-[y,z]+r\sigma(z))\leqslant -[y,0]+r\sigma(0)=0$.
(4) This follows from (2) and (3).
(5) Combining (4) and (1) we have:

$$\sup_{(y,r)\in\Omega}\chi(v,(y,r))\leqslant 0=\chi(v,(0,0)),$$

for $v\leqslant 0$. Hence $\max_{(y,r)\in\Omega}\chi(v,(y,r))=0$ for all for $v\leqslant 0$.

REMARK 3.1. If the augmenting function σ is positively homogeneous then the function $v\mapsto\chi(v,(y,r))$ is also positively homogeneous with respect to v for any fixed (y,r).

PROPOSITION 3.2. *Let $(y,r)\in\Omega$ and $\bar{v}\in\mathbb{R}^m$. Assume that there exist numbers $R=R(\bar{v})$ and $\varepsilon=\varepsilon(\bar{v})>0$ such that*

$$\inf_{z\leqslant -v}(-[y,z]+r\sigma(z))=\min_{z\leqslant -v,\|z\|\leqslant R}(-[y,z]+r\sigma(z)) \tag{3.4}$$

for all $v\in\mathbb{R}^m$ with $\|v-\bar{v}\|<\varepsilon$. Then the function $\chi_{(y,r)}(v):=\chi(v,(y,r))$ is continuous at $\bar{v}$.

Proof. We have

$$\chi_{(y,r)}(v)=\min_{z\leqslant -v,\|z\|\leqslant R}(-[y,z]+r\sigma(z))$$

for all v such that $\|v-\bar{v}\|\leqslant\varepsilon$. Denote $-[y,z]+r\sigma(z)=\psi(z)$. The function ψ is continuous. Let $v^k\to\bar{v}$ and let $\bar{z}\in\operatorname{argmin}\{\psi(z):z\leqslant-\bar{v},\|z\|\leqslant R\}$. Consider the vector z^k with components $z_i^k=\min(-v_i^k,\bar{z}_i)$, $i=1,...m$. Then $z^k\leqslant -v^k$, so

$$\psi(z^k)\geqslant \inf_{z\leqslant -v^k}\psi(z)=\chi_{(y,r)}(v^k).$$

Since $v^k\to\bar{v}$ it follows that $z^k\to\bar{z}$, so $\chi_{(y,r)}(\bar{v})=\psi(\bar{z})\geqslant\limsup_{k\to+\infty}\chi_{(y,r)}(v^k)$. Consider now a vector $\bar{z}_k\in\operatorname{argmin}\{\psi(z):z\leqslant -v^k,\|z\|\leqslant R\}$. Without loss of generality assume that there exists $\lim\bar{z}^k:=\bar{z}$. Then $\bar{z}\leqslant-\bar{v}$ and $\|\bar{z}\|\leqslant R$, so $\chi_{(y,r)}(\bar{v})\leqslant\psi(\bar{z})=\lim\psi(\bar{z}_k)=\lim\chi_{(y,r)}(v^k)$.

REMARK 3.2. The condition (3.4) holds for $(y,r)\in\Omega$ and all $\bar{v}\in\mathbb{R}^m$ if

$$\lim_{\|z\|\to+\infty}(-[y,z]+r\sigma(z))=+\infty. \tag{3.5}$$

4. Sharp Augmented Lagrangian for Problems with a Single Constraint

Let X be a metric space. We shall consider a problem $P(f,f_1)$ with a single constraint:

$$\text{minimize } f(x) \text{ subject to } x\in X,\quad f_1(x)\leqslant 0,$$

where f and f_1 are finite functions defined on X. Recall that the optimal value of $P(f,f_1)$ has been denoted by M: $M=\inf_{f_1(x)\leqslant 0}f(x)$ and the set of feasible solutions has been denoted by X_0: $X_0=\{x\in X:f_1(x)\leqslant 0\}$. We will also use the following notation:

$$X_1=\{x\in X:f_1(x)>0\}=\{x\in X:x\notin X_0\}. \tag{4.1}$$

In this section we shall study a sharp Lagrangian with $\sigma(z)=|z|$ in the case $m=1$. Thus

$$l(x,(y,r))=\inf_{z+f_1(x)\leqslant 0}(f(x)-yz+r|z|).$$

Let χ be a function defined by (3.3). Then

$$l(x,(y,r)) = f(x) + \chi(f_1(x),(y,r)). \tag{4.2}$$

Note that

$$\chi(v,(y,r)) = \inf_{z\leqslant -v}(-yz + r|z|).$$

Let us calculate the quantity $\chi(v,(y,r))$ for all $(y,r)\in\mathbb{R}^2$ and $v\in\mathbb{R}$, explicitly.

PROPOSITION 4.1. *Let* $\Omega = \{(y,r): y+r\geqslant 0\}$. *If* $(y,r)\in\Omega$ *then*

$$\chi(v,(y,r)) = \begin{cases} v(y+r) & \textit{if } v>0; \\ v(y-r) & \textit{if } v\leqslant 0, r<y \\ 0 & \textit{if } v\leqslant 0, r\geqslant y \end{cases} \tag{4.3}$$

If $(y,r)\notin\Omega$ *then* $\chi(v,(y,r)) = -\infty$.

Proof. We consider separately the following two cases:
(1) $v>0$. In this case $z\leqslant -v$ implies $z\leqslant 0$, and therefore for all $(y,r)\in\mathbb{R}^2$ we have:

$$\chi(v,(y,r)) = \inf_{z\leqslant -v}(-z)(y+r) = \begin{cases} -\infty & \text{if } y+r<0, \\ v(y+r) & \text{if } y+r\geqslant 0. \end{cases} \tag{4.4}$$

(2) $v\leqslant 0$. Then for all (y,r) we have:

$$\chi(v,(y,r)) = \min\big(\inf_{0\leqslant z\leqslant -v} z(r-y), \inf_{z<0}(-z)(r+y)\big).$$

Since

$$\inf_{0\leqslant z\leqslant -v} z(r-y) = \begin{cases} v(y-r) & \text{if } r<y, \\ 0 & \text{if } r\geqslant y \end{cases}$$

and

$$\inf_{z<0}(-z)(r+y) = \begin{cases} -\infty & \text{if } y+r<0, \\ 0 & y+r\geqslant 0 \end{cases}$$

we conclude that

$$\chi(v,(y,r)) = \begin{cases} -\infty & \text{if } y+r<0, \\ v(y-r) & \text{if } y+r\geqslant 0,\ r<y, \\ 0 & \text{if } y+r\geqslant 0,\ r\geqslant y, \end{cases} \tag{4.5}$$

for all $v\leqslant 0$. It follows from (4.4) and (4.5) that $\chi(v,(y,r)) = -\infty$ for $(y,r)\notin\Omega$ and all v. Applying again (4.4) and (4.5) we obtain (4.3).

COROLLARY 4.1. *A sharp Lagrangian $l: X \times \Omega \to \bar{\mathbb{R}}$ has the following form*:

$$l(x,(y,r)) = f(x) + \begin{cases} f_1(x)(y+r) & \text{if } x \in X_1; \\ f_1(x)(y-r) & \text{if } x \in X_0,\ r < y; \\ 0 & x \in X_0,\ r \geqslant y. \end{cases}$$

Proof. It follows directly from (4.2) and (4.3).

Let us compare the sharp Lagrangian $l(x,(y,r))$ with the classical penalty function $L^+(x,\lambda)$ defined on $X \times \mathbb{R}_+$. By definition L^+ is the Lagrange-type function corresponding to convolution function $h^+: \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+$, where

$$h^+(u,v,\lambda) = u + \lambda v^+ = u + \chi^+(v,\lambda) \tag{4.6}$$

with

$$\chi^+(v,\lambda) = \lambda v^+. \tag{4.7}$$

We have

$$L^+(x,\lambda) = f(x) + \begin{cases} \lambda f_1(x) & \text{if } x \in X_1, \\ 0 & \text{if } x \in X_0. \end{cases} \tag{4.8}$$

Let $(y,r) \in \Omega$. If $r \geqslant y$ then $l(x,(y,r)) = L^+(x,y+r)$ for all $x \in X$. If $r < y$ then $l(x,(y,r)) = L^+(x,y+r)$ for $x \in X_1$ and $l(x,(y,r)) \leqslant L^+(x,y+r)$ for $x \in X_0$. Thus

$$l(x,(y,r)) \leqslant L^+(x,y+r), \qquad x \in X,\ (y,r) \in \Omega. \tag{4.9}$$

Let

$$\Omega_1 = \{(y,r) \in \Omega : r \geqslant y\}. \tag{4.10}$$

It follows from the above considerations that the following assertion holds.

PROPOSITION 4.2. $l(x,(y,r)) = L^+(x,y+r)$ *for all* $x \in X$ *and* $(y,r) \in \Omega_1$.

5. Dual Functions for Sharp Lagrangians

In this section we calculate the dual function q corresponding to the sharp Lagrangian l of the problem $P(f,f_1)$. Recall that, by definition

$$q(y,r) = \inf_{x \in X} l(x,(y,r)), \qquad (y,r) \in \Omega := \{(y,r) : y+r \geqslant 0\}.$$

Let

$$q^e(y,r) = \inf_{x \in X}(f(x) + yf_1(x) + r|f_1(x)|), \qquad (y,r) \in \Omega. \tag{5.1}$$

be the dual function for the problem $P^e(f,f_1)$ with the equality constraint function f_1.

PROPOSITION 5.1. *We have*

$$q(y,r)=\min(M,q^e(y,r))\, for (y,r)\in\Omega.$$

Proof. Applying Corollary 4.1 we conclude that:

$$\inf_{x\in X_1} l(x,(y,r))=\inf_{x\in X_1}(f(x)+(y+r)f_1(x)), \tag{5.2}$$

$$\inf_{x\in X_0} l(x,(y,r))=\begin{cases}\inf_{x\in X_0}(f(x)+(y-r)f_1(x)) & \text{if } r<y\\ \inf_{x\in X_0} f(x):=M & \text{if } r\geqslant y.\end{cases} \tag{5.3}$$

By definition,

$$q(y,r)=\inf_{x\in X} l(x,(y,r))=\min(\inf_{x\in X_0} l(x,(y,r)),\inf_{x\in X_1} l(x,(y,r))).$$

Combining (5.2) and (5.3) we conclude that

$$q(y,r)=\begin{cases}\min(\inf_{x\in X_0}(f(x)+(y-r)f_1(x)),\inf_{x\in X_1}(f(x)+(y+r)f_1(x)) & \text{if } r<y,\\ \min(M,\inf_{x\in X_1}(f(x)+(y+r)f_1(x)) & \text{if } r\geqslant y.\end{cases} \tag{5.4}$$

Note that

$$\begin{aligned}q^e(y,r) &= \min(\inf_{x\in X_0}(f(x)+yf_1(x)+r|f_1(x)|),\\ &\quad \inf_{x\in X_1}(f(x)+yf_1(x)+r|f_1(x)|)).\end{aligned} \tag{5.5}$$

Since $f_1(x)\leqslant 0$ for $x\in X_0$ and $f_1(x)>0$ for $x\in X_1$ we deduce that

$$\begin{aligned}&\min(\inf_{x\in X_0}(f(x)+(y-r)f_1(x)),\inf_{x\in X_1}(f(x)+(y+r)f_1(x))=\\ &\min(\inf_{x\in X_0}(f(x)+yf_1(x)+r|f_1(x)|),\\ &\quad\inf_{x\in X_1}(f(x)+yf_1(x)+r|f_1(x)|))=q^e(y,r).\end{aligned} \tag{5.6}$$

Let $y>r$ and $x\in X_0$. Then $(y-r)f_1(x)\leqslant 0$ so $\inf_{x\in X_0}(f(x)-(r-y)f_1(x))\leqslant M$. It follows from (5.6) that $q^e(y,r)\leqslant m$, so $\min(M,q^e(y,r))=q^e(y,r)$. Combining this equality with (5.4) and (5.6) we have

$$q(y,r)=q^e(y,r)=\min(M,q^e(y,r)),\qquad y>r.$$

Assume now that $y\leqslant r$. First we prove that

$$\min(M,\inf_{x\in X_1}(f(x)+(y+r)f_1(x)))=\min(M,q^e(y,r)). \tag{5.7}$$

Indeed,

$$q^e(y,r) \leqslant \inf_{x \in X_1} (f(x) + yf_1(x) + r|f_1(x)|). \tag{5.8}$$

If the equality holds in (5.8) then (5.7) trivially holds. Otherwise we have, due to (5.5)

$$q^e(y,r) = \inf_{x \in X_0} (f(x) + yf_1(x) + r|f_1(x)|) < \inf_{x \in X_1} (f(x) + yf_1(x) + r|f_1(x)|). \tag{5.9}$$

Since $y \leqslant r$, it follows that:

$$f(x) + yf_1(x) + r|f_1(x)| = f(x) + (y - r)f_1(x) \geqslant f(x), \qquad x \in X_0.$$

So

$$M \leqslant \inf_{x \in X_0} (f(x) + yf_1(x) + r|f_1(x)|) = q^e(y,r) < \inf_{x \in X_1} (f(x) + yf_1(x) + r|f_1(x)|).$$

Thus (5.7) is valid. Combining (5.4) and (5.7) we conclude that

$$q(y,r) = \min(q^e(y,r), M), \qquad \text{if } y \leqslant r.$$

COROLLARY 5.1. *Assume that the constraint f_1 is active, that is, the problem $P(f, f_1)$ has a solution x_* such that $f_1(x_*) = 0$. (Then x_* is also a solution of $P^e(f, f_1)$.) Then we have $q(y,r) = q^e(y,r)$, $\quad (y,r) \in \Omega$.*

Proof. Indeed,

$$M = f(x_*) = f(x_*) + yf_1(x_*) + r|f_1(x_*)| \geqslant \inf_{x \in X} (f(x) + yf_1(x) + r|f_1(x)|) = q^e(y,r).$$

Hence, $q(y,r) = \min(M, q^e(y,r)) = q^e(y,r)$.

Consider the dual function q^+ corresponding to the penalty function L^+. By definition,

$$q^+(\lambda) = \inf_{x \in X} (f(x) + \lambda f_1^+(x)), \qquad \lambda > 0. \tag{5.10}$$

Since $l(x,(y,r)) \leqslant L^+(x, y+r)$ for all $x \in X$ and $(y,r) \in \Omega$ (see (4.9)), it follows that $q(y,r) \leqslant q^+(y+r)$ for $(y,r) \in \Omega$.

Consider now the set of parameters Ω_1 defined by (4.10): $\Omega_1 = \{(y,r) \in \Omega : r \geqslant y\}$. Since $l(x,(y,r)) = L^+(x, y+r)$ for all $x \in X$ and $(y,r) \in \Omega_1$ (see Proposition 4.2), it follows that

$$q(y,r) = q^+(y+r), \qquad (y,r) \in \Omega_1. \tag{5.11}$$

6. An Approach to Constructing Nonlinear Lagrangians

Proposition 5.1 establishes links between dual functions for the sharp Lagrangians for the problems $P(f, f_1)$ and $P^e(f, f_1)$ respectively. Namely,

$$q(y, r) = \min(q^e(y, r), M), \qquad (y, r) \in \Omega = \{(y', r') : y' + r' \geqslant 0\}. \tag{6.1}$$

Note that the weak duality property does not necessarily hold for the function q^e, if we consider this function as dual with respect to the inequality constrained problem $P(f, f_1)$. The simplest way to obtain this property is to use the construction from (6.1). In this section we shall examine an abstract version of this construction and then apply it to some augmented Lagrangians.

Let X be a metric space. Consider a problem $P(f, g)$ with $f : X \to \mathbb{R}$ and $g : X \to \mathbb{R}^m$. Let Ω be a set of parameters, and let $h_\Delta : \mathbb{R} \times \mathbb{R}^m \times \Omega \to \mathbb{R}$ be an arbitrary convolution function. Then the corresponding Lagrange-type function $L_\Delta : X \times \Omega \to \mathbb{R}$ is defined as

$$L_\Delta(x, \omega) = h_\Delta(f(x), g(x), \omega), \qquad x \in X, \omega \in \Omega.$$

Let q_Δ be the corresponding dual function:

$$q_\Delta(\omega) = \inf_{x \in X} L_\Delta(x, \omega).$$

Let M be the optimal value of the problem $P(f, g)$. Consider the following function:

$$q(\omega) = \min(q_\Delta(\omega), M), \qquad \omega \in \Omega$$

and corresponding dual problem:

$$(D): \qquad\qquad \text{maximize } q(\omega) \text{ subject to } \omega \in \Omega.$$

If the weak duality property holds for the function q_Δ, that is if $q_\Delta(\omega) \leqslant M$ for all $\omega \in \Omega$, then $q = q_\Delta$.

We now describe some properties of the function q and problem D.

LEMMA 6.1. (1) *Suppose that ω is not a solution of the dual problem. Then* $q(\omega) = q_\Delta(\omega)$.
(2) *Let $\omega_* \in \Omega$ be an element such that there exists a vector $x_* \in X_0$ with the property $f(x_*) = \min_{x \in X} L_\Delta(x, \omega_*)$. Then ω_* is a solution of the dual problem* (D).

Proof. (1) Since ω is not a solution of the dual problem, it follows that

$$q(\omega) = \min(M, q_\Delta(\omega)) < M.$$

Hence $q(\omega) = q_\Delta(\omega)$.

(2) We have $f(x_*) = \min_{x \in X} L_\Delta(x, \omega_*) = q_\Delta(\omega_*)$. Since $x_* \in X_0$ it follows that $f(x_*) \geqslant M$. Thus $q(\omega_*) = \min(q_\Delta(\omega_*), M) = M = \max_{\omega \in \Omega} q(\omega)$.

Let σ be an arbitrary augmenting function and $\Omega = \{(y, r) \in \mathbb{R}^2 : y + r \geqslant 0\}$. Consider a problem $P^e(f, f_1)$ and the augmented Lagrangian l^e of this problem:

$$l^e(x, (y, r)) = f(x) + y f_1(x) + r\sigma(f_1(x)), \qquad x \in X,\ \omega \in \Omega.$$

Consider now the problem $P(f, f_1)$ and the following nonlinear Lagrangian L of this problem:

$$L(x, (y, r)) = \begin{cases} l^e(x, (y, r)) & x \in X_1 \\ f(x) & x \in X_0 \end{cases}, \qquad \omega = (y, r) \in \Omega. \tag{6.2}$$

Note that $L = l^e_+$, where l^e_+ is the augmented Lagrangian of the problem $P^e(f, f_1^+)$ or $P(f, f_1^+)$. Later on we shall use the augmenting functions σ for a problem $P(f, f_1)$ with the following property

$$\sigma(z) \geqslant |z| \text{ for } z \in \{f_1(x) : x \in X\} \tag{6.3}$$

The inequality (6.3) holds for all $P(f, f_1)$ with the sharp augmenting function $\sigma(z) = |z|$. Notice that we use the constraint function f_1 only for describing the set of feasible solutions. Therefore, by replacing $f_1(x)$ with the function $\tilde{f}_1(x)$ defined by

$$\tilde{f}_1(x) = \begin{cases} -1 & \text{if } f_1(x) \leqslant -1, \\ f_1(x) & \text{if } -1 \leqslant f_1(x) \leqslant 1, \\ 1 & \text{if } f_1(x) \geqslant 1, \end{cases}$$

we can always assume that the values of the constraint function are contained between -1 and $+1$.

Let $|f_1(x)| \leqslant 1$ for all $x \in X$ and let $0 < k \leqslant 1$. Consider the function $\sigma_k(z) = |z|^k$. Then $\sigma_k(z) \geqslant |z|$ for all $z \in [-1, 1]$, so (6.3) also holds.

To confirm that the augmenting function $\sigma(z) = |z|^k$ with $0 < k \leqslant 1$ is of a certain interest we shall examine saddle points of corresponding nonlinear Lagrangians. First consider an arbitrary augmenting function σ and the Lagrangian L defined by (6.2) on $X \times \Omega$. Recall that a point $(x_*, (y^*, r^*)) \in X_0 \times \Omega$ is called a saddle point of L if

$$L(x_*, (y, r)) \leqslant L(x_*, (y^*, r^*)) \leqslant L(x, (y^*, r^*)), \qquad x \in X,\ (y, r) \in \Omega. \tag{6.4}$$

Since $x_* \in X_0$ it follows that $L(x_*, (y, r)) = f(x)$ for all $(y, r) \in \Omega$, so (6.4) is equivalent to

$$f(x_*) \leqslant \begin{cases} f(x) + y^* f_1(x) + r^* \sigma(f_1(x)) & \text{if } x \in X_1, \\ f(x) & \text{if } x \in X_0. \end{cases} \tag{6.5}$$

Let x_* be a solution of $P(f,f_1)$. It follows from (6.5) that $(x_*,(y^*,r^*))$ is a saddle point if and only if

$$f(x_*)-f(x)\leqslant y^*f_1(x)+r^*\sigma(f_1(x)),\qquad x\in X_1.$$

Without loss of generality assume that $|f_1(x)|\leqslant 1$ for all $x\in X$ and that $\sigma_k(z)=|z|^k$, $(k>0)$. Then a saddle point for some k remains a saddle point for all $0<k'<k$. It is easy to give examples, which demonstrate that a saddle point does not exist for a certain k and exists for some $k'<k$. Thus it is beneficial to consider small $k>0$, in particular $k\leqslant 1$. Condition (6.3) is valid for such k. If $P(f,f_1)$ is a convex problem then the use of function σ_k can destroy the nice structure of this problem. However, for nonconvex problems of global optimization the use of such functions is inevitable. This observation shows that sometimes the augmented Lagrangians with a concave augmenting function are more beneficial than those with a convex augmenting function, which as rule have been studied in literature (see, for example [6]).

Consider a pair of functions f,f_1 defined on X, problems $P(f,f_1)$ and $P_e(f,f_1)$, augmented Lagrangian l^e for the problem $P^e(f,f_1)$ and nonlinear Lagrangian L defined by (6.2). Let q^e be the dual function, corresponding to l^e and q be the dual function corresponding to L.

PROPOSITION 6.1. *Let σ be an augmenting function such that* (6.3) *is valid. Then*

$$q(y,r)=\min(\ m,q^e(y,r))$$

for all $(y,r)\in\Omega_1=\{(y,r)\in\Omega,r\geqslant y\}$.

Proof. We need to prove that

$$q(y,r):=\min(M,\inf_{x\in X_1}(f(x)+yf_1(x)+r\sigma(f_1(x)))=\min(M,q^e(y,r)).\tag{6.6}$$

We have

$$\begin{aligned}q^e(y,r)\ &=\ \min(\inf_{x\in X_0}(f(x)+yf_1(x)+r\sigma(f_1(x)),\\&\qquad\inf_{x\in X_1}(f(x)+yf_1(x)+r\sigma(f_1(x))).\end{aligned}$$

If $q^e(y,r)=\inf_{x\in X_1}(f(x)+yf_1(x)+r\sigma(f_1(x)))$ then (6.6) is valid. Otherwise

$$\begin{aligned}q^e(y,r)\ &=\ \inf_{x\in X_0}(f(x)+yf_1(x)+r\sigma(f_1(x)))\leqslant\\&\qquad\inf_{x\in X_1}(f(x)+yf_1(x)+r\sigma(f_1(x))).\end{aligned}\tag{6.7}$$

It follows from (6.3) that

$$\begin{aligned}q^e(y,r)\ &=\ \inf_{x\in X_0}(f(x)+yf_1(x)+r\sigma(f_1(x))\geqslant\\&\qquad\inf_{x\in X_0}(f(x)+yf_1(x)-rf_1(x)\geqslant\inf_{x\in X_0}f(x)=M.\end{aligned}$$

Hence $\min(q^e(y,r),M)=M$. Applying the inequality from (6.7) we also have

$$q(y,r)=\min(M,\inf_{x\in X_1}(f(x)+yf_1(x)+r\sigma(f_1(x)))=M.$$

7. Duality

Consider the problem $P(f,f_1)$. Let

$$\ell(x,\omega)=f(x)+\chi(f_1(x),\omega), \qquad x\in X, \omega\in\Omega$$

be a certain Lagrange-type function for this problem. The dual problem to $P(f,f_1)$ corresponding to ℓ has the form

$$(D) \qquad\qquad \max q(\omega) \text{ subject to } \omega\in\Omega,$$

where q is the dual function: $q(\omega)=\inf_{x\in X}\ell(x,\omega)$.

The equality $\sup_{\omega\in\Omega}q(\omega)=\inf_{x\in X_0}f(x)$ is called the zero duality gap property. If the zero duality gap property holds then solving $P(f,f_1)$ can be reduced to solving a series of unconstrained problems. The zero duality gap property holds under very mild conditions (see, for example [11] and references therein). In particular, assume that $\Omega=\{(y,r):y\in\mathbb{R},r\in\mathbb{R}_+,y+r\geqslant 0\}$ and $\chi(v,(y,r))$ is defined by (3.3): $\chi(v,(y,r)=\inf_{z+v\leqslant 0}(-yz+r\sigma(z)$, where $\sigma(0)=0$ and $\sigma(z)>0$ for $z\neq 0$. The latter holds if σ is defined by (6.3). In this case the lower semicontinuity of the perturbation function $\beta(y)=\inf_{f_1(x)+y\leqslant 0}f(x)$ at the origin implies the zero duality gap property.

An element $\bar{\omega}\in\Omega$ is called an exact parameter for the problem $P(f,f_1)$ if

$$M=\inf_{x\in X}\ell(x,\bar{\omega}).$$

We do not require that the solution set for $P(f,f_1)$ coincides with $\operatorname{argmin}_{x\in X}\ell(x,\bar{\omega})$. Assume that the weak duality holds, that is $M\geqslant\inf_{x\in X}\ell(x,\omega)$ for all $\omega\in\Omega$. Then $\bar{\omega}$ is an exact parameter if and only if

$$M\leqslant\ell(x,\bar{\omega})=f(x)+\chi(f_1(x),\bar{\omega}) \text{ for all } x\in X. \tag{7.1}$$

Consider now the classical (linear) penalty function $L^+(x,\lambda)$ and the quadratic penalty function $L_2^+(x,\lambda)$, where

$$L^+(x,\lambda)=f(x)+\lambda f_1^+(x), \qquad L_2^+(x,\lambda)=f(x)+\lambda(f_1^+(x))^2.$$

In some applications the quadratic penalty function is preferred due to its differentiability properties, however it is easy to determine the cases when an exact parameter does not exist for quadratic penalty function while it exists for linear penalty function. Indeed, since $f_1^+(x)=0$ for $x\in X_0=\{x\in X:f_1(x)\leqslant 0\}$

and $f_1^+(x) > 0$ for $x \in X_1 = X \setminus X_0$, it easily follows from (7.1) that the exact parameter for the linear penalty function exists if and only if

$$\bar{\lambda} \equiv \sup_{x \in X_1} \frac{M - f(x)}{f_1(x)} < +\infty. \tag{7.2}$$

It is also easy to check that the least exact parameter is equal to $\bar{\lambda}$. Due to (7.1) we can conclude that the exact parameter for the quadratic penalty function exists if and only if

$$\bar{\lambda}_2 \equiv \sup_{x \in X_1} \frac{M - f(x)}{(f_1(x))^2} < +\infty$$

and $\bar{\lambda}_2$ is the least exact parameter. Assume that $f_1(x) \leqslant 1$. Then $\bar{\lambda}_2 \geqslant \bar{\lambda}$. Thus the existence of an exact parameter for the quadratic penalty function implies that for the linear penalty function. The following example demonstrates the case, when the exact parameter for the linear penalty function exists and it does not exist for quadratic penalty function.

EXAMPLE 7.1. Let $X = \mathbb{R}_+$. Consider the problem

$$\min x^2 \text{ subject to } 1 - x \leqslant 0.$$

The value M of this problem is equal to 1. We have

$$\bar{\lambda} = \sup_{1-x>0} \frac{1-x^2}{1-x} = 2 < +\infty, \qquad \bar{\lambda}_2 = \sup_{1-x>0} \frac{1-x^2}{(1-x)^2} = +\infty.$$

Consider now the sharp augmented Lagrangian

$$l(x,(y,r)) = \inf_{z + f_1(x) \leqslant 0} (f(x) - yz + r|z|),$$
$$\omega = (y,r) \in \Omega_1 = \{(y,r) \in \Omega, r \geqslant y\}.$$

Due to Proposition 4.2 we have $l(x,(y,r)) = L^+(x, y+r)$. Thus the exact parameter (y,r) exists if and only if the quantity $\bar{\lambda}$ defined by (7.2) is finite. Note that the inequality $\bar{\lambda} < +\infty$ is a sufficient condition for the existence of an exact parameter, if the augmented Lagrangian with the augmenting function σ, satisfying (6.3), is used for constructing the dual problem.

8. Supergradient Method for Solving the Dual Problems

In this section we first consider a version of a supergradient method for solving the dual of the problem $P^e(f, f_1)$ with the equality constraint function f_1. Then by using the approach given in Section 6 we extend the obtained results to problems with an inequality constraint.

Let $\sigma:\mathbb{R}\to\mathbb{R}_+$ be an augmenting function, such that (6.3) holds. Then

$$\sigma(0)=0, \qquad \sigma(z)\geqslant|z|. \tag{8.1}$$

Consider a problem $P^e(f,f_1)$. Corresponding to σ the augmented Lagrangian l^e and the dual function q^e have the following form, respectively:

$$l^e(x,(y,r))=f(x)+yf_1(x)+r\sigma(f_1(x)), \qquad x\in X,(y,r)\in\Omega; \tag{8.2}$$

$$q^e(y,r)=\inf_{x\in X}(f(x)+yf_1(x)+r\sigma(f_1(x)), \qquad (y,r)\in\Omega. \tag{8.3}$$

Let

$$Q(y,r)=\operatorname{argmin}_{x\in X}(f_0(x)+yf_1(x)+r\sigma(f_1(x))), \text{for } (y,r)\in\Omega. \tag{8.4}$$

The dual function q^e is concave, so the dual problem is a convex program. The following simple result characterizes the solutions of primal and dual problems and allows one to calculate a supergradient of the dual function explicitly.

THEOREM 8.1. *Let $(\bar y,\bar r)\in\Omega$ be a point with nonempty $Q(\bar y,\bar r)$ and let $\bar x\in Q(\bar y,\bar r)$. Then the pair $(f_1(\bar x),\sigma(f_1(\bar x)))$ is a supergradient of the dual function q^e at $(\bar y,\bar r)$. If $\bar r+\bar y>0$ and $f_1(\bar x)=0$ then $\bar x$ is a solution of $P^e(f,f_1)$ and $(\bar y,\bar r)$ is a solution of the dual problem.*

Proof. For all $(y,r)\in\Omega$ we have

$$\begin{aligned} q^e(y,r) &= \min_{x\in X}(f(x)+yf_1(x)+r\sigma(f_1(x))\leqslant f(\bar x)+yf_1(\bar x)+r\sigma(f_1(\bar x)) \\ &= f(\bar x)+\bar y f_1(\bar x)+\bar r\sigma(f_1(\bar x))+(y-y_1)f_1(\bar x)+(r-\bar r)\sigma(f_1(\bar x)) \\ &= \qquad q^e(\bar y,\bar r)+(y-\bar y)f_1(\bar x)+(r-\bar r))\sigma(f_1(\bar x)). \end{aligned}$$

Thus $(f_1(\bar x),\sigma(f_1(\bar x))\in\partial q^e(\bar y,\bar r)$. If $f_1(\bar x)=0$ then $0\in\partial q^e(\bar y,\bar r)$. The inequality $\bar y+\bar r>0$ implies $(\bar y,\bar r)\in\operatorname{int}\Omega$ so $(\bar y,\bar r)$ is a solution of dual problem. Since $q^e(y,r)\leqslant f(x)$ for all x with $f_1(x)=0$ and $f(x)\geqslant M$ for all such x, the equality $f(\bar x)=q^e(\bar y,\bar r)$ implies $f(\bar x)=M$. □

Now by using this theorem, we present the modified supergradient algorithm for calculating the optimal value of the dual problem (D^e):

$$\max\ q^e(y,r) \ \text{ subject to } (y,r)\in\Omega.$$

We assume that the zero duality gap property holds (see Section 7), hence the optimal value of dual problem coincides with the optimal value of the primal one.

At each iteration we check the optimality condition presented in Theorem 7.1; in the case when the point x_k calculated at kth iteration is not optimal, we define the new values of dual variables by moving in the supergradient direction. In Theorem 7.2 below we shall show that the moving in the supergradient direction strongly improves the value of dual function at each iteration. Note that the classical supergradient method constructed by using the ordinary Lagrangian does not possess such a property. Theoretically, it is possible to choose the supergradient which improves the value of dual function, among all supergradients, at each iteration. In the case, when the dual function is nonsmooth, it is difficult to calculate the whole subdifferential at the given point.

ALGORITHM 8.1. Initialization Step. *Let* $y_0 \geqslant 0$, $r_0 \geqslant 0$, $k=0$, *and go to the main step.*

Main Step

(1) *Given* $(y_k, r_k) \in \Omega$, *solve the following subproblem*:

$$\text{minimize } (f(x) + y_k f_1(x) + r_k \sigma(f_1(x))) \text{ subject to } x \in X.$$

Let x_k *be any solution. If* $f_1(x_k)=0$, *then stop*; *by Theorem* 8.1, (y_k, r_k) *is a solution to the dual problem and* x_k *is a solution to* $P^e(f, f_1)$. *Otherwise, go to step* 2.

(2) *Let*

$$y_{k+1} = y_k + s_k f_1(x_k), \quad r_{k+1} = r_k + (s_k + \varepsilon_k)\sigma(f_1(x_k)), \tag{8.5}$$

where $s_k > 0$ *and* $\varepsilon_k > 0$ *are step-size parameters, replace* k *by* $k+1$, *and repeat step* 1.

REMARK 8.1. Since $\sigma(f_1(x_k)) \geqslant |f_1(x_k)|$, at each iteration we have

$$\begin{aligned} y_{k+1} + r_{k+1} = {} & y_k + s_k f_1(x_k) + r_k + (s_k + \varepsilon_k)\sigma(f_1(x_k)) \\ = {} & y_k + r_k + s_k(\sigma(f_1(x_k)) + f_1(x_k)) + \varepsilon_k \sigma(f_1(x_k)) \\ \geqslant {} & y_k + r_k + s_k(|f_1(x_k)| + f_1(x_k)) + \varepsilon_k |f_1(x_k)| \\ \geqslant {} & y_k + r_k, \end{aligned} \tag{8.6}$$

which implies that the new pair of dual variables (y_{k+1}, r_{k+1}) calculated by the algorithm at $(k+1)th$ iterate, also belongs to Ω.

The following theorem shows that in contrast with the supergradient methods developed for dual problems formulated by using ordinary Lagrangians, the new iterate strictly improves the value of the objective function q^e for all $s_k > 0$ and $\varepsilon_k > 0$.

THEOREM 8.2. *Consider the problem* $P^e(f, f_1)$ *with* $|f_1(x)| \leqslant 1$. *Let* $\sigma: \mathbb{R}_+ \to \mathbb{R}_+$ *be a function with properties* (7.1). *Assume the set* $Q(y, r)$ *is nonempty for each* $(y, r) \in \Omega$. *Consider a point* $(y_k, r_k) \in \Omega$ *that is not a solution of the dual problem and let* $x_k \in Q(y_k, r_k)$. *Then for a point* (y_{k+1}, r_{k+1}) *calculated from* (7.5) *for any positive step-size parameters* s_k *and* ε_k *we have*:

$$0 < q^e(y_{k+1}, r_{k+1}) - q^e(y_k, r_k) \leqslant s_k[f_1(x_k)]^2 + (s_k + \varepsilon_k)[\sigma(f_1(x_k))]^2. \tag{8.7}$$

Proof. Let $(y_k, r_k) \in \Omega$. Consider $l^e(x,(y_k,r_k)) = f(x) + y_k f_1(x) + r_k \sigma(f_1(x))$. Applying (7.1) we conclude that

$$\begin{aligned} q^e(y_{k+1}, r_{k+1}) &= \min_{x\in X}(l^e(x,(y_k,r_k)) + (s_k+\varepsilon_k)\sigma(f_1(x_k))\sigma(f_1(x)) + \qquad (8.8)\\ &\qquad s_k f_1(x_k) f_1(x)) \\ &\geqslant \min_{x\in X}(l^e(x,(y_k,r_k)) + (s_k+\varepsilon_k)\sigma(f_1(x_k))\sigma(f_1(x)) - \qquad (8.9)\\ &\qquad s_k |f_1(x_k)||f_1(x)|) \\ &\geqslant \min_{x\in X}(f(x) + y_k f_1(x) + (r_k + \varepsilon_k \sigma(f_1(x_k)))\sigma(f_1(x)) \\ &= q^e(y_k, r_k + \varepsilon_k\sigma(f_1(x_k))). \qquad (8.10) \end{aligned}$$

Let $\widetilde{x} \in Q(y_k, r_k + \varepsilon_k \sigma(f_1(x_k))$. If $f_1(\widetilde{x}) = 0$, then by Theorem 8.1, the pair $(y_k, r_k + \varepsilon_k\sigma(f_1(x_k))$ would be a solution to the dual problem. It follows from the inequality $q^e(y_{k+1}, r_{k+1}) \geqslant q^e(y_k, r_k + \varepsilon_k\sigma(f_1(x_k))$, that (y_{k+1}, r_{k+1}) is also a solution. Since (y_k, r_k) is not a solution, we conclude that $q^e(y_k, r_k) < q^e(y_{k+1}, r_{k+1})$.

When $f_1(\widetilde{x}) \neq 0$, we have $\sigma(f_1(\widetilde{x})) > 0$. Since (y_k, r_k) is not a solution of dual problem, we can apply Theorem 8.1, by which $f_1(x_k) \neq 0$. Hence $\sigma(f_1(x_k)) > 0$. We have

$$\begin{aligned} q^e(y_{k+1}, r_{k+1}) &\geqslant f(\widetilde{x}) + y_k f_1(\widetilde{x}) + (r_k + \varepsilon_k\sigma(f_1(x_k))\sigma(f_1(\widetilde{x})) \\ &> f(\widetilde{x}) + y_k f_1(\widetilde{x}) + r_k\sigma(f_1(\widetilde{x})) \geqslant q^e(y_k, r_k). \end{aligned}$$

Thus the left-hand side of (8.7) has been proved. We now prove the right-hand side. Let $x_k \in Q(y_k, r_k)$. Since the pair $(f_1(x_k), \sigma(f_1(x_k))$ is a supergradient of a concave function q at (y_k, r_k), we have

$$\begin{aligned} & q^e(y_{k+1}, r_{k+1}) - q^e(y_k, r_k) \\ & \leqslant (y_{k+1} - y_k) f_1(x_k) + (r_{k+1} - r_k)\sigma(f_1(x_k)) \\ & = s_k [f_1(x_k)]^2 + (s_k + \varepsilon_k)[\sigma(f_1(x_k))]^2. \end{aligned}$$

The theorem is proved.

REMARK 8.2. For discussion of condition (8.1) see Section 6.

REMARK 8.3. The algorithm presented above is a generalization of the modified supergradient method suggested by Gasimov [1] for solving the dual problems constructed via sharp augmented Lagrangian for problems with equality constraints. The convergence of this method has been proved in [1] under a certain choice of parameters s_k and ε_k. The following two theorems demonstrate that the similar assertions are true also for the supergradient algorithm presented in this paper. Because of the similarity of proofs of these theorems to those from [1], we present the following two theorems without proofs.

THEOREM 8.3. *Assume that* $\sigma(z) \geqslant |z|$, *for all* $z \in R$. *Let* (y_k, r_k) *be any iteration generated by the algorithm and let* $x_k \in Q(y_k, r_k)$. *Suppose that* (y_k, r_k) *is not a solution to the dual problem, so* $f_1(x_k) \neq 0$. *Then for any dual solution* $(\overline{y}, \overline{r})$, *we have*

$$\|(\overline{y}, \overline{r}) - (y_{k+1}, r_{k+1})\| < \|(\overline{y}, \overline{r}) - (y_k, r_k)\|$$

for all step-sizes s_k *such that*

$$0 < s_k < \frac{2[q^e(\overline{y}, \overline{r}) - q^e(y_k, r_k)]}{[f_1(x_k)]^2 + 4[\sigma(f_1(x_k))]^2}, \tag{8.11}$$

and $0 < \varepsilon_k < s_k$.

THEOREM 8.4. *Assume that the constraint function* f_1 *satisfies the boundedness condition* $|f_1(x)| \leqslant 1$ *and the augmenting function* σ *is continuous. Let* (y_k, r_k) *be any iteration of the supergradient method. Suppose that each new iteration* (y_{k+1}, r_{k+1}) *calculated from* (8.5) *for the step-size*

$$s_k = \frac{\overline{q} - q_k}{c_k} \text{ and } 0 < \varepsilon_k < s_k,$$

where $\overline{q} = q^e(\overline{y}, \overline{r})$ *denotes the optimal dual value,* $q_k = q^e(y_k, r_k)$ *and* $c_k = f_1^2(x_k) + 4\sigma^2(f_1(x_k))$. *Then* $q_k \to \overline{q}$.

The method under consideration can also be applied to problems $P(f, f_1)$ with an inequality constraint. We can use the technique described in Section 6.

Consider the problem $P(f, f_1)$. Let σ be an augmenting function such that (6.3) holds and let l^e be the augmented Lagrangian for the problem $P^e(f, f_1)$ corresponding to σ:

$$l^e(x, (y, r)) = f(x) + y f_1(x) + r\sigma(f_1(x)), \qquad x \in X, \ (y, r) \in \Omega.$$

Let q^e be the corresponding dual function. We assume that q^e is defined only on the set $\Omega_1 = \{(y, r) : y + r \geqslant 0, r \geqslant y\}$. Let L be the nonlinear Lagrangian defined by (6.2) and let q be the corresponding dual function. Due to Proposition 6.1 we have

$$q(y, r) = \min(M, q^e(y, r)), \qquad (y, r) \in \Omega_1.$$

The function q^e is concave and upper semicontinuous as the infimum of a set of affine functions $(y, r) \mapsto f(x) + y f_1(x) + r\sigma(f_1(x))$. We need the following result.

PROPOSITION 8.1. *Assume that the pair* $(\overline{y}, \overline{r}) \in \Omega_1 = \{(y, r) : -r \leqslant y \leqslant r, r \geqslant 0\}$ *is not a solution of the dual problem and that the set* $Q(\bar{y}, \bar{r})$, *defined by* (8.4), *is not empty. Let* $\bar{x} \in Q(\bar{y}, \bar{r})$. *Then the pair* $(f_1(\bar{x})), \sigma(f_1(\bar{x})))$ *is a supergradient of the dual function* q *at* $(\bar{y}, \bar{r})$.

Proof. By Proposition 6.1 we have:

$$q(y,r) \leqslant q^e(y,r), \text{ for all } (y,r) \in \Omega_1.$$

By Theorem 7.1,

$$q^e(y,r) \leqslant q^e(\overline{y},\overline{r}) + (y-\overline{y}) f_1(\overline{x}) + (r-\overline{r})\sigma(f_1(\overline{x})).$$

Since $(\overline{y},\overline{r})$ is not a solution, again, by Proposition 6.1 we have $q(\overline{y},\overline{r}) = q^e(\overline{y},\overline{r})$.

REMARK 8.4. For problems with the active constraint the above inequalities can be written for all $(y,r) \in \Omega$.

REMARK 8.5. In the case that the set $Q(y,r)$ is not empty for all $(y,r) \in \Omega_1$, we can apply the algorithm described above for maximization of the function q. If $(y_0,r_0) \in \Omega_1$, then $(y_k,r_k) \in \Omega_1$ for all k. Assume that (y_k,r_k) is not a solution of the dual problem and (y_{k+1},r_{k+1}) is generated by the algorithm: $y_{k+1} = y_k + s_k f_1(x_k), r_{k+1} = r_k + (s_k + \varepsilon_k)\sigma(f_1(x_k))$, where $x_k \in Q(y_k,r_k)$. Then it follows from (8.10) that

$$q(y_{k+1},r_{k+1}) \geqslant q(y_k, r_k + \varepsilon_k \sigma(f_1(x_k)).$$

Assume that $q_(y_{k+1},r_{k+1})$ is not a solution of dual problem. Then (see Lemma 6.1(1)),

$$q((y_k, r_k + \varepsilon_k \sigma(f_1(x_k)) = q^e((y_k, r_k + \varepsilon_k \sigma(f_1(x_k)).$$

Let $\widetilde{x} \in Q(y_k, r_k + \varepsilon_k \sigma(f_1(x_k))$. It follows from Lemma 6.1(2) that $\widetilde{x} \notin X_0$ so $f_1(\widetilde{x}) > 0$. Since (y_k,r_k) is not a solution of the dual problem, we have $f_1(x_k) > 0$. The same argument as in the proof of Theorem 8.2 shows that $q(y_{k+1},r_{k+1}) > q(y_k,r_k)$. If (y_{k+1},r_{k+1}) is a solution of dual problem, we also have $q(y_{k+1},r_{k+1}) > q(y_k,r_k)$.

REMARK 8.6. Some numerical methods for unconstrained global Lipschitz optimization have recently been developed, (for example, the cutting angle method, see [7], Ch. 9 and references therein). These methods can be used for solving the subproblem at the Main Step 1. The proposed supergradient algorithm allow us to use these methods for solving constraint problems.

9. Acknowledgement

The authors would like to thank Professor L.Qi and two anonymous referees for their comments which substantially improved the quality of the paper.

References

1. Gasimov, R. N. (2002), Augmented Lagrangian duality and nondifferentiable optimization methods in nonconvex programming, *Journal of Global Optimization* 24, 187–203.

2. Giannessi, F. (1984), Theorems of the alternative and optimality conditions, *J. Optimiz. Theory Appl.* 42, 331–365.
3. Huang, X.X. and Yang, X.Q. A unified augmented Lagrangian approach to duality and exact penalization, submitted paper.
4. Rockafellar, R.T. (1974), Augmented Lagrange multiplier functions and duality in nonconvex programming, *SIAM Journal on Control and Optimization*, 12, 268–285.
5. Rockafellar, R.T. (1993), Lagrange multipliers and optimality, *SIAM Review*, 35, 183–238.
6. Rockafellar, R. T. and Wets, J.-B. (1998), *Variational Analysis*, Springer, Berlin, Heidelberg.
7. Rubinov, A. M. (2000), *Abstract Convexity and Global Optimization*, Kluwer Academic Publishers, Dordrecht.
8. Rubinov, A.M. and Uderzo, A. (2001), On global optimality conditions via separation functions, J. Optimization Theory and Applications, 109, 345–370.
9. Rubinov, A. M., Glover, B.M. and Yang, X.Q. (1999), Decreasing functions with application to penalization, *SIAM Journal Optimization*, 10, 289–313.
10. Rubinov, A. M., Glover, B. M. and Yang, X. Q. (1999) Extended Lagrange and penalty functions in continuous optimization, *Optimization*, 46, 326–351.
11. Rubinov, A.M. and Yang, X.Q. *Lagrange-type functions in constrained non-convex optimization*, Kluwer Academic Publishers, Dordrecht, (in press).
12. Yang, X.Q. and Huang, X.X. (2001), A nonlinear Lagrangian approach to constrained optimization problems, *SIAM Journal Optimization,* 14, 1119–1144.