Journal of Global Optimization **27:** 473–484, 2003.

Nonlinear Lagrange Duality Theorems and Penalty Function Methods In Continuous Optimization

C. Y. WANG[1], X. Q. YANG[2] and X. M. YANG[3]
[1]*Department of Mathematics, Qufu Normal University, Qufu, P. R. China;* [2]*Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong, P.R. China;* [3]*Department of Mathematics, Chongqing Normal University, Chongqing, P. R. China*

(Received 6 September 2002; accepted in revised form 22 April 2003)

Abstract. We propose a general dual program for a constrained optimization problem via generalized nonlinear Lagrangian functions. Our dual program includes a class of general dual programs with explicit structures as special cases. Duality theorems with the zero duality gap are proved under very general assumptions and several important corollaries which include some known results are given. Using dual functions as penalty functions, we also establish that a sequence of approximate optimal solutions of the penalty function converges to the optimal solution of the original optimization problem.

Key words: Dual program, Nonlinear Lagrangian function, ϵ-optimal solution, penalty function, zero duality gap

1. Introduction

Continuous optimization problems have various applications in economics, engineering and management science. Consequently considerable attention has been devoted to the study of theory and efficient numerical methods for optimization problems. Based on the zero duality gap property between the primal convex optimization problem and its linear Lagrangian dual problem, some important algorithms have been proposed under regular assumptions (see, e.g., [2, 3, 8, 10]). For a generalized convex optimization problem and its linear Lagrangian dual program, the zero duality gap property has been widely studied and a number of important results have been obtained (see, e.g., [7, 4, 11]). For various Lagrangian methods, the zero duality gap property is essential. However, for nonconvex optimization problems, many examples demonstrate that nonzero duality gap exists if the dual problem is constructed with a linear Lagrangian function. Thus, it is difficult to develop effective algorithms for nonconvex programs via linear Lagrangian functions, which leads many researchers to construct nonlinear Lagrangian dual programs for nonconvex optimization problems (see, e.g., [15, 6, 5, 12, 13, 1, 9]). In [15, 6], nonlinear Lagrangian functions are constructed by maximizing the convolution function of the objective function and the constrained functions. In [5, 12, 13, 1], nonlinear Lagrangian functions are formulated via increasing functions or increasing and positively homogenuous functions. In [9] more general nonlinear Lagrangian functions, which include the above mentioned nonlinear Lagrangian functions as special cases, are proposed and relevant dual programs are studied. What is worth noting is that the zero duality property studied in [15, 6, 5, 12, 13, 1, 9] was proved under very strong conditions.

This paper improves and generalizes the nonlinear Lagrangian functions in [9]. Under very general assumptions, the zero duality property is established, and some important corollaries are also obtained. These corollaries include the corresponding results in [5, 12, 13, 1, 9] as special cases. Furthermore, using the dual functions as penalty functions, we prove that the ϵ-optimal solution of the penalty function approaches a solution of the original problem as $\epsilon \to 0$, which provides a theoretical foundation for the development of algorithms with a global convergence property. Finally, some special classes of nonlinear functions Lagrangian are studied to verify our above results.

2. Generalized Lagrangian Functions and Dual Programs

Consider the following general nonlinear optimization problem:

$$\text{(P)} \quad \min\{f_0(x) | x \in X_0 \subset X\},$$

where X is a metric space, $X_0 \subset X$ is a nonempty closed subset which denotes the feasible solution set of (P), and $f_0 : X \longrightarrow \mathbb{R}$ is a continuous function.

A dual for (P) can be constructed as follows: given a nonempty set R, the space of multipliers. Then a generalized Lagrangian for problem (P) is any function $L : X \times R \longrightarrow \mathbb{R}$ such that

$$\inf_{x \in X_0} f_0(x) = \inf_{x \in X} \sup_{d \in R} L(x, d). \tag{1}$$

The dual of (P) is defined as

$$\text{(D)} \quad \max\{q(d) | d \in R\},$$

where

$$q(d) := \inf_{x \in X} L(x, d), \quad d \in R.$$

Since, obviously,

$$\inf_{x \in X} \sup_{d \in R} L(x, d) \geqslant \sup_{d \in R} \inf_{x \in X} L(x, d),$$

it follows that

$$M_P := \inf_{x \in X_0} f_0(x) \geqslant \sup_{d \in R} q(d) =: M_D.$$

M_P is the primal optimal value, and M_D is the dual optimal value. And $M_P - M_D \geqslant 0$ is the duality gap. If $M_P = M_D$, (P) and (D) are said to have a zero duality gap. Since $M_P = -\infty$ implies $M_D = -\infty$, we always assumption that $M_P > 0$ in the sequel.

Now we make some hypotheses for $f_0(x)$ and $L(x, d)$. In Section 3, we will see that the zero duality gap is true under these hypotheses.

Let $\psi : X \longrightarrow \mathbb{R}$ be a continuous function such that

$$x \in X_0 \Longleftrightarrow \psi(x) \leqslant 0;$$

$$\overline{X}^* = \{x \in X_0 | f_0(x) = M_P\}, \text{ i.e., } \overline{X}^* \text{ is the set of optimal solutions of (P);}$$

$$\overline{X}^{*\delta} = \{x \in X | M_P - \delta < f_0(x) < M_P + \delta\};$$

$$\overline{X}_\psi^{<\delta} = \{x \in X | \psi(x) < \delta\};$$

$$\overline{X}_\psi^{\geqslant\delta} = \{x \in X | \psi(x) \geqslant \delta\};$$

$B(x_0, \rho) = \{x \in X | \inf_{\xi \in X_0} d(x, \xi) < \rho\}$, where $d(\cdot, \cdot)$ is the metric in X.

Some hypotheses are given as follows:

(H_1) $f_0(x)$ be uniformly continuous on an open set G containing X_0;

(H_2) for all $x \in X_0$, for all $d \in R$, $L(x, d) = f_0(x)$;

(H_3) for all $\epsilon > 0$, there exists an open set $G(\epsilon)$ containing X_0, such that

$$L(x, d) \geqslant f_0(x) - \epsilon, \quad \forall x \in G(\epsilon), \forall d \in R;$$

(H_4) for every $\gamma > 0$, there exists $\overline{d} \in R$ such that

$$L(x, \overline{d}) \geqslant \gamma \psi(x), \quad \forall x \in X \backslash X_0;$$

(H_5) the point-to-set map $\overline{X}_\psi^{<\delta}$ is upper semi continuous at $\delta = 0$.

The following are some results on weak duality and saddle point, whose proofs are elementary and omitted.

THEOREM 2.1. *(Weak duality). Assume that (H_2) and (H_4) hold. Then*

$$M_P \geqslant M_D.$$

THEOREM 2.2. *Assume that (H_2) and (H_4) hold. If (x^*, d^*) is a saddle point of $L(x, d)$, then x^* is an optimal solution of problem (P).*

3. Zero Duality Gap

Zero duality gap and weak duality provide a basic characterization of dual programs. Specially, zero duality gap is a basis for designing various efficient algorithms. Now we prove zero duality gap for (P) and (D).

THEOREM 3.1. *Suppose that* $(H_1) - (H_5)$ *hold, then* $M_P = M_D$.

Proof. By Theorem 2.1, $M_P \geqslant M_D$. We only need to prove $M_P \leqslant M_D$. For any $\epsilon > 0$, we have

$$M_P - \epsilon < M_P \leqslant f_0(x).$$

By (H_1), $\rho_0 > 0$ can be taken sufficiently small such that $B(X_0, \rho_0) \subseteq G$ and for any $x \in B(X_0, p_0)$

$$f_0(x) > M_P - 2\epsilon \tag{2}$$

holds. It follows from (H_3) that there exists an open set $G(\epsilon) \supseteq X_0$ such that, for any $x \in G(\epsilon)$, $d \in R$, there holds

$$L(x, d) \geqslant f_0(x) - \epsilon. \tag{3}$$

We assume, without loss of generality, that $G(\epsilon) \subseteq B(X_0, \rho_0)$ (otherwise $G(\epsilon) \bigcap B$ (X_0, ρ_0) can be used to replace $G(\epsilon)$). It follows from (2) and (3) that

$$L(x, d) \geqslant M_P - 3\epsilon \tag{4}$$

holds for any $\epsilon \in G(\epsilon)$, $d \in R$.

By (H_5), $\bar{X}_{\psi}^{<\delta}$ is upper semi-continuous at $\delta = 0$. Let δ_0 be sufficiently small so that

$$\bar{X}_{\psi}^{<\delta_0} \subseteq G(\epsilon) \quad \text{or } X\backslash G(\epsilon) \subseteq \bar{X}_{\psi}^{\geqslant\delta_0} \tag{5}$$

By (H_4), for $2M/\delta_0$, there exists $\bar{d} \in R$ such that, for any $x \in X\backslash X_0$, there holds

$$L(x, \bar{d}) \geqslant \frac{2M_P}{\delta_0}\psi(x) \tag{6}$$

The combination of (4), (5) and (6) yields

$$\begin{aligned}
M_D &= \sup_{d\in R} \inf_{x\in X} L(x, d) \geqslant \inf_{x\in X} L(x, \bar{d}) \\
&= \min\{\inf_{x\in G(\epsilon)} L(x, \bar{d}), \inf_{x\in X\backslash G(\epsilon)} L(x, \bar{d})\} \\
&\geqslant \min\{\inf_{x\in G(\epsilon)} L(x, \bar{d}), \inf_{x\in \bar{X}_{\psi}^{\geqslant\delta_0}} L(x, \bar{d})\} \\
&\geqslant \min\{M_P - 3\epsilon, \tfrac{2M_P}{\delta_0} \inf_{x\in \bar{X}_{\psi}^{\geqslant\delta_0}} \psi(x)\} \\
&\geqslant \min\{M_P - 3\epsilon, 2M_P\} = M_P - 3\epsilon.
\end{aligned}$$

Since $\epsilon > 0$ is arbitrary, we have

$$M_D \geqslant M_P. \qquad \square$$

From the proof of Theorem 3.1, function $\psi(x)$ in (H_4) and upper semi-continuity assumption of the level set for $\psi(x)$ in (H_5) are important conditions for the zero duality gap. But, in fact, it is difficult to verify (H_5). We will give some class of functions which satisfy (H_5) and several Corollaries of Theorem 3.1.

DEFINITION 3.1. The function $g : X \longrightarrow \mathbb{R}$ is said to have property (A) at $x_0 \in X' \subset X$, if, for sufficiency small $\rho > 0$, there exists $M > 0$, for any $x \in X \backslash B(X', \rho)$, there exists $y \in X \backslash B(X', \rho)$, such that

(i) $d(x_0, y) \leqslant M$,

(ii) $g(y) \leqslant \max\{g(x_0), g(x)\}$.

DEFINITION 3.2. The function $g : X \longrightarrow \mathbb{R}$ is said to have property (B) at $x_0 \in X' \subset X$, if there exists a finite $M > 0$, for any $x \in X \backslash X'$, there exists sequence $y_0(= x), y_1, \cdots, y_k, y_{k+1}(= x_0) \in X$, such that

(i) $\max\{d(y_i, y_{i+1}) | 0 \leqslant i \leqslant k\} \leqslant M$,

(ii) $\max\{g(y_i) | 1 \leqslant i \leqslant k\} \leqslant \max\{g(x_0, g(x)\}$.

It is obvious that any function g which is defined on a compact set has property (A) and property (B). If g is quasi convex, psedoconvex or arcwise quasiconvex at some $x_0 \in X'$, it is obvious that g has property (B).

Let $X' = \{x \in X | g(x) \leqslant \alpha\}$. Recall that $\overline{X}_g^{<\delta} = \{x \in X | g(x) < \delta\}$.

PROPOSITION 3.1. *Assume that $g(x)$ is continuous and has property (A). Then $\overline{X}_g^{<\delta}$ is upper continuous at $\delta = \alpha^+$ (from the right).*

Proof. If $\overline{X}_g^{<\delta}$ is not upper continuous at $\delta = \alpha^+$, then there exist sufficiently small $\rho_0 > 0$, $\delta_k \longrightarrow \alpha^+$ $(k \longrightarrow \infty)$, and $x_k \in \overline{X}_g^{<\delta_k}$, such that

$$x_k \notin B(X', \rho_0).$$

By definition 3.1, there exist $M > 0$ and $y_k \in X \backslash B(X', \rho)$

(i) $d(x_0, y_k) \leqslant M$,

(ii) $g(y_k) \leqslant \max\{g(x_0), g(x_k)\}$.

(i) yields that $\{y_k\}$ is bounded, hence there exist an infinite subset $N_0 \subset N = \{1, 2, \cdots, \}$, such that

$$\{y_k\}_{k \in N_0} \longrightarrow y^*.$$

From $x_0 \in X'$, $x_k \in \overline{X}_g^{<\delta_k}$ and $\delta_k > \alpha$, (ii) yields

$$g(y_k) \leqslant \max\{g(x_0), g(x_k)\} \leqslant \max\{\alpha, \delta_k\} = \delta_k.$$

By the continuity of g and noting that $\delta_k \longrightarrow \alpha^+$, we have

$$g(y^*) \leqslant \alpha.$$

Hence, $y^* \in X' \subset B(X', \rho)$. This is a contradiction.

PROPOSITION 3.2. *Let X' be compact and $g(x)$ be continuous and have property (B). Then $\overline{X}_g^{<\delta}$ is upper continuous at $\delta = \alpha^+$.*

Proof. From the boundedness of X', we have that

property (B) implies property (A).

By Proposition 3.1, we know that Proposition 3.2 is true. □

COROLLARY 3.1. *Assume that (H_1)-(H_4) hold and $\psi(x)$ has property (A) at $x_0 \in X_0$. Then $M_P = M_D$.*

Let $U, V \subset X$. Define by

$$U \oplus V = \{z \in X | z = x + y, x \in U, y \in V\}.$$

COROLLARY 3.2. *Assume that (H_1)-(H_4) hold, $\psi(x)$ is a concave function on X and satisfies*

$$\{x \in X | \psi(x) \geqslant 0\} = U \oplus V,$$

where U is a nonempty compact convex set, V is a nonempty closed convex cone. Then $M_P = M_D$.

Proof. By Corollary 3.1, it is sufficient to show that $\psi(x)$ has property (A). Since X is a closed set and X_0 is a closed set, $X \backslash X_0$ is an open set . Taking $\rho > 0$ sufficiently small such that $U \backslash B(X_0, \rho) \neq \emptyset$, then $X \backslash B(X_0, \rho)$ is a closed convex set (since $B(X_0, \rho)$ is an open set). From the conditions, we know that there exist a compact convex set U' and a closed convex cone V' such that

$$X \backslash B(X_0, \rho) = U' \oplus V'. \tag{7}$$

Noting that $\psi(x) \geqslant 0 (x \in X \backslash X_0)$, i.e., $\psi(x)$ has a lower bound on $X \backslash X_0$. It follows from the results in [14] that

$$\psi(y) \leqslant \psi(y + \lambda z), \quad \forall y \in U', \forall z \in V' \text{ and } \lambda \geqslant 0. \tag{8}$$

Hence, for $x \in X \backslash B(X_0, \rho)$, by (8), we know that there exist $y \in U'$, $z \in V'$ and $\lambda_0 \geqslant 0$, such that $x = y + \lambda_0 z$. By (9), we obtain

$$\psi(y) \leqslant \psi(x).$$

From Boundness of U', we see that $\psi(x)$ have property (A) for any $x_0 \in X_0$. □

COROLLARY 3.3. *Suppose that $(H_1) - (H_4)$ hold, X_0 is bound and $\psi(x)$ have property (B) at $x_0 \in X_0$. Then $M_P = M_D$.*

COROLLARY 3.4. *Suppose that $(H_1) - (H_4)$ hold, X_0 is bound and $\psi(x)$ is quasiconvex or pseudoconvex or arcwise quasiconvex at $x_0 \in X_0$. Then $M_P = M_D$.*

COROLLARY 3.5. *If* $(H_1) - (H_4)$ *hold and* X *is a compact set, then* $M_P = M_D$.

REMARK . Corollary 3.5 is the main result in [9] (see. Theorem 1). If (H_3) is instead of the follow assumption

$(H_3)'$ $\forall x \in X\backslash X_0$, $\forall d \in \mathbb{R}$, there holds $L(x, d) \geqslant f_0(x)$. Then our Lagrangian model is corresponding model in [5, 12, 13] and [9]. Hence, based on Corollary 3.5, we have follow Corollary.

COROLLARY 3.6. *If* $(H_1), (H_2), (H_3)', (H_4)$ *hold and there exists* $\alpha > M_P$ *such that level set* $\overline{X}_{f_0}^{\leqslant\alpha} = \{x \in X | f_0(x) \leqslant \alpha\}$ *has bound. Then* $M_P = M_D$.

Proof. From $(H_3)'$, we have

$$\inf_{x \in X\backslash\overline{X}_{f_0}^{\leqslant\alpha}} L(x, d) \geqslant \inf_{x \in X\backslash\overline{X}_{f_0}^{\leqslant\alpha}} f_0(x) \geqslant \alpha > M_P$$

$M_P \geqslant M_D$ yields

$$\inf_{x \in X} L(x, d) = \inf_{x \in \overline{X}_{f_0}^{\leqslant\alpha}} L(x, d), \forall d \in \mathbb{R}.$$

Hence, replacing X by $\overline{X}_{f_0}^{\leqslant\alpha}$ and applying Corollary 3.5, the Corollary follows. □

REMARK . *Corollary 3.6 is the main results in [5, 13] (see Theorem 4.2 and Theorem 3.1).*

4. Penalty Function Methods

In this section, we use the dual function $q(d)$ of (P), for designing a theoretical algorithm, and we also prove that the algorithm is globally convergent.

DEFINITION 4.1. *Let* $\epsilon > 0$ *and* $d \in R$. $x^*(d, \epsilon) \in X$ *is said to be an* ϵ*-optimal solution for the following problem* $(D(d))$*:*

$$q(d) = \inf\{L(x, d) | x \in X\},$$

if $L(x^*(d, \epsilon), d) \leqslant q(d) + \epsilon$.

From $L(x, d) \geqslant 0$ $(\forall x \in X, \forall d \in R)$*, it is obvious that for each* $d \in \mathbb{R}$, $(D(d))$ *always has an* ϵ*-optimal solution.*

Algorithm (A). Let $M \geqslant 2M_P + 1$, δ_1, $\epsilon_k \in (0, 1)$, $r_k \geqslant 1$, and $\lim_{k\longrightarrow\infty} \epsilon_k = 0$, $\lim_{k\longrightarrow\infty} r_k = \infty$.

Step 1. Let $k := 1$, and find $d_k \in R$ such that

$$\frac{2Mr_k}{\delta_k}\psi(x) \leqslant L(x, d_k), \quad \forall x \in X\backslash X_0.$$

Step 2. Solving problem $(D(d_k))$, let $x^*(d_k, \epsilon_k)$ be an ϵ-optimal solution.
Step 3. If $\psi(x^*(d_k, \epsilon_k)) \leqslant 0$, stop; otherwise, go to step 4.
Step 4. Taking $\delta_{k+1} = \psi(x^*(d_k, \epsilon_k))$, let $k := k + 1$, and return to step 1.
Each step of Algorithm (A) is well defined under assumptions (H_2) and (H_4).

LEMMA 4.1. *Assume that (H_2) and (H_4) hold. If algorithm (A) stops at kth iteration, then $x^*(d_k, \epsilon_k)$ is an ϵ_k-optimal solution for (P).*

Proof. Suppose that algorithm (A) stops at kth iteration. Then

$$\psi(x^*(d_k, \epsilon_k)) \leqslant 0,$$

i.e., $x^*(d_k, \epsilon_k) \in X^0$. Thus, from assumption (H_2),

$$\begin{aligned} M_P &\leqslant f_0(x^*(d_k, \epsilon_k)) = L(x^*(d_k, \epsilon_k), d_k) \\ &= \inf_{x \in X} L(x, d_k) + \epsilon_k \leqslant \inf_{x \in X_0} L(x, d_k) + \epsilon_k \\ &= M_P + \epsilon_k. \end{aligned}$$

□

LEMMA 4.2. *Assume that (H_2) and (H_4) hold. If the sequence $\{x^*(d_k, \epsilon_k)\}$ is produced by algorithm (A), then*
(a) $\psi(x^*(d_k, \epsilon_k))$ *is a decreasing sequence;*
(b) $\lim_{k \longrightarrow \infty} r_k \psi(x^*(d_k, \epsilon_k)) = 0.$

Proof. First we prove

$$r_k \psi(x^*(d_k, \epsilon_k)) < \delta_k, \forall k. \tag{9}$$

Otherwise, there exists k_0 such that

$$r_{k_0} \psi(x^*(d_{k_0}, \epsilon_{k_0})) \geqslant \delta_{k_0}.$$

From step 1 and assumption (H_2), we have

$$\begin{aligned} 2M_P + 1 &\leqslant \frac{2Mr_{k_0}}{\delta_{k_0}} \psi(x^*(d_{k_0}, \epsilon_{k_0})) \leqslant L(x^*(d_{k_0}, \epsilon_{k_0}), d_k) \\ &\leqslant \inf_{x \in X} L(x, d_{k_0}) + \epsilon_k \leqslant M_P + 1. \end{aligned}$$

This is a contradiction. Hence, step 4 and inequality (10) yield

$$r_{k+1} \psi(x^*(d_{k+1}, \epsilon_{k+1})) < \delta_{k+1} = \psi(x^*(d_k, \epsilon_k)). \tag{10}$$

By inequality (11) and $r_{k+1} \geqslant 1$, we know that $\{\psi(x^*(d_k, \epsilon_k))\}$ is decreasing sequence. Thus, $\{\psi(x^*(d_k, \epsilon_k))\}$ is a bounded set. Let C be an upper bound of $\{\psi(x^*(d_k, \epsilon_k))\}$. By inequality (11), we have

$$0 \leqslant \lim_{k \longrightarrow \infty} \psi(x^*(d_k, \epsilon_k)) \leqslant \lim_{k \longrightarrow \infty} \frac{C}{r_k} = 0.$$

Thus

$$0 \leqslant \lim_{k \longrightarrow \infty} \sup r_k \psi(x^*(d_k, \epsilon_k)) \leqslant \lim_{k \longrightarrow \infty} \psi(x^*(d_{k-1}, \epsilon_{k-1})) = 0.$$

So (b) holds. □

Now we give a global convergence result of algorithm (A) (see Theorems 4.1 -4.3).

THEOREM 4.1. *Assume that (H_2) -(H_4) hold and that the sequence $\{x^*(d_k, \epsilon_k)\}$ is produced by algorithm (A). If algorithm (A) stops at kth iteration, then $x^*(d_k, \epsilon_k)$ is an ϵ_k-optimal solution for (P). Otherwise the infinite sequence $\{x^*(d_k, \epsilon_k)\}$ produced by algorithm (A) satisfies*

(a) $\{\psi(x^*(d_k, \epsilon_k))\}$ is a decreasing sequence;

(b) $\lim_{k \longrightarrow \infty} r_k \psi(x^*(d_k, \epsilon_k)) = 0$;

(c) each cluster point x^* of $\{(x^*(d_k, \epsilon))\}$ belongs to X^*.

Proof. From Lemmas 4.1 and 4.2, it is sufficient to show that (c) is true. Suppose that there exists an infinite subset $N_0 \subset N$, such that $\{x^*(d_k, \epsilon_k)\}_{k \in N_0} \longrightarrow x^*$. Then, from (b) and the continuity of $\psi(x)$, we have $\psi(x^*) \leqslant 0$, i.e., $x^* \in X_0$. By assumption (H_3), for $\epsilon_k > 0$, there exists an open set $G(\epsilon_k) \supseteq X_0$, such that

$$L(x, d) \geqslant f_0(x) - \epsilon_k, \qquad \forall x \in G(\epsilon_k),\ \forall d \in \mathbb{R}. \tag{11}$$

Thus, there exists k_0, such that

$$x^*(d_k, \epsilon_k) \in G(\epsilon_k), \quad \forall k \geqslant k_0$$

Inequality (12) yields

$$f_0(x^*(d_k, \epsilon_k)) - \epsilon_k \leqslant L(x^*(d_k, \epsilon_k), d_k) \leqslant \inf_{x \in X} L(x, d_k) + \epsilon_k \leqslant M_P + \epsilon_k.$$

Let $k \longrightarrow \infty$, $\epsilon_k \longrightarrow 0$ and by the continuity of f_0, we obtain

$$f_0(x^*) \leqslant M_P.$$

Thus $x^* \in X^*$. □

We will further give some convergence property about algorithm (A).

LEMMA 4.3. *Assume that (H_1) -(H_5) hold. If the sequence $\{x^*(d_k, \epsilon_k)\}$ is produced by algorithm (A), then for any $\delta > 0$, there exists k_0, such that*

$$x^*(d_k, \epsilon_k) \in \overline{X}^*(\delta), \qquad \forall k \geqslant k_0.$$

Proof. For any $\epsilon > 0$, from (3)-(5), there exists an open set $G(\epsilon) \supseteq X_0$, such that

$$M_P - 2\epsilon \leqslant f_0(x) \leqslant L(x, d) + \epsilon, \quad \forall x \in G(\epsilon), \forall d \in \mathbb{R}. \tag{12}$$

By (H_5), we can take $\delta_0 > 0$ sufficiently small, such that $\overline{X}_{\psi}^{<\delta_0} \subseteq G(\epsilon)$. By Lemma 4.2 (b), we know that there exists k_0, such that

$$x^*(d_k, \epsilon_k) \in \overline{X}_{\psi}^{<\delta_0} \subseteq G(\epsilon), \quad \forall k \geqslant k_0. \tag{13}$$

Without loss of generality, we assume that $\epsilon_k \leqslant \epsilon$, it follows from (13) and (14) that

$$\begin{aligned} M_P - 2\epsilon &\leqslant f_0(x^*(d_k, \epsilon_k)) \leqslant L(x^*(d_k, \epsilon_k), d_k) + \epsilon \\ &\leqslant \inf_{x \in X} L(x, d_k) + 2\epsilon \leqslant M_P + 2\epsilon. \end{aligned} \tag{14}$$

Replacing 2ϵ by δ, the conclusion follows. □

THEOREM 4.2. *Assume that (H_1) -(H_5) hold. If the sequence $\{x^*(d_k, \epsilon_k)\}$ is produced by algorithm (A), then*

(a) $\lim_{k \longrightarrow \infty} f_0(x^*(d_k, \epsilon_k)) = M_P$;

(b) $\lim_{k \longrightarrow \infty} d(x^*(d_k, \epsilon_k), X_0) = 0.$

Proof. From Lemma 4.3, we know that (a) is true. Now from Lemma 4.2 (b) and assumption (H_5), for any open neighborhood $B(X_0, \rho) \supseteq X_0$, there exist $\delta > 0$ and $k_0 > 0$, such that

$$x^*(d_k, \epsilon_k) \in \overline{X}_{\psi}^{<\delta} \subseteq B(X_0, \rho), \quad \forall k \geqslant k_0.$$

By arbitrarity of $\rho > 0$, we know that (b) is true. □

COROLLARY 4.1. *Assume that X_0 is compact and that there is an unique solution for problem (P). If conditions of Theorem 4.2 are satisfied, then*

$$\lim_{k \longrightarrow \infty} x^*(d_k, \epsilon_k) = x^*.$$

THEOREM 4.3. *Assume that (H_1) -(H_5) hold and $\overline{X}^*(\delta)$ be upper semicontinuous at $\delta = 0$. If the sequence $\{x^*(d_k, \epsilon_k)\}$ is produced by algorithm (A), then*

$$\lim_{k \longrightarrow \infty} d(x^*(d_k, \epsilon_k), \overline{X}_0^*) = 0.$$

Proof. For any open neighborhood $B(\overline{X}_0^*, \rho) \supseteq \overline{X}_0^*$, by the conditions of Theorem, there exists $\delta > 0$ sufficiently small, such that $\overline{X}_0^* = \overline{X}^*(\delta) \subseteq B(\overline{X}_0^*, \rho)$. By Lemma 4.3, there exists $k_0 > 0$, such that

$$x^*(d_k, \epsilon_k) \in \overline{X}^*(\delta) \subseteq B(\overline{X}_0^*, \rho),$$

By arbitrarity of $\rho > 0$, we know that The conclusion is true. □

REMARK . If for any $d \in R$, there exists an optimal solution $x^*(d, 0)$ for $(D(d))$, then Theorem 4.3 is reduced to the main result (Theorem 2.2) in [1].

COROLLARY 4.2. *Under the conditions of Theorem 4.3, if there exists an unique optimal solution x^* for (P), then*

$$\lim_{k \longrightarrow \infty} x^*(d_k, \epsilon_k) = x^*.$$

5. Examples

In this section, we give some general Lagrangian functions $L(x, d)$ that satisfy assumptions $(H_2) - (H_4)$. Assumptions (H_1) and (H_5) are required for f_0 and ψ to assure a zero duality gap. It is not difficult to see that Lagrangian functions in [5, 12, 13] and [9] satisfy assumptions $(H_2) - (H_4)$.

Let $X_0 = \{x \in X | f_i(x) \leqslant 0, 1 \leqslant i \leqslant m\}$. $f_0(x)$ is a uniformly continuous function on an open neighborhood containing X_0. $\theta(x)$ satisfies

(1) $\theta(x) = 0 \Longleftrightarrow x \in X_0$;

(2) it is a uniformly continuous function on an open neighborhood containing X_0. In addition, we assume that $M_P > 0$.

1). $L(x, d) = |f_0(x) + \theta(x)| + \sum_{i=1}^m d_i \max\{f_i(x), 0\}, \quad d \in \mathbb{R}_+^m$;

2). $L(x, d) = (|f_0(x) + \theta(x)|^p + \sum_{i=1}^m (d_i \max\{f_i(x), 0\})^p)^{\frac{1}{p}}, \quad p > 0, \ d \in \mathbb{R}_+^m$;

3). $L(x, d) = \max\{f_0(x) + \theta(x), d_1 f_1(x), \cdots, d_m f_m(x)\}, \quad d \in \mathbb{R}_+^m$;

4). $L(x, d) = |f_0(x) + \theta(x)| + \sum_{i=1}^m (\exp(d_i \max\{f_i(x), 0\} - 1)) \quad d \in \mathbb{R}_+^m$;

5). $L(x, d) = \max\{f_0(x) + \theta(x), \exp(d_1 \max\{f_1(x), 0\}) - 1, \cdots, \exp(d_m \max \{f_m(x), 0\}) - 1\}, \quad d \in \mathbb{R}_+^m$;

6). $L(x, d) = \max\{f_0(x) + \theta(x), d_1 f_1(x), \cdots, d_m f_m(x)\} + \prod_{i=1}^m r_i \max\{f_i(x), 0\}, \quad r, d \in \mathbb{R}_+^m$;

7). $L(x, d) = f_0(x) + |\theta(x)| + \sum_{i=1}^m d_i \max\{f_i(x), 0\}, \quad d \in \mathbb{R}_+^m$.

It is easy to check that $L(x, d)$ in 1)-7) satisfy assumptions $(H_2) - (H_4)$. For example

$$L(x, d) = |f_0(x) + \theta(x)| + \sum_{i=1}^m d_i \max\{f_i(x), 0\}, \quad d \in \mathbb{R}_+^m$$

in 1) obviously satisfies (H_2). For given $r > 0$, let $d_i = r, 1 \leqslant i \leqslant m$. And suppose that $\psi(x) = \sum_{i=1}^m \max\{f_i(x), 0\}$. When $x \in X \backslash X_0$, we have

$$L(x, d) \geqslant r\psi(x), \quad \forall d \in \mathbb{R}^m,$$

i.e., $L(x, d)$ satisfying (H_4). From the assumptions on $f_0(x)$ and $\theta(x)$, for any $\epsilon > 0$, there exists an open set $G(\epsilon) \supseteq X_0$, such that

$$f_0(x) \geqslant 0, \quad |\theta(x)| < \epsilon, \qquad \forall x \in G(\epsilon).$$

Hence, for $x \in G(\epsilon)$, we have

$$L(x,d) \geqslant |f_0(x)+\theta(x)| \geqslant |f_0(x)| - |\theta(x)| \geqslant f_0(x) - \epsilon, \quad \forall d \in \mathbb{R}^m_+,$$

i.e., $L(x,d)$ satisfies (H_3).

Here $\theta(x)$ plays a perturbation role, its construction is very flexible. For example, we can take

$$\theta(x) = \begin{cases} \sum_{i=1}^m \alpha_i \max\{f_i(x), 0\}, & if\, x \in G \supset X_0 \\ +\infty, & if\, x \notin G, \end{cases}$$

where G is an open set, $f_i(x)$ $(1 \leqslant i \leqslant m)$ are uniformly continuous functions on G, $\alpha = (\alpha_1, \cdots, \alpha_m)$ belongs to a compact set in $\mathbb{R}^n$.

Acknowledgements

This research was partially supported by the National Natural Science Foundation of China (Grants 10171055 and 10171118) and the Internal Grant (A-PB49) of The Hong Kong Polytechnic University.

References

1. Andramonov, M. (1999), An approach to constructing generalized penalty functions. In: Rubinov, A. M. and Glover, B. M. (eds), *Optimization and Related Topics*, Kluwer Academic Publishers, *Appl. Optim.*, 47, 3–14.
2. Avriel, M. (1976), *Nonlinear Programming*, Prentice-Hall, Englewood Ciffs, NJ.
3. Bertsekas, D. P. (1976), Nondifferentiable optimization via approximation, *Mathematical Programming* 3, 1–25.
4. Glover, B. M., Rubinov, A. M. and Craven, B. D. (1995), Solvability theorems involving inf-convex functions, *Journal of Mathematical Analysis and Applications* 191, 303–330.
5. Goh, C. J. and Yang, X. Q. (2001), A nonlinear Lagrangian theory for nonconvex optimization problems, *Journal of Optimization Theory and Applications* 109, 99–121.
6. Grossman, Ch. and Kaplan, A. (1979), *Straffunktionen und Modifizierte Lagrange-funktionen in der Nichtlinearen optimierund*, Texte Zur Mathematik, Leipzig.
7. Jeyakumar, V. and Wolkowicz, H. (1990), Zero duality gaps in infinite-dimensional programming, *Journal of Optimization Theory and Applications* 67, 87–108.
8. Luenberger, D. G. (1984), *Linear and Nonlinear Programming*, Addison-Wesley, Reading, MA.
9. Oettli, W. and Yang, X. Q. (2002), Modified Lagrangian and least root approaches for general nonlinear optimization problems, *ACTA Mathematicae Applicatae Sinica*, English Series 18, 147–152.
10. Rockaffellar, R. T. (1974), *Conjugate Duality and Optimization*, CBMS Lecture Notes Series, Vol. 162, SIAM Publications, Philadephia, PA.
11. Thach, R. T. (1993), Global optimality criterion and a duality with zero gap in nonconvex optimization, *SIAM Journal on Mathematical Analysis* 24, 1537–1556.
12. Rubinov, A. M. and Glover, B. M. (1998), Duality for increasing positively homogeneous functions and normal sets, to appear, *RAIRO Operations Research* 32, 105–123.
13. Rubinov, A. M., Glover, B. M. and Yang, X. Q. (1999), Modified Lagrange and penalty functions in continuous optimization, *Optimization* 46, 327–351.
14. Wang, C. Y. and Zhao, Q. Z. (1989), Existence for nonlinear dual programming, (in Chinese) *Chinese Ann. Math. Ser. A* 10, 650–659.
15. Zabotin, Y. I. (1975), A minimax method for solving problems of mathematical programming, *Soviet Mathematics (Iz VUZ)* 19, 28–33.