Nonlinear Augmented Lagrangian and Duality Theory

C. Y. Wang
Institute of Operations Research, Qufu Normal University, Qufu, China, wcy0537@126.com

X. Q. Yang
Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong, China, mayangxq@polyu.edu.hk

X. M. Yang
Department of Mathematics, Chongqing Normal University, Chongqing 400047, China, yang@cqnu.edu.cn

In this paper, a unified framework of a nonlinear augmented Lagrangian dual problem is investigated for the primal problem of minimizing an extended real-valued function by virtue of a nonlinear augmenting penalty function. Our framework is more general than the ones in the literature in the sense that our nonlinear augmenting penalty function is defined on an open set and that our assumptions are presented in terms of a substitution of the dual variable, so our scheme includes barrier penalty functions and the weak peak at zero property as special cases. By assuming that the increment of the nonlinear augmenting penalty function with respect to the penalty parameter satisfies a generalized peak at zero property, necessary and sufficient conditions for the zero duality gap property are established and the existence of an exact penalty representation is obtained.

Key words: nonlinear augmented Lagrangian; generalized peak at zero; zero duality gap; exact penalty representation
MSC2000 subject classification: Primary: 90C30, 90C46, 90C26
OR/MS subject classification: Primary: nonlinear programming; secondary: optimization
History: Received February 2, 2011; revised August 25, 2011, January 3, 2013. Published online in Articles in Advance March 29, 2013.

1. Introduction. It is well known that duality theory is always an important issue in optimization. The zero duality gap property provides a theoretical foundation for establishing the global convergence of numerical algorithms for solving constrained optimization problems. For convex optimization problems, a zero duality gap property is established by using classical Lagrangian functions and assuming some constraint qualifications. For nonconvex optimization problems, the zero duality gap property can be obtained by using augmented Lagrangian functions, in which the augmenting penalty function is the product of the penalty parameter and an augmenting function, and is thus linear in the penalty parameter. Early augmented Lagrangian functions were usually defined by using convex and quadratic augmenting functions or convex separable augmenting functions (see Balder [1], Ben-Tal and Zibulevsky [2], Bertsekas [3], Golhshtein and Tretyakov [8], Hestenes [10], Powell [16], Rockafellar [17]). Based on these augmented Lagrangian functions (or their generalized form), many effective numerical algorithms have been designed for solving nonconvex optimization problems (see Ben-Tal and Zibulevsky [2], Bertsekas [3], Burachik et al. [7], Burachik and Kaya [4], Hestenes [10], Luo et al. [12], Polyak [15], Powell [16], Wang and Li [24]).

In the past decade, the study of augmented Lagrangian duality theory has been developed. In Rockafellar and Wets [18], Huang and Yang [11], Rubinov et al. [21], and Rubinov and Yang [20], the primal problem of minimizing an extended real-valued function, which is more general than a constrained optimization problem, was discussed. In Rockafellar and Wets [18], a nonnegative convex augmenting function and the corresponding augmented Lagrangian dual problem of the primal problem were introduced. A sufficient condition for the zero duality gap and a necessary and sufficient condition for the existence of an exact penalty representation were obtained under mild conditions. In Huang and Yang [11], the same results as Rockafellar and Wets [18] were obtained by replacing the convexity condition of the augmenting function with a level-boundedness condition. Furthermore, using the theory of abstract convexity (see Rubinov [19]), a family of augmenting functions with almost peak at zero property and a class of corresponding augmented Lagrangian dual problems were introduced in Rubinov et al. [21] and Rubinov and Yang [20]. And it was established that if the family of augmenting functions is almost peak at zero, then the lower semicontinuity of the perturbation function at zero guarantees the zero duality gap property. It is worth noting that the peak at zero property that was employed in Rubinov et al. [21] and Rubinov and Yang [20] is more general than the level-boundedness condition in Huang and Yang [11].

Recently, the augmented Lagrangian duality theory has been further advanced in the literature from two aspects. On one hand, in Nedich and Ozdaglar [14], a nonnegative augmenting function for a primal geometry problem was introduced and a geometry framework of the corresponding augmented Lagrangian dual problem was constructed. Under appropriate assumptions necessary and sufficient conditions were obtained for the zero
duality gap property. It is worth noting that the peak at zero property in Rubinov et al. [21] and Rubinov and Yang [20] was further extended in Nedich and Ozdaglar [14], that is, only the nonnegative part of the dual variable of the augmenting function is required to satisfy a peak at zero property. Obviously, this requirement is weaker than the usual peak at zero property. So this property is called a weak peak at zero property in the present paper. In addition, it is noticed that the augmenting penalty function is linear in the penalty parameter in Rockafellar and Wets [18], Huang and Yang [11], Rubinov et al. [21], Rubinov and Yang [20] and Nedich and Ozdaglar [14]. Thus, if an augmenting function satisfies the weak peak at zero property, then the augmenting penalty function satisfies the coercivity condition. It is worth noting that if convex augmenting functions in Nedich and Ozdaglar [14] are allowed to be negative, the zero duality gap property may not hold. In Nedich and Ozdaglar [13], a nonlinear augmenting penalty function, in which the convex augmenting function is allowed to be negative, was introduced and a geometry framework of the primal and dual problems was constructed. Under various coercivity conditions, necessary conditions and sufficient conditions for the zero duality gap property were obtained. An exact penalty representation was not discussed in Nedich and Ozdaglar [13, 14], but it was pointed out in Nedich and Ozdaglar [13] that it is significant to study such conditions.

On the other hand, in a Banach space, using the theory of abstract convexity, a framework of an augmented Lagrangian dual problem for the primal problem of minimizing an extended real-valued function was constructed in Burachik and Rubinov [5], in which the nonlinear augmenting penalty function is composite of a coupling function with certain monotonic property and augmenting/penalty terms. This framework includes many augmenting penalty functions (such as the ones in Rockafellar and Wets [18], Huang and Yang [11], Rubinov et al. [21], Rubinov and Yang [20]) as special cases. In Burachik and Rubinov [5], if the family of augmenting functions is a peak at zero and satisfies a coercivity condition, sufficient conditions for the zero duality gap property and the existence of an exact penalty representation were given. Later, in a Hausdorff topological space, a unified framework of the augmented Lagrangian dual problem was constructed in Burachik et al. [6], which is more general than the one in Burachik and Rubinov [5]. It was assumed that the increment of the nonlinear augmenting penalty function with respect to the penalty parameter satisfies a peak at zero property and a coercivity condition. In addition, the nonlinear augmenting penalty function was required to be semicontinuous at zero. Under appropriate conditions, a sufficient condition for the zero duality gap and a necessary and sufficient condition for the existence of an exact penalty representation were obtained in Burachik et al. [6].

Inspired by Burachik et al. [6] and Nedich and Ozdaglar [14], in a finite dimensional Euclidean space, we construct a unified framework of the augmented Lagrangian dual problem for the primal problem of minimizing an extended real-valued function by virtue of a nondecreasing nonlinear augmenting penalty function. This framework is more general than the ones in Burachik et al. [6] and Nedich and Ozdaglar [14] in a sense that our nonlinear augmenting penalty function is defined on an open set and that our assumptions are presented in terms of a substitution of the dual variable, so our scheme includes barrier penalty functions in, e.g., Luo et al. [12], Polyak [15], Sun et al. [22], and the weak peak at zero property used in Nedich and Ozdaglar [14] as special cases. In our assumptions, we need neither the coercivity condition of the increment of the nonlinear augmenting penalty function as in Burachik et al. [6] or that of the augmenting function as in Nedich and Ozdaglar [13], nor the semicontinuity condition of the nonlinear augmenting penalty function at zero as in Burachik et al. [6]. We only assume that the increment of the nonlinear augmenting penalty function with respect to the penalty parameter satisfies a generalized peak at zero property. We introduce a pair of $\varepsilon$-perturbation conditions and show that when $\varepsilon > 0$ this pair of perturbation conditions present necessary and sufficient conditions for the zero duality gap property; and when $\varepsilon = 0$ this pair of perturbation conditions present necessary and sufficient conditions for the existence of an exact penalty representation. Thus these results clearly reveal the close relationship between the zero duality gap property and the exact penalty representation property. To the best of our knowledge, these results have not been investigated in the literature. Furthermore, assuming a coercivity condition of the increment of the nonlinear augmenting penalty function, we obtain some simplified necessary conditions and sufficient conditions for the zero duality gap property and the exact penalty representation.

The outline of this paper is as follows. In §2, a nonlinear augmenting penalty function and the corresponding nonlinear augmented Lagrangian dual problem for a primal problem of minimizing an extended real-value function are introduced. In §3, several types of typical nonlinear augmenting penalty functions are presented and shown to be the special cases of the nonlinear augmenting penalty functions given in this paper. In §4, necessary and sufficient conditions for the zero duality gap property are established. In §5, necessary and sufficient conditions for the existence of an exact penalty representation are obtained.

2. Nonlinear augmented Lagrangian dual problem. Let $\bar{R} = R \cup \{+\infty\}$. Consider the primal problem of minimizing an extended real-valued function

$$
(P) \quad \min_{x \in \bar{R}^n} \varphi(x),
$$
where \( \phi: \mathbb{R}^n \to \mathbb{R} \) is a proper function. A function \( \tilde{f}: \mathbb{R}^n \times \mathbb{R}^{m_i} \to \mathbb{R} \) is said to be a dualizing parameterization function for \( \phi \), if \( \tilde{f} \) is proper and satisfies
\[
\tilde{f}(x, 0) = \phi(x), \quad \forall x \in \mathbb{R}^n.
\]

The perturbation function of (P) is defined as
\[
\beta_f(u) := \inf_{x \in \mathbb{R}^n} \tilde{f}(x, u), \quad \forall u \in \mathbb{R}^{m_i}.
\]

The optimal value of (P) is denoted by \( M_p = \inf_{x \in \mathbb{R}^n} \phi(x) \). Then \( M_p = \beta_f(0) \).

Let \( R_{++} = (0, +\infty) \). To introduce a nonlinear augmenting penalty function, we introduce a class of auxiliary set-valued mappings \( U: R_{++} \to \mathbb{R}^{m_i} \), which, depending on the penalty parameter \( r \), represents the effective domain of the nonlinear augmenting penalty function with respect to the dual variable \( u \). By means of the set-valued mapping \( U \), we are able to include many classical barrier penalty functions as special cases of our nonlinear augmenting penalty function. Assume that \( U \) satisfies the following conditions:

(a) for all \( r \in R_{++} \), \( U(r) \) is an open and convex set with \( 0 \in U(r) \); and
(b) for all \( r_1, r_2 \in R_{++}, r_2 \geq r_1 \Rightarrow U(r_2) \subseteq U(r_1) \).

The following set-valued mappings satisfy conditions (a) and (b): for \( r \in R_{++} \),
\[
U(r) = \mathbb{R}^{m_i};
\]
\[
U(r) = \{ u \in \mathbb{R}^{m_i} \mid \| u \| < 1/r \};
\]
\[
U(r) = \{ u \in \mathbb{R}^{m_i} \mid u_i < 1/r, 1 \leq i \leq m_i \}; \quad \text{and}
\]
\[
U(r) = \{ u \in \mathbb{R}^{m_i} \mid ru \in C \},
\]
where \( C \subseteq \mathbb{R}^{m_i} \) is an open convex set and \( 0 \in C \).

To include the weak peak at zero condition in Nedich and Ozdaglar [14] in condition (A3) of Definition 1 below, we introduce a variable substitution. That is, for a continuous, not identically equal to 0 and nondecreasing function \( e: \mathbb{R} \to \mathbb{R} \) with \( e(0) = 0 \), we define the variable substitution \( \alpha(u) \) for \( u \) as
\[
\alpha(u) = (e(u_1), \ldots, e(u_{m_i})), \quad u \in \mathbb{R}^{m_i}.
\]

Some examples of \( \alpha(u) \) follow:

(i) If \( e(t) = t \), then \( \alpha(u) = u = (u_1, \ldots, u_{m_i}) \).
(ii) If \( e(t) = t^+: = \max \{ 0, t \} \), then \( \alpha(u) = u^+ = (u_1^+, \ldots, u_{m_i}^+) \).
(iii) If \( e(t) = t^-: = \min \{ 0, t \} \), then \( \alpha(u) = u^- = (u_1^-, \ldots, u_{m_i}^-) \).

Obviously, \( u = u^+ + u^- \). We will use the two variable substitutions \( \alpha(u) = u \) and \( \alpha(u) = u^+ \), respectively, in our study.

Now we give the definition of a nonlinear augmenting penalty function. Let \( \rho: U(r) \times V \times R_{++} \to \mathbb{R} \), where \( U(r) \subseteq \mathbb{R}^{m_i} \) satisfies (a1) and (a2) and \( V \subseteq \mathbb{R}^{m_2} \) is a nonempty subset. For given \( r, \tilde{r} \in R_{++} \), we denote
\[
\Delta_r(u, v, r) := \rho(u, v, r) - \rho(u, v, \tilde{r}).
\]

DEFINITION 1. The function \( \rho: U(r) \times V \times R_{++} \to \mathbb{R} \) is called a nonlinear augmenting penalty function if the following conditions are satisfied:

(A1) \( \forall (v, r) \in V \times R_{++} \), \( \rho(0, v, r) = 0 \);

and there exists a function \( \alpha(\cdot): \mathbb{R}^{m_1} \to \mathbb{R}^{m_1} \) satisfying (A2) and (A3):

(A2) \( \forall (v, \tilde{r}) \in V \times R_{++} \),
\[
\inf \{ \Delta_r(u, v, r) \mid u \in U(r), \alpha(u) = 0 \} \geq 0, \quad \forall r \geq \tilde{r}.
\]

(A3) \( \forall \varepsilon > 0 \), and \( (v, \tilde{r}) \in V \times R_{++} \),
\[
\inf \{ \Delta_r(u, v, r) \mid u \in U(r), \| \alpha(u) \| \geq \varepsilon \} > 0, \quad \forall r > \tilde{r}.
\]

In (A3), we use the convention \( \inf \emptyset = +\infty \).
Remark 1. When \( \alpha(u) = u \), it follows from \((A_1)\) that \((A_2)\) is always true. Thus \( \rho \) is only required to satisfy the conditions \((A_1)\) and \((A_2)\) in this case. These two conditions are much weaker than the conditions \((C_1)\) and \((C_2)\) in Burachik et al. \([6]\). In fact, the lower semicontinuity of \( \rho(\cdot, v, r) \) at 0 in \((C_1)\) and the coercivity condition of \( \rho(u, v, \cdot) \) in \((C_2)\) are both relaxed in this paper. When \( \rho(u, v, r) = \langle u, v \rangle + r \sigma(u) \) and \( \alpha(u) = u^+ \), the condition \((A_3)\) becomes the “weak peak at zero condition” of a convex augmenting function in Nedich and Ozdaglar \([14, \text{Assumption 2(b)}]\), as follows:
\[
\inf \{ \sigma(u) \mid \|u^+\| \geq \varepsilon \} > 0.
\]
That is, for a given sequence \( \{u_k\} \subset \mathbb{R}^n \), the convergence of \( \sigma(u_k) \) to zero implies the convergence of the nonnegative part of the sequence \( \{u_k\} \) to zero. Therefore, the condition \((A_3)\) is a generalized peak at zero condition.

Remark 2. Since \( \{ u \in U(r) \mid \|u^+\| \geq \varepsilon \} \subset \{ u \in U(r) \mid \|u\| \geq \varepsilon \} \), any \( \rho \) that satisfies \((A_1)\) and \((A_3)\) for \( \alpha(u) = u \) also satisfies \((A_1)\), \((A_2)\), and \((A_3)\) for \( \alpha(u) = u^+ \). However the reverse is in general not true.

Using \( \tilde{f} \) and \( \rho \) given above, we can construct a nonlinear augmented Lagrangian function of \((P)\), that is,
\[
l(x, v, r) = \inf_{u \in U(r)} \{ \tilde{f}(x, u) + \rho(u, v, r) \}, \quad (x, v, r) \in \mathbb{R}^n \times \mathbb{V} \times \mathbb{R}_+.\]
The dual function of \((P)\) is
\[
\tilde{\psi}(v, r) = \inf_{x \in \mathbb{R}^n} l(x, v, r), \quad (v, r) \in \mathbb{V} \times \mathbb{R}_+.
\]
By virtue of the dual function \( \tilde{\psi}(\cdot, \cdot) \), we introduce the following dual problem of \((P)\):

\[
(\text{D}) \quad \sup_{(v, r) \in \mathbb{V} \times \mathbb{R}_+} \tilde{\psi}(v, r).
\]

It is easy to see that, for all \( (v, r) \in \mathbb{V} \times \mathbb{R}_+ \),
\[
\tilde{\psi}(v, r) = \inf_{u \in U(r)} \{ \beta_f(u) + \rho(u, v, r) \}.
\]

Let \( M_D = \sup_{(v, r) \in \mathbb{V} \times \mathbb{R}_+} \tilde{\psi}(v, r) \) be the optimal value of the dual problem \((\text{D})\). By the definition of \( \tilde{f} \) and conditions \((a_1)\) and \((A_1)\), it is easy to show that the weak duality property holds, that is,
\[
M_D \leq M_p \quad \text{or} \quad M_D \leq \beta_f(0).
\]

For functions \( l \) and \( \tilde{\psi} \), we have the following basic properties.

Proposition 1. For all \( v \in \mathbb{V} \), we have

(i) for all \( x \in \mathbb{R}^n \), \( l(x, v, \cdot) \) is monotonically nondecreasing on \( r > 0 \); and

(ii) \( \psi(v, \cdot) \) is monotonically nondecreasing on \( r > 0 \).

Proof. We only need to prove (i), because (ii) follows from (i) immediately. Let \( r_2 > r_1 > 0 \). From \((A_2)\), \((A_3)\), for any \( (u, v) \in U(r_2) \times \mathbb{V} \), we have
\[
\rho(u, v, r_1) \leq \rho(u, v, r_2).
\]
Noting that \( U(r_2) \subseteq U(r_1) \), for any \( (x, v) \in \mathbb{R}^n \times \mathbb{V} \), the above inequality implies that
\[
l(u, v, r_1) = \inf_{u \in U(r_1)} \{ \tilde{f}(x, u) + \rho(u, v, r_1) \}
\leq \inf_{u \in U(r_2)} \{ \tilde{f}(x, u) + \rho(u, v, r_1) \}
\leq \inf_{u \in U(r_2)} \{ \tilde{f}(x, u) + \rho(u, v, r_2) \}
= l(u, v, r_2).
\]
Thus \( l(x, v, \cdot) \) is monotonically nondecreasing on \( r > 0 \).  \( \square \)

Proposition 1 is still valid if the strict inequality in \((A_3)\) is changed to an inequality. However this strict inequality is needed in the establishment of the zero duality gap and exact penalty representation properties in \( \S \S 4 \) and 5.
Definition 2. We say that the zero duality gap property between (P) and its dual problem (D) holds if

\[ M_D = M_p \quad \text{or} \quad M_D = \beta \hat{\gamma}(0). \]  

Let us consider an important special example of (P), that is, the constrained optimization problem

\[ (\hat{P}) \quad \inf_{x \in X} f(x), \]

where \( X_0 = \{ x \in X \mid g_1(x) = 0, g_0(x) \leq 0 \} \neq \emptyset, X \subseteq \mathbb{R}^n \) is a nonempty open subset, \( g_i(x) = (g_1(x), \ldots, g_m(x))^T, g_0(x) = (g_{i+1}(x), \ldots, g_{m_0}(x))^T \) and \( f, g_j : X \to \mathbb{R}^j (j = 1, \ldots, m) \) are real-valued functions. It is easy to obtain the nonlinear augmented Lagrangian dual problem of (P). Indeed, let

\[ \varphi(x) = \begin{cases} f(x), & \text{if } x \in X_0, \\ +\infty, & \text{otherwise}. \end{cases} \]  

Obviously, the constrained optimization problem \((\hat{P})\) is equivalent to (P), where \( \varphi \) is defined by (3). Let the dualizing parameterization function of \((\hat{P})\) be

\[ \hat{f}(x, u) = \begin{cases} f(x), & \text{if } x \in X, \ g_i(x) = u_i, \ g_0(x) \leq u_{\Pi}, \ u \in \mathbb{R}^m, \\ +\infty, & \text{otherwise}, \end{cases} \]  

where \((x, u) \in \mathbb{R}^n \times \mathbb{R}^m, u = (u_1, \ldots, u_m)^T\) and \( u_{\Pi} = (u_{i+1}, \ldots, u_{m})^T \). So the corresponding perturbation function is

\[ \hat{\beta}_j(u) = \inf \{ f(x) \mid x \in X, g_i(x) = u_i, g_0(x) \leq u_{\Pi} \}, \ u \in \mathbb{R}^m. \]  

The corresponding nonlinear augmented Lagrangian penalty function of \((\hat{P})\) is

\[ \hat{I}(x, v, r) = \inf_{u \in U(r)} \{ \hat{f}(x, u) + \rho(u, v, r) \} \]

\[ = f(x) + \inf_{u \in U(r)} \{ \rho(u, v, r) \mid x \in X, g_i(x) = u_i, g_0(x) \leq u_{\Pi} \}, \]

where \( \rho \) is given in Definition 1. Thus the dual function of \((\hat{P})\) is

\[ \hat{\psi}(v, r) = \inf_{x \in X} \hat{I}(x, v, r), \ \forall (v, r) \in V \times R_++. \]  

The dual problem of \((\hat{P})\) is

\[ (\hat{D}) \quad \sup_{(v, r) \in V \times R_+} \hat{\psi}(v, r). \]

In addition, let \( M_p \) and \( M_D \) be the optimal values of \((\hat{P})\) and \((\hat{D})\), respectively. Thus all the results established for primal problem (P) in this paper can be obtained for the constrained optimization problem \((\hat{P})\) from (3)–(7). These details are omitted because of the page space limitation.

3. Several special types of nonlinear augmenting penalty functions. In this section, we consider several special types of the nonlinear augmenting penalty function introduced in Definition 1, which include many augmenting functions and penalty functions in the literature as special cases.

Before presenting examples of the nonlinear augmenting penalty function introduced in Definition 1, we need the following propositions.

Assume that \( C \subseteq \mathbb{R}^m \) is an open convex set with \( 0 \in C \), and \( V \subseteq \mathbb{R}^m \) is a nonempty subset. Let

\[ R_+^m = \{ u \in \mathbb{R}^m \mid u \geq 0 \}, \]

\[ U(r) = \{ u \in R^m \mid u \in C \}, \]

\[ B = \{ u \in \mathbb{R}^m \mid \| u \| = 1 \}, \]

\[ B^+ = \{ u \in \mathbb{R}^m_+ \mid \| u \| = 1 \}. \]
Proposition 2. Let $P: C \times V \to R$ be a real-valued function such that, for any $v \in V$, $P(\cdot, v)$ is a convex function on $C$ and $P(0, v) = 0$. For any $\varepsilon > 0$, $(v, \bar{r}) \in V \times R^+_+$ and $r > \bar{r}$, let $\varepsilon, \bar{r} \in (0, \varepsilon)$ satisfy $\varepsilon, B \subset U(r)$. Then

$$\inf_{u \in U(r), \|u\| \geq \varepsilon} \left\{ \frac{1}{r} P(ru, v) - \frac{1}{r} P(\bar{r}u, v) \right\} \geq \min_{u \in B} \left\{ \frac{1}{r} P(r\varepsilon, u^0, v) - \frac{1}{r} P(\bar{r}\varepsilon, u^0, v) \right\} \geq 0.$$  

Furthermore, if, for any $v \in V$, $P(\cdot, v)$ is a strictly convex function on $C$, then the second one in the above inequalities holds strictly.

Proof. Let $v \in V$. For any $r > \bar{r} > 0$ and $u \in U(r)$, by the convexity of $P(\cdot, v)$ and $P(0, v) = 0$, we have

$$\frac{1}{r} P(ru, v) - \frac{1}{r} P(\bar{r}u, v) \geq 0. \quad (9)$$

Particularly, if $P(\cdot, v)$ is strictly convex on $C$, the inequality (9) holds strictly.

In particular, if $P(\cdot, v)$ is strictly convex on $C$, the inequality (9) holds strictly.

For $\varepsilon > 0$ and $r > \bar{r} > 0$, let $u \in U(r)$ and $\varepsilon, \bar{r} \in (0, \varepsilon)$ satisfy $\|u\| \geq \varepsilon$ and $\varepsilon, B \subset U(r)$, respectively. Let $u = \|u\|u^0$ with $u^0 \in B$ and let $u_\varepsilon, \bar{r}u_\varepsilon$. It follows from the property of $U(\cdot)$ and (8) that $ru, \bar{r}u, ru_\varepsilon, \bar{r}u_\varepsilon \in C$. Using the convexity of $P(\cdot, v)$ on $C$, it can be obtained that the difference quotient

$$\frac{1}{r} \{ P(z + td, v) - P(z, v) \}$$

is monotonically nondecreasing in $t$ as long as $z, z + td \in C$ and $d \in R^m$. Thus using this property twice, we have

$$\frac{1}{r} \{ P(\bar{r}u, v) - P(\bar{r}u_\varepsilon, v) \} = \frac{\|u\| - \varepsilon_r}{r\|u\| - \varepsilon_r} \left\{ P(\bar{r}u + \bar{r}(\|u\| - \varepsilon_r)u^0, v) - P(\bar{r}u_\varepsilon, v) \right\} \leq \frac{\|u\| - \varepsilon_r}{r\|u\| - \varepsilon_r} \left\{ P(\bar{r}u_\varepsilon, v) - P(\bar{r}u_\varepsilon, v) \right\} = \frac{\|u\| - \varepsilon_r}{r\varepsilon_r - r\|u\|} \left\{ P(\bar{r}u_\varepsilon, v) - P(\bar{r}u_\varepsilon, v) \right\} \leq \frac{\|u\| - \varepsilon_r}{r\varepsilon_r - r\|u\|} \left\{ P(ru + (r\varepsilon_r - r\|u\|)u^0, v) - P(ru, v) \right\} \leq \frac{1}{r} \{ P(ru, v) - P(ru_\varepsilon, v) \}.$$

So when $u \in U(r)$ and $\|u\| \geq \varepsilon$, noting $u_\varepsilon = \varepsilon_ru_0^0$, it follows from rearranging the last inequality that

$$\frac{1}{r} P(ru, v) - \frac{1}{r} P(\bar{r}u, v) \geq \frac{1}{r} P(r\varepsilon_r, u_0^0, v) - \frac{1}{r} P(\bar{r}\varepsilon_r, u_0^0, v). \quad (10)$$

Now we analyze the right-hand term in (10). Let

$$\varphi(u_0^0, v) := \frac{1}{r} P(r\varepsilon_r, u_0^0, v) - \frac{1}{r} P(\bar{r}\varepsilon_r, u_0^0, v), \quad u_0^0 \in B.$$  

By (9), it is clear that $\varphi(u, v) \geq 0, \forall u \in B$. It follows from $\varepsilon, B \subset U(r)$ that for all $u_0^0 \in B$, we have $r\varepsilon_r, \bar{r}\varepsilon_r, u_0^0 \subset C$. From the continuity of convex function $P(\cdot, v)$ on the open set $C$ and the compactness of $B$, the minimum solution of $\varphi(\cdot, v)$ on $B$ exists and

$$\min_{u_0^0 \in B} \varphi(u_0^0, v) \geq 0, \quad v \in V.$$  

Then (10) implies that

$$\frac{1}{r} P(ru, v) - \frac{1}{r} P(\bar{r}u, v) \geq \min_{u_0^0 \in B} \varphi(u_0^0, v).$$

Since the point $u$ satisfying $u \in U(r)$ and $\|u\| \geq \varepsilon$ is arbitrary, we obtain that

$$\inf_{u \in U(r), \|u\| \geq \varepsilon} \left\{ \frac{1}{r} P(ru, v) - \frac{1}{r} P(\bar{r}u, v) \right\} \geq \min_{u_0^0 \in B} \varphi(u_0^0, v).$$

In particular, if $P(\cdot, v)$ is a strictly convex function on $C$, then by (9) we have $\varphi(u_0^0, v) > 0, \forall u_0^0 \in B$. Thus

$$\min_{u_0^0 \in B} \varphi(u_0^0, v) > 0, \quad v \in V.$$  

The proof is complete. □
PROPOSITION 3. Let $P$: $R^{m_1} \times V \to R$ be a real-valued function such that, for any $v \in V$, $P(\cdot, v)$ is a convex function on $R^{m_1}$ and $P(0, v) = 0$. Let $\varepsilon > 0$, $\bar{v} \in (0, \varepsilon)$. Then, for any $(v, \bar{r}) \in V \times R_{++}$ and $r > \bar{r}$,

$$\inf_{u \in R^{m_1}, \|u\| \geq \varepsilon} \left\{ \frac{1}{r} P(ru^+, v) - \frac{1}{\bar{r}} P(\bar{r}u^+, v) \right\} \geq \min_{\bar{u} \in R^{m_1}} \left\{ \frac{1}{r} P(r\bar{u}^0, v) - \frac{1}{\bar{r}} P(\bar{r}\bar{u}^0, v) \right\} \geq 0.$$  

Furthermore, if, for all $v \in V$, $P(\cdot, v)$ is a strictly convex function on $R^{m_1}$, then the last one in the above inequalities holds strictly.

PROOF. The results follow from Proposition 2 with $C = R^{m_1}$ and $u^+$ and $B^+$ instead of $u$ and $B$, respectively.  

Now we use Propositions 2 and 3 to study Examples 1 and 2.

EXAMPLE 1. Let $\rho(u, v, r) = (1/r)P(ru, v)$, $(u, v, r) \in U(r) \times V \times R_{++}$, be an augmenting penalty function, where for any $v \in V$, $P(\cdot, v)$ is strictly convex on the open convex set $C \subseteq R^{m_1}$ with $0 \in C$, $P(0, v) = 0$, $U(r)$ is given by (8), and $V \subseteq R^{m_1}$ is a nonempty subset. Let $\alpha(u) = u$.

We check that $\rho$ satisfies the conditions (A₁) and (A₃) in Definition 1. First, from $P(0, v) = 0$, it is easy to check that $\rho$ satisfies the condition (A₁). To verify condition (A₃), by the strict convexity of $P(\cdot, v)$ and Proposition 2, we have that, for $\varepsilon > 0$,

$$\inf \{ \Delta_i(u, v, r) \mid u \in U(r), \|u\| \geq \varepsilon \} > 0, \quad \forall (v, \bar{r}) \in V \times R_{++}, \quad \forall r > \bar{r}.$$  

Hence, $\rho$ satisfies (A₃). Therefore, this $\rho$ is a special case of the nonlinear augmenting penalty function introduced in Definition 1.

We will use (R₁)–(R₃) to denote various comparisons between the special cases of our nonlinear augmenting penalty function presented in Examples 1–3 and those in the literature.

(R₁) It is known that the nonlinear augmenting penalty function of the form in Example 1 and its special cases are widely used in the literature, in which, however apart from the convexity or strict convexity of $P(\cdot, v)$, some strong conditions (such as coercivity condition, continuous differentiability condition, and so on) are assumed. Notice that $P(\cdot, v)$ is only required to be strictly convex on $C$ in Example 1. So, $P(\cdot, v)$ in Example 1 is more general than other augmenting penalty functions in the literature.

In Nedich and Ozdaglar [13], an augmented Lagrangian dual problem of a geometric primal problem was introduced by the following nonlinear augmenting penalty function:

$$\frac{1}{r} \sigma(ru) \quad \forall (u, r) \in R^{m_1} \times R_{++},$$

(11)

where the augmenting function $\sigma$: $R^{m_1} \to \bar{R}$ is proper and convex on $R^{m_1}$, not identically equal to 0, and $\sigma(0) = 0$. When $\sigma$ satisfies three types of the coercivity condition (see Nedich and Ozdaglar [13, Assumptions 3.2, 3.3, and 3.6]), respectively, necessary and sufficient conditions for the zero duality gap were obtained.

In the following we show that (11) is actually a special case of Example 1. Let the effective domain $C$ of $\sigma$, denoted by dom($\sigma$), be open. It follows from the convexity of $\sigma$ and $\sigma(0) = 0$ that $C$ is an open convex set in $R^{m_1}$, and $0 \in C$. For example, the effective domain of the augmenting function $\sigma$ satisfying in Nedich and Ozdaglar [13, Assumption 3.6(a)] is such an open convex set. Let $U(\cdot)$ be given by (8). Thus (11) is a special case of Example 1, that is, $V = \{0\}$ and

$$\rho(u, 0, r) = \frac{1}{r} \sigma(ru) \quad \forall (u, r) \in U(r) \times R_{++}.$$  

(12)

The following augmenting functions satisfy in Nedich and Ozdaglar [13, Assumption 3.3]:

$$\sigma_i(u) = \sum_{i=1}^{m_1} \phi_i(u), \quad i = 1, 2, 3,$$

where

$$\phi_1(t) = \begin{cases} -\log(1-t), & \text{if } t < 1; \\ +\infty, & \text{otherwise}, \end{cases} \quad \text{dom}(\sigma_1) = \{u \in R^{m_1} \mid u_i < 1, 1 \leq i \leq m_1\},$$

$$\phi_2(t) = \begin{cases} t/(1-t), & \text{if } t < 1; \\ +\infty, & \text{otherwise}, \end{cases} \quad \text{dom}(\sigma_2) = \{u \in R^{m_1} \mid u_i < 1, 1 \leq i \leq m_1\},$$

$$\phi_3(t) = \begin{cases} t + \frac{1}{2} t^2, & \text{if } t \geq -\frac{1}{2}; \\ -\frac{1}{2} \log(-2t) - \frac{1}{8}, & \text{otherwise}, \end{cases} \quad \text{dom}(\sigma_3) = C = R^{m_1},$$

respectively.
The above three augmenting functions $\sigma_i (i = 1, 2, 3)$ are strictly convex on open sets $C$. Therefore, the augmenting penalty functions (12) constructed by using $\sigma_i (i = 1, 2, 3)$ are the special cases of Example 1. Moreover, there are still many strictly convex augmenting functions that do not belong to any of the three types of augmenting functions in Nedich and Ozdaglar [13]. Two strictly convex functions below (given in Gonzaga and Castillo [9]) are such examples:

\[
\phi_4(t) = \log \frac{1 + e^t}{2}, \quad t \in R,
\]

\[
\phi_5(t) = \frac{t + \sqrt{t^2 + 4}}{2} - 1, \quad t \in R.
\]

As such, under the strict convexity condition of $\sigma$, the nonlinear augmenting penalty function (11) introduced in Nedich and Ozdaglar [13] is a special case of Example 1. It is worth noting that the five nonlinear augmenting penalty functions ($1/r) \sigma_i(ru) (i = 1, 2, \ldots, 5$) mentioned above do not satisfy any of the coercivity conditions in Burachik and Rubinov [5] and Burachik et al. [6].

(R3) Let us consider the application of Example 1 to the optimization problem ($\hat{P}$). As will be shown, we will derive several important types of augmenting penalty functions by some special forms of Example 1. First take the following special case of Example 1:

\[
\rho(u, v, r) = \frac{1}{r} \sum_{i=1}^{m_1} \phi(ru_i), \quad (u, v, r) \in R^{m_1} \times R_{++} \times R_{++},
\]

where $m_1 = 1$, $\phi$ is monotonically nondecreasing and strictly convex on $R$ with $\phi(0) = 0$. Substituting (13) into (6), we obtain the following augmented Lagrangian function of ($\hat{P}$):

\[
\tilde{l}(x, v, r) = f(x) + \frac{1}{r} \sum_{i=1}^{m_1} \phi(rg_i(x)), \quad (x, v, r) \in X \times R_{++} \times R_{++}.
\]

If furthermore $\phi$ is continuously differentiable, $\lim_{t \rightarrow -\infty} \phi(t) = a \in R$ and $\lim_{t \rightarrow +\infty} \phi(t)/t = 1$ (it is easy to verify that $\phi_i(t) (i = 4, 5)$ satisfy these conditions), then the augmented Lagrangian function in (14) is the one given in Gonzaga and Castillo [9], which can smoothly approximate the $l_1$ exact penalty function, that is,

\[
\lim_{r \rightarrow +\infty} \frac{1}{v} \sum_{i=1}^{m_1} \phi(rg_i(x)) = v \sum_{i=1}^{m_1} \max \{0, g_i(x)\}.
\]

Now we take another special case of Example 1:

\[
\rho(u, v, r) = \frac{1}{r} \sum_{i=1}^{m_1} \phi(ru_i), \quad (u, v, r) \in R^{m_1} \times R_{++} \times R_{++},
\]

where $\phi$ is monotonically nondecreasing and strictly convex on $R$ with $\phi(0) = 0$. Substituting (15) into (6), we obtain another augmented Lagrangian function of ($\hat{P}$):

\[
\tilde{l}(x, v, r) = f(x) + \frac{1}{r} \sum_{i=1}^{m_1} \phi(rg_i(x)), \quad (x, v, r) \in X \times R_{++} \times R_{++}.
\]

In addition, if $\phi$ is twice continuously differentiable with $\phi'(0) = 1$, $\phi''(0) > 0$, $\lim_{t \rightarrow -\infty} \phi(t) > -\infty$, $\lim_{t \rightarrow -\infty} \phi'(t) = 0$, and $\lim_{t \rightarrow +\infty} \phi'(t) = +\infty$, then the augmented Lagrangian function (16) is a $\hat{P}_1$ penalty function in Bertsekas [3], a widely used exponential type augmenting penalty function in the literature (see Luo et al. [12], Sun et al. [22], Tseng and Bertsekas [23], and Wang et al. [25]).

Finally, we take the following special case of Example 1:

\[
\rho(x, v, r) = \langle u, v \rangle + \frac{1}{r} \sum_{i=1}^{m_1} \phi(ru_i), \quad (u, v, r) \in R^{m_1} \times R_{++} \times R_{++},
\]

where $\phi$ is strictly convex on $R$ with $\phi(0) = 0$. Substituting (17) into (6), it follows that

\[
\tilde{l}(x, v, r) = f(x) + \frac{1}{r} \sum_{i=1}^{m_1} \inf_{r \leq u_i} \{ru_i + \phi(ru_i)\}
\]

\[
= f(x) + \frac{1}{r} \sum_{i=1}^{m_1} \theta(rg_i(x), v_i), \quad (x, v, r) \in X \times R_{++} \times R_{++},
\]
where

\[ \theta(s, \tau) = \inf_{t \in I} \{ t \tau + \phi(t) \}. \]  

(19)

In addition, if \( \phi \) is continuously differentiable with \( \phi'(0) = 0 \), \( \lim_{t \to -\infty} \phi'(t) = -\infty \), and \( \lim_{t \to +\infty} \phi'(t) = +\infty \), then (19) can be expressed by

\[ \theta(s, \tau) = \begin{cases} s \tau + \phi(s), & \text{if } \tau + \phi'(s) \geq 0; \\ \min_{t \in \mathbb{R}} \{ t \tau + \phi(t) \}, & \text{otherwise}. \end{cases} \]

In this case, (18) is a \( P^+_E \) type penalty function in Bertsekas [3]. If \( \phi \) is also twice continuously differentiable with \( \phi''(0) > 0 \), then \( \hat{h}(x, v, r) \) in (18) is the essential twice augmenting penalty function widely used in the literature (see Luo et al. [12], Rockafellar [17], Rockafellar and Wets [18], Sun et al. [22], Wang et al. [25], and Wang and Li [24]).

Similarly, we can also use other special cases of Example 1 to derive exponential type (Bertsekas [3], Polyak [15], Sun et al. [22]) and barrier type (Luo et al. [12], Polyak [15], Sun et al. [22]) penalty functions for \( \hat{P} \). These details are omitted here.

**Example 2.** Let \( \rho(u, v, r) = (1/r)P(ru, v) \), \( (u, v, r) \in R^{m_1} \times V \times R_{++}, \) where \( P: R^{m_1} \times V \to R \) satisfies the following conditions:

\begin{itemize}
  \item [(p_1)] for any \( v \in V \), \( P(\cdot, v) \) is convex on \( R^{m_1} \) and strictly convex on \( R^{m_1}_+ \) and \( P(0, v) = 0 \);
  \item [(p_2)] for all \( (u, v, r) \in R^{m_1} \times V \times R_{++} \) and \( r > \tilde{r} \), there holds
    \[ \frac{1}{r} P(ru, v) - \frac{1}{r} P(\tilde{r}u, v) \geq \frac{1}{r} P(ru^+, v) - \frac{1}{r} P(\tilde{r}u^+, v). \]
\end{itemize}

Let \( \alpha(u) = u^+ \). We verify that the function \( \rho \) satisfies conditions (A_1), (A_2), and (A_3) of Definition 1. First, from \( P(0, v) = 0 \) and (9), it is easy to check that \( \rho \) satisfies (A_1) and (A_2). Next we verify that \( \rho \) satisfies (A_3). Let \( \varepsilon > 0 \). From the strict convexity of \( P(\cdot, v) \) on \( R^{m_1}_+ \), the assumption (p_2) and Proposition 3, it follows that, for all \( (\bar{v}, v, r) \in V \times R_{++} \) and \( r > \tilde{r} \),

\[ \inf \{ \Delta_j(u, v, r) \mid u \in R^{m_1}, \| u^+ \| \geq \varepsilon \} \geq \inf_{u \in R^{m_1}, \| u^+ \| \geq \varepsilon} \left\{ \frac{1}{r} P(ru^+, v) - \frac{1}{r} P(\tilde{r}u^+, v) \right\} > 0. \]

So (A_3) holds. Therefore, Example 2 is a special case of the nonlinear augmenting penalty function introduced in Definition 1.

Now we consider two special cases of Example 2.

(R_3) Special case 1 of Example 2. Let

\[ \rho(u, v, r) = \frac{1}{r} \sum_{i=1}^{m_1} \Phi(ru_i, v_i), \quad (u, v, r) \in R^{m_1} \times R^{m_1}_+ \times R_{++}, \]  

(20)

where \( P(u, v) = \sum_{i=1}^{m_1} \Phi(u_i, v_i) \), \( \Phi: R \times R_+ \to R \) satisfies the following condition: for all \( s \in R_+ \), \( \Phi(\cdot, s) \) is convex on \( R^{m_1} \) and strictly convex on \( R^{m_1}_+ \) with \( \Phi(0, s) = 0 \).

Obviously, \( P \) satisfies (p_1). Next we verify that \( P \) satisfies (p_2). Let \( (u, v, r) \in R^{m_1} \times R^{m_1}_+ \times R_{++} \) and \( r > \tilde{r} \).

By (9) and \( \Phi(0, s) = 0 \), we have,

\[ \frac{1}{r} P(ru, v) - \frac{1}{r} P(\tilde{r}u, v) \]

\[ = \sum_{i, j} \left\{ \frac{1}{r} \Phi(ru_i, v_i) - \frac{1}{r} \Phi(\tilde{r}u_i, v_i) \right\} \]

\[ \geq \sum_{i, j} \left\{ \frac{1}{r} \Phi(ru_i^+, v_i) - \frac{1}{r} \Phi(\tilde{r}u_i^+, v_i) \right\} \]

\[ = \frac{1}{r} P(ru^+, v) - \frac{1}{r} P(\tilde{r}u^+, v). \]

So (p_2) holds.
Now we consider the application of special case 1 of Example 2 (that is, \( \rho \) is defined by (20)) to optimization problem (\( \hat{P} \)). Let \( \Phi(\cdot, s) \) be monotonically nondecreasing for any \( s \in R_+ \). Substituting \( \rho \) of (20) into (6), we obtain the following augmented Lagrangian function of (\( \hat{P} \)):

\[
\hat{l}(x, v, r) = f(x) + \frac{1}{r} \sum_{i=1}^{m_1} \Phi(r g_i(x), v_i), \quad (x, v, r) \in X \times R^{m_1}_+ \times R_{++},
\]

where \( m_2 = m_1 \). In addition, if \( \Phi \) is continuously differentiable on \( R \times R_{++} \) and satisfies some coercivity conditions (see Bertsekas [3, pp. 305–308]), then \( \hat{P} \) is a \( \hat{P}_1 \) type penalty function in Bertsekas [3].

(R_4) Special case 2 of Example 2. Let \( V = \{0\} \) and

\[
\rho(u, 0, r) = \frac{1}{r} \sigma(ru), \quad (u, r) \in R^{m_1} \times R_{++},
\]

where \( P = \sigma; R^{m_1} \to R \) satisfies conditions \((p_1)\) and \((p_2)\). By (20), we know that one subclass of (21) is

\[
\sigma(u) = \sum_{i=1}^{m_1} \phi(u_i),
\]

where \( \phi: R \to R \) is convex on \( R \) and strictly convex on \( R_+ \) with \( \phi(0) = 0 \). Obviously (22) is more general than the case that \( \sigma \) is a strictly convex function on \( R \).

It is worth noting that the real-valued convex augmenting function \( \sigma \) in (21) contains some examples of the first type (satisfying Nedich and Ozdaglar [13, Assumption 3.2]) and the second type (satisfying Nedich and Ozdaglar [13, Assumption 3.3]) augmenting functions. For example, (see, e.g., Nedich and Ozdaglar [13, 14], Bertsekas [3]):

\[
\sigma_1(u) = (u^+)^T Qu^+, \text{ where } Q \text{ is a symmetric positive definite matrix;}
\]

\[
\sigma_2(u) = \max \{0, a_1 (e^{a_2} - 1), \ldots, a_{m_1} (e^{a_{m_1}} - 1)\}, \text{ where } a_i > 0, 1 \leq i \leq m_1;
\]

\[
\sigma_3(u) = \sum_{i=1}^{m_1} (\max \{0, u_i\})^{\beta_i}, \text{ where } \beta > 1;
\]

\[
\sigma_4(u) = \sum_{i=1}^{m_1} \phi(u_i), \quad \phi(t) = \begin{cases} t + \frac{1}{2} t^2, & \text{if } t \geq 0, \\ -\log(1-t), & \text{otherwise}. \end{cases}
\]

Of course, there are also augmenting functions \( \sigma \) of (21), which do not belong to the cases of Nedich and Ozdaglar [13]. Such examples are

\[
\sigma_5(u) = \max \{-\log(1 + u_1^+), \ldots, -\log(1 + u_{m_1}^+), \}
\]

\[
\sigma_6(u) = \sum_{i=1}^{m_1} \phi(u_i), \quad \phi(t) = \begin{cases} e^{-t} - 1, & \text{if } t \geq 0, \\ -t, & \text{otherwise}. \end{cases}
\]

It is worth noting that if the augmenting function \( \sigma \) in (21) satisfies either Nedich and Ozdaglar [13, Assumption 3.2(b)] or Nedich and Ozdaglar [13, Assumption 3.3(a)], then, by \((p_2)\) and Proposition 3, we can easily prove that, for all \( \varepsilon > 0 \),

\[
\inf_{u \in R^{m_1}, \|u^+\| \geq \varepsilon} \Delta_s(u, 0, r) := \inf_{u \in R^{m_1}, \|u^+\| \geq \varepsilon} \left\{ \frac{1}{r} \sigma(\tilde{r} u) - \frac{1}{r} \sigma(\tilde{r} u) \right\} \to +\infty \quad \text{as } r \to +\infty.
\]

That is to say, these two assumptions in Nedich and Ozdaglar [13] are too strong.

**Example 3.** Let \( \rho: R^{m_1} \times V \times R_{++} \to R, (u, v, r) \in R^{m_1} \times V \times R_{++} \), where \( V = V_1 \times V_2 \subseteq R^m \); satisfies the following conditions:

\[
\rho(0, v, r) = 0;
\]

\[
\rho(u, v_1, v_2, r) \in \rho(\rho(u, v_1, v_2, r) \in R^{m_1} \times V_1 \times V_2 \times R_{++} \}
\]

where \( \hat{\rho}: R_+ \times R_{++} \to R_+ \) is any function with the following properties:

\[
\hat{\rho}(r, \tilde{r}) \text{ when } r > \tilde{r}, \hat{\rho}(r, r) = 0;
\]

\[
\hat{\rho}(r, \tilde{r}) \text{ when } r > \tilde{r}, \hat{\rho}(r, \tilde{r}) \text{ is strictly increasing,}
\]

\[
\text{where } \hat{\rho}(r, \tilde{r}) = \begin{cases} r, & \text{if } r \leq \tilde{r}, \\ \tilde{r}, & \text{otherwise}. \end{cases}
\]
and the family of functions \{σ_{v_2}\}_{v_2∈V_2} has the following properties:

(σ_1) for all \((u, v_2) ∈ R^n_u × V_2\), \(σ_{v_2}(u) ≥ 0\) and \(σ_{v_2}(0) = 0\);

(σ_2) there exists a variable substitution \(α(u)\) for \(u\) such that, for all \(ε > 0\) and \(v_2 ∈ V_2\),

\[
σ_{v_2, ε} := \inf_{w ∈ R^n_u, \|w - u\| ≤ ε} σ_{v_2}(u) > 0.
\]

Based on the above conditions, we can easily verify that \(ρ\) satisfies conditions (A_1), (A_2), and (A_3) in Definition 1.

(R_3) Let \(ρ\) be given by the conditions (3.1)-(3.3), \((U_2), (U_1)\), and \((Y_1)\) in Burachik and Rubinov [5], \(α(u) = u\), and, when \(r ≥ R > 0\), \(Ψ(t, r, R) = Ψ((r - R)t)\), where \(Ψ\) satisfies condition (3.1) in Burachik and Rubinov [5]. Then \(-ρ\) is a special case of the nonlinear augmenting penalty function of Example 3.

In the following, we give a concrete example of Example 3, but it does not belong to the augmenting penalty function in Burachik and Rubinov [5]. Let

\[
ρ(u, v, r) = \langle u, v \rangle + \log(1 + r^q σ_{v_2}(u)), \quad (u, v, r) ∈ R^n_u × V × R_+^n,
\]

where \(V = V_1 × V_2 ⊆ R^n_u\), \(q ∈ (0, +∞)\), \{σ_{v_2}\}_{v_2∈V_2} satisfies (σ_1) and, for \(α(u) = u^+\), \{σ_{v_2}\}_{v_2∈V_2} satisfies (σ_2).

It is clear that \(ρ\) defined in (23) satisfies \((ρ_2)\). We only need to verify that the function \(ρ\) satisfies \((ρ_2)\). Let \((u, v, r) ∈ R^n_u × V × R_+^n\) and \(r > R\). By the mean-value theorem and the monotonicity of \(s/(1 + s)\) on \(s > 0\), we have

\[
Δ_L(u, v, r) = \log(1 + r^q σ_{v_2}(u)) - \log(1 + R^q σ_{v_2}(u))
\]

\[
= \frac{q R^{q - 1} σ_{v_2}(u)}{1 + r^q σ_{v_2}(u)} (R - r) \quad (R < r)
\]

\[
≥ \frac{q R^{q} σ_{v_2}(u)}{r(1 + r^q σ_{v_2}(u))} (R - r)
\]

\[
= Ψ(σ_{v_2}(u), r, R),
\]

where \(Ψ(t, r, R) = q(1 - R/r)R^{q} t/(1 + r^q t)\) obviously satisfies \((Ψ_1)\) and \((Ψ_2)\).

4. Zero duality gap property. In this section, we present some necessary and sufficient conditions for the zero duality gap property (2) to hold, and obtain some corollaries, some of which are known results, and some others provide a theoretical foundation for the global convergence of the augmented Lagrangian method. Let

\[
N(v, r, δ) = \{u ∈ U(r) | β_j(u) + ρ(u, v, r) ≤ δ\},
\]

where \(U(·)\) satisfies (a_1) and (a_2). Let

\[
W_η^n(0) = \{u ∈ R^n_u | \|α(u)\| ≤ η\}, \quad η > 0.
\]

It is clear from the nondecreasing property of \(ε\) that \(W_η^n(0)\) is a neighborhood of \(0 ∈ R^n_u\). In particular, if \(α(u) = u\), it is denoted by \(W_η^n(0)\). If \(α(u) = u^+\), it is denoted by \(W_η^n(0)\).

We now give necessary conditions for the zero duality gap property to hold.

Recall that \(Ψ(u, v)\) is the dual function of (P). Assume that, for any \(ε > 0\) and \(R > 0\), there exist \(ε_R > R\) and \(v_ε ∈ V\) and a neighborhood \(W_η^n(0)\) of \(0 ∈ R^n_u\) such that

\[
(B_j^+_1) \inf \{Δ(u, v_ε, r_ε) | u ∈ N(v_ε, r_ε, δ + ε)\} = 0, \quad ∀ δ ≥ Ψ(v_ε, r_ε);
\]

\[
(B_j^+_2) β_j(0) - ε ≤ β_j(u) + ρ(u, v_ε, r_ε), \quad ∀ u ∈ W_η^n(0) \cap U(r_ε);
\]

\[
(B_j^+_3) \text{ and } (B_j^+_4) \text{ are called a set of the } ε\text{-perturbation conditions.}
\]

THEOREM 1. Assume that the nonlinear augmenting penalty function \(ρ\) satisfies (A_1), (A_2), and (A_3). Then the zero duality gap property (2) holds if and only if \((B_j^+_1)\) and \((B_j^+_2)\) hold.

PROOF. Necessity. Let \(M_ρ = M_ρ\). Notice that \(Ψ(u, ·)\) is monotonically nondecreasing on \(r > 0\). Then, for any \(ε > 0\) and \(R > 0\), there exist \(ε_R > R\) and \(v_ε ∈ V\) such that

\[
β_j(0) - ε ≤ Ψ(v_ε, r_ε)
\]

\[
= \inf_{u ∈ U(r_ε)} \{β_j(u) + ρ(u, v_ε, r_ε)\}
\]

\[
≤ β_j(u) + ρ(u, v_ε, r_ε), \quad ∀ u ∈ U(r_ε).
\]
From (24), we know that \((B_3^0)\) holds. Let \(\delta > \tilde{\psi}(v_e, r_e)\). When \(u \in N(v_e, r_e, \delta)\), (A_1) and (24) imply that
\[
\beta_j(0) + \rho(0, v_e, r_e) = \beta_j(0) \leq \beta_j(u) + \rho(u, v_e, r_e) + \varepsilon \leq \delta + \varepsilon.
\]
That is, \(0 \in N(v_e, r_e, \delta + \varepsilon)\). By using (A_1), (A_2), and (A_3), we have
\[
\inf \{\Delta_j(u, v_e, r_e) \mid u \in N(v_e, r_e, \delta + \varepsilon)\} = 0, \quad \forall \, \delta > \tilde{\psi}(v_e, r_e).
\]
So \((B_3^0)\) holds.

Sufficiency. We assume that, for any \(\varepsilon > 0\) and \(\tilde{r} > 0\), there exist \(r_e > \tilde{r}\) and \(v_e \in V\) and a neighborhood \(W_0^\alpha(0)\) of \(0 \in R_{m_1}\) satisfying \((B_3^0)\) and \((B_3^1)\). Choose a sequence \(\{\delta_k\}\) such that
\[
\delta_k \downarrow \tilde{\psi}(v_e, r_e) \quad \text{(as } k \to +\infty). \tag{25}
\]
By condition \((B_3^0)\), there exists a sequence \(\{u_k\} \subset N(v_e, r_e, \delta_k + \varepsilon)\) such that
\[
\lim_{k \to +\infty} \Delta_j(u_k, v_e, r_e) = 0. \tag{26}
\]
By (26) and (A_3), we have that
\[
\lim_{k \to +\infty} \alpha(u_k) = 0.
\]
Hence, when \(k\) is sufficiently large, we have \(u_k \in W_0^\alpha(0) \cap U(r_e)\). Noting that \(u_k \in N(v_e, r_e, \delta_k + \varepsilon)\), together with \((B_3^0)\), we have
\[
\beta_j(0) - \varepsilon \leq \beta_j(u_k) + \rho(u_k, v_e, r_e) \leq \delta_k + \varepsilon.
\]
By the inequality above and (25), we have
\[
\beta_j(0) - \varepsilon \leq \lim_{k \to +\infty} \delta_k + \varepsilon \leq \tilde{\psi}(v_e, r_e) + \varepsilon \leq M_p + \varepsilon. \tag{27}
\]
Since \(\varepsilon > 0\) is arbitrary, (27) and the weak duality property yield that the zero duality gap property holds, that is, \(M_p = M_D\).

**Remark 3.** Theorem 1 illustrates that \((B_3^0)\) and \((B_3^1)\) are necessary and sufficient conditions for the zero duality gap property to hold. This set of perturbation conditions will be shown to be a necessary and sufficient condition for the exact penalty representation of \((P)\) when \(\varepsilon = 0\) (see Theorem 3). Thus this set of the perturbation conditions clearly reveals the close relation between the zero duality gap property and the exact penalty representation. To the best of our knowledge, no such kind of perturbation conditions have been investigated in the literature yet.

Let us introduce a coercivity assumption as follows:
\((A_4)\) for all \(\varepsilon > 0\) and \((v, \tilde{r}) \in V \times R_{++}\), we have
\[
\liminf_{r \to +\infty} \{\Delta_j(u, v, r) \mid u \in U(r), \|\alpha(u)\| \geq \varepsilon\} = +\infty.
\]
When \(U(r) = R_{m_1}\) and \(\alpha(u) = u\), \((A_4)\) reduces to Assumption \((C_2)\)-(ii) in Burachik et al. [6].

We have the following results.

**Lemma 1.** Let \(\tilde{\delta} > M_p\). Suppose that the assumption \((A_4)\) holds, and there exists \((\tilde{v}, \tilde{r}) \in V \times R_{++}\) satisfying
\[
\tilde{\psi}(\tilde{v}, \tilde{r}) > -\infty.
\]
Then, for any \(\eta > 0\), there exists \(r_\eta \geq \tilde{r}\) such that
\[
N(\tilde{v}, r_\eta, \delta) \subseteq W_0^\alpha(0) \cap U(r_\eta), \quad \forall \, \delta \in (\tilde{\psi}(\tilde{v}, r_\eta), \tilde{\delta}).
\]
Proof. Assume to the contrary that there exists \( \tilde{\eta} > 0 \), for all \( r > \tilde{r} \), we can choose \( \delta_r \in (\tilde{\psi}(\tilde{v}, r), \tilde{\delta}] \) such that

\[
u_r \in N(\tilde{v}, r, \delta_r),
\]

(28)

but \( u_r \not\in W^u_\tilde{\eta}(0) \). It follows from \( u_r \not\in W^u_\tilde{\eta}(0) \) that

\[\|\alpha(u_r)\| > \tilde{\eta}.\]

(29)

From (28), \( u_r \in U(r) \subseteq U(\tilde{r}) \), for all \( r > \tilde{r} \), and thus we have, for all \( r > \tilde{r} \),

\[
\tilde{\delta} \geq \delta_r \geq \beta_j(u_r) + \rho(u_r, \tilde{v}, r) = \beta_j(u_r) + \rho(u_r, \tilde{v}, \tilde{r}) + \Delta_j(u_r, \tilde{v}, r).
\]

Furthermore, from (1) and \( u_r \in U(\tilde{r}) \), we have

\[
\tilde{\delta} \geq \tilde{\psi}(\tilde{v}, \tilde{r}) + \Delta_j(u_r, \tilde{v}, r).
\]

This, together with (29) and \( u_r \in U(r) \), we have

\[
\tilde{\delta} \geq \tilde{\psi}(\tilde{v}, \tilde{r}) + \inf\{\Delta_j(u, \tilde{v}, r) \mid u \in U(r), \|\alpha(u)\| \geq \tilde{\eta}\}.
\]

(30)

As \( \tilde{\psi}(\tilde{v}, \tilde{r}) > -\infty \), by (A4), the formula (30) deduces a contradiction. \( \square \)

Corollary 1. Suppose that the assumptions of Theorem 1 hold, (A4) holds, and there exists \( (\tilde{v}, \tilde{r}) \in V \times R_{++} \) such that

\[
\tilde{\psi}(\tilde{v}, \tilde{r}) < -\infty.
\]

If, for all \( \varepsilon > 0 \), there exists a neighborhood \( W^u_\eta(0) \) of \( 0 \in R^m \) such that

\[
\beta_j(0) - \varepsilon \leq \beta_j(u) + \rho(u, \tilde{v}, \tilde{r}), \quad \forall u \in W^u_\eta(0) \cap U(\tilde{r}),
\]

(31)

then the zero duality gap (2) holds.

Proof. We only need to prove that (B1) and (B2) hold. For all \( \varepsilon > 0 \), there exists a neighborhood \( W^u_\eta(0) \) of \( 0 \in R^m \) such that (31) holds. Notice that \( \eta \) depends on \( \varepsilon \) in (31). Thus, for this \( \eta > 0 \), Lemma 1 implies that there exists \( r_\varepsilon \geq \tilde{r} \) such that

\[
N(\tilde{v}, r_\varepsilon, \delta) \subseteq W^u_\eta(0) \cap U(r_\varepsilon), \quad \forall \delta \in (\tilde{\psi}(\tilde{v}, r_\varepsilon), \tilde{\delta}]\]

(32)

From \( U(r_\varepsilon) \subseteq U(\tilde{r}) \), the monotonic nondecreasing property of \( \rho \) on \( r > 0 \) and (31), we have

\[
\beta_j(0) - \varepsilon \leq \beta_j(u) + \rho(u, \tilde{v}, \tilde{r}), \quad \forall u \in W^u_\eta(0) \cap U(r_\varepsilon).
\]

(33)

This shows that (B1) holds. From (32) and (33), we have, for all \( u \in N(\tilde{v}, r_\varepsilon, \delta) \),

\[
\beta_j(0) - \varepsilon \leq \beta_j(u) + \rho(u, \tilde{v}, r_\varepsilon) \leq \delta, \quad \forall \delta \in (\tilde{\psi}(\tilde{v}, r_\varepsilon), \tilde{\delta}]\]

(34)

It follows from (A1), (A2), and (A3), we have

\[
\inf\{\Delta_j(u, \tilde{v}, r_\varepsilon) \mid u \in N(\tilde{v}, r_\varepsilon, \delta + \varepsilon)\} = 0, \quad \forall \delta \in (\tilde{\psi}(\tilde{v}, r_\varepsilon), \tilde{\delta}]\]

Thus (B2) holds. \( \square \)

Definition 3. Let \( \alpha(u) \) be a variable substitution for \( u \). Then \( f : R^m_1 \to R \) is said to be \( \alpha \)-lower semicontinuous at \( u = 0 \) if

\[
\liminf_{\alpha(u) \to 0} f(u) \geq f(0).
\]

Similarly, the \( \alpha \)-upper semicontinuity of \( f \) at \( u = 0 \) can be defined. We say that the function \( f \) is \( \alpha \)-continuous at \( u = 0 \), if

\[
\lim_{\alpha(u) \to 0} f(u) = f(0).
\]
Obviously, when \( \alpha(u) = u \), the \( \alpha \)-continuity (\( \alpha \)-semicontinuity) of \( f \) at \( u = 0 \) is equivalent to continuity (semicontinuity) of \( f \) at \( u = 0 \). When \( \alpha(u) = u^m \), the \( \alpha \)-continuity (\( \alpha \)-semicontinuity) of \( f \) at \( u = 0 \) implies that \( f \) is continuous (semicontinuous) at \( u = 0 \). But the reverse is in general not true. Simple counterexamples can be constructed easily. Details are not given here.

**Corollary 2.** Suppose that the assumptions of Theorem 1 hold, \((A_4)\) holds, and there exists \((\tilde{v}, \tilde{r}) \in V \times R_{++}\) such that

\[
\tilde{\psi}(\tilde{v}, \tilde{r}) > -\infty.
\]

Then

(i) if \( \rho(\cdot, \tilde{v}, \tilde{r}) \) and \( \beta_f(\cdot) \) are \( \alpha \)-lower semicontinuous at \( u = 0 \), then the zero duality gap (2) holds;

(ii) let the zero duality gap (2) hold. If, for all \((u, r) \in V \times R_{++}, \rho(\cdot, v, r) \) is \( \alpha \)-continuous at \( u = 0 \), then \( \beta_f(\cdot) \) is \( \alpha \)-lower semicontinuous at \( u = 0 \).

**Proof.** (i) Assume that \( \rho(\cdot, \tilde{v}, \tilde{r}) \) and \( \beta_f(\cdot) \) are \( \alpha \)-lower semicontinuous at \( u = 0 \). Then for all \( \varepsilon > 0 \), there exists a neighborhood \( W^u_\varepsilon(0) \) of \( 0 \in R^n \) such that \((31)\) holds. By Corollary 1, we know that (i) holds.

(ii) Now, let \( M_p = M_p \). By Theorem 1, for all \( \varepsilon > 0 \), \((B^\varepsilon_1)\) holds. By the \( \alpha \)-continuity of \( \rho(\cdot, v, r) \) at \( u = 0 \) and \((B^\varepsilon_2)\), we get that \( \beta_f(\cdot) \) is \( \alpha \)-lower semicontinuous at \( u = 0 \).

From Corollary 2(i) and the fact that \((A_3)\) always holds when \( \alpha(u) = u \), we have the following result.

**Corollary 3.** Assume that there exists \((\tilde{v}, \tilde{r}) \in V \times R_{++}\) such that

\[
\tilde{\psi}(\tilde{v}, \tilde{r}) > -\infty,
\]

that \( \rho \) satisfies \((A_1), (A_3), \) and \((A_4)\) with \( \alpha(u) = u \) and that \( \rho(\cdot, \tilde{v}, \tilde{r}) \) and \( \beta_f(\cdot) \) are lower semicontinuous at \( u = 0 \). Then the zero duality gap property (2) holds.

**Remark 4.** Corollary 3 becomes Theorem 3.2 of Burachik et al. [6], when \( U(r) = R^m \) and \( -\rho \) replaces \( \rho \), apart from the fact that \( \rho(\cdot, \tilde{v}, \tilde{r}) \) is only required to be lower semicontinuous at \( 0 \) in our corollary. However, \( -\rho(\cdot, y, r) \) was required to be lower semicontinuous at \( 0 \) for any \((y, r) \in Y \times R_+\) in \((C_1)\) of Theorem 3.2 of Burachik et al. [6].

By Proposition 1, the dual function \( \tilde{\psi}(v, \cdot) \) of \((P)\) is monotonically nondecreasing on \( r > 0 \). Hence, \( \lim_{r \to +\infty} \tilde{\psi}(v, r) \) exists (finite or +\( \infty \)). In the following, we present a necessary and sufficient condition for this limit to coincide with the optimal value \( \beta_f(0) \) of \((P)\). This result provides a theoretical foundation for the global convergence of an augmented Lagrangian method of \((P)\) by using the function \( l \).

**Lemma 2.** Suppose that the assumptions of Theorem 1 hold, \((A_4)\) holds, and there exists \((\tilde{v}, \tilde{r}) \in V \times R_{++}\) such that

\[
\tilde{\psi}(\tilde{v}, \tilde{r}) > -\infty.
\]

If \( \rho(\cdot, \tilde{v}, \tilde{r}) \) is \( \alpha \)-lower semicontinuous at \( u = 0 \), then

\[
\liminf_{\alpha(u) = 0} \beta_f(u) \leq \lim_{r \to +\infty} \tilde{\psi}(\tilde{v}, r).
\]

(35)

Furthermore, if for all \( r > 0 \), \( \rho(\cdot, \tilde{v}, \tilde{r}) \) is \( \alpha \)-continuous at \( u = 0 \), then (35) holds as an equality.

**Proof.** By the assumptions and the monotonic nondecreasing property of \( \tilde{\psi}(v, \cdot) \) on \( r > 0 \), we have

\[
\tilde{\psi}(\tilde{v}, r) > -\infty, ~ \forall r \geq \tilde{r}.
\]

Noting that

\[
\tilde{\psi}(\tilde{v}, r) = \inf_{u \in U(r)} \{ \beta_f(u) + \rho(u, \tilde{v}, \tilde{r}) \},
\]

then, from Lemma 1, for \( \eta_k \downarrow 0 \) (as \( k \to +\infty \)), there exist \( r_k \uparrow +\infty \) (as \( k \to +\infty \)) and \( u_k \in U(r_k) \) such that

\[
\beta_f(u_k) + \rho(u_k, \tilde{v}, r_k) \leq \tilde{\psi}(\tilde{v}, r_k) + \eta_k := \delta_k,
\]

(36)

and

\[
u_k \in N(\tilde{v}, r_k, \delta_k) \subseteq W^u_\eta(0) \cap U(r_k).
\]
By (37), we have
\[ \lim_{k \to +\infty} \alpha(u_k) = 0. \tag{38} \]

For \( r_k \geq \bar{r} \), we know that \( U(r_k) \subseteq U(\bar{r}) \). Hence, from \((A_1), (A_2), \) and \((A_3) \), we have
\[ \rho(u_k, \bar{v}, r_k) \leq \rho(u_k, \bar{v}, \bar{r}). \tag{39} \]

By (36) and (39), we have
\[ \bar{\psi}(\bar{v}, r_k) + \eta_k \geq \beta_j(u_k) + \rho(u_k, \bar{v}, r_k) \]
\[ \geq \beta_j(u_k) + \rho(u_k, \bar{v}, \bar{r}). \]

Since \( \rho(\cdot, \bar{v}, \bar{r}) \) is \( \alpha \)-lower semicontinuous at \( u = 0 \), by \((A_4), (38) \), and the inequality above, we have
\[ \liminf_{\alpha(u) \to 0} \beta_j(u) \leq \liminf_{\alpha(u_k) \to 0} \beta_j(u_k) \]
\[ \leq \liminf_{\alpha(u_k) \to 0} \beta_j(u_k) + \liminf_{\alpha(u_k) \to 0} \rho(u_k, \bar{v}, \bar{r}) \]
\[ \leq \liminf_{\alpha(u) \to 0} \{ \beta_j(u_k) + \rho(u_k, \bar{v}, \bar{r}) \} \]
\[ \leq \lim_{r \to +\infty} \bar{\psi}(\bar{v}, r). \]

So (35) holds. Assume now that, for all \( r > 0 \), \( \rho(\cdot, \bar{v}, r) \) is \( \alpha \)-continuous at \( u = 0 \). As such, by
\[ \bar{\psi}(\bar{v}, r) \leq \beta_j(u) + \rho(u, \bar{v}, r), \quad \forall u \in U(r), \]

combining the \( \alpha \)-continuity of \( \rho(\cdot, \bar{v}, r) \) at \( u = 0 \) with \((A_4) \), we have that, for all \( r > 0 \),
\[ \bar{\psi}(\bar{v}, r) \leq \liminf_{\alpha(u) \to 0} \{ \beta_j(u) + \rho(u, \bar{v}, r) \} \]
\[ = \liminf_{\alpha(u) \to 0} \beta_j(u). \]

Hence,
\[ \lim_{r \to +\infty} \bar{\psi}(\bar{v}, r) \leq \liminf_{\alpha(u) \to 0} \beta_j(u). \]

By (35) again, the equality in (35) holds. \( \Box \)

In the following, we present necessary and sufficient conditions for the following equality to hold:
\[ \lim_{r \to +\infty} \bar{\psi}(\bar{v}, r) = \beta_j(0). \tag{40} \]

**Theorem 2.** Suppose that the assumptions of Theorem 1 hold, \((A_4) \) holds, and there exists \( (\bar{v}, \bar{r}) \in V \times R_{++} \) such that
\[ \bar{\psi}(\bar{v}, \bar{r}) > -\infty. \]

Then
(i) if \( \rho(\cdot, \bar{v}, \bar{r}) \) and \( \beta_j(\cdot) \) are \( \alpha \)-lower semicontinuous at \( u = 0 \), then (40) holds;
(ii) assume that (40) holds. If, for all \((v, r) \in V \times R_{++}, \rho(\cdot, v, r) \) is \( \alpha \)-continuous at \( u = 0 \), then \( \beta_j(\cdot) \) is \( \alpha \)-lower semicontinuous at \( u = 0 \).

**Proof.** (i) Assume that \( \rho(\cdot, \bar{v}, \bar{r}) \) and \( \beta_j(\cdot) \) are \( \alpha \)-lower semicontinuous at \( u = 0 \). Then Lemma 2 implies that (37) holds. Hence, we have
\[ \beta_j(0) \leq \liminf_{\alpha(u) \to 0} \beta_j(u) \leq \lim_{r \to +\infty} \bar{\psi}(\bar{v}, r) \leq M_D. \tag{41} \]

Then by the weak duality property and (41), (40) holds.

(ii) Now we assume that (40) holds and, for all \((v, r) \in V \times R_{++}, \rho(\cdot, v, r) \) is \( \alpha \)-continuous at \( u = 0 \). By the weak duality property and (40), the zero duality gap property (2) holds. Thus Corollary 2(ii) implies that \( \beta_j(\cdot) \) is \( \alpha \)-lower semicontinuous at \( u = 0 \). \( \Box \)
5. Exact penalty representation. In this section, we give a necessary and sufficient condition for an exact penalty representation and show that some corollaries of this condition include several recent results in the literature as special cases.

**Definition 4.** A vector \( \bar{u} \in V \) is said to support an exact penalty representation of (P) if there exists \( \bar{r} \in R_{++} \) such that

\[
\beta_j(0) = \inf_{x \in R^n} l(x, \bar{u}, r), \quad \forall \; r \geq \bar{r},
\]

\[
\arg \min_{x \in R^n} \varphi(x) = \arg \min_{x \in R^n} l(x, \bar{u}, r), \quad \forall \; r \geq \bar{r}.
\]

A necessary and sufficient condition for \( \bar{u} \) to support an exact penalty representation of (P) can be obtained from the \( \varepsilon \)-perturbation conditions \((B_1^\varepsilon)\) and \((B_2^\varepsilon)\) when setting \( \varepsilon = 0 \), \( \bar{u} = u_0 \) and \( \bar{r} = r_0 \). Hence, these conditions are presented as follows.

There exist \( \bar{r} \in R_{++} \) and a neighborhood \( W^{\alpha}_\eta(0) \) of \( 0 \in R^n \) such that

\((B_1^\varepsilon)\) \( \inf \{ \Delta_\varepsilon(u, \bar{u}, r) \mid u \in N(\bar{u}, r, \delta) \} = 0, \forall \; r \geq \bar{r}, \delta > \tilde{\psi}(\bar{u}, r) \)

\((B_2^\varepsilon)\) \( \beta_j(0) \leq \beta_j(u) + \rho(u, \bar{u}, \bar{r}), \forall \; u \in W^{\alpha}_\eta(0) \cap U(\bar{r}) \).

In the following, we establish that \((B_1^\varepsilon)\) and \((B_2^\varepsilon)\) are the necessary and sufficient condition for an exact penalty representation.

**Lemma 3.** Suppose that the assumptions of Theorem 1 hold and that there exists \( (\bar{u}, \bar{r}) \in V \times R_{++} \) such that \((42)\) holds. If, for all \( x \in R^n \) and \( r \geq \bar{r} \), \( \bar{f}(x, \cdot) \) and \( \rho(\cdot, \bar{u}, \bar{r}) \) are \( \alpha \)-lower semicontinuous at \( u = 0 \). Then \((43)\) holds for all \( r > \bar{r} \).

**Proof.** Let \( (\bar{u}, \bar{r}) \in V \times R_{++} \) satisfy \((42)\). If \( \beta_j(0) = -\infty \), then

\[
\arg \min_{x \in R^n} \varphi(x) = \arg \min_{x \in R^n} l(x, \bar{u}, r) = \emptyset, \quad \forall \; r \geq \bar{r}.
\]

Now we assume that \( \beta_j(0) > -\infty \). From \((42)\) and the weak duality property, it is easy to show that, for all \( r \geq \bar{r} \),

\[
\arg \min_{x \in R^n} \varphi(x) \subseteq \arg \min_{x \in R^n} l(x, \bar{u}, r).
\]

Let \( r > \bar{r} \). We now show that the reverse inclusion relation of \((44)\) holds. Let \( x^*(r) \in \arg \min_{x \in R^n} l(x, \bar{u}, r) \).

It follows from \((42)\) that

\[
\beta_j(0) = \min_{x \in R^n} l(x, \bar{u}, r)
\]

\[
= \bar{f}(x^*(r), \bar{u}, r)
\]

\[
= \inf_{u \in U(\bar{r})} \{ \bar{f}(x^*(r), u) + \rho(u, \bar{u}, \bar{r}) \}.
\]

Choose \( u_k \in U(\bar{r}) \) such that

\[
\lim_{k \to +\infty} \{ \bar{f}(x^*(r), u_k) + \rho(u_k, \bar{u}, \bar{r}) \} = \inf_{u \in U(\bar{r})} \{ \bar{f}(x^*(r), u) + \rho(u, \bar{u}, \bar{r}) \}.
\]

The above equation, together with \((45)\), \((46)\), and \( U(\bar{r}) \subseteq U(\bar{r}) \), yields that

\[
\beta_j(0) = \lim_{k \to +\infty} \{ \bar{f}(x^*(r), u_k) + \rho(u_k, \bar{u}, \bar{r}) \}
\]

\[
= \lim_{k \to +\infty} \{ \bar{f}(x^*(r), u_k) + \rho(u_k, \bar{u}, \bar{r}) + \Delta_\varepsilon(u_k, \bar{u}, \bar{r}) \}
\]

\[
\geq \limsup_{k \to +\infty} \{ \inf_{u \in U(\bar{r})} \{ \bar{f}(x^*(r), u) + \rho(u, \bar{u}, \bar{r}) \} + \Delta_\varepsilon(u_k, \bar{u}, \bar{r}) \}
\]

\[
\geq \limsup_{k \to +\infty} \{ \bar{\psi}(\bar{u}, \bar{r}) + \Delta_\varepsilon(u_k, \bar{u}, \bar{r}) \}
\]

\[
= \bar{\psi}(\bar{u}, \bar{r}) + \limsup_{k \to +\infty} \Delta_\varepsilon(u_k, \bar{u}, \bar{r})
\]

\[
= \beta_j(0) + \limsup_{k \to +\infty} \Delta_\varepsilon(u_k, \bar{u}, \bar{r}).
\]
Noting that $\Delta_r(u_k, \tilde{v}, r) \geq 0 \forall k$, it follows from the relations above that

$$\lim_{k \to +\infty} \Delta_r(u_k, \tilde{v}, r) = 0.$$ 

From (A3), we obtain that

$$\lim_{k \to +\infty} \alpha(u_k) = 0. \quad (47)$$

By the fact that $\tilde{f}(x^*(r), \cdot)$ and $\rho(\cdot, \tilde{v}, r)$ are $\alpha$-lower semicontinuous at $u = 0$, it follows from (45)–(47) and condition (A1) that

$$\begin{align*}
\inf_{x \in \mathbb{R}^n} \varphi(x) &= \lim_{k \to +\infty} \{\tilde{f}(x^*(r), u_k) + \rho(u_k, \tilde{v}, r)\} \\
&\geq \lim_{\alpha(u_k) \to 0} \inf_{x \in \mathbb{R}^n} \tilde{f}(x^*(r), u_k) + \lim_{\alpha(u_k) \to 0} \inf_{x \in \mathbb{R}^n} \rho(u_k, \tilde{v}, r) \\
&\geq \tilde{f}(x^*(r), 0) = \varphi(x^*(r)).
\end{align*}$$

That is, $x^*(r) \in \arg\min_{x \in \mathbb{R}^n} \varphi(x)$. Thus (43) holds. □

**Theorem 3.** Suppose that the assumptions of Theorem 1 hold. Then the following two statements hold.

(i) If $\tilde{v} \in V$ supports an exact penalty representation of (P), then there exists $\tilde{r} \in R_{++}$ such that conditions (B1) and (B2) hold;

(ii) Assume that there exists $(\tilde{v}, \tilde{r}) \in V \times R_{++}$ such that, for all $x \in \mathbb{R}^n$ and $r \geq \tilde{r}$, $\tilde{f}(x, \cdot)$ and $\rho(\cdot, \tilde{v}, r)$ are $\alpha$-lower semicontinuous at $u = 0$. If conditions (B1) and (B2) hold, then $\tilde{v}$ supports an exact penalty representation of (P).

**Proof.** (i) Let $\tilde{v}$ support an exact penalty representation of (P). Then there exists $\tilde{r} \in R_{++}$ such that (42) holds, that is,

$$\beta_f(0) = \tilde{\psi}(\tilde{v}, r), \quad \forall r \geq \tilde{r}.\tag{48}$$

Thus, for all $r \geq \tilde{r}$, we have

$$\beta_f(0) \leq \beta_f(u) + \rho(u, \tilde{v}, r), \quad \forall u \in U(r). \tag{49}$$

In particular, when $r = \tilde{r}$, (48) implies that condition (B2) holds for $W_u^\alpha(0) \cap U(\tilde{r})$. From (48) and condition (A1), we obtain that $0 \in N(\tilde{v}, r, \delta)$, for all $\delta > \tilde{\psi}(\tilde{v}, r)$. Thus, for all $r \geq \tilde{r}$, it follows from (A1), (A2), and (A3) that

$$\inf\{\Delta_r(u, \tilde{v}, r) \mid u \in N(\tilde{v}, r, \delta)\} = 0, \quad \forall \delta > \tilde{\psi}(\tilde{v}, r).$$

Then condition (B1) holds.

(ii) Now we assume that conditions (B1) and (B2) hold. Let $r > \tilde{r}$, $\delta_k \downarrow \tilde{\psi}(\tilde{v}, r)$ (as $k \to +\infty$). From condition (B1), there exists $\{u_k\}$ satisfying

$$\begin{align*}
u_k &\in N(\tilde{v}, r, \delta_k), \quad (49) \\
\lim_{k \to +\infty} \Delta_r(u_k, \tilde{v}, r) &= 0. \quad (50)
\end{align*}$$

From (50) and condition (A3), we obtain that

$$\lim_{k \to +\infty} \alpha(u_k) = 0. \quad (51)$$

Thus, from (51) and $U(r) \subseteq U(\tilde{r})$, when $k$ is sufficiently large, $u_k \in W_u^\alpha(0) \cap U(\tilde{r})$. As such, it follows from (B2), (49), and the monotonic nondecreasing property of $\rho$ on $r > 0$ that

$$\begin{align*}
\delta_k &\geq \beta_f(u_k) + \rho(u_k, \tilde{v}, r) \\
&\geq \beta_f(u_k) + \rho(u_k, \tilde{v}, r) \\
&\geq \beta_f(0).
\end{align*}$$

Then

$$\tilde{\psi}(\tilde{v}, r) = \lim_{k \to +\infty} \delta_k \geq \beta_f(0).$$

So combining with the weak duality property, we obtain (42). Again, by the assumptions of (ii), (42) and Lemma 3, we see that (43) holds. □
From Theorem 3 and under the assumptions $(A_1)$, $(A_2)$, $(A_3)$, and $(A_4)$, we can obtain a necessary and sufficient condition for an exact penalty representation of $(P)$.

**Corollary 4.** Assume that $\rho$ satisfies conditions $(A_1)$, $(A_2)$, $(A_3)$, and $(A_4)$ and that there exists $(\bar{v}, \bar{r}) \in V \times R_{++}$ such that 
\[
\psi(\bar{v}, \bar{r}) > -\infty.
\]
If furthermore, for all $x \in R^n$ and $r \geq \bar{r}$, $f(x, \cdot)$ and $\rho(\cdot, \bar{v}, r)$ are $\alpha$-lower semicontinuous at $u = 0$, then $\bar{v}$ supports an exact penalty representation of $(P)$ if and only if condition $(B_2^\alpha)$ holds.

**Proof.** It is obvious that the necessity can be obtained from Theorem 3(i). So we only need to prove the sufficiency. Without loss of generality, assume that $(B_2^\alpha)$ holds for $\bar{r}$ and a neighborhood $W^\alpha_\eta(0)$ of $0 \in R^m$. That is,
\[
\beta_j(0) \leq \beta_j(u) + \rho(u, \bar{v}, r), \quad \forall u \in W^\alpha_\eta(0) \cap U(\bar{r}).
\]
For $\eta > 0$, by Lemma 1, there exists $r_\eta \geq \bar{r}$ such that 
\[
N(\bar{v}, r_\eta, \delta) \subseteq W^\alpha_\eta(0) \cap U(r_\eta), \quad \forall \delta \in (\psi(\bar{v}, r_\eta), \bar{\delta}],
\]
where $\bar{\delta} > M_\rho$. Noticing that $U(r_\eta) \subseteq U(\bar{r})$ and $\rho$ is monotonically nondecreasing on $r > 0$, by (52), we have 
\[
\beta_j(0) \leq \beta_j(u) + \rho(u, \bar{v}, r)
\]
\[
\leq \beta_j(u) + \rho(u, \bar{v}, r_\eta), \quad \forall u \in W^\alpha_\eta(0) \cap U(r_\eta).
\]
Inequality (54) indicates that condition $(B_2^\alpha)$ holds for $(\bar{v}, r_\eta)$. Now let $r \geq r_\eta$. It follows from $\psi(\bar{v}, r) \geq \psi(\bar{v}, r_\eta)$, the monotonically nondecreasing of $\rho$ on $r > 0$ and (53) that 
\[
N(\bar{v}, r, \delta) \subseteq N(\bar{v}, r_\eta, \delta) \subseteq W^\alpha_\eta(0) \cap U(r_\eta), \quad \forall \delta \in (\psi(\bar{v}, r), \bar{\delta}],
\]
where $\bar{\delta} > M_\rho$. Let $r \geq r_\eta$, $\delta > \psi(\bar{v}, r)$, and $\delta > \psi(\bar{v}, r)$ at $u = 0$. From Theorem 3(ii), the conclusion can be obtained. □

From Corollary 4, we have the following results.

**Corollary 5.** Assume that $U(r) = R^m$, $\rho$ satisfies conditions $(A_1)$, $(A_3)$, and $(A_4)$ at $u(a) = u$ and there exists $(\bar{v}, \bar{r}) \in V \times R_{++}$ such that 
\[
\psi(\bar{v}, \bar{r}) > -\infty.
\]
If, for all $x \in R^n$ and $\forall r \geq \bar{r}$, $f(x, \cdot)$ and $\rho(\cdot, \bar{v}, r)$ are both lower semicontinuous at $u = 0$, then $\bar{v}$ supports an exact penalty representation of $(P)$ if and only if there exist $r' \in R_{++}$ and a neighborhood $W^\alpha_\eta(0)$ of $0 \in R^m$ such that 
\[
\beta_j(0) \leq \beta_j(u) + \rho(u, \bar{v}, r'), \quad \forall u \in W^\alpha_\eta(0).
\]

**Remark 5.** In a Euclidean space, Burachik et al. [6, Theorem 3.3] can be derived from Corollary 5 by replacing $\rho$ with $-\rho$. It is worth noting that perturbation function $\beta$ is assumed to be lower semicontinuous at $u = 0$ in Burachik et al. [6, Theorem 3.3], however, this hypothesis is not required in Corollary 5.

**Remark 6.** In a Euclidean space, Burachik and Rubinov [5, Theorem 6.2], which is a special case of Corollary 5, missed a hypothesis, that is, the dualizing parameterization function $f(\cdot, \cdot)$ is lower semicontinuous on the whole space for all $x \in X$, because this hypothesis was used in the proof of Theorem 6.2 of Burachik and Rubinov [5]. Similarly, the special cases, Huang and Yang [11, Theorem 3.2] and Rubinov and Yang [20, Theorem 5.9], of Corollary 5 also missed this hypothesis. It is obvious that this hypothesis is stronger than the corresponding hypothesis of Corollary 5, where it is only required that $f(\cdot, \cdot)$ is lower semicontinuous at $u = 0$. 

---

*Wang, Yang, and Yang: Nonlinear Augmented Lagrangian and Duality Theory*  
In the study of an augmented Lagrangian dual problem in Nedich and Ozdaglar [13], conditions for the existence of an exact penalty representation were not given. But in Nedich and Ozdaglar [13] the authors pointed out that it is significant to further study this problem (see Nedich and Ozdaglar [13, p. 605]). In the following, we will study the existence of an exact penalty representation in the framework of the augmented Lagrangian dual of \( P \).

The nonlinear augmenting penalty function in Nedich and Ozdaglar [13] is \( (1/r)\sigma(ru) \), see (12). We know that another nonlinear augmenting penalty function can be given as

\[
\rho(u, v, r) = \langle u, v \rangle + \frac{1}{r} \sigma(ru), \quad (u, v, r) \in U(r) \times \{0\} \times \mathbb{R}_+, 
\]

where \( 0 \in \mathbb{R}^m, \sigma \) is convex on an open set \( C \subseteq \mathbb{R}^m \) with \( \sigma(0) = 0 \), and \( U(\cdot) \) is defined by (8). Thus, the exact penalty representation of \( P \) with \( \rho(u, r) = (1/r)\sigma(ru) \) (that is, (42) and (43) hold with \( \tilde{v} \) not appearing explicitly) is equivalent to that \( \tilde{v} = 0 \) supports an exact penalty representation. In this case,

\[
N(0, r, \delta) = \left\{ u \in U(r) \middle| \beta_f(u) + \frac{1}{r} \sigma(ru) \leq \delta \right\}.
\]

We will make our discussions in two cases as follows.

Case 1. The function \( \sigma \) is strictly convex on \( C \).

From Example 1 (R1), we have that \( \rho \) satisfies \((A_1)\) and \((A_2)\) for \( \alpha(u) = u \). By the convexity of \( \sigma \), it follows that \( \rho(\cdot, 0, r) \) is continuous on \( U(r) \) for all \( r \in \mathbb{R}_+ \). Therefore, a necessary and sufficient condition for an exact penalty representation of \( P \) with this type of function \( \rho(u, 0, r) = (1/r)\sigma(ru) \) can be obtained by Theorem 3 as follows.

**Corollary 6.** Let \( \rho(u, 0, r) = (1/r)\sigma(ru) \), where \( (u, r) \in U(r) \times \mathbb{R}_+ \), \( \sigma \) is a strictly convex function on \( C \) with \( \sigma(0) = 0 \) and \( C \subseteq \mathbb{R}^m \) is an open convex set with \( 0 \in C \), \( U(\cdot) \) is defined by (8). If \( \tilde{f}(x, \cdot) \) is lower semicontinuous at \( u = 0 \), then \( \tilde{v} = 0 \) supports an exact penalty representation of \( P \) if and only if there exist \( \tilde{r} \in \mathbb{R}_+ \) and a neighborhood \( W^\rho(0) \) of \( 0 \) in \( \mathbb{R}^m \) such that

(i) \( \inf \{(1/r)\sigma(ru) - (1/r)\sigma(\tilde{r}u) \mid u \in N(0, r, \delta)\} = 0 \), \( \forall r \geq \tilde{r}, \delta > \tilde{\psi}(0, r) \);

(ii) \( \beta_f(0) \leq \beta_f(u) + (1/\tilde{r})\sigma(\tilde{r}u), \forall u \in W^\rho(0) \cap U(\tilde{r}) \).

Case 2. The function \( \sigma \) is nonnegative on \( C = \mathbb{R}^m \) and satisfies conditions \((p_1)\) and \((p_2)\).

From Example 2 (R2), we have that \( \rho \) satisfies \((A_1)\), \((A_2)\), and \((A_3)\) for \( \alpha(u) = u^+ \). By the nonnegativity of \( \sigma \) and \( \sigma(0) = 0 \), it follows that \( \rho(\cdot, 0, r) \) is \( u^+ \)-lower continuous at \( 0 \in \mathbb{R}^m \) for all \( r \in \mathbb{R}_+ \). Therefore, a necessary and sufficient condition for an exact penalty representation of \( P \) with this type of functions can be obtained by Theorem 3 as follows.

**Corollary 7.** Let \( \rho(u, 0, r) = (1/r)\sigma(ru) \), where \( (u, r) \in \mathbb{R}^m \times \mathbb{R}_+ \), \( \sigma \) is nonnegative on \( \mathbb{R}^m \) and satisfies the following conditions:

\((q_1)\) \( \sigma \) is convex on \( \mathbb{R}^m \) and strictly convex on \( \mathbb{R}^m^+ \);

\((q_2)\) for all \( (u, \tilde{r}) \in \mathbb{R}^m \times \mathbb{R}_+ \) and \( r > \tilde{r} \), there holds

\[
\frac{1}{r} \sigma(ru) - \frac{1}{\tilde{r}} \sigma(\tilde{r}u) \geq \frac{1}{r} \sigma(ru^+) - \frac{1}{\tilde{r}} \sigma(\tilde{r}u^+).
\]

Then \( \tilde{v} = 0 \) supports an exact penalty representation of \( P \) if and only if there exist \( \tilde{r} \in \mathbb{R}_+ \) and a neighborhood \( W^\rho(0) \) of \( 0 \) in \( \mathbb{R}^m \) such that

(i) \( \inf \{(1/r)\sigma(ru) - (1/\tilde{r})\sigma(\tilde{r}u) \mid u \in N(0, r, \delta)\} = 0 \), \( \forall r \geq \tilde{r}, \delta > \tilde{\psi}(0, r) \);

(ii) \( \beta_f(0) \leq \beta_f(u) + (1/\tilde{r})\sigma(\tilde{r}u), \forall u \in W^\rho(0) \).

Similarly, no condition for the exact penalty representation was studied in Nedich and Ozdaglar [14]. Now we present such a condition for \( P \) using Corollary 4. The augmenting penalty function in Nedich and Ozdaglar [14] is

\[
\rho(u, v, r) = \langle u, v \rangle + r\sigma(u), \quad (u, v, r) \in \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}_+.
\]

This is a linear function in \( r \). Assume that the augmenting function \( \sigma \) is nonnegative and convex on \( \mathbb{R}^m_+ \) with \( \sigma(0) = 0 \), and \( \sigma \) satisfies the weak peak at zero condition. It is easy to show that \( \rho \) satisfies \((A_1)\), \((A_2)\), \((A_3)\), and \((A_4)\) for \( \alpha(u) = u^+ \). If \( \tilde{v} \leq 0 \), by the definition of \( \rho \) and the property of \( \sigma \), \( \rho(\cdot, \tilde{v}, r) \) is \( u^+ \)-lower
semicontinuous, for all $r \in R_{++}$. So Corollary 4 implies a necessary and sufficient condition for $\bar{v}$ to support an exact penalty representation of $(P)$ as follows.

**Corollary 8.** Let $p(u, v, r) = (u, v) + r\sigma(u)$, $(u, v, r) \in R^{m_1} \times R^{m_2} \times R_{++}$, where $\sigma$ is nonnegative and convex on $R^{m_1}$ with $\sigma(0) = 0$, and $\bar{v} \leq 0$. If

1. there exists $\bar{r} \in R_{++}$ such that $\psi(\bar{v}, \bar{r}) > -\infty$;
2. the weak peak at zero condition holds; and
3. for any $x \in R^{m_2}$, $f(x, \cdot)$ is $u^\top$-lower semicontinuous at $u = 0$, then $\bar{v}$ supports an exact penalty representation of $(P)$ if and only if there exist $r' \in R_{++}$ and a neighborhood $W^+_n(0)$ of $0 \in R^{m_1}$ such that

$$\beta_f(0) \leq \beta_f(u) + r'\sigma(u), \quad \forall u \in W^+_n(0).$$

**Remark 7.** If we remove the hypothesis of $\sigma$ being convex on $R^{m_1}$, then Corollary 8 still holds.

6. **Conclusions.** In this paper, we investigated an augmented Lagrangian dual framework for a primal problem of minimizing an extended real-valued function, and obtained necessary and sufficient conditions for the zero duality gap and the exact penalty representation. We defined the nonlinear augmenting penalty function on an open set and used a variable substitution for the dual variable. As such our framework includes some barrier penalty functions and the weak peak at zero condition in the literature as special cases. We assumed neither a coercivity assumption nor a lower semicontinuous at zero assumption of the nonlinear augmenting penalty function.

**Acknowledgments.** The authors would like to thank the two referees for their constructive comments and suggestions on the early versions of the paper, which have significantly improved the presentation of the paper. The work in the present paper was partially supported by the Research Grants Council of the Hong Kong Special Administrative Region, China, the National Natural Science Foundation of China [Grants 10831009, 10971118, 10901096, 11271226, 11271233] and the Natural Science Foundation of Chongqing [Grant CSTC, 2011BA0030].

**References**


