J Glob Optim (2008) 40:545–573
DOI 10.1007/s10898-006-9122-0

ORIGINAL PAPER

A geometric framework for nonconvex optimization duality using augmented lagrangian functions

Angelia Nedich · Asuman Ozdaglar

Received: 31 October 2006 / Accepted: 2 November 2006 / Published online: 6 December 2006

Abstract We provide a unifying geometric framework for the analysis of general classes of duality schemes and penalty methods for nonconvex constrained optimization problems. We present a separation result for nonconvex sets via general concave surfaces. We use this separation result to provide necessary and sufficient conditions for establishing strong duality between geometric primal and dual problems. Using the primal function of a constrained optimization problem, we apply our results both in the analysis of duality schemes constructed using augmented Lagrangian functions, and in establishing necessary and sufficient conditions for the convergence of penalty methods.

Keywords Augmented Lagrangian functions · Duality · Penalty

1 Introduction

Duality theory for convex optimization problems using ordinary Lagrangian functions has been well-established. There are many works that treat convex optimization duality, including the books by Rockafellar [12], Hiriart-Urruty and Lemarechal [8], Bonnans and Shapiro [6], Borwein and Lewis [7], Bertsekas et al. [3], and Auslender and Teboulle [2].

It is well known that convex optimization duality is related to the closedness of the epigraph of the primal (or perturbation) function of the optimization problem (see, e.g. [2, 3, 7, 12]). Furthermore, convex optimization duality can be visualized in terms

A. Nedich (✉)
Department of Industrial and Enterprise Systems Engineering, University of Illinois, Urbana, IL, USA
e-mail: angelia@uiuc.edu

A. Ozdaglar
Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, USA
e-mail: asuman@mit.edu

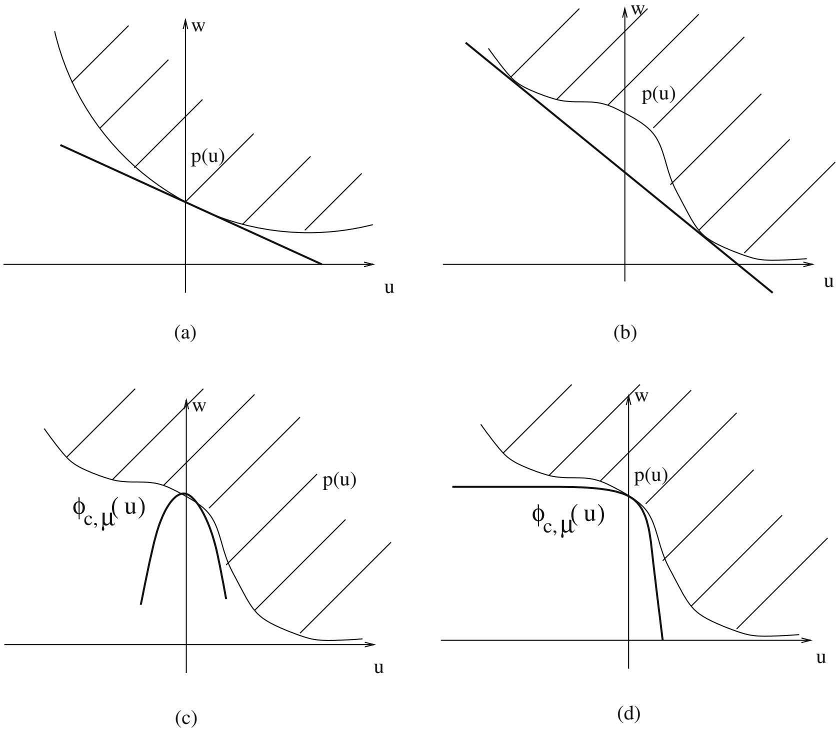

Fig. 1 Supporting the epigraph of the primal function $p(u)$ with affine and nonlinear functions. Parts (**a**) and (**b**) illustrate, respectively, supporting the epigraph of a convex p and a nonconvex p with an affine function. Part (**c**) illustrates supporting the epigraph with a nonlinear level-bounded augmenting function. Part (**d**) illustrates supporting the epigraph with a nonlinear augmenting function with unbounded level sets

of hyperplanes supporting the closure of the convex epigraph of the primal function, as shown in Fig. 1a (see [4]). A duality gap exists when the affine function cannot be pushed all the way up to the minimum common intercept of the epigraph of the primal function with the vertical axis. This situation may clearly arise in the presence of nonconvexities in the "bottom-shape" of the epigraph of the primal function, as shown in Fig. 1b. A duality gap may be avoided if nonlinear functions are used instead of affine functions. Nonlinear functions can penetrate a "possible dent" at the bottom of the epigraph, as shown in Fig. 1c, d.

This idea is introduced by Rockafellar and Wets in their seminal book [13]. In particular, Rockafellar and Wets use convex, nonnegative, and "level-bounded" augmenting functions (i.e., all lower level sets are bounded) to construct augmented Lagrangian functions and to show, under coercivity assumptions, that there is no duality gap between the nonconvex primal problem and the corresponding augmented dual problem. This analysis is extended by Huang and Yang [9] to nonnegative augmenting functions on which no convexity requirement is imposed, again under coercivity assumptions. Rubinov et al. [14] study the zero duality gap property for an augmented dual problem constructed using a family of augmenting functions. In particular, this work considers a family of augmenting functions that satisfy an *almost peak at zero* property, a property less restrictive than the level-boundedness. Neces-

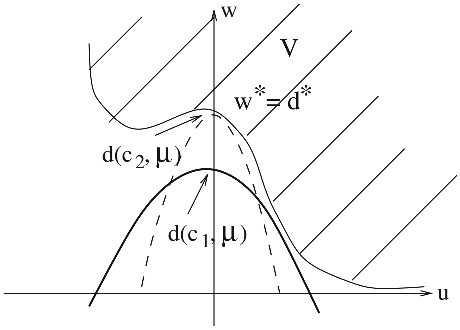

Fig. 2 Geometric primal and dual problems. The figure illustrates the maximum intercept of a concave surface supporting the set V, $d(c, \mu)$ for $c_2 \geq c_1$

sary and sufficient conditions for no duality gap are provided using tools from abstract convexity (see [15, 16]), under the assumption that the augmenting family contains an augmenting function minorizing the primal function.

In this paper, we present a geometric framework to analyze duality for nonconvex optimization problems using nonlinear/augmented Lagrangian functions. We consider convex augmenting functions that need not satisfy level-boundedness assumptions. To capture the essential aspects of nonconvex optimization duality, we consider two simple geometric optimization problems that are defined in terms of a nonempty set $V \subset \mathbb{R}^m \times \mathbb{R}$ that intersects the vertical axis $\{(0, w) \mid w \in \mathbb{R}\}$, i.e., the w-axis. In particular, we consider the following:

- *Geometric primal problem:* We would like to find the minimum value intercept of the set V and the w-axis.
- *Geometric dual problem:* Consider surfaces $\left\{\left(u, \phi_{c,\mu}(u)\right) \mid u \in \mathbb{R}^m\right\}$ that lie below the set V and $\phi_{c,\mu}: \mathbb{R}^m \mapsto \mathbb{R}$ has the following form

$$\phi_{c,\mu}(u) = -c\sigma(u) - \mu' u + \xi,$$

where σ is a convex augmenting function, $c \geq 0$ is a scaling (or penalty) parameter, μ is a vector in $\mathbb{R}^m$, and ξ is a scalar. We would like to find the maximum intercept of such a surface with the w-axis (see Fig. 1c, d).

Figure 1 suggests that the optimal value of the dual problem is no higher than the optimal value of the primal problem, i.e., a primal–dual relation known as the *weak duality*. We are interested in characterizing the primal–dual problems for which the two optimal values are equal, i.e., the problems for which a relation known as the *strong duality* holds. In particular, our objective is to establish necessary and sufficient conditions for strong duality for geometric primal and dual problems. Also, we are interested in the implications for the constrained nonconvex optimization problems from the aspects of Lagrangian relaxation and penalty methods. We show that our geometric approach provides a unified framework for studying both of these aspects.

We consider applications of the strong duality results of the geometric framework to (nonconvex) optimization duality. Given a primal optimization problem, we define the augmented Lagrangian function, which is used to construct the augmented dual problem. The augmenting functions we consider do not require level-bounded assumptions as in Rockafellar and Wets [13] and Huang and Yang [9]. They are related to the peak at zero functions studied by Rubinov et al. [14] (see Sect. 4). Using our results from the geometric framework, we provide necessary and sufficient conditions for the strong duality between the primal optimization problem and the augmented dual problem.

Finally, we consider a general class of penalty methods for constrained optimization problems. We show that our geometric approach can be used to establish the convergence of the optimal values of a sequence of penalized optimization problems to the optimal value of the constrained optimization problem. The convergence behavior of penalty methods is typically analyzed under the assumption that the optimal solution set of the penalized problem is nonempty and compact (see [1, 5, 10, 11). Here, we consider a larger class of penalty functions that need not be continuous, real-valued, or identically equal to 0 over the feasible set (see [10]), and provide necessary and sufficient conditions for convergence without imposing existence or compactness assumptions.

The paper is organized as follows: in Sect. 2, we introduce our geometric approach in terms of a nonempty set V and separating augmenting functions. In Sect. 3, we present some properties of the set V and the augmenting functions, and analyze implications related to separation properties. In Sect. 4, we present our separation theorem. In particular, we discuss sufficient conditions on augmenting functions and the set V that guarantee the separation of this set and a vector $(0, w_0)$ that does not belong to the closure of the set V. In Sect. 5, we use the separation result to provide necessary and sufficient conditions for strong duality for the geometric primal and dual problems. In Sects. 6 and 7, we apply our results to nonconvex optimization duality and penalty methods using the primal function of the constrained optimization problem.

1.1 Notation, terminology, and basics

For a scalar sequence $\{\gamma_k\}$ approaching the zero value monotonically from above, we write $\gamma_k \downarrow 0$. We view a vector as a column vector, and we denote the inner product of two vectors x and y by $x'y$.

For any vector $u \in \mathbb{R}^n$, we can write

$$u = u^+ + u^- \quad \text{with } \ u^+ \geq 0 \quad \text{and } u^- \leq 0,$$

where the vector u^+ is the component-wise maximum of u and the zero vector, i.e.,

$$u^+ = (\max\{0, u_1\}, \dots, \max\{0, u_n\})'$$

and the vector u^- is the component-wise minimum of u and the zero vector, i.e.,

$$u^- = (\min\{0, u_1\}, \dots, \min\{0, u_n\})'.$$

For a function $f : \mathbb{R}^n \mapsto [-\infty, \infty]$, we denote the domain of f by $\mathrm{dom}(f)$, i.e.,

$$\mathrm{dom}(f) = \{x \in \mathbb{R}^n \mid f(x) < \infty\}.$$

We denote the epigraph of f by $\mathrm{epi}(f)$, i.e.,

$$\mathrm{epi}(f) = \{(x, w) \in \mathbb{R}^n \times \mathbb{R} \mid f(x) \leq w\}.$$

For any scalar γ, we denote the (lower) γ-level set of f by $L_f(\gamma)$, i.e.,

$$L_f(\gamma) = \{x \in \mathbb{R}^n \mid f(x) \leq \gamma\}.$$

We say that the function f is *level-bounded* when the set $L_f(\gamma)$ is bounded for every scalar γ.

We denote the closure of a set X by $\mathrm{cl}(X)$. We define a cone K as a set of all vectors x such that $\lambda x \in K$ whenever $x \in K$ and $\lambda \geq 0$. For a given nonempty set X, the cone generated by the set X is denoted by $\mathrm{cone}(X)$ and is given by

$$\mathrm{cone}(X) = \{y \mid y = \lambda x \text{ for some } x \in X \text{ and } \lambda \geq 0\}.$$

When establishing separation results, we use the notion of a recession cone of a set. In particular, the recession cone of a set C is denoted by C^∞ and is defined as follows.

Definition 1 (*Recession cone*) The recession cone C^∞ of a nonempty set C is given by

$$C^\infty = \{d \mid \lambda_k x_k \to d \;\text{ for some } \{x_k\} \subset C \text{ and } \{\lambda_k\} \subset \mathbb{R} \text{ with } \lambda_k \geq 0,\ \lambda_k \to 0\}.$$

A direction $d \in C^\infty$ is referred to as a *recession direction* of the set C.

Some basic properties of a recession cone are given in the following lemma (cf. [13], Sect. 3B).

Lemma 1 (Recession cone properties) *The recession cone C^∞ of a nonempty set C is a closed cone. When C is a cone, we have $C^\infty = (\mathrm{cl}(C))^\infty = \mathrm{cl}(C)$.*

2 Geometric approach

In this section, we introduce primal and dual problems in our geometric framework. We also establish the weak duality relation between the primal optimal and the dual optimal values.

2.1 Geometric primal and dual problems

Consider a nonempty set $V \subset \mathbb{R}^m \times \mathbb{R}$ that intersects the w-axis, i.e., contains at least one point of the form $(0, w)$ with $w \in \mathbb{R}$. The geometric primal problem consists of determining the minimum value intercept of V and the w-axis, i.e.,

$$\inf_{(0,w)\in V} w.$$

We denote the primal optimal value by w^*.

To define the geometric dual problem, we consider the following class of convex augmenting functions.

Definition 2 A function $\sigma\colon \mathbb{R}^m \mapsto (-\infty, \infty]$ is called an *augmenting function* if it is convex, not identically equal to 0, and taking zero value at the origin,

$$\sigma(0) = 0.$$

This definition of augmenting function is motivated by the one introduced by Rockafellar and Wets [13] (see there Definition 11.55). Note that an augmenting function is a proper function.

The geometric dual problem considers surfaces $\left\{\left(u, \phi_{c,\mu}(u)\right) \mid u \in \mathbb{R}^m\right\}$ that lie below the set V. The function $\phi_{c,\mu}\colon \mathbb{R}^m \mapsto \mathbb{R}$ has the following form

$$\phi_{c,\mu}(u) = -c\sigma(u) - \mu' u + \xi,$$

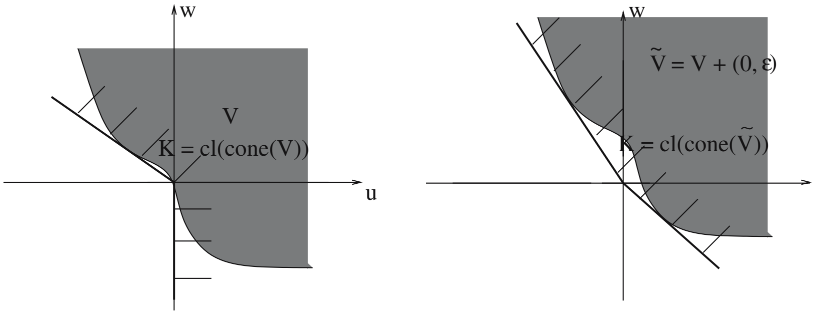

Fig. 3 Illustration of the closure of the cones generated, respectively, by the set V and by the set $\tilde{V}$, which is an upward translation of the set V

where σ is an augmenting function, $c \geq 0$ is a scaling (or penalty) parameter, $\mu \in \mathbb{R}^m$, and ξ is a scalar. This surface can be expressed as $\{(u, w) \in \mathbb{R}^m \times \mathbb{R} \mid w + c\sigma(u) + \mu' u = \xi\}$ and thus intercepts the vertical axis $\{(0, w) \mid w \in \mathbb{R}\}$ at the level ξ. It is below V if and only if

$$w + c\sigma(u) + \mu' u \geq \xi \quad \text{for all} \quad (u, w) \in V.$$

Therefore, among all surfaces that are defined by an augmenting function σ, a scalar $c \geq 0$, and a vector $\mu \in \mathbb{R}^m$ and that support the set V from below, the maximum intercept with the w-axis is given by

$$d(c, \mu) = \inf_{(u,w) \in V} \left\{ w + c\sigma(u) + \mu' u \right\}.$$

The dual problem consists of determining the maximum intercept of such surfaces with the w-axis over $c \geq 0$ and $\mu \in \mathbb{R}^m$, i.e.,

$$\sup_{c \geq 0,\ \mu \in \mathbb{R}^m} d(c, \mu).$$

We denote the dual optimal value by d^*. From the construction of the dual problem, we can see that the dual optimal value d^* does not exceed w^* (see Fig. 2), i.e., the weak duality relation

$$d^* \leq w^*$$

holds. We say that *there is zero duality gap* when $d^* = w^*$, and *there is a duality gap* when $d^* < w^*$.

The weak duality relation is formally established in the following proposition.

Proposition 1 (Weak duality) *The dual optimal value does not exceed the primal optimal value, i.e.,*

$$d^* \leq w^*.$$

Proof For any augmenting function σ, scalar $c \geq 0$, and vector $\mu \in \mathbb{R}^m$, we have

$$d(c, \mu) = \inf_{(u,w) \in V} \{w + c\sigma(u) + \mu' u\} \leq \inf_{(0,w) \in V} w = w^*.$$

Therefore

$$d^* = \sup_{c \geq 0,\ \mu \in \mathbb{R}^m} d(c, \mu) \leq w^*.$$

□

2.2 Separating surfaces

We consider a nonempty set $V \subset \mathbb{R}^m \times \mathbb{R}$ and a vector $(0, w_0)$ that does not belong to the closure of the set V. We are interested in conditions guaranteeing that $\mathrm{cl}(V)$ and the given vector can be separated by a surface $\{(u, \phi_{c,\mu}(u)) \mid u \in \mathbb{R}^m\}$, where the function $\phi_{c,\mu}$ has the form

$$\phi_{c,\mu}(u) = -c\sigma(u) - \mu' u + \xi$$

for an augmenting function σ, vector $\mu \in \mathbb{R}^m$, and scalars $c > 0$ and ξ. In the following sections, we will provide conditions under which the separation can be realized using the augmenting function only, i.e., μ can be taken equal to 0 in the separating concave surface $\{(u, \phi_{c,\mu}(u)) \mid u \in \mathbb{R}^m\}$. In particular, we say that the augmenting function σ *separates* the set V and the vector $(0, w_0) \notin \mathrm{cl}(V)$ when for some $c \geq 0$,

$$w + c\sigma(u) \geq w_0 \quad \text{for all } (u, w) \in V.$$

We also say that the augmenting function σ *strongly separates* the set V and the vector $(0, w_0) \notin \mathrm{cl}(V)$ when for some $c \geq 0$ and $\xi \in \mathbb{R}$,

$$w + c\sigma(u) \geq \xi > w_0 \quad \text{for all } (u, w) \in V.$$

3 Preliminary results

Our focus is on the separation of a given set V from a vector $(0, w_0)$ that does not belong to the closure of the set V, where the separation is realized through some augmenting function. In this section, we establish some properties of the set V and augmenting function σ, which will be essential in our subsequent separation results.

3.1 Properties of the set V

We first present some properties of the set V and establish their implications.

Definition 3 We say that a set $V \subset \mathbb{R}^m \times \mathbb{R}$ *is extending upward in w-space* if for every vector $(\bar{u}, \bar{w}) \in V$, the half-line $\{(\bar{u}, w) \mid w \geq \bar{w}\}$ is contained in V.

Definition 4 We say that a set $V \subset \mathbb{R}^m \times \mathbb{R}$ is *extending upward in u-space* if for every vector $(\bar{u}, \bar{w}) \in V$, the cone $\{(u, \bar{w}) \mid u \geq \bar{u}\}$ is contained in V.

These two properties are satisfied, for example, when the set V is the epigraph of a nonincreasing function. For the sets that are extending upward in w-space, we have the following lemma.

Lemma 2 *Let $V \subset \mathbb{R}^m \times \mathbb{R}$ be a nonempty set. Assume that the set V is extending upward in w-space. Then, we have:*

(a) *The closure* cl(V) *of the set V is also extending upward in w-space.*
(b) *For any vector* $(0, w_0) \notin \mathrm{cl}(V)$, *we have*

$$w_0 < \bar{w}^* = \inf_{(0,w)\in \mathrm{cl}(V)} w \le w^*.$$

Proof

(a) Let $(\bar{u}, \bar{w})$ be a vector that belongs to cl(V), and let w be a scalar such that $w > \bar{w}$. We prove that the vector $(\bar{u}, w)$ belongs to cl(V), which implies that the half-line $\{(\bar{u}, w) \mid w \ge \bar{w}\}$ lies in cl(V).
Let $\{(\bar{u}_k, \bar{w}_k)\}$ be a sequence of vectors in V such that

$$(\bar{u}_k, \bar{w}_k) \to (\bar{u}, \bar{w}).$$

Consider a sequence

$$\{(\bar{u}_k, w_k)\} \quad \text{with} \quad w_k = \bar{w}_k + w - \bar{w} \quad \text{for all} \quad k$$

and note that

$$(\bar{u}_k, w_k) \to (\bar{u}, w).$$

Since $w > \bar{w}$, we have that

$$w_k > \bar{w}_k \quad \text{for all} \quad k.$$

Because $(\bar{u}_k, \bar{w}_k) \in V$ for all k and the half-line $\{(\bar{u}, w) \mid w \ge \bar{w}\}$ lies in the set V for every $(\bar{u}, \bar{w}) \in V$, we further have that

$$(\bar{u}_k, w_k) \in V \quad \text{for all} \quad k.$$

This relation and $(\bar{u}_k, w_k) \to (\bar{u}, w)$ imply that $(\bar{u}, w) \in \mathrm{cl}(V)$, thus showing that the set cl(V) extends upward in w-space.
(b) We show that

$$w_0 < \bar{w}^* = \inf_{(0,w)\in \mathrm{cl}(V)} w \le w^*$$

for any vector $(0, w_0) \notin \mathrm{cl}(V)$. In particular, because $V \subset \mathrm{cl}(V)$, we have $w^* \ge \bar{w}^*$. Furthermore, for a vector $(0, w_0) \notin \mathrm{cl}(V)$, we must have $w_0 \ne \bar{w}^*$. Suppose that $\bar{w}^* < w_0$ for some vector $(0, w_0) \notin \mathrm{cl}(V)$. Then, because the set cl(V) extends upward in w-space, we would have $(0, w_0) \in \mathrm{cl}(V)$—a contradiction. Hence, we must have $w_0 < \bar{w}^*$. □

For sets that are extending upward in u-space, we have the following lemma.

Lemma 3 *Let* $V \subset \mathbb{R}^m \times \mathbb{R}$ *be a nonempty set extending upward in u-space. Then, the closure of the cone generated by the set V is also extending upward in u-space, i.e., for any* $(\bar{u}, \bar{w}) \in \mathrm{cl}(\mathrm{cone}(V))$, *the cone* $\{(u, \bar{w}) \mid u \ge \bar{u}\}$ *is contained in* cl(cone(V)).

Proof Let $(\bar{u}, \bar{w})$ be an arbitrary vector that belongs to cl(cone(V)), and let $u \ge \bar{u}$. Also, let e be an m-dimensional vector with all components equal to 1. We show that for any $\delta > 0$, we have $(u + \delta e, \bar{w}) \in \mathrm{cl}(\mathrm{cone}(V))$, thus implying that $(u, \bar{w}) \in \mathrm{cl}(\mathrm{cone}(V))$ and showing that cl(cone(V)) extends upward in u-space.

Since $(\bar{u}, \bar{w}) \in \mathrm{cl}(\mathrm{cone}(V))$, there exists a sequence of vectors $\{(\bar{u}_k, \bar{w}_k)\} \subset \mathrm{cone}(V)$ converging to $(\bar{u}, \bar{w})$, i.e.,

$$(\bar{u}_k, \bar{w}_k) \to (\bar{u}, \bar{w}) \quad \text{with } (\bar{u}_k, \bar{w}_k) \neq (\bar{u}, \bar{w}) \quad \text{for all} \quad k. \tag{1}$$

Because $\bar{u}_k \to \bar{u}$ and $u + \delta e > \bar{u}$ for any $\delta > 0$, we may assume without loss of generality that

$$\bar{u}_k < (u + \delta e) \quad \text{for all } k.$$

Since $\{(\bar{u}_k, \bar{w}_k)\} \subset \mathrm{cone}(V)$, for all k we have

$$(\bar{u}_k, \bar{w}_k) = \lambda_k (u_k, w_k) \quad \text{for some } (u_k, w_k) \in V \quad \text{and} \quad \lambda_k \geq 0. \tag{2}$$

In view of $(\bar{u}_k, \bar{w}_k) \neq (\bar{u}, \bar{w})$ for all k (cf. Eq. 1), we can assume without loss of generality that $\lambda_k > 0$ for all k. Hence, from the preceding two relations, we obtain

$$u_k = \frac{1}{\lambda_k} \bar{u}_k < \frac{1}{\lambda_k} (u + \delta e) \quad \text{for all } k.$$

The set V extends upward in u-space and therefore, the relations $(u_k, w_k) \in V$ and $u_k < \frac{1}{\lambda_k} (u + \delta e)$ for all k imply that

$$\left(\frac{1}{\lambda_k} (u + \delta e), w_k \right) \in V \quad \text{for all } k.$$

By multiplying this relation with λ_k and using $\lambda_k w_k = \bar{w}_k$ (cf. Eq. 2), we obtain

$$(u + \delta e, \lambda_k w_k) = (u + \delta e, \bar{w}_k) \in \mathrm{cone}(V) \quad \text{for all } k.$$

By letting $k \to \infty$ in the preceding relation and by using $\bar{w}_k \to \bar{w}$ (cf. Eq. 1), we further obtain

$$(u + \delta e, \bar{w}_k) \to (u + \delta e, \bar{w}),$$

implying that $(u + \delta e, \bar{w}) \in \mathrm{cl}(\mathrm{cone}(V))$. Finally, by taking $\delta \to 0$, we obtain $(u, \bar{w}) \in \mathrm{cl}(\mathrm{cone}(V))$. □

The proofs of our separation results require the separation of the half-line $\{(0, w) \mid w \leq \bar{w}\}$ for some $\bar{w} < 0$ and the cone generated by the set V. However, complications may arise when the set V has an infinite slope around the origin, in which case the cone generated by the set V cannot be separated from the half-line $\{(0, w) \mid w \leq \bar{w}\}$ with $\bar{w} < 0$. To avoid such complications without imposing additional restrictions on the set V, we consider a cone generated by a slightly upward translation of the set V in w-space, a set we denote by $\tilde{V}$ (see Fig. 3).

For the separation results, another important characteristic of the set V is the "bottom-shape" of V. In particular, it is desirable that the changes in w are commensurate with the changes in u for $(u, w) \in V$, i.e., the ratio of $\|u\|$ and w-values is asymptotically finite, as w decreases. To characterize this, we use the notion of recession directions and recession cone of a nonempty set (see Sect. 1.1).

In the next lemma, we study the implications of the recession directions of set V on the properties of the cone generated by the set $\tilde{V}$. This result plays a key role in the subsequent development.

Lemma 4 *Let $V \subset \mathbb{R}^m \times \mathbb{R}$ be a nonempty set. Assume that $w^* = \inf_{(0,w)\in V} w$ is finite, and that V extends upward in u-space and w-space. Assume that $(0, -1)$ is not a recession direction of V, i.e.,*

$$(0, -1) \notin V^{\infty}.$$

Let $(0, w_0)$ be a vector that does not belong to $\mathrm{cl}(V)$. For a given $\epsilon > 0$, consider the set $\tilde{V}$ given by

$$\tilde{V} = \{(u, w) \mid (u, w - \epsilon) \in V\} \tag{3}$$

and the cone generated by $\tilde{V}$, denoted by K. Then, the vector $(0, w_0)$ does not belong to the closure of the cone K generated by $\tilde{V}$, i.e.,

$$(0, w_0) \notin \mathrm{cl}(K).$$

Proof According to Lemma 2(b), we have that

$$w_0 < \bar{w}^* = \inf_{(0,w)\in \mathrm{cl}(V)} w \le w^*.$$

Since by assumption w^* is finite, it follows that $\bar{w}^*$ is finite. By using the translation of space along w-axis if necessary, without loss of generality, we may assume that

$$\bar{w}^* = 0, \tag{4}$$

so that

$$w_0 < 0. \tag{5}$$

For a given $\epsilon > 0$, consider the set $\tilde{V}$ defined in Eq. 3 and let K be the cone generated by $\tilde{V}$. To obtain a contradiction, assume that $(0, w_0) \in \mathrm{cl}(K)$. Since $\mathrm{cl}(K)$ is a cone, by Recession Cone Properties (cf. Lemma 1) it follows that $\mathrm{cl}(K) = K^{\infty}$. Therefore, the vector $(0, w_0)$ is a recession direction of K, and therefore, there exist a vector sequence $\{(u_k, w_k)\}$ and a scalar sequence $\{\lambda_k\}$ such that

$$\lambda_k(u_k, w_k) \to (0, w_0) \quad \text{with } \{(u_k, w_k)\} \subset K, \quad \lambda_k \ge 0, \quad \lambda_k \to 0.$$

Since the cone K is generated by the set $\tilde{V}$, for the sequence $\{(u_k, w_k)\} \subset K$ we have

$$(u_k, w_k) = \beta_k(\bar{u}_k, \bar{w}_k), \quad (\bar{u}_k, \bar{w}_k) \in \tilde{V}, \quad \beta_k \ge 0 \quad \text{for all } k.$$

Consider now the sequence $\{\lambda_k\beta_k(\bar{u}_k, \bar{w}_k)\}$. From the preceding two relations, we have $\lambda_k\beta_k(\bar{u}_k, \bar{w}_k) = \lambda_k(u_k, w_k)$ and

$$\lambda_k\beta_k(\bar{u}_k, \bar{w}_k) \to (0, w_0) \quad \text{with } (\bar{u}_k, \bar{w}_k) \in \tilde{V} \quad \text{and} \quad \lambda_k\beta_k \ge 0. \tag{6}$$

Because the set V extends upward in w-space and the set $\tilde{V}$ is a translation of V up w-axis [cf. Eq. (3)], we have that $(u, w) \in V$ for any $(u, w) \in \tilde{V}$. In particular, since $(\bar{u}_k, \bar{w}_k) \in \tilde{V}$ for all k, it follows that

$$(\bar{u}_k, \bar{w}_k) \in V \quad \text{for all } k. \tag{7}$$

In view of $\lambda_k\beta_k \geq 0$, we have $\liminf_{k\to\infty} \lambda_k\beta_k \geq 0$, and there are three cases to consider.

Case 1 $\liminf_{k\to\infty} \lambda_k\beta_k = 0$.

We can assume without loss of generality that $\lambda_k\beta_k \to 0$. Then, from Eq. 6 and the fact $(\bar{u}_k, \bar{w}_k) \in V$ for all k (cf. Eq. 7), it follows that the vector $(0, w_0)$ is a recession direction of the set V. Since $w_0 < 0$ (cf. Eq. 16), this implies that the direction $(0, -1)$ is a recession direction of $V-$ a contradiction.

Case 2 $\liminf_{k\to\infty} \lambda_k\beta_k = \infty$.

In this case, we have $\lim_{k\to\infty} \lambda_k\beta_k = \infty$, and by using this relation in Eq. 6, we obtain

$$\lim_{k\to\infty} (\bar{u}_k, \bar{w}_k) = (0, 0).$$

Therefore $(0, 0) \in \mathrm{cl}(\tilde{V})$, and by the definition of the set $\tilde{V}$ (cf. Eq. 3), it follows that $(0, -\epsilon) \in \mathrm{cl}(V)$. But this contradicts the relation $\bar{w}^* = 0$ (cf. Eq. 4).

Case 3 $\liminf_{k\to\infty} \lambda_k\beta_k = \xi > 0$.

We can assume without loss of generality that $\lambda_k\beta_k \to \xi > 0$ and $\lambda_k\beta_k > 0$ for all k. Then, from Eq. 6 we have for the sequence $\{(\bar{u}_k, \bar{w}_k)\}$,

$$\lim_{k\to\infty} (\bar{u}_k, \bar{w}_k) = \lim_{k\to\infty} \frac{(0, w_0)}{\lambda_k\beta_k} = \frac{1}{\xi}(0, w_0) \quad \text{with} \quad \xi > 0. \tag{8}$$

From the preceding relation, it follows that

$$(\bar{u}_k, \bar{w}_k) \to (0, \tilde{w}) \quad \text{with} \quad \tilde{w} = \frac{w_0}{\xi} < 0.$$

Because $(\bar{u}_k, \bar{w}_k) \in V$ for all k (cf. Eq. 7) it follows that the vector $(0, \tilde{w})$ with $\tilde{w} < 0$ lies in the closure of the set V. But this contradicts the relation $\bar{w}^* = 0$ (cf. Eq. 4). □

Note that the preceding result cannot be extended to the cone generated by the set V. In particular, the relation $(0, w_0) \notin \mathrm{cl}(V)$ for some $w_0 < 0$ need not imply $(0, w_0) \notin \mathrm{cl}(\mathrm{cone}(V))$. To see this, let the set V be the epigraph of the function $f(u) = -\sqrt{u}$ for $u \geq 0$ and $f(u) = +\infty$ otherwise. Here, a vector $(0, w_0)$ with $w_0 < 0$ does not belong to the closure of the set V, while the half-line $\{(0, w) \mid w \leq 0\}$ lies in the closure of the cone generated by V.

The preceding result will be essential for proving our main result in Sect. 4 for separating a set from the half-line $\{(0, w) \mid w \leq \bar{w}\}$ for some $\bar{w} < 0$. This result may be of independent interest in the context of abstract convexity (see [15]).

3.2 Separation properties of augmenting functions

In this section, we analyze some properties of nonnegative augmenting functions. These properties are crucial for our proof of the separation result in Sect. 4.

We consider a given nonempty and convex set, related to a level set of σ, and a nonempty cone $C \subset \mathbb{R}^m \times \mathbb{R}$. We show that, when the given set has no vector in common with the cone C, we can find a larger convex set $X \subset \mathbb{R}^m \times \mathbb{R}$ having no vector in common with the cone C.

Lemma 5 *Let $\sigma : \mathbb{R}^m \mapsto (-\infty, \infty]$ be an augmenting function taking nonnegative values, i.e.,*

$$\sigma(u) \geq 0 \quad \textit{for all } u.$$

Let$C \subset \mathbb{R}^m \times \mathbb{R}$ be a nonempty cone, and let $\tilde{w}$ be a scalar with $\tilde{w} < 0$. Furthermore, let $\gamma > 0$ be a scalar such that the set $\{(u, \tilde{w}) \mid u \in L_\sigma(\gamma)\}$ has no vector in common with the cone C, i.e.,

$$\{(u, \tilde{w}) \mid u \in L_\sigma(\gamma)\} \cap C = \emptyset. \tag{9}$$

Then, the set X defined by

$$X = \left\{ (u, w) \in \mathbb{R}^m \times \mathbb{R} \,\middle|\, w \leq -\frac{|\tilde{w}|}{\gamma}\, \sigma(u) + \tilde{w} \right\} \tag{10}$$

has no vector in common with the cone C.

Proof To obtain a contradiction, assume that there exists a vector $(\hat{u}, \hat{w})$ such that

$$(\hat{u}, \hat{w}) \in X \cap C. \tag{11}$$

By the definition of X (cf. Eq. 10), and the relations $\tilde{w} < 0$ and $\sigma(u) \geq 0$ for all u, it follows that

$$\hat{w} \leq \tilde{w} < 0. \tag{12}$$

Note that we can view the set X as the zero-level set of the function $F(u, w) = w - \tilde{w} + \frac{|\tilde{w}|}{\gamma}\, \sigma(u)$, i.e.,

$$X = \left\{ (u, w) \in \mathbb{R}^m \times \mathbb{R} \;\mid\; F(u, w) \leq 0 \right\}.$$

The function $F(u, w)$ is convex by the convexity of σ (cf. Definition 2), and therefore the set X is convex. Furthermore, X contains the vector $(0, \tilde{w})$ since $\sigma(0) = 0$ by the definition of the augmenting function.

Consider now the scalar

$$\hat{\alpha} = \frac{\tilde{w}}{\hat{w}},$$

which satisfies $\hat{\alpha} \in (0, 1]$ by Eq. 12. From the convexity of set X and the relations $(0, \tilde{w}) \in X$ and $(\hat{u}, \hat{w}) \in X$, it follows that

$$(1 - \hat{\alpha})(0, \tilde{w}) + \hat{\alpha}(\hat{u}, \hat{w}) = (\hat{\alpha}\hat{u}, (1 - \hat{\alpha})\tilde{w} + \tilde{w}) \in X.$$

Using the definition of set X (cf. Eq. 10), we obtain

$$(1 - \hat{\alpha})\tilde{w} \leq -\frac{|\tilde{w}|}{\gamma}\, \sigma(\hat{\alpha}\hat{u}).$$

Since $\tilde{w} < 0$, it follows that

$$\frac{|\tilde{w}|}{\gamma}\, \sigma(\hat{\alpha}\hat{u}) \leq -(1 - \hat{\alpha})\tilde{w} < -\tilde{w} = |\tilde{w}|,$$

implying that

$$\sigma(\hat{\alpha}\hat{u}) \leq \gamma.$$

Hence, the vector $\hat{\alpha}\hat{u}$ belongs to level set $L_\sigma(\gamma)$, and therefore

$$(\hat{\alpha}\hat{u}, \tilde{w}) \in \{(u, \tilde{w}) \mid u \in L_\sigma(\gamma)\}.$$

The vector $(\hat{u}, \hat{w})$ lies in the cone C (cf. Eq. 11) and, because $\hat{\alpha} > 0$, the vector $\hat{\alpha}(\hat{u}, \hat{w}) = (\hat{\alpha}\hat{u}, \tilde{w})$ also lies in C. Therefore

$$(\hat{\alpha}\hat{u}, \hat{w}) \in \{(u, \tilde{w}) \mid u \in L_\sigma(\gamma)\} \cap C,$$

contradicting the assumption that the set $\{(u, \tilde{w}) \mid u \in L_\sigma(\gamma)\}$ and the cone C have no common vectors (cf. Eq. 9). $\square$

The implication of Lemma 5 is that the set $S = \{(u, 2\tilde{w}) \mid u \in L_\sigma(\gamma)\}$ and the cone C can be separated by a concave function

$$\phi(u) = -\frac{|\tilde{w}|}{\gamma}\sigma(u) + \tilde{w} \quad \text{for all} \quad u \in \mathbb{R}^m.$$

In particular, Lemma 5 asserts that

$$w \le \phi(u) < z \quad \text{for all} \quad (u, w) \in X \text{ and } (u, z) \in C.$$

Furthermore, it can be seen that the set S is contained in the set X for $\tilde{w} < 0$, and therefore

$$w \le \phi(u) < z \quad \text{for all} \quad (u, w) \in S \text{ and } (u, z) \in C,$$

thus showing that ϕ separates the set S and the cone C.

4 Separation theorem

In this section, we discuss some sufficient conditions on augmenting functions and the set V that guarantee the separation of this set and a vector $(0, w_0)$ not belonging to the closure of the set V. In particular, throughout this section, we consider a set V that has a nonempty intersection with w-axis, extends upward both in u-space and w-space, and does not have $(0, -1)$ as its recession direction. These properties of V are formally imposed in the following assumption.

Assumption 1 Let $V \subset \mathbb{R}^m \times \mathbb{R}$ be a nonempty set that satisfies the following:

(a) The primal optimal value is finite, i.e., $w^* = \inf_{(0,w)\in V} w$ is finite.
(b) The set V extends upward in u-space and w-space.
(c) The vector $(0, -1)$ is not a recession direction of V, i.e.,

$$(0, -1) \notin V^\infty.$$

Assumption 1(a) states the requirement that the set V intersects the w-axis. Assumption 1(b) is satisfied, for example, when V is the epigraph of a nonincreasing function. Assumption 1(c) formalizes the requirement that the changes in w are commensurate with the changes in u for $(u, w) \in V$. It can be viewed as a requirement on the rate of decrease of w with respect to u. In particular, suppose that $-w \approx O\left(\|u\|^\beta\right)$ for some scalar β and all $(u, w) \in V$. Then, we have

$$\frac{1}{|w|}(u, w) \approx \left(O\left(\|u\|^{1-\beta}\right), -1\right).$$

Hence, for $\beta < 0$, as $u \to 0$, we have $w \to -\infty$ and $\frac{1}{|w|}(u, w) \to (0, -1)$, implying that the direction $(0, -1)$ is a recession direction of V (see Definition 1). However, for

$\beta > 0$, we have $w \to -\infty$ as $u \to \infty$, and $\frac{1}{|w|}(u, w) \to (0, -1)$ as $u \to \infty$ if and only if $1 - \beta < 0$. Thus, the vector $(0, -1)$ is again a recession direction of V for $\beta > 1$, and it is not a recession direction of V for $0 \le \beta \le 1$.

We present a separation result for nonnegative augmenting functions that have an additional property related to their behavior in the vicinity of the set of vectors u with $u \le 0$. In particular, we consider augmenting functions that satisfy the following assumption.

Assumption 2 Let σ be an augmenting function with the following properties:

(a) The function σ is nonnegative,

$$\sigma(u) \ge 0 \qquad \text{for all } u.$$

(b) Given a sequence $\{u_k\} \subset \mathbb{R}^m$, the convergence of $\sigma(u_k)$ to zero implies the convergence of the nonnegative part of the sequence $\{u_k\}$ to zero, i.e.,

$$\sigma(u_k) \to 0 \;\; \Rightarrow \;\; u_k^+ \to 0,$$

where $u^+ = (\max\{0, u_1\}, \dots, \max\{0, u_m\})'$.

To provide intuition for the property stated in Assumption 2(b), consider the nonpositive orthant $\mathbb{R}^m_-$ given by

$$\mathbb{R}^m_- = \{u \in \mathbb{R}^m \mid u \le 0\}.$$

It can be seen that the vector v^- defined by

$$v^- = (\min\{0, v_1\}, \dots, \min\{0, v_m\})'$$

is the projection of a vector v on the set $\mathbb{R}^m_-$, i.e.,

$$\min_{u \in \mathbb{R}^m_-} \|v - u\| = \|v - v^-\|.$$

Since $v = v^+ + v^-$, we have

$$\operatorname{dist}(v, \mathbb{R}^m_-) = \min_{u \in \mathbb{R}^m_-} \|v - u\| = \|v - v^-\| = \|v^+\|. \tag{13}$$

Thus, Assumption 2(b) is equivalent to the following: for any sequence $\{u_k\}$, the convergence of $\sigma(u_k)$ to zero implies that the distance between u_k and $\mathbb{R}^m_-$ converges to zero, i.e.,

$$\sigma(u_k) \to 0 \quad \Rightarrow \quad \operatorname{dist}(u_k, \mathbb{R}^m_-) \to 0.$$

It can be further seen that Assumption 2(b) is equivalent to the following condition: for all $\delta > 0$, there holds

$$\inf_{\{u \mid \operatorname{dist}(u, \mathbb{R}^m_-) \ge \delta\}} \sigma(u) > 0. \tag{14}$$

To see this, assume first that Assumption 2(b) holds and assume to arrive at a contradiction that there exists some $\delta > 0$ such that

$$\inf_{\{u \mid \operatorname{dist}(u, \mathbb{R}^m_-) \ge \delta\}} \sigma(u) = 0.$$

This implies that there exists a sequence $\{u_k\}$ such that $\sigma(u_k) \to 0$ and $\|u_k^+\| \geq \delta$ for all k (cf. Eq. 13), contradicting Assumption 2(b). Conversely, assume that condition (14) holds. Let $\{u_k\}$ be a sequence with $\sigma(u_k) \to 0$, and assume that $\limsup_{k\to\infty} \|u_k^+\| > 0$. This implies the existence of some $\delta > 0$ such that along a subsequence, we have $\text{dist}(u_k, \mathbb{R}^m_-) > \delta$ for all k sufficiently large. Since $\sigma(u_k) \to 0$, this contradicts condition (14).

Assumption 2(b) is related to the peak at zero condition (see [14]) which can be expressed as follows: for all $\delta > 0$, there holds

$$\inf_{\{u \mid \|u\| \geq \delta\}} \sigma(u) > 0.$$

This condition was studied by Rubinov et al. [14] to provide zero duality gap results for arbitrary dualizing parametrizations under the assumption that there exists an augmenting function minorizing the primal function. In this paper, we focus on the weaker condition, Assumption 2(b), which seems suitable for the study of parametrizations that yield nondecreasing primal functions (see Sect. 5).

The following are some examples of the functions $\sigma(u)$ for $u = (u_1, \ldots, u_m)$ that satisfy Assumption 2:

$$\sigma(u) = \max\{0, u_1, \ldots, u_m\},$$

$$\sigma(u) = \sum_{i=1}^{m} (\max\{0, u_i\})^\beta \quad \text{with } \beta > 0,$$

(cf. [10]), where $\beta = 1$ and $\beta = 2$ are among the most popular choices (e.g. see [11]);

$$\sigma(u) = u'Qu,$$

$$\sigma(u) = (u^+)'Qu^+ \quad \text{with } u_i^+ = \max\{0, u_i\},$$

(cf. [10]), where Q is a symmetric positive definite matrix;

$$\sigma(u) = \max\{0, a_1(e^{u_1} - 1), \ldots, a_m(e^{u_m} - 1)\},$$

$$\sigma(u) = \sum_{i=1}^{m} \max\{0, a_i(e^{u_i} - 1)\},$$

(cf. [17]), where $a_i > 0$ for all i.

Moreover, Assumption 2 is satisfied by any nonnegative convex function with $\sigma(0) = 0$ and with the set of minima over $\mathbb{R}^m$ consisting of the zero vector only,

$$\arg\min_{u\in\mathbb{R}^m} \sigma(u) = \{0\},$$

i.e., level-bounded augmenting functions studied by Rockafellar and Wets [13], and Huang and Yang [9].

Proposition 2 (Nonnegative augmenting function) *Let $V \subset \mathbb{R}^m \times \mathbb{R}$ be a nonempty set satisfying Assumption 1. Let σ be an augmenting function satisfying Assumption 2. Then, the set V and a vector $(0, w_0)$ that does not belong to the closure of V can be strongly separated by the function σ, i.e., there exist scalars $c > 0$ and ξ such that*

$$w + c\sigma(u) \geq \xi > w_0 \quad \textit{for all } (u, w) \in V.$$

Proof By Lemma 2(a), we have that

$$w_0 < \bar{w}^* = \inf_{(0,w)\in \mathrm{cl}(V)} w \le w^*.$$

Since w^* is finite (cf. Assumption 1(a)), it follows that $\bar{w}^*$ is finite. By using the translation of space along w-axis if necessary, without loss of generality, we may assume that

$$\bar{w}^* = 0, \tag{15}$$

so that

$$w_0 < 0. \tag{16}$$

Consider an upward translation of set V given by

$$\tilde{V} = \left\{(u,w) \;\middle|\; \left(u, w + \frac{w_0}{4}\right) \in V\right\} \tag{17}$$

and the cone generated by $\tilde{V}$, denoted by K. The proof relies on constructing a convex set X using the augmenting function σ that contains the vector $(0, w_0/2)$, extends downward w-axis, and does not have any vector in common with the closure of K. The set X is actually a surface that separates $(0, w_0/2)$ and the closure of K, and it also separates $(0, w_0)$ and V.

In particular, the proof is given in the following steps:

Step 1 We first show that there exists some $\gamma > 0$ such that

$$\{(u, w_0/2) \mid u \in L_\sigma(\gamma)\} \cap \mathrm{cl}(K) = \emptyset.$$

To arrive at a contradiction, suppose that the preceding relation does not hold. Then, there exist a sequence $\{\gamma_k\}$ with $\gamma_k \downarrow 0$ and a sequence $\{u_k\}$ such that

$$\sigma(u_k) \le \gamma_k, \qquad (u_k, w_0/2) \in \mathrm{cl}(K). \tag{18}$$

Since $\sigma(u) \ge 0$ for all u (cf. Assumption 2(a)) and $\gamma_k \downarrow 0$, it follows that

$$\lim_{k\to\infty} \sigma(u_k) = 0,$$

implying by Assumption 2(b) that

$$u_k^+ \to 0.$$

Note that we have

$$u_k = u_k^+ + u_k^- \le u_k^+ \qquad \text{for all } k.$$

By Assumption 1, the set V is extending upward in u-space, and so does the set $\tilde{V}$, an upward translation of the set V (cf. Eq. 17). Consequently, by Lemma 3, the closure $\mathrm{cl}(K)$ of the cone K generated by $\tilde{V}$ is also extending upward in u-space. Since $(u_k, w_0/2) \in \mathrm{cl}(K)$ (cf. Eq. 18) and $u_k \le u_k^+$ for all k, it follows that

$$(u_k^+, w_0/2) \in \mathrm{cl}(K) \qquad \text{for all } k.$$

Furthermore, because $u_k^+ \to 0$, we obtain $(0, w_0/2) \in \mathrm{cl}(K)$, and therefore

$$(0, w_0) \in \mathrm{cl}(K).$$

On the other hand, since $(0, w_0) \notin \mathrm{cl}(V)$, by Lemma 4 we have that $(0, w_0) \notin \mathrm{cl}(K)$, contradicting the preceding relation. Thus, the set $\{(u, w_0/2) \mid u \in L_\sigma(\gamma)\}$ and the cone $\mathrm{cl}(K)$ do not have any point in common, i.e.,

$$\{(u, w_0/2) \mid u \in L_\sigma(\gamma)\} \cap \mathrm{cl}(K) = \emptyset. \tag{19}$$

Step 2 We consider the set X given by

$$X = \left\{ (u, w) \in \mathbb{R}^m \times \mathbb{R} \;\middle|\; w \le -\frac{|w_0|}{2\gamma}\sigma(u) + \frac{w_0}{2} \right\},$$

and we prove that this set has no vector in common with the cone $\mathrm{cl}(K)$. We show this by using Lemma 5 with the identification as follows:

$$\tilde{w} = \frac{w_0}{2}, \qquad C = \mathrm{cl}(K). \tag{20}$$

By Assumption 2(a), the augmenting function σ is nonnegative. Furthermore, note that $\tilde{w} = w_0/2 < 0$ because $w_0 < 0$. In view of Eq. (19), we have that the set $\{(u, w_0/2) \mid u \in L_\sigma(\gamma)\}$ and the cone $\mathrm{cl}(K)$ have no vector in common. Thus, Lemma 5 implies that the set X has no vector in common with the cone $\mathrm{cl}(K)$, i.e., $X \cap \mathrm{cl}(K) = \emptyset$. From this and the definition of X, it follows that

$$w > -\frac{|w_0|}{2\gamma}\sigma(u) + \frac{w_0}{2} \quad \text{for all} \quad (u, w) \in \mathrm{cl}(K).$$

The cone K is generated by the set $\tilde{V}$, so that the preceding relation holds for all $(u, w) \in \tilde{V}$, implying that

$$w - \frac{w_0}{4} + c\sigma(u) > \frac{w_0}{2} \qquad \text{for all} \quad (u, w) \in V \quad \text{and} \quad c = \frac{|w_0|}{2\gamma}.$$

Furthermore, since $w_0 < 0$, it follows that

$$w + c\sigma(u) \ge \xi > w_0 \qquad \text{for all} \quad (u, w) \in V \quad \text{and} \quad \xi = \frac{3w_0}{4},$$

thus completing the proof. □

Note that Proposition 2 shows that for a set V satisfying Assumption 1 and an augmenting function σ satisfying Assumption 2, we can take $\mu = 0$ in the separating concave surface $\{(u, \phi_{c,\mu}(u)) \mid u \in \mathbb{R}^m\}$, where

$$\phi_{c,\mu}(u) = -c\sigma(u) - \mu' u + \xi,$$

i.e., separation of V from a vector that does not belong to the closure of V can be realized through the augmenting function only.

5 Application to constrained optimization duality

In this section, we use the separation result of Sect. 4 to provide necessary and sufficient conditions that guarantee that the optimal values of the geometric primal and dual problems are equal. We then discuss the implications of these conditions for constrained (nonconvex) optimization problems.

5.1 Necessary and sufficient conditions for geometric zero duality gap

Recall that for a given nonempty set $V \subset \mathbb{R}^m \times \mathbb{R}$, we define the geometric primal problem as

$$\inf_{(0,w)\in V} w, \tag{21}$$

and denote the primal optimal value by w^*. For a given augmenting function σ, scalar $c \geq 0$, and vector $\mu \in \mathbb{R}^m$, we define a dual function $d(c, \mu)$ as

$$d(c,\mu) = \inf_{(u,w)\in V}\{w + c\sigma(u) + \mu' u\}.$$

We consider the geometric dual problem

$$\sup_{c\geq 0,\ \mu\in\mathbb{R}^m} d(c,\mu) \tag{22}$$

and denote the dual optimal value by d^*.

In what follows, we use an additional assumption on the augmenting function.

Assumption 3 (*Continuity at the origin*) Let σ be an augmenting function. The augmenting function σ is continuous at $u = 0$.

Since an augmenting function σ is convex by definition, the assumption that σ is continuous at $u = 0$ holds when 0 is in the relative interior of the domain of σ (see [3, 12]). Note that all examples of augmenting functions we have considered in Sect. 4 satisfy this condition.

We now present a necessary condition for zero duality gap.

Proposition 3 (Necessary conditions for zero duality gap) *Let $V \subset \mathbb{R}^m \times \mathbb{R}$ be a nonempty set. Let σ be an augmenting function that satisfies Assumption 3. Consider the geometric primal and dual problems defined in Eqs. (21) and (22). Assume that there is zero duality gap, i.e., $d^* = w^*$. Then, for any sequence $\{(u_k, w_k)\} \subset V$ with $u_k \to 0$, we have*

$$\liminf_{k\to\infty} w_k \geq w^*.$$

Proof Let $\{(u_k, w_k)\} \subset V$ be a sequence such that $u_k \to 0$. By definition, the augmenting function $\sigma(u)$ satisfies $\sigma(0) = 0$ (cf. Definition 2). Using this and the continuity of $\sigma(u)$ at $u = 0$ (cf. Assumption 3), we obtain for all $c \geq 0$ and $\mu \in \mathbb{R}^m$,

$$\begin{aligned}\liminf_{k\to\infty} w_k &= \liminf_{k\to\infty} w_k + c\sigma(0) + \mu' 0\\ &= \liminf_{k\to\infty}\{w_k + c\sigma(u_k) + \mu' u_k\}\\ &\geq \inf_{(u,w)\in V}\{w + c\sigma(u) + \mu' u\}\\ &= d(c,\mu).\end{aligned}$$

Hence

$$\liminf_{k\to\infty} w_k \geq \sup_{c\geq 0,\ \mu\in\mathbb{R}^m} d(c,\mu) = d^*$$

and since $d^* = w^*$, it follows that $\liminf_{k\to\infty} w_k \geq w^*$. □

We next provide sufficient conditions for zero duality gap using the assumptions of Sect. 4 on set V and augmenting function σ.

Proposition 4 (Sufficient conditions for zero duality gap) *Let $V \subset \mathbb{R}^m \times \mathbb{R}$ be a nonempty set. Consider the geometric primal and dual problems defined in Eqs.* (21) *and* (22). *Assume that for any sequence $\{(u_k, w_k)\} \subset V$ with $u_k \to 0$, we have*

$$\liminf_{k\to\infty} w_k \geq w^*.$$

Assume that the set V satisfies Assumption 1 and the augmenting function σ satisfies Assumption 2. Then, there is zero duality gap, i.e., $d^* = w^*$.

Proof By Assumption 1(a), we have that w^* is finite. Let $\epsilon > 0$ be arbitrary, and consider the vector $(0, w^* - \epsilon)$. We show that $(0, w^* - \epsilon)$ does not belong to the closure of the set V. To obtain a contradiction, assume that $(0, w^* - \epsilon) \in \mathrm{cl}(V)$. Then, there exists a sequence $\{(u_k, w_k)\} \subset V$ with $u_k \to 0$ and $w_k \to w^* - \epsilon$, contradicting the assumption that $\liminf_{k\to\infty} w_k \geq w^*$. Hence, $(0, w^* - \epsilon)$ does not belong to the closure of V.

Consider the set V that satisfies Assumption 1 and the augmenting function σ that satisfies Assumption 2. Since $(0, w^* - \epsilon) \notin \mathrm{cl}(V)$, by Proposition 2, it follows that there exist scalars $c \geq 0$ and ξ such that

$$\inf_{(u,w)\in V} \{w + c\sigma(u)\} \geq \xi > w^* - \epsilon.$$

Therefore

$$d(c, 0) > w^* - \epsilon,$$

implying that

$$d^* = \sup_{c\geq 0,\ \mu\in\mathbb{R}^m} d(c, \mu) > w^* - \epsilon.$$

By letting $\epsilon \to 0$, we obtain

$$d^* \geq w^*,$$

which together with the weak duality relation ($d^* \leq w^*$) implies that $d^* = w^*$. □

5.2 Constrained optimization duality

We consider the following constrained optimization problem

$$\begin{aligned} &\min f_0(x) \\ &\text{s.t. } x \in X,\ f(x) \leq 0, \end{aligned} \tag{23}$$

where X is a nonempty subset of $\mathbb{R}^n$,

$$f(x) = (f_1(x), \ldots, f_m(x))$$

and $f_i : \mathbb{R}^n \mapsto (-\infty, \infty]$ for $i = 0, 1, \ldots, m$. We refer to this as the *primal problem*, and denote its optimal value by f^*.

For the primal problem, we define a *dualizing parametrization function* $\bar{f} : \mathbb{R}^n \times \mathbb{R}^m \mapsto (-\infty, \infty]$ as

$$\bar{f}(x,u) = \begin{cases} f_0(x), & \text{if } f(x) \leq u, \\ +\infty, & \text{otherwise.} \end{cases} \tag{24}$$

Given an augmenting function σ, we define the *augmented Lagrangian function* as

$$l(x,c,\mu) = \inf_{u \in \mathbb{R}^m} \{\bar{f}(x,u) + c\sigma(u) + \mu' u\}$$

and the *augmented dual function* as

$$q(c,\mu) = \inf_{x \in X} l(x,c,\mu).$$

We consider the problem

$$\begin{aligned} &\max\ q(c,\mu) \\ &\text{s.t. } c \geq 0, \quad \mu \in \mathbb{R}^m. \end{aligned} \tag{25}$$

We refer to this as the *augmented dual problem*, and denote its optimal value by q^*. We say that *there is zero duality gap* when $q^* = f^*$, and we say that *there is a duality gap* when $q^* < f^*$.

5.3 Weak duality

The following proposition shows the weak duality relation between the optimal values of the primal problem and the augmented dual problem.

Proposition 5 (Weak duality) *The augmented dual optimal value does not exceed the primal optimal value, i.e.,*

$$q^* \leq f^*.$$

Proof Using the definition of the dualizing parametrization function (cf. Eq. 24), we can write $q(c,\mu)$ as

$$q(c,\mu) = \inf_{u \in \mathbb{R}^m} \inf_{\substack{x \in X \\ f(x) \leq u}} \{f_0(x) + c\sigma(u) + \mu' u\} \quad \text{for all } c \geq 0, \quad \mu \in \mathbb{R}^m.$$

Substituting $u = 0$ in the preceding relation and using the assumption that $\sigma(0) = 0$ (cf. Definition 2), we have

$$q(c,\mu) \leq \inf_{\substack{x \in X \\ f(x) \leq 0}} \{f_0(x) + c\sigma(0)\} = \inf_{\substack{x \in X \\ f(x) \leq 0}} f_0(x) = f^*.$$

Therefore

$$q^* = \sup_{c \geq 0,\ \mu \in \mathbb{R}^m} q(c,\mu) \leq f^*.$$

□

5.4 Zero duality gap

We next provide necessary and sufficient conditions for zero duality gap. In our analysis, a critical role is played by the geometric framework of Sect. 2 and the results of Sect. 5.1, and by the *primal function* $p : \mathbb{R}^m \mapsto [-\infty, \infty]$ of the optimization problem (23), defined as

$$p(u) = \inf_{x \in X,\, f(x) \leq u} f_0(x).$$

The primal function $p(u)$ is clearly nonincreasing in u, i.e.,

$$p(u) \leq p(\tilde{u}) \qquad \text{for } u \geq \tilde{u}. \tag{26}$$

Furthermore, $p(u)$ is related to the dualizing parametrization function $\bar{f}$ (cf. Eq. 24) as follows:

$$p(u) = \inf_{x \in X} \bar{f}(x, u). \tag{27}$$

We next discuss some properties of the primal function and its connection to the existence of a duality gap. The connection to the existence of a duality gap is established through the geometric framework of Sect. 2. In particular, let V be the epigraph of the primal function,

$$V = \text{epi}(p).$$

Then, the geometric primal value w^* is equal to $p(0)$,

$$w^* = p(0) = f^*.$$

The corresponding geometric dual problem is

$$\begin{aligned} &\max\ d(c, \mu) \\ &\ \text{s.t. } c \geq 0,\ \mu \in \mathbb{R}^m, \end{aligned}$$

where

$$d(c, \mu) = \inf_{(u,w) \in V} \{w + c\sigma(u) + \mu' u\} = \inf_{\{(u,w) \in \mathbb{R}^m \times \mathbb{R} | p(u) \leq w\}} \{w + c\sigma(u) + \mu' u\}.$$

Using the relation between the primal function p and the dualizing parametrization function $\bar{f}$ (cf. Eq. 27), we can see that

$$d(c, \mu) = \inf_{u \in \mathbb{R}^m} \inf_{x \in X} \{\bar{f}(x, u) + c\sigma(u) + \mu' u\} = q(c, \mu) \quad \text{for all } c \geq 0, \quad \mu \in \mathbb{R}^m. \tag{28}$$

We now use the results of Sect. 5.1 with the identification $V = \text{epi}(p)$ to provide necessary and sufficient conditions for zero duality gap. These conditions involve the lower semicontinuity of the primal function at 0 (see [3, Chap. 6, and 16, Sect. 3.1.6] for discussions on the lower semicontinuity of the primal function). For the necessary conditions, we assume that the augmenting function σ is continuous at $u = 0$, i.e., it satisfies Assumption 3.

Proposition 6 (Necessary conditions for zero duality gap) *Let σ be an augmenting function that satisfies Assumption 3. Consider the primal problem and the augmented dual problem defined in Eqs. 23 and 25. Assume that there is zero duality gap, i.e.,*

$q^* = f^*$. Then, the primal function $p(u)$ is lower semicontinuous at $u = 0$, i.e., for all sequences $\{u_k\} \subset \mathbb{R}^m$ with $u_k \to 0$, we have

$$p(0) \leq \liminf_{k\to\infty} p(u_k).$$

Proof We apply Proposition 3 where the set V is the epigraph of p, i.e.,

$$V = \mathrm{epi}(p).$$

From the definition of p, we have

$$w^* = p(0) = f^*.$$

From Eq. (28), it also follows that $d^* = q^*$. Hence, the assumption that $q^* = f^*$ is equivalent to the assumption that $d^* = w^*$. Let $\{u_k\} \subset \mathbb{R}^m$ be a sequence with $u_k \to 0$. Since $\{(u_k, p(u_k))\} \subset \mathrm{epi}(p)$, it follows by Proposition 3 that

$$\liminf_{k\to\infty} p(u_k) \geq p(0),$$

completing the proof. □

We next present sufficient conditions for zero duality gap.

Proposition 7 (Sufficient conditions for zero duality gap) *Assume that the primal problem* (23) *is feasible and bounded from below, i.e., f^* is finite. Assume that the direction* $(0, -1)$ *is not a recession direction of the epigraph of the primal function, i.e.,*

$$(0, -1) \notin (\mathrm{epi}(p))^\infty.$$

Assume that $p(u)$ is lower semicontinuous at $u = 0$. Furthermore, assume that the augmenting function σ satisfies Assumption 2. Then, there is zero duality gap, i.e., $q^* = f^*$.

Proof We apply Proposition 4 where the set V is the epigraph of the function p, i.e.,

$$V = \mathrm{epi}(p).$$

From the definition of p, we have

$$w^* = p(0) = f^*.$$

The assumption that the primal problem (23) is feasible and bounded from below implies that f^* is finite, and therefore w^* is finite. Since V is the epigraph of a function, the set V is extending upward in w-space. Moreover, V is extending upward in u-space because the primal function p is nonincreasing (cf. Eq. 26). Together with the assumption that $(0, -1) \notin (\mathrm{epi}(p))^\infty$, we have that the set V satisfies Assumption 1.

From Eq. (28), we have $d^* = q^*$. By the choice of set V, the condition

$$p(0) \leq \liminf_{k\to\infty} p(u_k)$$

for all sequences $\{u_k\}$ with $u_k \to 0$ is equivalent to the condition that for every sequence $\{(u_k, w_k)\} \subset V$ with $u_k \to 0$, there holds $w^* \leq \liminf_{k\to\infty} w_k$. The result then follows from Proposition 4. □

It can be seen that the assumption of Proposition 7 that $(0, -1)$ is not a direction of recession of epi(p) is satisfied, for example, when $\inf_{x\in X} f_0(x) > -\infty$. This relation holds, for example, when X is compact and f_0 is lower semicontinuous, or when f_0 is bounded from below, i.e., $f(x) \geq b_0$ for some scalar b and for all $x \in \mathbb{R}^n$.

Generally speaking, we can view the condition $(0, -1) \notin \text{epi}(p)^\infty$ as a requirement on the rate of decrease of $p(u)$ with respect to u. For example, suppose that for some scalar β and all $(u, w) \in \text{epi}(p)$, we have $p(u) \approx O\left(-\|u\|^\beta\right)$. Then, similar to the discussion in the beginning of Section 4, we can see that the vector $(0, -1) \notin \text{epi}(p)^\infty$ when

$$p(u) \approx O\left(-\|u\|^\beta\right) \quad \text{as } \|u\| \to \infty \quad \text{and} \quad 0 \leq \beta \leq 1.$$

For $\beta > 1$, we have $(0, -1) \in \text{epi}(p)^\infty$. By viewing β as a rate of decrease of $p(u)$ to minus infinity, we see that when $p(u)$ decreases to infinity at a rate faster than the linear rate ($\beta = 1$), the vector $(0, -1)$ is a recession direction of epi(p). When the rate of decrease is slower than linear but at least constant ($\beta = 0$), then the vector $(0, -1)$ is not a recession direction of epi(p).

6 Application to penalty methods

In this section, we use the separation result of Sect. 4 to provide necessary and sufficient conditions for the convergence of penalty methods. In particular, we define a slightly different geometric dual problem suitable for the analysis of penalty methods. Based on the separation result of Sect. 4, we provide necessary and sufficient conditions for the convergence of the geometric dual optimal values to the geometric primal optimal value. We then discuss the implications of these conditions for the penalty methods for constrained (nonconvex) optimization problems.

6.1 Necessary and sufficient conditions for convergence in geometric penalty framework

Here, we consider a geometric framework similar to the framework of Sect. 2. In particular, given a nonempty set $V \subset \mathbb{R}^m \times \mathbb{R}$, the geometric primal problem is the same as in Sect. 2, i.e., we want to determine the value w^* where

$$w^* = \inf_{(0,w)\in V} w. \tag{29}$$

Given an augmenting function σ, we define a slightly different geometric problem, which we refer to as *geometric penalized problem*. In the geometric penalized problem, we want to determine the value $\tilde{d}^*$ where

$$\tilde{d}^* = \sup_{c\geq 0} \tilde{d}(c) \tag{30}$$

with

$$\tilde{d}(c) = \inf_{(u,w)\in V} \{w + c\sigma(u)\}. \tag{31}$$

Note that, here, we consider concave surfaces that support the set V from below and do not involve a linear term (see Sect. 2.2).

It is straightforward to establish the weak duality relation between the optimal value of the geometric primal problem and penalized problem. The proof for this result is the same as that of Proposition 1 (where $\mu = 0$) and, therefore, it is omitted.

Proposition 8 (Weak duality) *The penalized optimal value does not exceed the primal optimal value, i.e.,*

$$\tilde{d}^* \leq w^*.$$

We next consider necessary conditions for the convergence of the optimal values $\tilde{d}(c)$ of the geometric penalized problem to the geometric primal optimal value w^*. These conditions require the continuity of the augmenting (penalty) function σ at $u = 0$ (cf. Assumption 3). The proof of these necessary conditions is the same as that of Proposition 3 (where $\mu = 0$) and, therefore, it is omitted.

Proposition 9 (Necessary conditions for geometric penalty convergence) *Let $V \subset \mathbb{R}^m \times \mathbb{R}$ be a nonempty set. Let σ be an augmenting function that satisfies Assumption 3. Consider the geometric primal and penalized problems defined in Eqs. 29 and 30. Assume that the optimal values of the geometric penalized problem converge to the optimal value of the primal problem, i.e., $\lim_{c\to\infty} \tilde{d}(c) = w^*$. Then, for any sequence $\{(u_k, w_k)\} \subset V$ with $u_k \to 0$, we have*

$$\liminf_{k\to\infty} w_k \geq w^*.$$

We next provide sufficient conditions for zero duality gap using the assumptions of Sect. 4 on set V and augmenting function σ.

Proposition 10 (Sufficient conditions for geometric penalty convergence) *Let $V \subset \mathbb{R}^m \times \mathbb{R}$ be a nonempty set. Consider the geometric primal and penalized problems defined in Eqs. (29) and (30). Assume that for any sequence $\{(u_k, w_k)\} \subset V$ with $u_k \to 0$, we have*

$$\liminf_{k\to\infty} w_k \geq w^*.$$

Assume further that the set V satisfies Assumption 1 and the augmenting function σ satisfies Assumption 2. Then, the optimal values of the geometric penalized problem $\tilde{d}(c)$ converge to the optimal value of the primal problem w^* as $c \to \infty$, i.e.,

$$\lim_{c\to\infty} \tilde{d}(c) = w^*.$$

Proof Let $\{c_k\}$ be a nonnegative scalar sequence with $c_k \to \infty$. By the weak duality, we have $\tilde{d}(c) \leq f^*$ for all $c \geq 0$, implying that $\limsup_{k\to\infty} \tilde{d}(c_k) \leq w^*$. Thus, it suffices to show that

$$\liminf_{k\to\infty} \tilde{d}(c_k) \geq w^*.$$

To arrive at a contradiction, suppose that the preceding relation does not hold. Then, for some $\epsilon > 0$,

$$\liminf_{k\to\infty} \tilde{d}(c_k) \leq w^* - 2\epsilon. \tag{32}$$

By the definition of $\tilde{d}(c)$ (cf. Eq. 31), for each k, there exists $(u_k, w_k) \in V$ such that

$$w_k + c_k\sigma(u_k) \leq \tilde{d}(c_k) + \epsilon.$$

Hence, by considering an appropriate subsequence in relation (32), we can assume without loss of generality that

$$w_k + c_k\sigma(u_k) \leq w^* - \epsilon \qquad \text{for all } k. \tag{33}$$

By the assumption that, for every sequence $\{(u_k, w_k)\} \subset V$ with $u_k \to 0$, there holds

$$w^* \leq \liminf_{k\to\infty} w_k, \tag{34}$$

it follows that $(0, w^* - \epsilon)$ does not belong to the closure of the set V, i.e.,

$$(0, w^* - \epsilon) \notin \mathrm{cl}(V).$$

Since $(0, w^* - \epsilon) \notin \mathrm{cl}(V)$, by Proposition 2, there exists some scalar $\bar{c} > 0$ such that

$$w_k + \bar{c}\sigma(u_k) > w^* - \epsilon.$$

Because $c_k \to \infty$ and the function σ is nonnegative (cf. Assumption 2), there exists some sufficiently large $\bar{k}$ such that $c_k \geq \bar{c}$ for all $k \geq \bar{k}$. Therefore,

$$w_k + c_k\sigma(u_k) \geq w_k + \bar{c}\sigma(u_k) > w^* - \epsilon \qquad \text{for all} \quad k \geq \bar{k},$$

which contradicts Eq. (33). □

6.2 Penalty methods for constrained optimization

In this section, we make the connection between the geometric framework and the analysis of penalty methods for constrained optimization problems. In particular, we consider the following constrained optimization problem

$$\begin{aligned} &\min f_0(x) \\ &\text{ s.t. } x \in X, \quad f(x) \leq 0, \end{aligned} \tag{35}$$

where X is a nonempty subset of $\mathbb{R}^n$,

$$f(x) = (f_1(x), \ldots, f_m(x))$$

and $f_i : \mathbb{R}^n \mapsto (-\infty, \infty]$ for $i = 0, 1, \ldots, m$. We denote the optimal value of this problem by f^*.

We are interested in penalty methods for the solution of problem (35), which involve solving a sequence of less constrained optimization problems of the form

$$\begin{aligned} &\min \{f_0(x) + c\sigma(f(x))\} \\ &\text{ s.t. } x \in X. \end{aligned} \tag{36}$$

Here, $c \geq 0$ is a penalty parameter that will ultimately increase to $+\infty$ and σ is a penalty function. For a given $c \geq 0$, we denote the optimal value of problem (36) by $\tilde{f}(c)$.

To make the connection to the geometric framework, we use the primal function $p : \mathbb{R}^m \mapsto [-\infty, \infty]$ of the optimization problem (35), defined as

$$p(u) = \inf_{x\in X,\, f(x)\leq u} f_0(x). \tag{37}$$

We consider the set V of the geometric framework to be the epigraph of the primal function,

$$V = \mathrm{epi}(p).$$

Then, the geometric primal value w^* is equal to $p(0)$,

$$w^* = p(0) = f^*.$$

Moreover, the objective function of the geometric penalized problem is given by

$$\tilde{d}(c) = \inf_{\{(u,w)\ |\ p(u)\leq w\}} \{w + c\sigma(u)\} = \inf_{u\in\mathbb{R}^m} \inf_{\substack{x\in X\\ f(x)\leq u}} \{f_0(x) + c\sigma(u)\}. \tag{38}$$

In what follows, we use an additional assumption on the augmenting function α.

Assumption 4 Let σ be an augmenting function. The augmenting function $\sigma(u)$ is nondecreasing in u.

Note that all examples of augmenting functions we provided throughout Sect. 4 are nondecreasing in u, except for the nonnegative augmenting functions involving a symmetric positive definite matrix Q and level-bounded augmenting functions. However, even functions that involve a symmetric positive definite matrix Q are nondecreasing in u for the special case when Q is diagonal.

In the following lemma, we show that for any c, the optimal value of the penalized problem (36) is equal to the objective function of the geometric penalized problem, i.e., $\tilde{f}(c) = \tilde{d}(c)$ for all $c \geq 0$.

Lemma 6 *Let σ be an augmenting function and α be a scaling function that satisfy Assumption 4. Then,*

$$\tilde{d}(c) = \tilde{f}(c) \qquad \text{for all } c \geq 0.$$

Proof From the definition of the primal function in Eq. 37, it can be seen that

$$p(u) = \sup_{z\geq 0}\{f_0(x) + z'(f(x) - u)\}.$$

Therefore,

$$\tilde{d}(c) = \inf_{x\in X} \inf_{u\in\mathbb{R}^m} \sup_{z\geq 0}\{f_0(x) + z'(f(x) - u) + c\sigma(u)\}.$$

For every $x \in X$, by setting $u = f(x)$, we obtain

$$\inf_{u\in\mathbb{R}^m} \sup_{z\geq 0}\{f_0(x) + z'(f(x) - u) + c\sigma(u)\} \leq f_0(x) + c\sigma(f(x)) \qquad \text{for all } x \in X.$$

By taking the infimum over $x \in X$ of both sides, we see that

$$\tilde{d}(c) \leq \inf_{x\in X}\{f_0(x) + c\sigma(f(x))\} = \tilde{f}(c) \qquad \text{for all } c \geq 0.$$

By using Eq. 38 and the assumption that $\sigma(u)$ is nondecreasing in u, we obtain for all $c \geq 0$,

$$\tilde{d}(c) = \inf_{u\in\mathbb{R}^m} \inf_{\substack{x\in X\\ f(x)\leq u}} \{f_0(x) + c\sigma(u)\} \geq \inf_{x\in X}\{f_0(x) + c\sigma(f(x))\} = \tilde{f}(c),$$

establishing the result. □

6.3 Necessary and sufficient conditions for convergence of penalty methods

In this section, we consider a penalty method for which penalty (augmenting) function σ satisfies the assumptions of Sect. 4 We provide necessary and sufficient conditions for the convergence of the optimal values $\tilde{f}(c)$ of the penalized problems (36) to the optimal value f^* of the original constrained problem (35) when $c \to \infty$.

In the next proposition, we present the necessary conditions. These conditions require the continuity at $u = 0$ of the augmenting function $\sigma(u)$ (cf. Assumption 3).

Proposition 11 (Necessary conditions for penalty convergence) *Let σ be an augmenting function that satisfies Assumption 3. Consider the constrained problem (35) and the penalized problems (36) for $c \geq 0$. Assume that the optimal values $\tilde{f}(c)$ of the penalized problems converge to the optimal value f^* of the constrained problem as $c \to \infty$, i.e., $\lim_{c\to\infty} \tilde{f}(c) = f^*$. Then, the primal function $p(u)$ is lower semicontinuous at $u = 0$, i.e., for all sequences $\{u_k\} \subset \mathbb{R}^m$ with $u_k \to 0$, we have*

$$p(0) \leq \liminf_{k\to\infty} p(u_k).$$

Proof The result follows from Lemma 6 and Proposition 9 with $V = \mathrm{epi}(p)$. □

In the next proposition, we present sufficient conditions for the convergence of the optimal values $\tilde{f}(c)$ of the penalized problems to f^* as c tends to ∞.

Proposition 12 (Sufficient conditions for penalty convergence) *Assume that the constrained optimization problem (35) is feasible and bounded from below, i.e., f^* is finite. Assume that the direction $(0, -1)$ is not a recession direction of the epigraph of the primal function, i.e.,*

$$(0, -1) \notin (\mathrm{epi}(p))^{\infty}.$$

Assume that $p(u)$ is lower semicontinuous at $u = 0$. Furthermore, assume that the augmenting function σ satisfies Assumptions 2 and 4. Then, the optimal values $\tilde{f}(c)$ of the penalized problems (36) converge to the optimal value f^* of the constrained problem as $c \to \infty$, i.e.,

$$\lim_{c\to\infty} \tilde{f}(c) = f^*.$$

Proof We use Lemma 6 and Proposition 10 where the set V is the epigraph of the function p, i.e.,

$$V = \mathrm{epi}(p).$$

From the definition of p, we have

$$w^* = p(0) = f^*.$$

The assumption that the constrained optimization problem (35) is feasible and bounded from below implies that f^* is finite, and therefore w^* is finite. Since V is the epigraph of a function, the set V is extending upward in w-space. Moreover, V is extending upward in u-space because the primal function p is nonincreasing.

Together with the assumption that $(0,-1) \notin (\mathrm{epi}(p))^{\infty}$, we have that the set V satisfies Assumption 1.

By Assumption 4 and Lemma 6, we have $\tilde{d}(c) = \tilde{f}(c)$ for all $c \geq 0$. By the choice of set V, the condition

$$p(0) \leq \liminf_{k\to\infty} p(u_k)$$

for all sequences $\{u_k\}$ with $u_k \to 0$ is equivalent to the condition that for every sequence $\{(u_k, w_k)\} \subset V$ with $u_k \to 0$, there holds $w^* \leq \liminf_{k\to\infty} w_k$. The result then follows from Proposition 10. □

7 Conclusions

In this paper, we provided a unifying geometric framework that can be used to analyze optimization duality with nonlinear Lagrangian functions and to establish convergence behavior of general classes of penalty methods. We introduced two simple geometric optimization problems that are dual to each other and studied conditions under which the optimal values of these problems are equal. To establish this, we show that we can use general concave surfaces to separate nonconvex sets with certain properties.

We used our results to study both optimization duality and penalty methods for nonconvex constrained optimization problems. We first considered augmented dual problems constructed by general convex augmenting functions, with possibly unbounded level sets, and provided necessary and sufficient conditions for zero duality gap. We then considered penalty methods for which the associated penalty function need not be continuous, real-valued, or identically equal to zero over the feasible region. Without assuming any coercivity conditions, we provided necessary and sufficient conditions for the convergence of the penalized optimal values to the optimal value of the constrained optimization problem.

The zero duality gap results established here have potential use in the development of dual algorithms for solving nonconvex constrained optimization problems. In particular, for such problems, one may consider relaxing some or all of the constraints by using an augmented Lagrangian scheme or a penalty approach. Our results provide sufficient conditions guaranteeing the convergence of dual values to the primal optimal value.

Acknowledgements We would like to thank two anonymous referees for useful comments.

References

1. Auslender, A., Cominetti, R., Haddou, M.: Asymptotic analysis for penalty and barrier methods in convex and linear programming. Math. Oper. Res. **22**(1), 43–62 (1997)
2. Auslender, A., Teboulle, M.: Asymptotic Cones and Functions in Optimization and Variational Inequalities. Springer-Verlag, New York (2003)
3. Bertsekas, D.P., Nedić, A., Ozdaglar, A.: Convex Analysis and Optimization. Athena Scientific, Belmont, MA (2003)
4. Bertsekas D.P., Nedić A., Ozdaglar A.: Min common/max crossing duality: a simple geometric framework for convex optimization and minimax theory. Report LIDS-P-2536 (2002)
5. Bertsekas, D.P.: Nonlinear Programming. Athena Scientific, Belmont, MA (1999)

6. Bonnans, J.F., Shapiro, A.: Perturbation Analysis of Optimization Problems. Springer-Verlag, New York, NY (2000)
7. Borwein, J.M., Lewis, A.S.: Convex Analysis and Nonlinear Optimization. Springer-Verlag, New York, NY (2000)
8. Hiriart-Urruty, J.-B., Lemarechal, C.: Convex Analysis and Minimization Algorithms, vols I and II. Springer-Verlag, Berlin, NY (1993)
9. Huang, X.X., Yang, X.Q.: A unified augmented Lagrangian approach to duality and exact penalization. Math. Oper. Res. **28**(3), 533–552 (2003)
10. Luenberger, D.G.: Linear and Nonlinear Programming, 2nd edn. Kluwer Academic Publishers, MA (2004)
11. Polyak, B.T.: Introduction to Optimization. Optimization Software Inc., NY (1987)
12. Rockafellar, R.T.: Convex Analysis. Princeton University Press, Princeton NJ (1970)
13. Rockafellar, R.T., Wets, R.J.-B.: Variational Analysis. Springer-Verlag, New York (1998)
14. Rubinov, A.M., Huang, X.X., Yang, X.Q.: The zero duality gap property and lower semicontinuity of the perturbation function. Math. Oper. Res. **27**(4), 775–791 (2002)
15. Rubinov, A.M.: Abstract Convexity and Global Optimization. Kluwer Academic Publishers, Dordrecht (2000)
16. Rubinov, A.M., Yang, X.Q.: Lagrange-type Functions in Constrained Nonconvex Optimization. Kluwer Academic Publishers, MA (2003)
17. Tseng, P., Bertsekas, D.P.: On the convergence of the exponential multiplier method for convex programming. Math. Program. **60**, 1–19 (1993)