GAUGE OPTIMIZATION, DUALITY, AND APPLICATIONS*

MICHAEL P. FRIEDLANDER[†], IVES MACÊDO[†], AND TING KEI PONG[†]

Abstract. Gauge functions significantly generalize the notion of a norm, and gauge optimization, as defined by Freund (1987), seeks the element of a convex set that is minimal with respect to a gauge function. This conceptually simple problem can be used to model a remarkable array of useful problems, including a special case of conic optimization, and related problems that arise in machine learning and signal processing. The gauge structure of these problems allows for a special kind of duality framework. This paper explores the duality framework proposed by Freund, and proposes a particular form of the problem that exposes some useful properties of the gauge optimization framework (such as the variational properties of its value function), and yet maintains most of the generality of the abstract form of gauge optimization.

Key words. gauges, duality, convex optimization, nonsmooth optimization

AMS subject classifications. 90C15, 90C25

1. Introduction. One approach to solving linear inverse problems is to optimize a regularization function over the set of admissible deviations between the observations and the forward model. Although regularization functions come in a wide range of forms depending on the particular application, they often share some common properties. The aim of this paper is to describe the class of gauge optimization problems, which neatly captures a wide variety of regularization formulations that arise in fields such as machine learning and signal processing. We explore the duality and variational properties particular to this problem class, and consider some possible applications to relevant problems.

All of the problems that we consider can be expressed as

$$\text{(P)} \qquad \underset{x\in\mathcal{X}}{\text{minimize}} \quad \kappa(x) \quad \text{subject to} \quad x \in \mathcal{C},$$

where $\mathcal{X}$ is a finite-dimensional Euclidean space, $\mathcal{C} \subseteq \mathcal{X}$ is a closed convex set, and $\kappa : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ is a *gauge* function, i.e., a nonnegative, positively homogeneous convex function that vanishes at the origin. (We assume that $0 \notin \mathcal{C}$, since otherwise the origin is trivially a solution of the problem.) This class of problems admits a duality relationship that is different from Lagrange duality, and is founded on the gauge structure of its objective. Indeed, Freund (1987) defines the dual counterpart

$$\text{(D)} \qquad \underset{y\in\mathcal{X}}{\text{minimize}} \quad \kappa^\circ(y) \quad \text{subject to} \quad y \in \mathcal{C}',$$

where the set

$$\mathcal{C}' = \{\, y \mid \langle y, x\rangle \ge 1 \text{ for all } x \in \mathcal{C} \,\} \tag{1.1}$$

is the antipolar of $\mathcal{C}$ (in contrast to the better-known polar of a convex set), and the polar κ° is the function that best satisfies the Cauchy-Schwartz-like inequality

$$\langle x, y\rangle \le \kappa(x)\,\kappa^\circ(y), \qquad \forall x \in \operatorname{dom}\kappa,\ \forall y \in \operatorname{dom}\kappa^\circ. \tag{1.2}$$

*Research partially supported by NSERC Discovery grant 312104, and NSERC Collaborative Research and Development grant 375142-08.

[†]Department of Computer Science, University of British Columbia, Vancouver, BC, Canada. E-mail: {mpf,ijamj,tkpong}@cs.ubc.ca. MPF was also supported as a visiting scholar at the Simons Institute on Theoretical Computer Science at UC Berkeley; TKP was also supported as a PIMS Postdoctoral Fellow. **October 9, 2013**

It follows directly from this inequality and the definition of $\mathcal{C}'$ that all primal-dual feasible pairs (x, y) satisfy the weak-duality relationship

$$1 \le \kappa(x)\,\kappa^\circ(y), \qquad \forall x \in \mathcal{C},\ \forall y \in \mathcal{C}'. \tag{1.3}$$

This duality relationship stands in contrast to the more usual Lagrange framework, where the primal and dual objective values bound each other in an additive sense.

1.1. A roadmap. Freund's analysis of gauge duality is mainly concerned with specialized linear and quadratic problems that fit into the gauge framework, and with the pair of abstract problems (P) and (D).

Our treatment in this paper considers the particular formulation of (P) given by

$$\text{(P}_\rho\text{)} \qquad \underset{x\in\mathcal{X}}{\text{minimize}} \quad \kappa(x) \quad \text{subject to} \quad \rho(b - Ax) \le \sigma,$$

where ρ is also a gauge. Typical applications might use ρ to measure the mismatch between the model Ax and the measurements b, and in that case, it is natural to assume that ρ vanishes only at the origin, which implies that $Ax = b$ if $\sigma = 0$. This formulation is only very slightly less general than (P) because any closed convex set can be represented as $\{\, x \mid \rho(b - x) \le 1 \,\}$ for some vector b and gauge ρ; cf. §2.2. However, it is sufficiently concrete that it allows us to develop a toolkit for computing gauge duals for a wide range of existing problems. (Conic side constraints and a linear map in the objective can be easily accommodated; this is covered in §7.)

The special structure of the functions in the gauge program (P$_\rho$) leads to a duality framework that is analogous to the classical Lagrange-duality framework. The gauge dual program of (P$_\rho$) is

$$\text{(D}_\rho\text{)} \qquad \underset{y\in\mathcal{X}}{\text{minimize}} \quad \kappa^\circ(A^*y) \quad \text{subject to} \quad \langle y, b\rangle - \sigma\rho^\circ(y) \ge 1,$$

which bears a striking similarity to the Lagrange dual problem

$$\text{(D}_L\text{)} \qquad \underset{y\in\mathcal{X}}{\text{maximize}} \quad \langle y, b\rangle - \sigma\rho^\circ(y) \quad \text{subject to} \quad \kappa^\circ(A^*y) \le 1.$$

Note that the objective and constraints between the two duals play different roles. (These two duals are derived in §4 under suitable assumptions.) A significant practical difference between these two formulations is when ρ is a simple Euclidean norm and κ is a more complicated function (such as one described by Example 1.2 below). The result is that the Lagrange dual optimizes a "simple" objective function over a potentially "complicated" constraint; in contrast, the situation is reversed in the gauge optimization formulation.

We develop in §3 an antipolar calculus for computing the antipolars of sets such as $\{\, x \mid \rho(b - Ax) \le \sigma \,\}$, which corresponds to the constraint in our canonical formulation (P$_\rho$). This calculus is applied in §4 to derive the gauge dual (D$_\rho$).

The formal properties of the polar and antipolar operations are described in §§2–3. In §5 we develop conditions sufficient for strong duality, i.e., for there to exist a primal-dual pair that satisfies (1.3) with equality. Our derivation parts with the "ray-like" assumption used by Freund, and in certain cases further relaxes the required assumptions by leveraging connections with established results from Fenchel duality.

1.2. Examples. The following examples illustrate the versatility of the gauge optimization formulation.

Example 1.1 (Norms and minimum-length solutions). Norms are special cases of gauge functions that are finite everywhere, symmetric, and zero only at the origin. (Semi-norms drop the last requirement, and allow the function to be zero at other points.) Let $\kappa(x) = \|x\|$ be any norm, and $\mathcal{C} = \{\, x \mid Ax = b \,\}$ describe the solutions to an underdetermined linear system. Then (P) yields a minimum-length solution to the linear system $Ax = b$. In this case, ρ is any function such that $\rho^{-1}(0) = \{\, 0 \,\}$, and $\sigma = 0$. The polar $\kappa^\circ = \|\cdot\|_D$ is the norm dual to $\|\cdot\|$, and $\mathcal{C}' = \{\, A^*y \mid \langle b, y\rangle \geq 1 \,\}$; cf. Corollary 4.2. The corresponding gauge dual (D) is then

$$\underset{y\in\mathcal{X}}{\text{minimize}} \quad \|A^*y\|_D \quad \text{subject to} \quad \langle b, y\rangle \geq 1.$$

Example 1.2 (Sparse optimization and atomic norms). In his thesis, van den Berg (2009) describes a framework for sparse optimization based on the formulation where κ is a gauge, and the function ρ is differentiable away from the origin. The nonnegative regularization parameter σ influences the degree to which the linear model Ax fits the observations b. This problem is specialized by van den Berg and Friedlander (2011) to the particular case in which ρ is the 2-norm. In that case, $\mathcal{C} = \{\, x \mid \|Ax - b\|_2 \leq \sigma \,\}$ and

$$\mathcal{C}' = \{\, A^*y \mid \langle b, y\rangle - \sigma\|y\|_2 \geq 1 \,\};$$

cf. Corollary 4.1. Teuber, Steidl, and Chan (2013) consider a related case where the misfit between the model and the observations is measured by the Kullback-Leibler divergence.

Chandrasekaran, Recht, Parrilo, and Willsky (2012) describe how to construct regularizers that generalize the notion of sparsity in linear inverse problems. In particular, they define the gauge

$$\|x\|_{\mathcal{A}} := \inf\{\, \lambda \geq 0 \mid x \in \lambda \operatorname{conv}\mathcal{A} \,\} \tag{1.4}$$

over the convex hull of a set of canonical atoms in the set $\mathcal{A}$. If $0 \in \operatorname{int}\operatorname{conv}\mathcal{A}$ and $\mathcal{A}$ is bounded and symmetric, i.e., $\mathcal{A} = -\mathcal{A}$, then the definition (1.4) yields a norm. For example, if $\mathcal{A}$ consists of the set of unit n-vectors that contain a single nonzero element, then (1.4) is the 1-norm; if $\mathcal{A}$ consists of the set of rank-1 matrices with unit spectral norm, then (1.4) is the Schatten 1-norm. The polar $\kappa^\circ(y) = \sup\{\, \langle y, a\rangle \mid a \in \operatorname{conv}(\{0\} \cup \mathcal{A}) \,\}$ is the support function of the closure of $\operatorname{conv}(\{0\} \cup \mathcal{A})$. Jaggi (2013) catalogs various sets of atoms that yield commonly used gauges in machine learning.

Example 1.3 (Conic gauge optimization). Let $\mathcal{K}$ be a closed convex cone, and let $\mathcal{K}^*$ denote its dual. Then if $c \in \mathcal{K}^*$, the conic optimization problem

$$\underset{x}{\text{minimize}} \quad \langle c, x\rangle \quad \text{subject to} \quad Ax = b,\ x \in \mathcal{K} \tag{1.5}$$

has a nonnegative objective value for all feasible points, and can be formulated as a gauge optimization problem by defining

$$\kappa(x) = \langle c, x\rangle + \delta_{\mathcal{K}}(x) \quad \text{and} \quad \mathcal{C} = \{\, x \mid Ax = b \,\}, \tag{1.6}$$

where $\delta_{\mathcal{K}}$ is the indicator function on the set $\mathcal{K}$. This is a generalization of the nonnegative LP discussed by Freund, and we refer to it as conic gauge optimization.

The generalization captures some important problem classes, such as trace minimization of positive semidefinite matrices, which arises in the phase-retrieval problem (Candes, Strohmer, and Voroninski, 2012).

A concrete example of this general formulation is the semidefinite programming (SDP) relaxation of the max-cut problem studied by Goemans and Williamson (1995). Let $G = (V, E)$ be an undirected graph, and $D = \operatorname{diag}\big((d_v)_{v\in V}\big)$, where d_v denotes the degree of vertex $v \in V$. The max-cut problem can be formulated as

$$\underset{x}{\text{maximize}} \quad \tfrac{1}{4}\langle D - A, xx^T\rangle \quad \text{subject to} \quad x \in \{-1, 1\}^V,$$

where A denotes the adjacency matrix associated with G. The SDP relaxation for this problem is derived by "lifting" xx^T into a positive semidefinite matrix:

$$\underset{X}{\text{maximize}} \quad \tfrac{1}{4}\langle D - A, X\rangle \quad \text{subject to} \quad \operatorname{diag} X = e,\ X \succcurlyeq 0,$$

where e is the vector of all ones. The constraint $\operatorname{diag} X = e$ implies that $\langle D, X\rangle = \sum_{v\in V} d_v = 2|E|$ is constant. Thus, the optimal value is equal to

$$|E| - \tfrac{1}{4} \cdot \min_X \{\, \langle D + A, X\rangle \mid \operatorname{diag} X = e,\ X \succcurlyeq 0 \,\} \tag{1.7}$$

and the solution can be obtained by solving this latter problem. Note that $D + A$ is positive semidefinite because it has nonnegative diagonal and is diagonally dominant. (In fact, it is possible to reduce the problem in linear time to one where $D + A$ is positive definite by identifying its bipartite connected components.) Thus, (1.7) falls into the class of conic gauge problems defined by (1.5).

Example 1.4 (Submodular functions). Let $V = \{\,1, \dots, n\,\}$, and consider the set-function $f : 2^V \to \mathbb{R}$, where $f(\emptyset) = 0$. The Lovàsz (1983) extension $\widehat{f} : \mathbb{R}^n \to \mathbb{R}$ of f is given by

$$\widehat{f}(x) = \sum_{k=1}^{n} x_{j_k}\big[f(\{\, j_1, \dots, j_k \,\}) - f(\{\, j_1, \dots, j_{k-1} \,\})\big],$$

where $x_{j_1} \ge x_{j_2} \ge \cdots \ge x_{j_n}$ are the sorted elements of x. Clearly, the extension is positively homogeneous and vanishes at the origin. As shown by Lovász, the extension is convex if and only if f is *submodular*, i.e.,

$$f(A) + f(B) \ge f(A \cup B) + f(A \cap B) \quad \text{for all} \quad A, B \subset V;$$

see also (Bach, 2011, Proposition 2.3). If f is additionally non-decreasing, i.e.,

$$A, B \subset V \text{ and } A \subset B \quad \Longrightarrow \quad f(A) \le f(B),$$

then the extension is nonnegative. Thus, the extension $\widehat{f}$ of a submodular and non-decreasing set function is a gauge. Bach (2011) surveys the properties of submodular functions and their application in machine learning.

2. Background and notation. In this section we review known facts about polar sets, gauges and their polars, and introduce results that are useful for our subsequent analysis. We mainly follow Rockafellar (1970): see §14 in that text for a discussion of polarity operations on convex sets, and §15 for a discussion of gauges and the corresponding polarity operations.

We use the following notation throughout. For a closed convex set $\mathcal{D}$, $\mathcal{D}_\infty$ denotes the recession cone of $\mathcal{D}$ (Auslender and Teboulle, 2003, Definition 2.1.2), $\operatorname{ri}\mathcal{D}$ and $\operatorname{cl}\mathcal{D}$ denote, respectively, the relative interior and the closure of $\mathcal{D}$. The indicator function of $\mathcal{D}$ is denoted by $\delta_{\mathcal{D}}$.

For a gauge $\kappa : \mathcal{X} \to \mathbb{R} \cup \{\infty\}$, its domain is denoted by $\operatorname{dom}\kappa = \{ x \mid \kappa(x) < \infty \}$, and its epigraph is denoted by $\operatorname{epi}\kappa = \{ (x, \mu) \mid \kappa(x) \le \mu \}$. A function is called closed if its epigraph is closed, which is equivalent to the function being lower semi-continuous (Rockafellar, 1970, Theorem 7.1). Let $\operatorname{cl}\kappa$ denote the gauge whose epigraph is $\operatorname{cl}\operatorname{epi}\kappa$, which is the largest lower semi-continuous function smaller than κ (Rockafellar, 1970, p. 52). Finally, for any $x \in \operatorname{dom}\kappa$, the subdifferential of κ at x is denoted $\partial\kappa(x) = \{ y \mid \kappa(u) - \kappa(x) \ge \langle y, u - x\rangle,\ \forall u \}$.

We make the following blanket assumptions throughout. Let $\mathcal{C}$ denote a nonempty closed convex set that does not contain the origin; let $\mathcal{D}$ denote a nonempty convex set that may or may not contain the origin, depending on the context. The gauge function $\rho : \mathcal{X} \to \mathbb{R} \cup \{\infty\}$, used in ($\mathrm{P}_\rho$), is closed; when $\sigma = 0$, we additionally assume that $\rho^{-1}(0) = \{0\}$.

2.1. Polar sets. The polar $\mathcal{D}^\circ$ of a nonempty closed convex set $\mathcal{D}$ is defined by

$$\mathcal{D}^\circ := \{ y \mid \langle x, y\rangle \le 1,\ \forall x \in \mathcal{D} \},$$

which is necessarily closed convex, and contains the origin. The bipolar theorem states that if $\mathcal{D}$ is closed, then it contains the origin if and only if $\mathcal{D} = \mathcal{D}^{\circ\circ}$ (Rockafellar, 1970, Theorem 14.5).

When $\mathcal{D} = \mathcal{K}$ is a closed convex cone, the polar is equivalently given by

$$\mathcal{K}^\circ := \{ y \mid \langle x, y\rangle \le 0,\ \forall x \in \mathcal{K} \}.$$

The positive polar cone (also known as the dual cone) of $\mathcal{D}$ is given by

$$\mathcal{D}^* := \{ y \mid \langle x, y\rangle \ge 0,\ \forall x \in \mathcal{D} \}.$$

The polar and positive polar are related via the closure of the conic hull, i.e.,

$$\mathcal{D}^* = (\operatorname{cl}\operatorname{cone}\mathcal{D})^* = -(\operatorname{cl}\operatorname{cone}\mathcal{D})^\circ, \quad \text{where} \quad \operatorname{cone}\mathcal{D} = \bigcup_{\lambda\ge 0} \lambda\mathcal{D}.$$

2.2. Gauge functions. All gauges can be represented in the form of a Minkowski function $\gamma_{\mathcal{D}}$ of some nonempty convex set $\mathcal{D}$, i.e.,

$$\kappa(x) = \gamma_{\mathcal{D}}(x) := \inf \{ \lambda \ge 0 \mid x \in \lambda\mathcal{D} \}. \tag{2.1}$$

In particular, one can always choose $\mathcal{D} = \{ x \mid \kappa(x) \le 1 \}$, and the above representation holds. The polar κ° of the gauge κ is defined as

$$\kappa^\circ(y) := \inf \{ \mu > 0 \mid \langle x, y\rangle \le \mu\kappa(x),\ \forall x \},$$

which gives the inequality (1.2). Because κ is a proper convex function, one can also define its convex conjugate:

$$\kappa^*(y) := \sup_x \{ \langle x, y\rangle - \kappa(x) \}. \tag{2.2}$$

It is well known that κ^* is a proper closed convex function (Rockafellar, 1970, Theorem 12.2). The following proposition collects properties that relate the polar and conjugate of a gauge.

Proposition 2.1.
(i) κ° is a closed gauge function;
(ii) $\kappa^{\circ\circ} = \operatorname{cl}\kappa = \kappa^{**}$;
(iii) $\kappa^\circ(y) = \sup_x \{\, \langle x, y\rangle \mid \kappa(x) \le 1 \,\}$ for all y;
(iv) $\kappa^*(y) = \delta_{\kappa^\circ(\cdot)\le 1}(y)$ for all y;
(v) $\operatorname{dom}\kappa^\circ = \mathcal{X}$ if κ is closed and $\kappa^{-1}(0) = \{\, 0 \,\}$.

Proof. The first two items are proved in Theorems 15.1 and 12.2 of Rockafellar (1970). Item (iii) follows directly from the definition (2.2) of the polar gauge. Item (iv) follows by applying Theorem 15.3 of Rockafellar (1970) with $g(t) = t$, and κ^{**} in place of f in that theorem, and noting that $\kappa^{***} = \kappa^*$ and $\kappa^{**\circ} = \kappa^{\circ\circ\circ} = \kappa^\circ$. To prove item (v), note that the assumptions together with Proposition 3.1.3 of Auslender and Teboulle (2003) show that $0 \in \operatorname{int}\operatorname{dom}\kappa^*$. This together with item (iv) and the positive homogeneity of κ° shows that $\operatorname{dom}\kappa^\circ = \mathcal{X}$. ∎

In some interesting applications, the objective in (P) is the composition $\kappa \circ A$, where κ is a gauge and A is a linear map. Clearly, $\kappa \circ A$ is also a gauge. The next result gives the polar of this composition.

Proposition 2.2. Let A be a linear map. Suppose that either
(i) $\operatorname{epi}\kappa$ is polyhedral; or
(ii) $\operatorname{ri}\operatorname{dom}\kappa \cap \operatorname{range} A \ne \emptyset$.
Then

$$(\kappa \circ A)^\circ(y) = \inf_u \{\, \kappa^\circ(u) \mid A^* u = y \,\}.$$

Moreover, the infimum is attained when the value is finite.

Proof. Since $\kappa \circ A$ is a gauge, we have from Proposition 2.1(iii) that

$$(\kappa \circ A)^\circ(y) = \sup_x \{\, \langle y, x\rangle \mid \kappa(Ax) \le 1 \,\} = -\inf_x \{\, \langle -y, x\rangle + \delta_{\mathcal{D}}(Ax) \,\},$$

where $\mathcal{D} = \{\, x \mid \kappa(x) \le 1 \,\}$. Since κ is positively homogeneous, we have $\operatorname{dom}\kappa = \bigcup_{\lambda\ge 0} \lambda\mathcal{D}$. Hence, $\operatorname{ri}\operatorname{dom}\kappa = \bigcup_{\lambda>0} \lambda \operatorname{ri}\mathcal{D}$ from Rockafellar (1970, p. 50). Thus, assumption (ii) implies that $\operatorname{ri}\mathcal{D} \cap \operatorname{range} A \ne \emptyset$. On the other hand, assumption (i) implies that $\mathcal{D}$ is polyhedral. Using these and Rockafellar (1970, Corollary 31.2.1), we conclude that

$$\begin{aligned}(\kappa \circ A)^\circ(y) &= -\sup_u \{\, -(\langle -y, \cdot\rangle)^*(-A^* u) - (\delta_{\mathcal{D}})^*(u) \,\} \\ &= -\sup_u \{\, -\kappa^\circ(u) \mid A^* u = y \,\},\end{aligned}$$

where the second equality follows from the definition of conjugate functions and Proposition 2.1(iii). Moreover, the supremum is attained when finite, again by Rockafellar (1970, Theorem 31.1). This completes the proof. ∎

Suppose that a gauge is given as the Minkowski function of a nonempty convex set that may not necessarily contain the origin. The following proposition summarizes some properties concerning this representation.

Proposition 2.3. Suppose that $\mathcal{D}$ is a nonempty convex set. Then
(i) $(\gamma_{\mathcal{D}})^\circ = \gamma_{\mathcal{D}^\circ}$;
(ii) $\gamma_{\mathcal{D}} = \gamma_{\operatorname{conv}(\{0\}\cup\mathcal{D})}$;

(iii) $\gamma_{\mathcal{D}}$ is closed if $\operatorname{conv}(\{0\} \cup \mathcal{D})$ is closed.
(iv) If $\kappa = \gamma_{\mathcal{D}}$, $\mathcal{D}$ is closed, and $0 \in \mathcal{D}$, then $\mathcal{D}$ is the unique closed convex set containing the origin such that $\kappa = \gamma_{\mathcal{D}}$; indeed, $\mathcal{D} = \{ x \mid \kappa(x) \le 1 \}$.

Proof. Item (i) is proved in Rockafellar (1970, Theorem 15.1). Item (ii) follows directly from the definition. To prove (iii), we first notice from item (ii) that we may assume without loss of generality that $\mathcal{D}$ contains the origin. Notice also that $\gamma_{\mathcal{D}}$ is closed if and only if $\gamma_{\mathcal{D}} = \gamma_{\mathcal{D}}^{**}$. Moreover, $\gamma_{\mathcal{D}}^{**} = \gamma_{\mathcal{D}^{\circ\circ}} = \gamma_{\operatorname{cl}\mathcal{D}}$, where the first equality follows from Proposition 2.1(ii) and item (i), while the second equality follows from the bipolar theorem. Thus, $\gamma_{\mathcal{D}}$ is closed if and only if $\gamma_{\mathcal{D}} = \gamma_{\operatorname{cl}\mathcal{D}}$. The latter holds when $\mathcal{D} = \operatorname{cl}\mathcal{D}$. Finally, the conclusion in item (iv) was stated on Rockafellar (1970, p. 128); indeed, the relation $\mathcal{D} = \{ x \mid \kappa(x) \le 1 \}$ can be verified directly from definition. ∎

From Proposition 2.1(iv) and Proposition 2.3(iv), it is not hard to prove the following formula on the polar of the sum of two gauges of independent variables.

PROPOSITION 2.4. Let κ_1 and κ_2 be gauges. Then $\kappa(x_1, x_2) := \kappa_1(x_1) + \kappa_2(x_2)$ is a gauge, and its polar is given by

$$\kappa^\circ(y_1, y_2) = \max \{ \kappa_1^\circ(y_1),\ \kappa_2^\circ(y_2) \}.$$

Proof. It is clear that κ is a gauge. Moreover,

$$\kappa^*(y_1, y_2) = \kappa_1^*(y_1) + \kappa_2^*(y_2) = \delta_{\mathcal{D}_1 \times \mathcal{D}_2}(y_1, y_2),$$

where $\mathcal{D}_i = \{ x \mid \kappa_i^\circ(x) \le 1 \}$ for $i = 1, 2$; the first equality follows from the definition of the convex conjugate and the fact that y_1 and y_2 are decoupled, and the second equality follows from Proposition 2.1(iv). This together with Proposition 2.3(iv) implies that

$$\begin{aligned} \kappa^\circ(y_1, y_2) &= \inf \{ \lambda \ge 0 \mid y_1 \in \lambda\mathcal{D}_1,\ y_2 \in \lambda\mathcal{D}_2 \} \\ &= \max \{ \inf \{ \lambda \ge 0 \mid y_1 \in \lambda\mathcal{D}_1 \},\ \inf \{ \lambda \ge 0 \mid y_2 \in \lambda\mathcal{D}_2 \} \} \\ &= \max \{ \gamma_{\mathcal{D}_1}(y_1),\ \gamma_{\mathcal{D}_2}(y_2) \} = \max \{ \kappa_1^\circ(y_1),\ \kappa_2^\circ(y_2) \}. \end{aligned}$$

This completes the proof. ∎

The following corollary is immediate from Proposition 2.2 and Proposition 2.4.

COROLLARY 2.5. Let κ_1 and κ_2 be gauges. Suppose that either
(i) $\operatorname{epi}\kappa_1$ and $\operatorname{epi}\kappa_2$ are polyhedral; or
(ii) $\operatorname{ri}\operatorname{dom}\kappa_1 \cap \operatorname{ri}\operatorname{dom}\kappa_2 \ne \emptyset$.
Then the polar of $\kappa := \kappa_1 + \kappa_2$ is

$$\kappa^\circ(y) = \inf_{u_1, u_2} \{ \max \{ \kappa_1^\circ(u_1),\ \kappa_2^\circ(u_2) \} \mid u_1 + u_2 = y \}. \tag{2.3}$$

Moreover, the infimum is attained when finite.

Proof. Apply Proposition 2.2 with $Ax = (x, x)$ and the gauge $\kappa_1(x_1) + \kappa_2(x_2)$, whose polar is given by Proposition 2.4. ∎

For a nonempty convex set $\mathcal{D}$, the support function is defined as

$$\sigma_{\mathcal{D}}(y) = \sup_{x \in \mathcal{D}} \langle x, y \rangle.$$

It is easy to check that if $\mathcal{D}$ contains the origin, then the support function is a (closed) gauge function. Indeed, we have the following relationship between support and Minkowski functions; see Rockafellar (1970, Corollary 15.1.2).

PROPOSITION 2.6. Let $\mathcal{D}$ be a closed convex set that contains the origin. Then $\gamma_{\mathcal{D}}^{\circ} = \sigma_{\mathcal{D}}$ and $\sigma_{\mathcal{D}}^{\circ} = \gamma_{\mathcal{D}}$.

2.3. Antipolar sets. The antipolar $\mathcal{C}'$, defined by (1.1), is nonempty as a consequence of the separation theorem. Freund's 1987 derivations are largely based on the following definition of a ray-like set. (As Freund mentions, the terms *antipolar* and *ray-like* are not universally used.)

DEFINITION 2.7. A set $\mathcal{D}$ is *ray-like* if for any $x, y \in \mathcal{D}$,

$$x + \alpha y \in \mathcal{D} \quad \text{for all} \quad \alpha \geq 0.$$

It is easy to check that the antipolar $\mathcal{C}'$ of a (not necessarily ray-like) set $\mathcal{C}$ must be ray-like.

The following result is analogous to the bipolar theorem for antipolar operations; see McLinden (1978, p.176) and Freund (1987, Lemma 3).

THEOREM 2.8 (Bi-antipolar theorem). $\mathcal{C} = \mathcal{C}''$ if and only if $\mathcal{C}$ is ray-like.

The following proposition, stated by McLinden (1978, p.176), is not hard to show using the above theorem.

PROPOSITION 2.9. $\mathcal{C}'' = \bigcup_{\lambda \geq 1} \lambda \mathcal{C}$.

The next lemma relates the positive polar of a convex set, its antipolar and the recession cone of its antipolar.

LEMMA 2.10. $\operatorname{cl}\operatorname{cone}(\mathcal{C}') = \mathcal{C}^* = (\mathcal{C}')_\infty$.

Proof. It is evident that $\operatorname{cl}\operatorname{cone}(\mathcal{C}') \subseteq \mathcal{C}^*$. To show the converse inclusion, take any $x \in \mathcal{C}^*$ and fix an $x_0 \in \mathcal{C}'$. Then for any $\tau > 0$, we have

$$\langle c, x + \tau x_0 \rangle \geq \tau \langle c, x_0 \rangle \geq \tau \quad \text{for all} \quad c \in \mathcal{C},$$

which shows that $x + \tau x_0 \in \operatorname{cone}\mathcal{C}'$. Taking the limit as τ goes to 0 shows that $x \in \operatorname{cl}\operatorname{cone}(\mathcal{C}')$. This proves the first equality.

Next we show the second equality, and begin with the observation that $\mathcal{C}^* \subseteq (\mathcal{C}')_\infty$. Conversely, suppose that $x \in (\mathcal{C}')_\infty$ and fix any $x_0 \in \mathcal{C}'$. Then, by Auslender and Teboulle (2003, Proposition 2.1.5), $x_0 + \tau x \in \mathcal{C}'$ for all $\tau > 0$. Hence, for any $c \in \mathcal{C}$,

$$\frac{1}{\tau}\langle c, x_0 \rangle + \langle c, x \rangle = \frac{1}{\tau}\langle c, x_0 + \tau x \rangle \geq \frac{1}{\tau}.$$

Since this is true for all $\tau > 0$, we must have $\langle c, x \rangle \geq 0$. Since $c \in \mathcal{C}$ is arbitrary, we conclude that $x \in \mathcal{C}^*$. ∎

3. Antipolar calculus. In general, it may not always be easy to obtain an explicit formula for the Minkowski function of a given closed convex set $\mathcal{D}$. Hence, we derive some elements of an antipolar calculus that allows us to express the antipolar of a more complicated set in terms of the antipolars of its constituents. These rules are

TABLE 3.1
The main rules of the antipolar calculus; the required assumptions are made explicit in the specific references.

Result	Reference
$(A\mathcal{C})' = (A^*)^{-1}\mathcal{C}'$	Proposition 3.3
$(A^{-1}\mathcal{C})' = \mathrm{cl}(A^*\mathcal{C}')$	Corollaries 3.4 and 3.5
$(\mathcal{C}_1 \cup \mathcal{C}_2)' = \mathcal{C}_1' \cap \mathcal{C}_2'$	Proposition 3.6
$(\mathcal{C}_1 \cap \mathcal{C}_2)' = \mathrm{cl\,conv}(\mathcal{C}_1' \cup \mathcal{C}_2')$	Proposition 3.7

useful for writing down the explicit gauge duals of problems such as (P_ρ). Table 3.1 summarizes the main elements of the calculus.

As a first step, the following formula gives an expression for the antipolar of a set defined via a gauge. The formula follows directly from the definition of polar functions.

PROPOSITION 3.1. Let $\mathcal{C} = \{\, x \mid \rho(b - x) \le \sigma \,\}$ with $0 < \sigma < \rho(b)$. Then

$$\mathcal{C}' = \{\, y \mid \langle b, y\rangle - \sigma\rho^\circ(y) \ge 1 \,\}.$$

Proof. Note that $y \in \mathcal{C}'$ is equivalent to $\langle x, y\rangle \ge 1$ for all $x \in \mathcal{C}$. Thus, for all x such that $\rho(b - x) \le \sigma$,

$$\langle x - b, y\rangle \ge 1 - \langle b, y\rangle \iff \langle b - x, y\rangle \le \langle b, y\rangle - 1 \iff \sigma\rho^\circ(y) \le \langle b, y\rangle - 1,$$

where the last equivalence follows from Proposition 2.1(iii). This completes the proof. ∎

Proposition 3.1 is very general since any closed convex set $\mathcal{D}$ containing the origin can be represented in the form of $\{\, x \mid \rho(x) \le 1 \,\}$, where $\rho(x) = \inf\{\, \lambda \ge 0 \mid x \in \lambda\mathcal{D} \,\}$; cf. (2.1). The following corollary demonstrates the generality of the last result.

COROLLARY 3.2. Let $\mathcal{C} = \{\, x \mid x \in b + \mathcal{K} \,\}$ for some closed convex cone $\mathcal{K}$ and a vector $b \notin -\mathcal{K}$. Then

$$\mathcal{C}' = \{\, y \in \mathcal{K}^* \mid \langle b, y\rangle \ge 1 \,\}.$$

Proof. Apply Proposition 3.1 with $\rho(x) = \delta_{-\mathcal{K}}(x)$, the indicator function of the cone $-\mathcal{K}$, and notice that in this case, $\rho^\circ(y) = \delta_{\mathcal{K}^*}(y)$. ∎

Note that Proposition 3.1 excludes the potentially important case $\sigma = 0$; however, Corollary 3.2 can instead be applied by defining $\mathcal{K} = \rho^{-1}(0) = \{\, 0 \,\}$.

3.1. Linear transformations. We now consider the antipolar of the image of $\mathcal{C}$ under a linear map A.

PROPOSITION 3.3. It holds that

$$(A\mathcal{C})' = (A^*)^{-1}\mathcal{C}'.$$

Furthermore, if $\mathrm{cl}(A\mathcal{C})$ does not contain the origin, then both sets above are nonempty.

Proof. Note that $y \in (A\mathcal{C})'$ is equivalent to

$$\langle y, Ac\rangle = \langle A^*y, c\rangle \geq 1 \quad \text{for all} \quad c \in \mathcal{C}.$$

The last relation is equivalent to $A^*y \in \mathcal{C}'$. Hence, we have $(A\mathcal{C})' = (A^*)^{-1}\mathcal{C}'$. Furthermore, the assumption $\mathrm{cl}(A\mathcal{C})$ does not contain the origin implies that $(A\mathcal{C})'$ is nonempty. This completes the proof. □

As a corollary, we get the following result concerning the pre-image of $\mathcal{C}$.

COROLLARY 3.4. Suppose that $A^{-1}\mathcal{C} \neq \emptyset$. Then

$$(A^{-1}\mathcal{C})' = \mathrm{cl}(A^*\mathcal{C}'),$$

and both sets are nonempty.

Proof. First, it is clear that $\mathrm{cl}(A^*\mathcal{C}')$ is nonempty. Moreover, since $\mathcal{C}$ does not contain the origin, $A^{-1}\mathcal{C}$ is also a closed convex set that does not contain the origin. Hence, $(A^{-1}\mathcal{C})'$ is also nonempty.

We next show that $\mathrm{cl}(A^*\mathcal{C}')$ does not contain the origin. Suppose that $y \in A^*\mathcal{C}'$ so that $y = A^*u$ for some $u \in \mathcal{C}'$. Then for any $x \in A^{-1}\mathcal{C}$, we have $Ax \in \mathcal{C}$ and thus

$$\langle x, y\rangle = \langle x, A^*u\rangle = \langle Ax, u\rangle \geq 1,$$

which shows that $y \in (A^{-1}\mathcal{C})'$. Thus, we have $A^*\mathcal{C}' \subseteq (A^{-1}\mathcal{C})'$ and consequently that $\mathrm{cl}(A^T\mathcal{C}) \subseteq (A^{-1}\mathcal{C})'$. Since the set $A^{-1}\mathcal{C}$ is nonempty, $(A^{-1}\mathcal{C})'$ does not contain the origin. Hence, it follows that $\mathrm{cl}(A^*\mathcal{C}')$ also does not contain the origin.

Now apply Proposition 3.3 with A^* in place of A, and $\mathcal{C}'$ in place of $\mathcal{C}$, to obtain

$$(A^*\mathcal{C}')' = A^{-1}\mathcal{C}''.$$

Taking the antipolar on both sides of the above relation, we arrive at

$$(A^*\mathcal{C}')'' = (A^{-1}\mathcal{C}'')'. \tag{3.1}$$

Since $\mathcal{C}'$ is ray-like, it follows that $\mathrm{cl}(A^*\mathcal{C}')$ is also ray-like. Since $\mathrm{cl}(A^*\mathcal{C}')$ does not contain the origin, we conclude that $(A^*\mathcal{C}')'' = \mathrm{cl}(A^*\mathcal{C}')$. Moreover, we have

$$(A^{-1}\mathcal{C}'')' = \left(\bigcup_{\lambda\geq 1} \lambda A^{-1}\mathcal{C}\right)' = \left(A^{-1}\mathcal{C}\right)',$$

where the first equality follows from Proposition 2.9, and the second equality can be verified directly from definition. The conclusion now follows from the above discussion and (3.1). □

We have the following further corollary.

COROLLARY 3.5. Suppose that $A^{-1}\mathcal{C} \neq \emptyset$, and either $\mathcal{C}$ is polyhedral or $\mathrm{ri}\,\mathcal{C} \cap \mathrm{range}\, A \neq \emptyset$. Then $(A^{-1}\mathcal{C})'$ is nonempty and

$$\left(A^{-1}\mathcal{C}\right)' = A^*\mathcal{C}'.$$

Proof. We will show that $A^*\mathcal{C}'$ is closed under the assumption of this corollary. Then the conclusion follows immediately from Corollary 3.4.

Abrams's theorem (Berman, 1973, Lemma 3.1) asserts that $A^*\mathcal{C}'$ is closed if and only if $\mathcal{C}' + \ker A^*$ is closed. We will thus establish the closedness of the latter set.

Suppose that $\mathcal{C}$ is a polyhedral. Then it is routine to show that $\mathcal{C}'$ is also a polyhedral and thus $\mathcal{C}' + \ker A^*$ is closed. Hence, the conclusion of the corollary holds under this assumption.

Finally, suppose that $\operatorname{ri}\mathcal{C} \cap \operatorname{range} A \neq \emptyset$. From Auslender and Teboulle (2003, Theorem 2.2.1) and the bipolar theorem, we have $\operatorname{cl}\operatorname{dom}(\sigma_{\mathcal{C}'}) = [(\mathcal{C}')_\infty]^\circ$, where $(\mathcal{C}')_\infty$ is the recession cone of $\mathcal{C}'$, which turns out to be just $\mathcal{C}^*$ by Lemma 2.10. From this and the bipolar theorem, we see further that

$$\operatorname{cl}\operatorname{dom}(\sigma_{\mathcal{C}'}) = [\mathcal{C}^*]^\circ = -\operatorname{cl}\operatorname{cone}\mathcal{C},$$

and hence $\operatorname{ri}\operatorname{dom}\sigma_{\mathcal{C}'} = -\operatorname{ri}\operatorname{cone}\mathcal{C}$, thanks to Rockafellar (1970, Theorem 6.3). Furthermore, the assumption that $\operatorname{ri}\mathcal{C} \cap \operatorname{range} A \neq \emptyset$ is equivalent to $\operatorname{ri}\operatorname{cone}\mathcal{C} \cap \operatorname{range} A \neq \emptyset$, since $\operatorname{ri}\operatorname{cone}\mathcal{C} = \bigcup_{\lambda>0} \lambda \operatorname{ri}\mathcal{C}$; see Rockafellar (1970, p. 50). Thus, the assumption $\operatorname{ri}\mathcal{C} \cap \operatorname{range} A \neq \emptyset$ together with Rockafellar (1970, Theorem 23.8) imply that

$$\mathcal{C}' + \ker A^* = \partial\sigma_{\mathcal{C}'}(0) + N_{\operatorname{range} A}(0) = \partial(\sigma_{\mathcal{C}'} + \delta_{\operatorname{range} A})(0).$$

In particular, $\mathcal{C}' + \ker A^*$ is closed. □

3.2. Unions and intersections. Other important set operations are union and intersection, which we discuss here. Ruys and Weddepohl (1979, Appendix A.1) outline additional rules.

PROPOSITION 3.6. Let $\mathcal{C}_1$ and $\mathcal{C}_2$ be nonempty closed convex sets. Then

$$(\mathcal{C}_1 \cup \mathcal{C}_2)' = \mathcal{C}_1' \cap \mathcal{C}_2'.$$

If $0 \notin \operatorname{cl}\operatorname{conv}(\mathcal{C}_1 \cup \mathcal{C}_2)$, then the sets above are nonempty.

Proof. Note that $y \in (\mathcal{C}_1 \cup \mathcal{C}_2)'$ is equivalent to $\langle y, x\rangle \geq 1$ for all $x \in \mathcal{C}_1$ as well as $x \in \mathcal{C}_2$. This is equivalent to $y \in \mathcal{C}_1' \cap \mathcal{C}_2'$. Moreover, if we assume further that $0 \notin \operatorname{cl}\operatorname{conv}(\mathcal{C}_1 \cup \mathcal{C}_2)$, then $(\mathcal{C}_1 \cup \mathcal{C}_2)' = [\operatorname{cl}\operatorname{conv}(\mathcal{C}_1 \cup \mathcal{C}_2)]'$ is nonempty. This completes the proof. □

We now consider the antipolar of intersections. Note that it is necessary to assume that both $\mathcal{C}_1$ and $\mathcal{C}_2$ are ray-like, which was missing from Ruys and Weddepohl (1979, Property A.5). (The necessity of this assumption is demonstrated by Example 3.1, which follows the proposition.)

PROPOSITION 3.7. Let $\mathcal{C}_1$ and $\mathcal{C}_2$ be nonempty ray-like closed convex sets not containing the origin. Suppose further that $\mathcal{C}_1 \cap \mathcal{C}_2 \neq \emptyset$. Then

$$(\mathcal{C}_1 \cap \mathcal{C}_2)' = \operatorname{cl}\operatorname{conv}(\mathcal{C}_1' \cup \mathcal{C}_2'),$$

and both sets are nonempty.

Proof. First, it is clear that $\operatorname{cl}\operatorname{conv}(\mathcal{C}_1' \cup \mathcal{C}_2')$ is nonempty. Moreover, since $\mathcal{C}_1 \cap \mathcal{C}_2$ does not contain the origin, $(\mathcal{C}_1 \cap \mathcal{C}_2)'$ is also nonempty.

We first show that $\operatorname{cl}\operatorname{conv}(\mathcal{C}_1' \cup \mathcal{C}_2')$ does not contain the origin. To this end, let $y \in \mathcal{C}_1' \cup \mathcal{C}_2'$. For any $x \in \mathcal{C}_1 \cap \mathcal{C}_2$, we have $\langle y, x\rangle \geq 1$, which shows that $\mathcal{C}_1' \cup \mathcal{C}_2' \subseteq (\mathcal{C}_1 \cap \mathcal{C}_2)'$, and hence $\operatorname{cl}\operatorname{conv}(\mathcal{C}_1' \cup \mathcal{C}_2') \subseteq (\mathcal{C}_1 \cap \mathcal{C}_2)'$. Since $\mathcal{C}_1 \cap \mathcal{C}_2$ is nonempty, $(\mathcal{C}_1 \cap \mathcal{C}_2)'$ does not contain the origin. Consequently, $\operatorname{cl}\operatorname{conv}(\mathcal{C}_1' \cup \mathcal{C}_2')$ does not contain the origin, as claimed.

Now apply Proposition 3.6, with $\mathcal{C}_1'$ in place of $\mathcal{C}_1$ and $\mathcal{C}_2'$ in place of $\mathcal{C}_2$, to obtain

$$(\mathcal{C}_1' \cup \mathcal{C}_2')' = \mathcal{C}_1'' \cap \mathcal{C}_2'' = \mathcal{C}_1 \cap \mathcal{C}_2.$$

Taking antipolar on both sides, we obtain further that

$$(\mathcal{C}_1 \cap \mathcal{C}_2)' = (\mathcal{C}_1' \cup \mathcal{C}_2')'' = [\operatorname{cl}\operatorname{conv}(\mathcal{C}_1' \cup \mathcal{C}_2')]'' = \operatorname{cl}\operatorname{conv}(\mathcal{C}_1' \cup \mathcal{C}_2'),$$

where the second equality follows from the definition of antipolar, and the third equality follows from the observation that $\operatorname{cl}\operatorname{conv}(\mathcal{C}_1' \cup \mathcal{C}_2')$ is a nonempty ray-like closed convex set not containing the origin. This completes the proof. □

The following counter-example shows that the requirement that $\mathcal{C}_1$ and $\mathcal{C}_2$ are ray-like cannot be removed from Proposition 3.7.

Example 3.1 (Set intersection and the ray-like property). Consider the sets

$$\mathcal{C}_1 = \{\, (x_1, x_2) \mid 1 - x_1 \le x_2 \le x_1 - 1 \,\} \quad \text{and} \quad \mathcal{C}_2 = \{\, (x_1, x_2) \mid x_1 = 1 \,\}.$$

Define $H_1 = \{\, (x_1, x_2) \mid x_1 + x_2 \ge 1 \,\}$ and $H_2 = \{\, (x_1, x_2) \mid x_1 - x_2 \ge 1 \,\}$ so that $C_1 = H_1 \cap H_2$. Clearly the set $\mathcal{C}_2$ is not ray-like, while the sets $\mathcal{C}_1$, H_1, and H_2 are. Moreover, all four sets do not contain the origin. Furthermore, $\mathcal{C}_1 \cap \mathcal{C}_2$ is the singleton $\{\, (1, 0) \,\}$, and hence a direct computation shows that $(\mathcal{C}_1 \cap \mathcal{C}_2)' = \{\, (y_1, y_2) \mid y_1 \ge 1 \,\}$.

Next, we have directly from definition that $\mathcal{C}_2' = \{\, (y_1, 0) \mid y_1 \ge 1 \,\}$. On the other hand, note that $H_1 = L_1^{-1} I$, where $L_1(x_1, x_2) = x_1 + x_2$ and $I = \{\, u \mid u \ge 1 \,\}$. Thus, by Corollary 3.5, $H_1' = \{\, (y_1, y_1) \mid y_1 \ge 1 \,\}$. Similarly, $H_2' = \{\, (y_1, -y_1) \mid y_1 \ge 1 \,\}$. Since H_1 and H_2 are ray-like, we obtain from Proposition 3.7 that

$$\mathcal{C}_1' = (H_1 \cap H_2)' = \operatorname{cl}\operatorname{conv}(H_1' \cup H_2'),$$

which contains $\mathcal{C}_2'$. Thus,

$$\operatorname{cl}\operatorname{conv}(\mathcal{C}_1' \cup \mathcal{C}_2') = \mathcal{C}_1' \subsetneq \{\, (y_1, y_2) \mid y_1 \ge 1 \,\} = (\mathcal{C}_1 \cap \mathcal{C}_2)'.$$

4. Duality derivations. We derive in this section the gauge and Lagrange duals of the primal problem (P_ρ). Let

$$\mathcal{C} = \{\, x \mid \rho(b - Ax) \le \sigma \,\}, \tag{4.1}$$

where ρ is a closed gauge and $0 \le \sigma < \rho(b)$, denote the constraint set. We generally do not assume that $\rho^{-1}(0) = \{\, 0 \,\}$, though we must do so in the case in which $\sigma = 0$.

4.1. The gauge dual. We consider two approaches for deriving the gauge dual of (P_ρ). The first uses explicitly the abstract definition of the gauge dual (D). The second approach redefines the objective function to also contain an indicator for the nonlinear gauge ρ where $\mathcal{C}$ is simply affine. This alternative approach is instructive, because it illustrates the modeling choices that are available when working with gauge functions.

4.1.1. First approach. The following combines Corollary 3.5 with Proposition 3.1, and gives an explicit expression for the antipolar of $\mathcal{C}$ when $\sigma > 0$.

COROLLARY 4.1. Suppose that $\mathcal{C}$ is given by (4.1), where $\sigma \in (0, \rho(b))$. If $\mathcal{C}$ is polyhedral, or $\operatorname{ri}\{\, u \mid \rho(b-u) \le \sigma \,\} \cap \operatorname{range} A \neq \emptyset$, then

$$\mathcal{C}' = \{\, A^* y \mid \langle b, y\rangle - \sigma\rho^\circ(y) \ge 1 \,\}.$$

As an aside, we consider the special case $\sigma = 0$. The following result follows from Corollaries 3.5 and 3.2, and allows for the case in which $\rho^{-1}(0)$ is not trivially $\{\, 0 \,\}$.

COROLLARY 4.2. Suppose that $\mathcal{C} = \{\, x \mid Ax - b \in \mathcal{K} \,\}$ for some closed convex cone $\mathcal{K}$ and $b \notin -\mathcal{K}$. If $\mathcal{C}$ is polyhedral, or $(b + \operatorname{ri}\mathcal{K}) \cap \operatorname{range} A \neq \emptyset$, then

$$\mathcal{C}' = \{\, A^* y \in \mathcal{K}^* \mid \langle b, y\rangle \ge 1 \,\}.$$

To obtain an explicit representation of the gauge dual problem, we assume that

$$\mathcal{C} \text{ is polyhedral,} \quad \text{or} \quad \operatorname{ri}\{\, u \mid \rho(b-u) \le \sigma \,\} \cap \operatorname{range} A \neq \emptyset. \tag{4.2}$$

We rely on the antipolar calculus developed in §3. Consider separately the cases $\sigma > 0$ and $\sigma = 0$.

Case 1: $\sigma > 0$. Apply Corollary 4.1 to derive the antipolar set

$$\mathcal{C}' = \{\, A^* y \mid \langle b, y\rangle - \sigma\rho^\circ(y) \ge 1 \,\}. \tag{4.3}$$

Case 2: $\sigma = 0$. Here we need to assume that $\rho^{-1}(0) = \{\, 0 \,\}$, and in that case, $\mathcal{C} = \{\, x \mid Ax = b \,\}$. We can now apply Corollary 4.2 with $\mathcal{K} = \{\, 0 \,\}$ and obtain

$$\mathcal{C}' = \{\, A^* y \mid \langle b, y\rangle \ge 1 \,\}. \tag{4.4}$$

Since $\rho^{-1}(0) = \{\, 0 \,\}$ and ρ is closed, we conclude from Proposition 2.1(v) that $\operatorname{dom}\rho^\circ = \mathcal{X}$. Hence, (4.4) can be seen as a special case of (4.3) with $\sigma = 0$.

These two cases can be combined, and we see that when (4.2) holds, the gauge dual problem (D) for (P_ρ) can be expressed as (D_ρ). If the assumptions (4.2) are not satisfied, then in view of Corollary 3.4, it still holds that (D) is equivalent to

$$\operatorname*{minimize}_{u,y} \quad \kappa^\circ(u) \quad \text{subject to} \quad u \in \operatorname{cl}\{\, A^* y \mid \langle y, b\rangle - \sigma\rho^\circ(y) \ge 1 \,\}.$$

This optimal value can in general be less than or equal to that of (D_ρ).

4.1.2. Second approach. This approach does not rely on assumptions (4.2). Define the function $\xi(x, r, \tau) := \kappa(x) + \delta_{\operatorname{epi}\rho}(r, \tau)$, which is a gauge because $\operatorname{epi}\rho$ is a cone. Then (P_ρ) can be equivalently reformulated as

$$\operatorname*{minimize}_{x,r,\tau} \quad \xi(x, r, \tau) \quad \text{subject to} \quad Ax + r = b,\ \tau = \sigma. \tag{4.5}$$

Invoke Proposition 2.4 to establish that

$$\begin{aligned} \xi^\circ(z, y, \alpha) &= \max\{\, \kappa^\circ(z),\ \delta_{(\operatorname{epi}\rho)^\circ}(y, \alpha) \,\} \\ &= \kappa^\circ(z) + \delta_{(\operatorname{epi}\rho)^\circ}(y, \alpha) \\ &= \kappa^\circ(z) + \delta_{\operatorname{epi}(\rho^\circ)}(y, -\alpha). \end{aligned}$$

As Freund (1987, Section 2) shows for gauge programs with linear constraints, the gauge dual is given by

$$\operatorname*{minimize}_{y,\alpha} \quad \xi^\circ(A^* y, y, \alpha) \quad \text{subject to} \quad \langle y, b\rangle + \sigma\alpha \ge 1,$$

which can be rewritten as

$$\underset{y,\alpha}{\text{minimize}} \quad \kappa^\circ(A^*y) \quad \text{subject to} \quad \langle y, b\rangle + \sigma\alpha \geq 1 \text{ and } \rho^\circ(y) \leq -\alpha.$$

(The gauge dual for problems with linear constraints also follows directly from Corollary 4.2 with $\mathcal{K} = \{0\}$.) Further simplification leads to the gauge dual program (D_ρ).

Note that the transformation used to derive (4.5) is very flexible. For example, if (P_ρ) contained the additional conic constraint $x \in \mathcal{K}$, then ξ could be defined to contain an additional term given by the indicator of $\mathcal{K}$.

Even though this approach does not require the assumptions (4.2) used in §4.1, and thus appears to apply more generally, it is important to keep in mind we have yet to impose conditions that imply strong duality. In fact, as we show in §5, the assumptions required there imply (4.2).

4.2. Lagrange duality. To derive the Lagrange dual of (P_ρ), we reformulate it by introducing an artificial "residual" variable r, which leads to

$$\underset{x,r}{\text{minimize}} \quad \kappa(x) \quad \text{subject to} \quad Ax + r = b \quad \text{and} \quad \rho(r) \leq \sigma. \tag{4.6}$$

Define the Lagrangian function

$$\begin{aligned} L(x,r,y,\lambda) &= \kappa(x) + \langle y, b - Ax - r\rangle + \lambda\big(\rho(r) - \sigma\big) \\ &= \big(\langle y, b\rangle - \lambda\sigma\big) - \big(\langle A^*y, x\rangle - \kappa(x)\big) - \big(\langle y, r\rangle - \lambda\rho(r)\big). \end{aligned}$$

The Lagrange dual problem is given by

$$\underset{y,\lambda}{\text{maximize}} \quad \inf_{x,r} L(x,r,y,\lambda) \quad \text{subject to} \quad \lambda \geq 0.$$

Consider the (concave) dual function

$$\begin{aligned} \ell(y,\lambda) &= \inf_{x,r} L(x,r,y,\lambda) \\ &= \inf_{x,r}\Big\{\big(\langle y, b\rangle - \lambda\sigma\big) - \big(\langle A^*y, x\rangle - \kappa(x)\big) - \big(\langle y, r\rangle - \lambda\rho(r)\big)\Big\} \\ &= \big(\langle y, b\rangle - \lambda\sigma\big) - \sup_x\Big\{\langle A^*y, x\rangle - \kappa(x)\Big\} - \sup_r\Big\{\langle y, r\rangle - \lambda\rho(r)\Big\} \\ &= \big(\langle y, b\rangle - \lambda\sigma\big) - \delta_{\kappa^\circ(\cdot)\leq 1}(A^*y) - \delta_{\rho^\circ(\cdot)\leq\cdot}(y,\lambda), \end{aligned}$$

where the first conjugate on the right-hand side follows from Proposition 2.1(iv), while the last conjugate follows from (Rockafellar and Wets, 1998, Theorem 11.21). The Lagrange dual problem is obtained by maximizing ℓ, i.e.,

$$\underset{y,\lambda}{\text{maximize}} \quad \langle y, b\rangle - \lambda\sigma \quad \text{subject to} \quad \kappa^\circ(A^*y) \leq 1,\ \rho^\circ(y) \leq \lambda. \tag{4.7}$$

This last expression can be simplified by maximizing over $\lambda \geq 0$, which leads to (D_L).

Strictly speaking, the Lagrangian primal-dual pair of programs that we have derived is given by (4.6) and (4.7), but it is easily verifiable that this pair is equivalent to (P) and (D_L) in the sense that the respective optimal values are the same and that solutions to either pair readily lead to solutions for the other. For that reason, without loss of generality, we refer to (D_L) as the Lagrange dual to the primal problem (P).

5. Strong duality. Freund's 1987 analysis of the gauge dual pair is mainly based on the classical separation theorem. It relies on the ray-like property of the constraint set $\mathcal{C}$. Our study of the gauge dual pairs allows us to relax the ray-like assumption. By establishing connections with the Fenchel duality framework, we can develop strong duality conditions that are analogous to those required for Lagrange duality theory.

5.1. Preliminaries. Our main tool of analysis is the *antigauge* function

$$g(y) := \sup \{ \alpha \geq 0 \mid y \in \alpha C' \},$$

of $\mathcal{C}'$, modeled after a similar definition stated by McLinden (1978). This function was used by Freund to relate his gauge duality theory to the duality theory of Gwinner (1985). The following result collects some useful properties of the antigauge function.

LEMMA 5.1. The following properties hold for the function g:
(i) $\operatorname{dom} g = \operatorname{cone} \mathcal{C}' \neq \emptyset$, on which g is finite;
(ii) the supremum in the definition of g is attained for any $y \in \operatorname{dom} g$;
(iii) $y \in \mathcal{C}'$ if and only if $g(y) \geq 1$;
(iv) g is positively homogeneous;
(v) $y \in \mathcal{C}'$ implies that $y \in g(y)\mathcal{C}'$.

Proof. It is clear that if $y \notin \operatorname{cone} \mathcal{C}'$, then $g(y) = -\infty$. On the other hand, if $y \in \operatorname{cone} \mathcal{C}'$, then for any fixed $c \in \mathcal{C}$, we have $\langle c, y\rangle \geq \alpha\langle c, w\rangle \geq \alpha$ whenever $y = \alpha w$ for some $w \in \mathcal{C}'$ and $\alpha \geq 0$. This implies that $0 \leq g(y) \leq \langle c, y\rangle < \infty$, i.e., $y \in \operatorname{dom} g$ and $g(y)$ is finite.

To see that the supremum is attained, first notice that for any nonzero $y \in \operatorname{dom} g$, there exists $\underline{\alpha} > 0$ and $w \in \mathcal{C}'$ so that $y = \underline{\alpha} w$. This shows that one can restrict attention to any maximizing sequence with $\inf \alpha_t \geq \underline{\alpha}$. Then, along this maximizing sequence, there exists $w^t \in \mathcal{C}'$ with $w^t = y/\alpha_t$. Since $\{\alpha_t\}$ is uniformly bounded away from zero, the sequence $\{w^t\}$ is bounded and hence one can obtain a convergent subsequence. The attainment of the supremum now follows from a standard argument. On the other hand, if $y = 0$, the supremum is clearly attained.

To prove (iii), we note first that $y \in \mathcal{C}'$ clearly implies that $g(y) \geq 1$. On the other hand, if $g(y) \geq 1$, then $y = g(y)w = w + (g(y) - 1)w$ for some $w \in \mathcal{C}'$. Since $\mathcal{C}'$ is ray-like, we conclude that $y \in \mathcal{C}'$.

Positive homogeneity follows directly from definition.

Finally, if $y \in \mathcal{C}'$, then $g(y) \geq 1 > 0$ and so $g(\frac{y}{g(y)}) = 1$ by positive homogeneity. The result (v) now follows from (iii). ∎

The next lemma gives an alternative representation of the function g. The case when $\mathcal{C}$ is ray-like was discussed in McLinden (1978, Corollary 14B).

LEMMA 5.2. For any y, we have $\inf_{c\in\mathcal{C}} \langle c, y\rangle \geq g(y)$. Moreover,

$$\inf_{c\in\mathcal{C}} \langle c, y\rangle \begin{cases} = g(y) & \text{if } y \in \operatorname{dom} g, \\ = 0 & \text{if } y \in (\operatorname{cl} \operatorname{dom} g)\backslash(\operatorname{dom} g), \\ < 0 & \text{if } y \notin \operatorname{cl} \operatorname{dom} g. \end{cases} \tag{5.1}$$

If, in addition, $\mathcal{C}$ is ray-like, then $\inf_{c\in\mathcal{C}} \langle c, y\rangle = g(y) = -\infty$ for all $y \notin \operatorname{cl} \operatorname{dom} g$.

Proof. For any $y \notin \operatorname{dom} g$, the inequality $\inf_{c\in\mathcal{C}} \langle c, y\rangle \geq g(y) = -\infty$ holds trivially. On the other hand, if $y \in \operatorname{dom} g = \operatorname{cone} \mathcal{C}'$, then for any fixed $c \in \mathcal{C}$, we have

$\langle c, y\rangle \geq \alpha\langle c, w\rangle \geq \alpha$ whenever $y = \alpha w$ for some $\alpha \geq 0$ and $w \in \mathcal{C}'$, from which the inequality follows.

We now prove (5.1). Suppose that $y \in (\operatorname{cl}\operatorname{dom} g)\backslash(\operatorname{dom} g)$. It then follows from Lemma 2.10 that $\operatorname{cl}\operatorname{dom} g = \mathcal{C}^*$ and so $\inf_{c\in\mathcal{C}}\langle c, y\rangle \geq 0$. If $\inf_{c\in\mathcal{C}}\langle c, y\rangle \geq \lambda > 0$, then it follows from definition that $y \in \lambda\mathcal{C}' \subseteq \operatorname{cone}\mathcal{C}' = \operatorname{dom} g$ in view of Lemma 5.1(i), a contradiction. Hence, we must have $\inf_{c\in\mathcal{C}}\langle c, y\rangle = 0$.

Suppose now that $y \in \operatorname{dom} g$. Then $g(y) \geq 0$. If $\lambda := \inf_{c\in\mathcal{C}}\langle c, y\rangle = 0$, then equality holds. On the other hand, if $\lambda > 0$, then $\langle c, \frac{y}{\lambda}\rangle \geq 1$ for all $c \in \mathcal{C}$, meaning that $\frac{y}{\lambda} \in \mathcal{C}'$. Using properties (iii) and (iv) of g in Lemma 5.1, we conclude that $g(y) \geq \lambda$. Thus, $\inf_{c\in\mathcal{C}}\langle c, y\rangle = g(y)$.

Next, suppose that $y \notin \operatorname{cl}\operatorname{dom} g$. If $\inf_{c\in\mathcal{C}}\langle c, y\rangle \geq 0$, then $y \in \mathcal{C}^* = \operatorname{cl}\operatorname{cone}\mathcal{C}' = \operatorname{cl}\operatorname{dom} g$, where the equalities follow from Lemma 2.10 and Lemma 5.1(i). This is a contradiction. Thus, $\langle c, y\rangle < 0$ for some $c \in \mathcal{C}$, or equivalently, $\inf_{c\in\mathcal{C}}\langle c, y\rangle < 0$.

Finally, suppose in addition that $\mathcal{C}$ is ray-like. Consider any $y \notin \operatorname{cl}\operatorname{dom} g$. This implies that $g(y) = -\infty$. Moreover, it follows from the above discussion that $\langle c, y\rangle < 0$ for some $c \in \mathcal{C}$. Since the set $\mathcal{C}$ is ray-like, we immediately have $\inf_{c\in\mathcal{C}}\langle c, y\rangle = -\infty = g(y)$. Hence, equality also holds in this case. □

5.2. From Fenchel to gauge duality. We now draw a link between Fenchel and gauge duality. The Fenchel dual to (P) is given by

$$\underset{y}{\text{maximize}} \quad -\sigma_{\mathcal{C}}(-y) \quad \text{subject to} \quad \kappa^\circ(y) \leq 1, \tag{5.2}$$

where we use Proposition 2.1(iv) for the conjugate of κ. Let v_p, v_g, and v_f, respectively, denote the optimal values of (P), (D) and (5.2). The following result relates their optimal values and dual solutions.

THEOREM 5.3 (Weak duality). Suppose that $\operatorname{dom}\kappa^\circ \cap \mathcal{C}' \neq \emptyset$. Then

$$v_p \geq v_f = 1/v_g > 0$$

Furthermore,
(i) if y^* solves (5.2), then $y^* \in \operatorname{cone}\mathcal{C}'$ and y^*/v_f solves (D);
(ii) if y^* solves (D) and $v_g > 0$, then $v_f y^*$ solves (5.2).

Proof. The fact that $v_p \geq v_f$ follows from standard Fenchel duality theory. We now show that $v_f = 1/v_g$.

To this end, note that

$$-\sigma_{\mathcal{C}}(-y) = -\sup_{c\in\mathcal{C}}\langle c, -y\rangle = \inf_{c\in\mathcal{C}}\langle c, y\rangle. \tag{5.3}$$

Since $\operatorname{dom}\kappa^\circ \cap \mathcal{C}' \neq \emptyset$, there exists y_0 such that $\kappa^\circ(y_0) \leq 1$ and $y_0 \in \tau\mathcal{C}'$ for some $\tau > 0$. In particular, we have

$$v_f \geq \inf_{c\in\mathcal{C}}\langle c, y_0\rangle \geq \tau > 0. \tag{5.4}$$

Moreover, using (5.3) and the definition of v_f, we have

$$\begin{aligned} v_f &= \sup_y \{ -\sigma_{\mathcal{C}}(-y) \mid \kappa^\circ(y) \le 1 \} \\ &\overset{(i)}{=} \sup_y \{ -\sigma_{\mathcal{C}}(-y) \mid \kappa^\circ(y) \le 1,\ y \in \operatorname{dom} g \} \\ &\overset{(ii)}{=} \sup_y \{ g(y) \mid \kappa^\circ(y) \le 1 \}, \end{aligned} \tag{5.5}$$

where equality (i) follows from the fact that $\inf_{c\in\mathcal{C}}\langle c, y\rangle \le 0$ when $y \notin \operatorname{dom} g$ by Lemma 5.2, and the positivity of v_f from (5.4); while equality (ii) holds because of Lemma 5.2.

Furthermore,

$$v_f = \sup_{y,\lambda} \{ \lambda \mid \kappa^\circ(y) \le 1,\ g(y) \ge \lambda \} = \sup_{y,\lambda} \{ \lambda \mid \kappa^\circ(y) \le 1,\ g(y) \ge \lambda,\ \lambda > 0 \},$$

where the second equality follows from (5.4). From this, we have further that

$$\begin{aligned} v_f &= \sup_{y,\lambda} \{ \lambda \mid \kappa^\circ(y/\lambda) \le 1/\lambda,\ g(y/\lambda) \ge 1, 1/\lambda > 0 \} \\ &= \sup_{y,\mu} \{ 1/\mu \mid \kappa^\circ(\mu y) \le \mu,\ g(\mu y) \ge 1,\ \mu > 0 \}. \end{aligned}$$

Inverting both sides of this equation gives

$$\begin{aligned} 1/v_f &= \inf_{y,\mu} \{ \mu \mid \kappa^\circ(\mu y) \le \mu,\ g(\mu y) \ge 1,\ \mu > 0 \} \\ &= \inf_{w,\mu} \{ \mu \mid \kappa^\circ(w) \le \mu,\ g(w) \ge 1,\ \mu > 0 \} \\ &\overset{(i)}{=} \inf_{w,\mu} \{ \mu \mid \kappa^\circ(w) \le \mu,\ w \in \mathcal{C}',\ \mu > 0 \} \\ &= \inf_{w,\mu} \{ \mu \mid \kappa^\circ(w) \le \mu,\ w \in \mathcal{C}' \} \\ &= \inf_{w} \{ \kappa^\circ(w) \mid w \in \mathcal{C}' \} = v_g, \end{aligned} \tag{5.6}$$

where equality (i) follows from Lemma 5.1(iii). This proves $v_f = 1/v_g$.

We now prove item (i). Assume that y^* solves (5.2). Then v_f is nonzero (by (5.4)) and finite, and so is $v_g = 1/v_f$. Then we see immediately from (5.6) that $y^* \in \operatorname{cone} C'$ because $g(y^*) = v_f > 0$, and y^*/v_f solves (D). We now prove item (ii). Note that if y^* solves (D) and $v_g > 0$, then $\kappa^\circ(y^*) > 0$. One can then observe similarly from (5.6) that $y^*/v_g = v_f y^*$ solves (5.2). This completes the proof. ∎

Fenchel duality theory allows us to use Theorem 5.3 to obtain several sufficient conditions that guarantee strong duality, i.e., $v_p v_g = 1$, and the attainment of the gauge dual problem (D). For example, applying Rockafellar (1970, Theorem 31.1)) yields the following corollary.

Corollary 5.4 (Strong duality I). Suppose that $\operatorname{dom}\kappa^\circ \cap \mathcal{C}' \neq \emptyset$ and $\operatorname{ri}\operatorname{dom}\kappa \cap \operatorname{ri}\mathcal{C} \neq \emptyset$. Then $v_p v_g = 1$ and the gauge dual (D) attains its optimal value.

Proof. From $\operatorname{ri}\operatorname{dom}\kappa \cap \operatorname{ri}\mathcal{C} \neq \emptyset$ and (Rockafellar, 1970, Theorem 31.1), we see that $v_p = v_f$ and v_f is attained. The conclusion of the corollary now follows immediately from Theorem 5.3. ∎

We would also like to guarantee *primal attainment*. Note that the gauge dual of the gauge dual problem (D) (i.e., the bidual of (P)) is given by

$$\underset{x}{\text{minimize}} \quad \kappa^{\circ\circ}(x) \quad \text{subject to} \quad x \in \mathcal{C}'', \tag{5.7}$$

which is not the same as (P) unless $\mathcal{C}$ is ray-like and κ is closed; see Theorem 2.8 and Proposition 2.1(ii). However, we show in the next proposition that (5.7) and (P) always have the same optimal value when κ is closed (even if $\mathcal{C}$ is not ray-like), and that if the optimal value is attained in one problem, it is also attained in the other.

PROPOSITION 5.5. Suppose that κ is closed. Then the optimal values of (P) and (5.7) are the same. Moreover, if the optimal value is attained in one problem, it is also attained in the other.

Proof. From Proposition 2.9, we see that (5.7) is equivalent to

$$\underset{\lambda,x}{\text{minimize}} \quad \lambda\kappa(x) \quad \text{subject to} \quad x \in \mathcal{C},\ \lambda \ge 1,$$

which clearly gives the same optimal value as (P). This proves the first conclusion. The second conclusion now also follows immediately. ∎

Hence, we obtain the following corollary, which generalizes Freund (1987, Theorem 2A) by dropping the ray-like assumption on $\mathcal{C}$.

COROLLARY 5.6 (Strong duality II). Suppose that κ is closed, and that $\operatorname{ri}\operatorname{dom}\kappa \cap \operatorname{ri}\mathcal{C} \neq \emptyset$ and $\operatorname{ri}\operatorname{dom}\kappa^\circ \cap \operatorname{ri}\mathcal{C}' \neq \emptyset$. Then $v_p v_g = 1$ and both values are attained.

Proof. The conclusion follows from Corollary 5.4, Proposition 5.5, the fact that $\kappa = \kappa^{\circ\circ}$ for closed gauge functions, and the observation that $\operatorname{ri}\operatorname{dom}\kappa \cap \operatorname{ri}\mathcal{C} \neq \emptyset$ if and only if $\operatorname{ri}\operatorname{dom}\kappa \cap \operatorname{ri}\mathcal{C}'' \neq \emptyset$, since $\operatorname{ri}\mathcal{C}'' = \bigcup_{\lambda>1}\lambda\operatorname{ri}\mathcal{C}$ (see Rockafellar (1970, p. 50)) and $\operatorname{dom}\kappa$ is a cone. ∎

Before closing this section, we specialize Theorem 5.3 to study the relationship between the Lagrange (D_L) and gauge (D_ρ) duals. We let v_l denote the optimal value of (D_L).

COROLLARY 5.7. Suppose that $\mathcal{C}$ is given by (4.1), where $\sigma \in [0, \rho(b))$, assumption (4.2) holds, and $\operatorname{dom}\kappa^\circ \cap \mathcal{C}' \neq \emptyset$. Then $v_l = v_f > 0$. Moreover,
(i) if y^* solves (D_L), then y^*/v_l solves (D_ρ);
(ii) if y^* solves (D_ρ) and $v_g > 0$, then $v_l y^*$ solves (D_L).

Proof. Recall from Lemma 5.2 that for any $y \in \operatorname{dom} g$, we have $g(y) = \inf_{c\in\mathcal{C}}\langle c, y\rangle$. Following a similar derivation in §4.2, we can show that $\sup_{y=A^*u}\{\langle b, u\rangle - \sigma\rho^\circ(u)\}$ is a Lagrange dual problem of $\inf_{c\in\mathcal{C}}\langle c, y\rangle$. Due to assumption (4.2), it holds that

$$g(y) = \inf_{c\in\mathcal{C}}\langle c, y\rangle = \sup_{y=A^*u}\{\langle b, u\rangle - \sigma\rho^\circ(u)\} \tag{5.8}$$

and the supremum is attained, a consequence of Rockafellar (1970, Corollary 28.2.2). On the other hand, for any $y \notin \operatorname{dom} g$, we have from weak duality that

$$\sup_{y=A^*u}\{\langle b, u\rangle - \sigma\rho^\circ(u)\} \le \inf_{c\in\mathcal{C}}\langle c, y\rangle \le 0, \tag{5.9}$$

because $\inf_{c\in\mathcal{C}}\langle c, y\rangle \le 0$ when $y \notin \operatorname{dom} g$, by Lemma 5.2.

Since $\operatorname{dom}\kappa^\circ \cap \mathcal{C}' \neq \emptyset$, we can plug (5.8) into (5.5) and obtain

$$\begin{aligned} 0 < v_f &= \sup\{\, \langle b, u\rangle - \sigma\rho^\circ(u) \mid \kappa^\circ(A^*u) \le 1,\ A^*u \in \operatorname{dom} g \,\} \\ &= \sup\{\, \langle b, u\rangle - \sigma\rho^\circ(u) \mid \kappa^\circ(A^*u) \le 1 \,\} = v_l, \end{aligned}$$

where the second equality follows from (5.9) and the positivity of v_f. This completes the first part of the proof. In particular, the Fenchel dual problem (5.2) has the same optimal value as the Lagrange dual problem (D_L), and $y^* = A^*u^*$ solves (5.2) if and only if u^* solves (D_L). Moreover, since assumption 4.2 holds, §4.1 shows that (D) is equivalent to (D_ρ). The conclusion now follows from these and Theorem 5.3. ∎

We next state a strong duality result concerning the primal-dual gauge pair (P_ρ) and (D_ρ).

Corollary 5.8. Suppose that $\mathcal{C}$ is given by (4.1), where $\sigma \in [0, \rho(b))$ and define $\mathcal{D} = \{\, u \mid \rho(b-u) \le \sigma \,\}$. Suppose also that κ is closed,

$$\operatorname{ri}\operatorname{dom}\kappa \cap A^{-1}\operatorname{ri}\mathcal{D} \neq \emptyset \quad \text{and} \quad \operatorname{ri}\operatorname{dom}\kappa^\circ \cap A^*\operatorname{ri}\mathcal{D}' \neq \emptyset. \tag{5.10}$$

Then the optimal values of (P_ρ) and (D_ρ) are attained, and their product is 1.

Proof. Since $A^{-1}\operatorname{ri}\mathcal{D} \neq \emptyset$, A satisfies the assumption in (4.2). Then §4.1 shows that (D) is equivalent to (D_ρ). Moreover, from Rockafellar (1970, Theorem 6.6, Theorem 6.7), we see that $\operatorname{ri}\mathcal{C} = A^{-1}\operatorname{ri}\mathcal{D}$ and $\operatorname{ri}\mathcal{C}' = A^*\operatorname{ri}\mathcal{D}'$. The conclusion now follows from Corollary 5.6. ∎

This last result also holds if $\mathcal{C}$ were polyhedral; in that case, the assumptions (5.10) could be replaced with $\operatorname{ri}\operatorname{dom}\kappa \cap \mathcal{C} \neq \emptyset$ and $\operatorname{ri}\operatorname{dom}\kappa^\circ \cap \mathcal{C}' \neq \emptyset$.

6. Variational properties of the gauge value function. Thus far, our analysis has focused on the relationship between the optimal values of the primal-dual pair (P_ρ) and (D_ρ). As with Lagrange duality, however, there is also a fruitful view of dual solutions as providing sensitivity information on the primal optimal value. Here we provide a corresponding variational analysis of the gauge optimal-value function with respect to perturbations in b and σ.

Sensitivity information is captured in the subdifferential of the value function

$$v(b,\sigma) = \inf_x \; f(x,b,\sigma), \tag{6.1}$$

with

$$f(x,b,\sigma) = \kappa(x) + \delta_{\operatorname{epi}\rho}(b - Ax, \sigma).$$

Interestingly, both f and v are gauges. Following the discussion in Aravkin, Burke, and Friedlander (2013, Section 4), we start by computing the conjugate of f, which can be done as follows:

$$\begin{aligned} f^*(z,y,\tau) &= \sup_{x,b,\sigma} \{\, \langle z, x\rangle + \langle y, b\rangle + \tau\sigma - \kappa(x) - \delta_{\operatorname{epi}\rho}(b - Ax, \sigma) \,\} \\ &= \sup_{x,w,\sigma} \{\, \langle z + A^*y, x\rangle - \kappa(x) + \langle y, w\rangle + \tau\sigma - \delta_{\operatorname{epi}\rho}(w,\sigma) \,\} \\ &= \kappa^*(z + A^*y) + \delta^*_{\operatorname{epi}\rho}(y,\tau). \end{aligned}$$

Use Proposition 2.1(iv) and the definition of support function and convex conjugate to further transform this as

$$\begin{aligned} f^*(z,y,\tau) &= \delta_{\kappa^\circ(\cdot)\le 1}(z+A^*y) + \sigma_{\operatorname{epi}\rho}(y,\tau) \\ &\overset{(i)}{=} \delta_{\kappa^\circ(\cdot)\le 1}(z+A^*y) + \delta_{(\operatorname{epi}\rho)^\circ}(y,\tau) \\ &\overset{(ii)}{=} \delta_{\kappa^\circ(\cdot)\le 1}(z+A^*y) + \delta_{\operatorname{epi}(\rho^\circ)}(y,-\tau) \\ &= \delta_{\kappa^\circ(\cdot)\le 1}(z+A^*y) + \delta_{\rho^\circ(\cdot)\le\cdot}(y,-\tau), \end{aligned}$$

where equality (i) follows from Proposition 2.6 and Proposition 2.3(i), and equality (ii) follows from the definition of the polar of gauges; see also Rockafellar (1970, p. 137). Combining this with the definition of the value function $v(b,\sigma)$,

$$v^*(y,\tau) = f^*(0,y,\tau) = \delta_{\kappa^\circ(\cdot)\le 1}(A^*y) + \delta_{\rho^\circ(\cdot)\le\cdot}(y,-\tau). \tag{6.2}$$

In view of Rockafellar and Wets (1998, Theorem 11.39), under a suitable constraint qualification, the set of subgradients of v is nonempty and is given by

$$\begin{aligned} \partial v(b,\sigma) &= \operatorname*{argmax}_{y,\tau} \{\, \langle b,y\rangle + \sigma\tau - v^*(u,\tau) \,\} \\ &= \operatorname*{argmax}_{y,\tau} \{\, \langle b,y\rangle + \sigma\tau \mid \kappa^\circ(A^*y)\le 1,\ \rho^\circ(y)\le -\tau \,\} \\ &= \left\{ (y,-\rho^\circ(y)) \;\middle|\; y\in \operatorname*{argmax}_{y} \{\, \langle b,y\rangle - \sigma\rho^\circ(y) \mid \kappa^\circ(A^*y)\le 1 \,\} \right\}, \end{aligned} \tag{6.3}$$

in terms of the solution set of (D_L) and the corresponding function value of $\rho^\circ(y)$. We state formally this result, which is a consequence of the above discussion and Corollary 5.7.

PROPOSITION 6.1. For fixed (b,σ),

$$\operatorname{dom} f(\cdot,b,\sigma)\ne\emptyset \iff 0\in A\operatorname{dom}\kappa - [\rho(b-\cdot)\le\sigma],$$

and hence

$$(b,\sigma)\in\operatorname{int}\operatorname{dom} v \iff 0\in\operatorname{int}(A\operatorname{dom}\kappa - [\rho(b-\cdot)<\sigma])$$

If $(b,\sigma)\in\operatorname{int}\operatorname{dom} v$ and $v(b,\sigma)>0$, then $\partial v(b,\sigma)\ne\emptyset$ with

$$\begin{aligned} \partial v(b,\sigma) &= \left\{ (y,-\rho^\circ(y)) \;\middle|\; y\in \operatorname*{argmax}_{y} \{\, \langle b,y\rangle - \sigma\rho^\circ(y) \mid \kappa^\circ(A^*y)\le 1 \,\} \right\} \\ &= \left\{ v(b,\sigma)\cdot(y,-\rho^\circ(y)) \;\middle|\; y\in \operatorname*{argmax}_{y} \{\, \kappa^\circ(A^*y) \mid \langle b,y\rangle - \sigma\rho^\circ(y)\ge 1 \,\} \right\}. \end{aligned}$$

Proof. It is routine to verify the properties of the domain of $f(\cdot,b,\sigma)$ and the interior of the domain of v. Suppose that $(b,\sigma)\in\operatorname{int}\operatorname{dom} v$. Then the value function is continuous at (b,σ) and hence $\partial v(b,\sigma)\ne\emptyset$. The first expression of $\partial v(b,\sigma)$ follows directly from Rockafellar and Wets (1998, Theorem 11.39) and the discussions preceding this proposition.

We next derive the second expression of $\partial v(b,\sigma)$. Since $(b,\sigma)\in\operatorname{int}\operatorname{dom} v$ implies $0\in\operatorname{int}(A\operatorname{dom}\kappa - [\rho(b-\cdot)<\sigma])$, the linear map A satisfies assumption 4.2. Moreover,

$(b,\sigma) \in \operatorname{int}\operatorname{dom} v$ also implies that $v(b,\sigma)$ equals the optimal value of the Lagrange dual problem (D_L), another consequence of Rockafellar and Wets (1998, Theorem 11.39). Furthermore, $v(b,\sigma)$ being finite and nonzero together with the definition of (D_L) and (4.3) implies that $\operatorname{dom}\kappa^\circ \cap \mathcal{C}' \neq \emptyset$. The second expression of $\partial v(b,\sigma)$ now follows from these three observations and Corollary 5.7. □

7. Extensions. The following examples illustrate how to extend the canonical formulation (P_ρ) to accommodate related problems. It also provides an illustration of the techniques that can be used to pose problems in gauge form and how to derive their corresponding gauge duals.

7.1. Composition and conic side constraints. A useful generalization of (P_ρ) is to allow the gauge objective to be composed with a linear map, and for the addition of conic side constraints, i.e.,

$$\underset{x}{\text{minimize}} \quad \kappa(Dx) \quad \text{subject to} \quad \rho(b-Ax) \le \sigma,\ x \in \mathcal{K}, \tag{7.1}$$

where D is a linear map and $\mathcal{K}$ is a convex cone. The composite objective can be used to capture, for example, problems such as weighted basis pursuit (e.g., Candés, Wakin, and Boyd (2008); Friedlander, Mansour, Saab, and Yilmaz (2012)), or together with the conic constraint, problems such as nonnegative total variation (Krishnan, Lin, and Yip, 2007).

One way to fit (7.1) into the gauge framework is to introduce additional variables, and lift both the cone $\mathcal{K}$ and the epigraph $\operatorname{epi}\rho$ into the objective by means of their indicator functions: use the function $f(x,s,r,\tau) := \delta_{\mathcal{K}}(x) + \kappa(s) + \delta_{\operatorname{epi}\rho}(r,\tau)$ to define the equivalent gauge optimization problem

$$\underset{x,s,r,\tau}{\text{minimize}} \quad f(x,s,r,\tau) \quad \text{subject to} \quad Dx = s,\ Ax + r = b,\ \tau = \sigma.$$

As with §4.1, observe that f is a sum of gauges on disjoint variables. Thus, we invoke Proposition 2.4 to deduce the polar of the above objective:

$$\begin{aligned} f^\circ(u,z,y,\alpha) &= \max\{\, \delta_{\mathcal{K}}^\circ(u),\ \kappa^\circ(z),\ \delta_{\operatorname{epi}\rho}^\circ(y,\alpha) \,\} \\ &= \max\{\, \delta_{\mathcal{K}^\circ}(u),\ \kappa^\circ(z),\ \delta_{\operatorname{epi}\rho^\circ}(y,-\alpha) \,\} \\ &= \delta_{\mathcal{K}^*}(-u) + \kappa^\circ(z) + \delta_{\operatorname{epi}\rho^\circ}(y,-\alpha). \end{aligned}$$

We now use Corollary 4.2, with $\mathcal{K} = \{\,0\,\}$, to deduce the antipolar constraint set, and arrive at the gauge program

$$\underset{y,z,\alpha}{\text{minimize}} \quad \delta_{\mathcal{K}^*}(D^*z - A^*y) + \kappa^\circ(z) + \delta_{\operatorname{epi}\rho^\circ}(y,-\alpha) \quad \text{subject to} \quad \langle b,y\rangle + \sigma\alpha \ge 1.$$

Bringing the indicator functions down to the constraints leads to

$$\underset{y,z,\alpha}{\text{minimize}} \quad \kappa^\circ(z) \quad \text{subject to} \quad \langle y,b\rangle + \sigma\alpha \ge 1,\ \rho^\circ(y) \le -\alpha,\ D^*z - A^*y \in \mathcal{K}^*,$$

which can be further simplified by eliminating α, which yields the gauge dual problem

$$\underset{y,z}{\text{minimize}} \quad \kappa^\circ(z) \quad \text{subject to} \quad \langle y,b\rangle - \sigma\rho^\circ(y) \ge 1,\ D^*z - A^*y \in \mathcal{K}^*. \tag{7.2}$$

Note that the canonical dual gauge problem (D_ρ) is recovered if $D = I$ and $\mathcal{K} = \mathcal{X}$.

7.2. Nonnegative conic optimization. Conic optimization subsumes a large class of convex optimization problems that ranges from linear, to second-order, to semidefinite programming, among others. Problem (1.5) gives the general statement of a conic program; the additional requirement that $c \in \mathcal{K}^*$ is sufficient to guarantee that the objective in (1.6) is a gauge. (More generally, it is evident that any function of the form $\gamma_{\mathcal{D}} + \delta_{\mathcal{K}}$ is a gauge if $\mathcal{D}$ is a non-empty convex set and $\mathcal{K}$ is a convex cone.)

We can easily accommodate a generalization of (1.6) by embedding it within the formulation defined by (1.5), and define

$$\underset{x}{\text{minimize}} \quad \langle c, x\rangle + \delta_{\mathcal{K}}(x) \quad \text{subject to} \quad \rho(b - Ax) \le \sigma, \tag{7.3}$$

with $c \in \mathcal{K}^*$, as the conic gauge optimization problem. The following result describes its gauge dual.

PROPOSITION 7.1. Suppose that $\mathcal{K} \subset \mathcal{X}$ is a convex cone and $c \in \mathcal{K}^*$. Then the gauge

$$\kappa(x) = \langle c, x\rangle + \delta_{\mathcal{K}}(x)$$

has the polar

$$\kappa^\circ(u) = \inf\{\, \alpha \ge 0 \mid \alpha c \in \mathcal{K}^* + u \,\}.$$

If $\mathcal{K}$ is closed and $c \in \operatorname{int} \mathcal{K}^*$, then κ has compact level sets, and $\operatorname{dom} \kappa^\circ = \mathcal{X}$.

Proof. To derive the polar of κ, use Proposition 2.1 to obtain

$$\begin{aligned}
\kappa^\circ(u) &= \sup_x \{\, \langle u, x\rangle \mid \kappa(x) \le 1 \,\} \\
&\overset{(i)}{=} \sup_x \{\, \langle u, x\rangle \mid \langle c, x\rangle \le 1,\ x \in \mathcal{K} \,\} \\
&\overset{(ii)}{=} \inf\{\, \alpha \ge 0 \mid (\alpha c - u) \in \mathcal{K}^* \,\},
\end{aligned}$$

where (ii) follows from strong duality between the conic program (i) and its dual (ii), which is guaranteed because the assumptions guarantee that Slater's condition holds (Rockafellar, 1970, Corollary 28.2.2).

To prove compactness of the level sets of κ, let $\gamma := \inf_x \{\, \langle c, x\rangle \mid \|x\| = 1,\ x \in \mathcal{K} \,\}$. Because $\mathcal{K}$ is closed and $c \in \operatorname{int} \mathcal{K}^*$, compactness of the feasible set in the infimum implies that the infimum is attained and that $\gamma > 0$. Thus, for any $x \in \mathcal{K} \setminus \{0\}$, $\langle c, x\rangle \ge \gamma\|x\| > 0$ and, consequently, that $\{\, x \in \mathcal{X} \mid \kappa(x) \le \alpha \,\} = \{\, x \in \mathcal{K} \mid \langle c, x\rangle \le \alpha \,\} \subset \{\, x \in \mathcal{X} \mid \|x\| \le \alpha/\gamma \,\}$. This guarantees that the level sets of κ are bounded, which establishes the compactness of its level sets. From this and Proposition 2.1(iii), we see that $\kappa^\circ(y)$ is finite for any $y \in \mathcal{X}$. ◻

This last result on the polar of the conic gauge objective, together with Corollary 4.1 allows us to derive the gauge dual of (7.3). As a concrete example of the application of Proposition 7.1 to conic gauge programming, consider the SDP

$$\underset{X}{\text{minimize}} \quad \langle C, X\rangle \quad \text{subject to} \quad \mathcal{A}X = b,\ X \succeq 0, \tag{7.4}$$

where $C \succ 0$, and $\mathcal{A} : \mathcal{S}^n \to \mathbb{R}^m$ is a linear map from symmetric n-by-n matrices to m-vectors. Define the gauge objective $\kappa(X) = \langle C, X\rangle + \delta_{\cdot \succeq 0}(X)$, set $\sigma = 0$, and let

$\rho = \|\cdot\|$, i.e., the constraint set is $\mathcal{C} = \{ X \mid \mathcal{A}X = b \}$. Proposition 7.1, with $\mathcal{K}$ equal to the (self-dual) semidefinite cone, gives the gauge polar:

$$\kappa^\circ(U) = \inf \{ \alpha \geq 0 \mid \alpha C \succeq U \} = \inf \{ \alpha \geq 0 \mid C^{-\frac{1}{2}} U C^{-\frac{1}{2}} \preceq \alpha I \} .$$

In particular,

$$\kappa^\circ(U) = \lambda_{\max}(U, C) \quad \text{if} \quad U \succeq 0,$$

where $\lambda_{\max}(U, C)$ is the largest generalized eigenvalue corresponding to the eigenvalue problem $Ux = \lambda Cx$. Together with Corollary 4.2, which gives the antipolar of $\mathcal{C}$, and Theorem 5.3, which asserts that the optimal dual value is positive, the gauge dual problem can then be written as

$$\underset{y}{\text{minimize}} \quad \lambda_{\max}(\mathcal{A}^* y, C) \quad \text{subject to} \quad \langle b, y \rangle \geq 1.$$

To make this example even more specific, consider the case $C = I$. In that case, (7.4) is the problem of minimizing the trace of a positive semidefinite matrix X that satisfies a set of linear constraints. This problem arises, for example, in the phase-retrieval problem (Candes et al., 2012); the above gauge dual is a maximum eigenvalue problem subject to a single linear constraint.

8. Discussion. Our focus in this paper has been mainly on the duality aspects of gauge optimization. The structure particular to gauge optimization allows for an alternative to the usual Lagrange duality, and this may be useful for providing new avenues of exploration for modeling and algorithm developments. Depending on the particular application, it may prove computationally convenient or more efficient to use existing algorithms to solve the gauge dual rather than the Lagrange dual problem. For example, some variation of the projected subgradient method might be used to exploit the relative simplicity of the gauge dual constraints in (D_ρ). As with methods that solve the Lagrange dual problem, some procedure would be needed to recover the primal solution. Although this is difficult to do in general, for specific problems it is possible to develop a primal-from-dual recovery procedure via the optimality conditions.

More generally, an important question left unanswered is whether there exists a class of algorithms that can leverage this special structure. We are intrigued by the possibility of developing a primal-dual algorithm specific to the primal-dual gauge pair.

The sensitivity analysis presented in §6 relied on existing results from Lagrange duality. We would prefer, however, to develop a line of analysis that is self-contained and based entirely on gauge duality theory and some form of "gauge multipliers". Because the value function (6.1) is a gauge, it is conceivable that sensitivity analysis could be carried out based on studying the polar

$$v^\circ(y, \tau) = \inf \{ \mu \geq 0 \mid (y, \tau) \in \mu \mathcal{D} \} = \kappa^\circ(A^* y) + \delta_{\rho^\circ(\cdot) \leq \cdot}(y, -\tau)$$

of the value function, where $\mathcal{D} = \{ (y, \tau) \mid \kappa^\circ(A^* y) \leq 1, \rho^\circ(y) \leq -\tau \}$, which is a consequence of Proposition 2.1(iv) and (6.2). This approach would be in contrast to the usual sensitivity analysis, which is based on studying a certain (convex) value function and its conjugate.

REFERENCES

A. Y. Aravkin, J. Burke, and M. P. Friedlander. Variational properties of value functions. *SIAM J. Optim.*, 2013. To appear.

A. Auslender and M. Teboulle. *Asymptotic Cones and Functions in Optimization and Variational Inequalities*. Springer Monographs in Mathematics. Springer-Verlag, New York, 2003.

F. Bach. Learning with submodular functions: A convex optimization perspective. *arXiv preprint arXiv:1111.6453*, 2011.

A. Berman. *Cones, Matrices and Mathematical Programming*, volume 79 of *Lecture Notes in Economics and Mathematical Systems*. Springer-Verlag, 1973.

E. J. Candés, M. B. Wakin, and S. P. Boyd. Enhancing sparsity by reweighted L1 minimization. *J. Fourier Anal. Appl.*, 14(5):877–905, 2008. DOI: 10.1007/s00041-008-9045-x.

E. J. Candes, T. Strohmer, and V. Voroninski. Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming. *Commun. Pur. Appl. Ana.*, 2012.

V. Chandrasekaran, B. Recht, P. Parrilo, and A. Willsky. The convex geometry of linear inverse problems. *Found. Comput. Math.*, 12(6):805–849, 2012. DOI: 10.1007/s10208-012-9135-7.

R. M. Freund. Dual gauge programs, with applications to quadratic programming and the minimum-norm problem. *Math. Program.*, 38(1):47–67, 1987. DOI: 10.1007/BF02591851.

M. P. Friedlander, H. Mansour, R. Saab, and O. Yilmaz. Recovering compressively sampled signals using partial support information. *IEEE Trans. Inform. Theory*, 58(2):1122–1134, 2012.

M. X. Goemans and D. P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *J. ACM*, 42(6):1115–1145, November 1995. DOI: 10.1145/227683.227684.

J. Gwinner. An extension lemma and homogeneous programming. *J. Optim. Theory Appl.*, 47: 321–336, 1985.

M. Jaggi. Revisiting Frank-Wolfe: Projection-free sparse convex optimization. In *Proc. 30th Intern. Conf. Machine Learning (ICML-13)*, pages 427–435, 2013.

D. Krishnan, P. Lin, and A. M. Yip. A primal-dual active-set method for non-negativity constrained total variation deblurring problems. *IEEE Trans. Image Process.*, 16(11):2766–2777, 2007.

L. Lovàsz. Submodular functions and convexity. In A. Bachem, B. Korte, and M. GrÃűtschel, editors, *Mathematical Programming The State of the Art*, pages 235–257. Springer, 1983. DOI: 10.1007/978-3-642-68874-4_10.

L. McLinden. Symmetric duality for structured convex programs. *Trans. Amer. Math. Soc.*, 245: 147–181, 1978.

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

R. T. Rockafellar and R. J.-B. Wets. *Variational Analysis*. Springer-Verlag, New York, 1998.

P. H. Ruys and H. M. Weddepohl. Economic theory and duality. In J. Kriens, editor, *Convex analysis and mathematical economics*, pages 1–72. Springer, Berlin, 1979.

T. Teuber, G. Steidl, and R. H. Chan. Minimization and parameter estimation for seminorm regularization models with I-divergence constraints. *Inverse Probl.*, 29(3), 2013.

E. van den Berg. *Convex optimization for generalized sparse recovery*. PhD thesis, University of British Columbia, December 2009.

E. van den Berg and M. P. Friedlander. Sparse optimization with least-squares constraints. *SIAM J. Optim.*, 21(4):1201–1229, 2011.