GAUGE OPTIMIZATION AND DUALITY[*]

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 [R. M. Freund, *Math. Programming*, 38 (1987), pp. 47–67], 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

DOI. 10.1137/130940785

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 family of 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

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

[†]Department of Mathematics, University of California, Davis, CA 95616 (mpfriedlander@ucdavis.edu). This author's research was also supported as a visiting scholar at the Simons Institute on Theoretical Computer Science at UC Berkeley.

[‡]Department of Computer Science, University of British Columbia, Vancouver, BC V6T 1Z4, Canada (ijamj@cs.ubc.ca).

[§]Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong (tk.pong@polyu.edu.hk). This author's research was also supported as a PIMS postdoctoral fellow at Department of Computer Science, University of British Columbia, Vancouver, Canada, during the early stage of the preparation of this manuscript.

where the set

$$(1.1) \qquad \mathcal{C}' = \{\, y \mid \langle y, x \rangle \geq 1 \,\forall\, x \in \mathcal{C} \,\}$$

is the antipolar of $\mathcal{C}$ (in contrast to the better-known polar of a convex set), and the polar κ° (also a gauge) is the function that satisfies the Cauchy–Schwarz-like inequality most tightly:

$$(1.2) \qquad \langle x, y \rangle \leq \kappa(x)\,\kappa^\circ(y) \qquad \forall x \in \mathrm{dom}\,\kappa, \ \forall y \in \mathrm{dom}\,\kappa^\circ;$$

see (2.3) for the precise definition. 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

$$(1.3) \qquad 1 \leq \kappa(x)\,\kappa^\circ(y) \qquad \forall x \in \mathcal{C} \cap \mathrm{dom}\,\kappa, \ \forall y \in \mathcal{C}' \cap \mathrm{dom}\,\kappa^\circ.$$

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

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

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, so that the constraint reduces to $Ax = b$ when $\sigma = 0$. This formulation is only very slightly less general than (P) because any closed convex set can be represented as $\{\, x \mid \rho(b - x) \leq 1 \,\}$ for some vector b and gauge ρ; cf. section 2.3. However, it is sufficiently concrete that it allows us to develop a calculus for computing gauge duals for a wide range of existing problems. (Conic side constraints and a linear map in the objective are easily accommodated, and this is covered in section 7.)

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

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

which bears a striking similarity to the Lagrange dual problem

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

(These two duals are derived in section 4 under suitable assumptions.) Note that the objective and constraints between the two duals play different roles. A significant practical difference between these two formulations occurs when ρ is simply the Euclidean norm and κ is a more complicated function (such as one described by Example 1.2). 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.

In section 3 we develop an antipolar calculus for computing the antipolars of sets such as $\{\, x \mid \rho(b - Ax) \leq \sigma \,\}$, which corresponds to the constraint in our canonical formulation (P_ρ). This calculus is applied in section 4 to derive the gauge dual (D_ρ).

The formal properties of the polar and antipolar operations are described in sections 2–3. In section 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. (Seminorms drop the last requirement, and allow the function to be zero at other points.) Let $\kappa(x) = \|x\|$ be any norm, and let $\mathcal{C} = \{\, x \mid Ax = b \,\}$ describe the solutions to an underdetermined linear system. Then (P) yields a minimum-length solution to the linear system $Ax = b$. This problem can be modeled as an instance of (P_ρ) by letting ρ be any gauge function for which $\rho^{-1}(0) = \{\, 0 \,\}$ and setting $\sigma = 0$. 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 formulation 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 et al. (2012) describe how to construct regularizers that generalize the notion of sparsity in linear inverse problems. In particular, they define the gauge

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

over the convex hull of the set of canonical atoms given by 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). In this example we demonstrate that it is possible to cast any convex conic optimization problem in the gauge framework.

Let $\mathcal{K}$ be a closed convex cone, and let $\mathcal{K}^*$ denote its dual. Consider the primal-dual pair of feasible conic problems:

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

$$(1.5b) \qquad \underset{y}{\text{maximize}} \quad \langle b, y \rangle \quad \text{subject to} \quad c - A^*y \in \mathcal{K}^*.$$

Suppose that $\widehat{y}$ is a dual-feasible point, and define $\widehat{c} = c - A^*\widehat{y}$. Because $\widehat{c} \in \mathcal{K}^*$, it follows that $\langle \widehat{c}, x \rangle \geq 0$ for all $x \in \mathcal{K}$. In particular, the primal problem can be equivalently formulated as a gauge optimization problem by defining

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

where $\delta_{\mathcal{K}}$ is the indicator function on the set $\mathcal{K}$. (More generally, it is evident that any function of the form $\gamma + \delta_{\mathcal{K}}$ is a gauge if γ is a gauge.) This formulation is a generalization of the nonnegative linear program 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 (PSD) matrices, which arises in the phase-retrieval problem Candès, Strohmer, and Voroninski (2012). This is an example where $c \in \mathcal{K}^*$, in which case the dual-feasible point $\widehat{y} = 0$ is trivially available for the gauge reformulation; cf. section 7.2.1.

A concrete example of the simple case where $c \in \mathcal{K}^*$ is the semidefinite programming relaxation of the max-cut problem studied by Goemans and Williamson (1995). Let $G = (V, E)$ be an undirected graph, and let $D = \text{diag}\left((d_v)_{v \in V}\right)$, 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 semidefinite programming relaxation for this problem is derived by "lifting" xx^T into a PSD matrix:

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

where e is the vector of all ones, and $X \succeq 0$ denotes the the PSD constraint on X. The constraint $\text{diag}\, X = e$ implies that $\langle D, X \rangle = \sum_{v \in V} d_v = 2|E|$ is constant. Thus, the optimal value is equal to

$$(1.7) \qquad |E| - \tfrac{1}{4} \cdot \underset{X}{\min} \{\, \langle D + A, X \rangle \mid \text{diag}\, X = e, \ X \succeq 0 \,\},$$

and the solution can be obtained by solving this latter problem. Note that $D + A$ is PSD because it has nonnegative diagonals 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.) Under the trace inner product, the cone of PSD matrices is self-dual; therefore, the trace inner product between PSD matrices is nonnegative, and (1.7) falls into the class of conic gauge problems defined by (1.5a).

Example 1.4 (submodular functions). Let $V = \{\, 1, \ldots, n \,\}$, 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

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

where $x_{j_1} \geq x_{j_2} \geq \cdots \geq x_{j_n}$ 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.,

$$f(A) + f(B) \geq f(A \cup B) + f(A \cap B) \quad \forall A, B \subset V;$$

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

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

then the extension is nonnegative over $\mathbb{R}^n_+$. Thus, when f is a submodular and nondecreasing set function, that function plus the indicator on the nonnegative orthant, i.e., $\widehat{f} + \delta_{\mathbb{R}^n_+}$, is a gauge. Bach (2011) surveys the properties of submodular functions and their application in machine learning; see Proposition 3.7 therein.

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): in that text, see section 14 for a discussion of polarity operations on convex sets, and section 15 for a discussion of gauge functions and their corresponding polarity operations.

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

For a convex function $f : \mathcal{X} \to \mathbb{R} \cup \{\infty\}$, its domain is denoted by $\operatorname{dom} f = \{x \mid f(x) < \infty\}$, and its epigraph by $\operatorname{epi} f = \{(x, \mu) \mid f(x) \leq \mu\}$. The function is proper if $\operatorname{dom} f \neq \emptyset$, and it is closed if its epigraph is closed, which is equivalent to the function being lower semi-continuous (Rockafellar, 1970, Theorem 7.1). Let $\operatorname{cl} f$ denote the convex function whose epigraph is $\operatorname{cl} \operatorname{epi} f$, which is the largest lower semi-continuous function smaller than f (Rockafellar, 1970, p. 52). Also, for any $x \in \operatorname{dom} f$, the subdifferential of f at x is denoted $\partial f(x) = \{y \mid f(u) - f(x) \geq \langle y, u - x \rangle \ \forall u\}$. Finally, the convex conjugate of a proper convex function f is

$$(2.1) \qquad f^*(y) := \sup_x \{\langle x, y \rangle - f(x)\},$$

which is a proper closed convex function (Rockafellar, 1970, Theorem 12.2).

2.1. Assumptions. We make the following blanket assumptions throughout. The set $\mathcal{C}$ is a nonempty closed convex set that does not contain the origin; the set $\mathcal{D}$ is 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 (P_ρ), is closed; when $\sigma = 0$, we additionally assume that $\rho^{-1}(0) = \{0\}$.

2.2. Polar sets. The polar of a nonempty convex set $\mathcal{D}$ is defined as

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

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

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

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

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

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

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

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

$$(2.2) \qquad \kappa(x) = \gamma_{\mathcal{D}}(x) := \inf\{\, \lambda \geq 0 \mid x \in \lambda \mathcal{D} \,\}.$$

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

$$(2.3) \qquad \kappa^\circ(y) := \inf\{\, \mu > 0 \mid \langle x, y \rangle \leq \mu \kappa(x) \; \forall x \,\},$$

which explains the inequality (1.2). The following proposition collects properties that relate the polar and conjugate of a gauge.

PROPOSITION 2.1. *For the gauge $\kappa : \mathcal{X} \to \mathbb{R} \cup \{\infty\}$, it holds that*
  (i) *κ° is a closed gauge function;*
 (ii) *$\kappa^{\circ\circ} = \operatorname{cl} \kappa = \kappa^{**}$;*
(iii) *$\kappa^\circ(y) = \sup_x \{\, \langle x, y \rangle \mid \kappa(x) \leq 1 \,\}$ for all y;*
 (iv) *$\kappa^*(y) = \delta_{\kappa^\circ(\cdot) \leq 1}(y)$ for all y;*
  (v) *$\operatorname{dom} \kappa^\circ = \mathcal{X}$ if κ is closed and $\kappa^{-1}(0) = \{\, 0 \,\}$;*
 (vi) *$\operatorname{epi} \kappa^\circ = \{\, (y, \lambda) \mid (y, -\lambda) \in (\operatorname{epi} \kappa)^\circ \,\}$.*

Proof. The first two items are proved in Theorems 15.1 and 12.2 of Rockafellar (1970). Item (iii) follows directly from the definition (2.3) of the polar gauge. To prove item (iv), we note that if $g(t) = t$, $t \in \mathbb{R}$, then the so-called monotone conjugate g^+ is

$$g^+(s) = \sup_{t \geq 0}\{\, st - t \,\} = \delta_{[0,1]}(s),$$

where $s \geq 0$. Now, apply Rockafellar (1970, Theorem 15.3) with $g(t) = t$, and κ^{**} in place of f in that theorem to obtain that $\kappa^{***}(y) = \delta_{[0,1]}(\kappa^{**\circ}(y))$. The conclusion in item (iv) now follows by noting that $\kappa^{***} = \kappa^*$ and $\kappa^{**\circ} = \kappa^{\circ\circ\circ} = \kappa^\circ$. To prove item (v), note that the assumptions together with Auslender and Teboulle (2003, Proposition 3.1.3) show that $0 \in \operatorname{int} \operatorname{dom} \kappa^*$. This together with item (iv) and the positive homogeneity of κ° shows that $\operatorname{dom} \kappa^\circ = \mathcal{X}$. Finally, item (vi) is stated on Rockafellar (1970, p. 137) and can also be verified directly from the definition. ∎

In many 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) *$\operatorname{ri} \operatorname{dom} \kappa \cap \operatorname{range} A \neq \emptyset$.*
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

$$(\kappa \circ A)^{\circ}(y) = \sup_{x} \left\{ \langle y, x \rangle \mid \kappa(Ax) \leq 1 \right\} = -\inf_{x} \left\{ \langle -y, x \rangle + \delta_{\mathcal{D}}(Ax) \right\},$$

where $\mathcal{D} = \{ x \mid \kappa(x) \leq 1 \}$. Since κ is positively homogeneous, we have $\operatorname{dom} \kappa = \bigcup_{\lambda \geq 0} \lambda \mathcal{D}$. Hence, $\operatorname{ri} \operatorname{dom} \kappa = \bigcup_{\lambda > 0} \lambda \operatorname{ri} \mathcal{D}$ from Rockafellar (1970, p. 50). Thus, assumption (ii) implies that $\operatorname{ri} \mathcal{D} \cap \operatorname{range} A \neq \emptyset$. On the other hand, assumption (i) implies that $\mathcal{D}$ is polyhedral; and $\mathcal{D} \cap \operatorname{range} A \neq \emptyset$ because they both contain the origin. Use these conclusions and apply Rockafellar (1970, Corollary 31.2.1(a)), with $f(x) := \langle -y, x \rangle$ and $g(Ax) := -\delta_{\mathcal{D}}(Ax)$ (see also Rockafellar's remark right after that corollary for the case when $\mathcal{D}$ is polyhedral), to conclude that

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

where the second equality follows from the definition of conjugate functions and Proposition 2.1(iii). Moreover, from that same corollary, the supremum is attained when finite. (Note that Rockafeller's statement of that corollary is formulated for the difference between convex and concave function, and must be appropriately adapted to our case.) This completes the proof. $\square$

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,* $\mathcal{D} = \{ x \mid \kappa(x) \leq 1 \}$.

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 $\mathcal{D} = \{ x \mid \kappa(x) \leq 1 \}$ can be verified directly from definition. $\square$

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*

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

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 $\mathcal{D}_i = \{ x \mid \kappa_i^\circ(x) \leq 1 \}$ 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

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

This completes the proof. □

The following corollary is immediate from Propositions 2.2 and 2.4.

COROLLARY 2.5. *Let κ_1 and κ_2 be gauges. Suppose that either*
 (i) *epi κ_1 and epi κ_2 are polyhedral; or*
 (ii) *ri dom $\kappa_1 \cap$ ri dom $\kappa_2 \neq \emptyset$.*
Then

$$(2.4) \qquad (\kappa_1 + \kappa_2)^\circ(y) = \inf_{u_1, u_2} \{ \max \{ \kappa_1^\circ(u_1), \ \kappa_2^\circ(u_2) \} \mid u_1 + u_2 = y \}.$$

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. □

The support function for a nonempty convex set $\mathcal{D}$ is defined as

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

see Rockafellar (1970, section 13) for properties of support functions. It is immediate from the definition of the convex conjugate that $\sigma_{\mathcal{D}} = \delta_{\mathcal{D}}^*$. It is also straightforward to verify that if $\mathcal{D}$ contains the origin, then the corresponding support function $\sigma_{\mathcal{D}}$ is a (closed) gauge function. Moreover, we have the following relationship between support and Minkowski functions (Rockafellar, 1970, Corollary 15.1.2).

PROPOSITION 2.6. *Let $\mathcal{D}$ be a closed convex set that contains the origin. Then*

$$\gamma_{\mathcal{D}}^\circ = \sigma_{\mathcal{D}} \quad and \quad \sigma_{\mathcal{D}}^\circ = \gamma_{\mathcal{D}}.$$

2.4. 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 $x, y \in \mathcal{D}$,*

$$x + \alpha y \in \mathcal{D} \quad \forall \alpha \geq 0.$$

Note 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), follows from the bi-antipolar 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. *$\operatorname{cl cone}(\mathcal{C}') = \mathcal{C}^* = (\mathcal{C}')_\infty$.*

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

Proof. It is evident that $\mathrm{cl}\,\mathrm{cone}(\mathcal{C}') \subseteq \mathcal{C}^*$. 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

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

which shows that $x + \tau x_0 \in \mathrm{cone}\,\mathcal{C}'$. Taking the limit as τ goes to 0 shows that $x \in \mathrm{cl}\,\mathrm{cone}(\mathcal{C}')$. 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 $x_0 + \tau x \in \mathcal{C}'$ for all $\tau > 0$ (see, e.g., Auslender and Teboulle (2003, Proposition 2.1.5)). 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 $x \in \mathcal{C}^*$. $\square$

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 useful for writing down the explicit gauge duals of problems such as (P_ρ). Table 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 $\mathcal{C} = \{\, x \mid \rho(b - x) \leq \sigma \,\}$ with $0 < \sigma < \rho(b)$. Then*

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

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

$$\langle x - b, y \rangle \geq 1 - \langle b, y \rangle \iff \langle b - x, y \rangle \leq \langle b, y \rangle - 1.$$

From Proposition 2.1(iii), this is further equivalent to $\sigma\rho^\circ(y) \leq \langle b, y \rangle - 1$. $\square$

Proposition 3.1 is very general since any closed convex set $\mathcal{D}$ containing the origin can be represented in the form of $\{\, x \mid \rho(x) \leq 1 \,\}$, where $\rho(x) = \inf\{\, \lambda \geq 0 \mid x \in \lambda\mathcal{D} \,\}$; cf. (2.2). For conic constraints in particular, one obtains the following corollary by setting $\rho(x) = \delta_{-\mathcal{K}}(x)$.

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*

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

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

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

The last relation is equivalent to $A^*y \in \mathcal{C}'$. Hence, $(A\mathcal{C})' = (A^*)^{-1}\mathcal{C}'$. Furthermore, the assumption that $\mathrm{cl}(A\mathcal{C})$ does not contain the origin, together with an argument using supporting hyperplanes, implies $(A\mathcal{C})'$ is nonempty. This completes the proof. $\square$

We have the following result concerning the preimage of $\mathcal{C}$.

PROPOSITION 3.4. *If the set $\mathcal{C}$ is closed convex and does not contain the origin, and $A^{-1}\mathcal{C} \neq \emptyset$, then*

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

and both sets are nonempty.

Proof. It follows from the assumptions that $\mathrm{cl}(A^*\mathcal{C}')$ is nonempty. Moreover, $A^{-1}\mathcal{C}$ is also closed convex and 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 $\mathrm{cl}(A^*\mathcal{C}) \subseteq (A^{-1}\mathcal{C})'$. 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

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

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 from the bi-antipolar theorem that $(A^*\mathcal{C}')'' = \mathrm{cl}(A^*\mathcal{C}')$. Moreover, we have

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

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 consequence.

PROPOSITION 3.5. *Suppose that $A^{-1}\mathcal{C} \neq \emptyset$, and either $\mathcal{C}$ is polyhedral or $\operatorname{ri}\mathcal{C} \cap$ 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 proposition. The conclusion then follows immediately from Proposition 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 polyhedral and thus $\mathcal{C}' + \ker A^*$ is closed. Hence, the conclusion of the proposition 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

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

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), implies that

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

where we use the fact that $\mathcal{C}' = \partial\sigma_{\mathcal{C}'}(0) \equiv \{y \mid \langle y, x\rangle \leq \sigma_{\mathcal{C}'}(x)\ \forall x\}$ for the first equality; see Rockafellar (1970, Theorem 13.1). 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. From the fact that both $\mathcal{C}_1$ and $\mathcal{C}_2$ are closed convex sets not containing the origin, it follows that $\mathcal{C}_1'$ and $\mathcal{C}_2'$ are nonempty and hence $\mathrm{cl\,conv}(\mathcal{C}_1' \cup \mathcal{C}_2') \neq \emptyset$. Moreover, because $\mathcal{C}_1 \cap \mathcal{C}_2$ does not contain the origin, $(\mathcal{C}_1 \cap \mathcal{C}_2)'$ is also nonempty.

We first show that $\mathrm{cl\,conv}(\mathcal{C}_1' \cup \mathcal{C}_2')$ 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 $\mathrm{cl\,conv}(\mathcal{C}_1' \cup \mathcal{C}_2') \subseteq (\mathcal{C}_1 \cap \mathcal{C}_2)'$. Since $\mathcal{C}_1 \cap \mathcal{C}_2$ is nonempty, $(\mathcal{C}_1 \cap \mathcal{C}_2)'$ does not contain the origin. Consequently, $\mathrm{cl\,conv}(\mathcal{C}_1' \cup \mathcal{C}_2')$ 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

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

Take the antipolar of both sides to obtain

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

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

The following counterexample 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

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

Define $H_1 = \{ (x_1, x_2) \mid x_1 + x_2 \geq 1 \}$ and $H_2 = \{ (x_1, x_2) \mid x_1 - x_2 \geq 1 \}$ so that $\mathcal{C}_1 = H_1 \cap H_2$. 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 $(\mathcal{C}_1 \cap \mathcal{C}_2)' = \{ (y_1, y_2) \mid y_1 \geq 1 \}$.

Next, it follows directly from the antipolar definition that $\mathcal{C}_2' = \{ (y_1, 0) \mid y_1 \geq 1 \}$. Also note that $H_1 = L_1^{-1} I$, where $L_1(x_1, x_2) = x_1 + x_2$ and $I = \{ u \mid u \geq 1 \}$. Thus, by Proposition 3.5, $H_1' = \{ (y_1, y_1) \mid y_1 \geq 1 \}$. Similarly, $H_2' = \{ (y_1, -y_1) \mid y_1 \geq 1 \}$. Because H_1 and H_2 are ray-like, it follows from Proposition 3.7 that

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

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

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

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

$$(4.1) \qquad \qquad \mathcal{C} = \{ x \mid \rho(b - Ax) \leq \sigma \}$$

denote the constraint set, where ρ is a closed gauge and $0 \leq \sigma < \rho(b)$. We also consider the associated set

$$(4.2) \qquad \qquad \mathcal{C}_0 = \{ u \mid \rho(b - u) \leq \sigma \},$$

and note that $\mathcal{C} = A^{-1} \mathcal{C}_0$. Recall from our blanket assumption in section 2 that when $\sigma = 0$, we consider only closed gauges ρ with $\rho^{-1}(0) = \{ 0 \}$.

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 an affine set. 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 Proposition 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 $0 < \sigma < \rho(b)$, and $\mathcal{C}_0$ is given by (4.2). If $\mathcal{C}_0$ is polyhedral, or $\operatorname{ri}\mathcal{C}_0 \cap \operatorname{range} A \neq \emptyset$, then*

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

As an aside, we present the following result, which follows from Corollary 3.2 and Proposition 3.5. It concerns a general closed convex cone $\mathcal{K}$.

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 $\mathcal{K}$ is polyhedral, or $(b + \operatorname{ri}\mathcal{K}) \cap \operatorname{range} A \neq \emptyset$, then*

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

These results can be used to obtain an explicit representation of the gauge dual problem. We rely on the antipolar calculus developed in section 3. Assume that

$$(4.3) \qquad \mathcal{C}_0 \text{ is polyhedral} \quad \text{or} \quad \operatorname{ri}\mathcal{C}_0 \cap \operatorname{range} A \neq \emptyset.$$

Consider separately the cases $\sigma > 0$ and $\sigma = 0$.

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

$$(4.4) \qquad \mathcal{C}' = \{\, A^*y \mid \langle b, y \rangle - \sigma\rho^\circ(y) \geq 1 \,\}.$$

Case 2: $\sigma = 0$. Here we use the blanket assumption (see section 2) that $\rho^{-1}(0) = \{0\}$, and in that case, $\mathcal{C} = \{\, x \mid Ax = b \,\}$. Apply Corollary 4.2 with $\mathcal{K} = \{0\}$ to obtain

$$(4.5) \qquad \mathcal{C}' = \{\, A^*y \mid \langle b, y \rangle \geq 1 \,\}.$$

Since $\rho^{-1}(0) = \{0\}$ and ρ is closed, we conclude from Proposition 2.1(v) that $\operatorname{dom}\rho^\circ = \mathcal{X}$. Hence, (4.5) can be seen as a special case of (4.4) with $\sigma = 0$.

These two cases can be combined, and we see that when (4.3) holds, the gauge dual problem (D) for (P_ρ) can be expressed as (D_ρ). If the assumptions (4.3) are not satisfied, then in view of Proposition 3.4, it still holds that (D) is equivalent to

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

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 (4.3). 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

$$(4.6) \qquad \underset{x,r,\tau}{\operatorname{minimize}} \quad \xi(x, r, \tau) \quad \text{subject to} \quad Ax + r = b, \ \tau = \sigma.$$

Invoke Proposition 2.4 to obtain

$$\xi^\circ(z, y, \alpha) = \max\left\{ \kappa^\circ(z),\ (\delta_{\mathrm{epi}\,\rho})^\circ(y, \alpha) \right\}$$

$$\overset{(i)}{=} \max\left\{ \kappa^\circ(z),\ \delta_{(\mathrm{epi}\,\rho)^\circ}(y, \alpha) \right\}$$

$$\overset{(ii)}{=} \kappa^\circ(z) + \delta_{(\mathrm{epi}\,\rho)^\circ}(y, \alpha) \overset{(iii)}{=} \kappa^\circ(z) + \delta_{\mathrm{epi}(\rho^\circ)}(y, -\alpha),$$

where (i) follows from Proposition 2.3(i), (ii) follows from the definition of indicator function, and (iii) follows from Proposition 2.1(vi). As Freund (1987, section 2) shows for gauge programs with linear constraints, the gauge dual is given by

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

which can be rewritten as

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

(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 (4.6) 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 (4.3) used in section 4.1, and thus appears to apply more generally, it is important to keep in mind that we have yet to impose conditions that imply strong duality. In fact, as we show in section 5, the assumptions required there imply (4.3).

4.2. Lagrange duality. Our derivation of the Lagrange dual problem (D_ℓ) is standard, and we include here as a counterpoint to the corresponding gauge dual derivation. We begin by reformulating (P_ρ) by introducing an artificial variable r, and deriving the dual of the equivalent problem

$$(4.7) \qquad \underset{x,r}{\mathrm{minimize}} \quad \kappa(x) \quad \text{subject to} \quad Ax + r = b,\ \rho(r) \leq \sigma.$$

Define the Lagrangian function

$$L(x, r, y) = \kappa(x) + \langle y, b - Ax - r \rangle.$$

The Lagrange dual problem is given by

$$\underset{y}{\mathrm{maximize}} \quad \underset{x,\,\rho(r) \leq \sigma}{\inf} L(x, r, y).$$

Consider the (concave) dual function

$$\ell(y) = \underset{x,\,\rho(r) \leq \sigma}{\inf} L(x, r, y)$$

$$= \underset{x,\,\rho(r) \leq \sigma}{\inf} \left\{ \langle y, b \rangle - \langle y, r \rangle - \left(\langle A^*y, x \rangle - \kappa(x) \right) \right\}$$

$$= \langle y, b \rangle - \underset{\rho(r) \leq \sigma}{\sup} \langle y, r \rangle - \underset{x}{\sup}\left\{ \langle A^*y, x \rangle - \kappa(x) \right\}$$

$$= \langle y, b \rangle - \sigma\rho^\circ(y) - \delta_{\kappa^\circ(\cdot) \leq 1}(A^*y),$$

where the first conjugate on the right-hand side follows from Proposition 2.1(iii) when $\sigma > 0$, and when $\sigma = 0$, it is a direct consequence of the assumption that $\rho^{-1}(0) = \{0\}$ so that $\operatorname{dom}\rho^\circ = \mathcal{X}$ from Proposition 2.1(v); the last conjugate follows from Proposition 2.1(iv). The Lagrange dual problem is obtained by maximizing ℓ, leading to (D_ℓ).

Strictly speaking, the Lagrangian primal-dual pair of problems that we have derived is given by (4.7) and (D_ℓ), but it is easy to see that (P_ρ) is equivalent to (4.7), in the sense that the respective optimal values are the same, and that solutions to one problem readily lead to solutions for the other. Thus, without loss of generality, we refer to (D_ℓ) as the Lagrange dual to the primal problem (P_ρ).

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.

The Fenchel dual (Rockafellar, 1970, section 31) of (P) is given by

$$(5.1) \qquad \underset{y}{\text{maximize}} \quad -\sigma_\mathcal{C}(-y) \quad \text{subject to} \quad \kappa^\circ(y) \le 1,$$

where we use $(\delta_\mathcal{C})^* = \sigma_\mathcal{C}$ and Proposition 2.1(iv) to obtain $\kappa^* = \delta_{[\kappa^\circ \le 1]}$. Let v_p, v_g, and v_f, respectively, denote the optimal values of (P), (D), and (5.1). The following result relates their optimal values and dual solutions.

THEOREM 5.1 (weak duality). *Suppose that* $\operatorname{dom}\kappa^\circ \cap \mathcal{C}' \ne \emptyset$. *Then*

$$v_p \ge v_f = 1/v_g > 0.$$

Furthermore,
(i) *if* y^* *solves* (5.1), *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* (5.1).

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

Because $\operatorname{dom}\kappa^\circ \cap \mathcal{C}' \ne \emptyset$, there exists y_0 such that $\kappa^\circ(y_0) \le 1$ and $y_0 \in \tau\mathcal{C}'$ for some $\tau > 0$. In particular, because $-\sigma_\mathcal{C}(-y) = \inf_{c \in \mathcal{C}}\langle c, y\rangle$ for all y, it follows from the definition of v_f that

$$(5.2) \qquad v_f = \sup_y \left\{ \inf_{c \in \mathcal{C}}\langle c, y\rangle \mid \kappa^\circ(y) \le 1 \right\} \ge \inf_{c \in \mathcal{C}}\langle c, y_0\rangle \ge \tau > 0.$$

Hence

$$(5.3) \qquad v_f = \sup_{y,\lambda} \left\{ \lambda \mid \kappa^\circ(y) \le 1, \ -\sigma_\mathcal{C}(-y) \ge \lambda, \ \lambda > 0 \right\}.$$

From this, we have further that

$$v_f = \sup_{y,\lambda} \left\{ \lambda \mid \kappa^\circ(y/\lambda) \le 1/\lambda, \ -\sigma_\mathcal{C}(-y/\lambda) \ge 1, 1/\lambda > 0 \right\}$$
$$= \sup_{y,\mu} \left\{ 1/\mu \mid \kappa^\circ(\mu y) \le \mu, \ -\sigma_\mathcal{C}(-\mu y) \ge 1, \ \mu > 0 \right\}.$$

Inverting both sides of this equation gives

$$1/v_f = \inf_{y,\mu} \{ \mu \mid \kappa^\circ(\mu y) \le \mu, \ -\sigma_{\mathcal{C}}(-\mu y) \ge 1, \ \mu > 0 \}$$

$$= \inf_{w,\mu} \{ \mu \mid \kappa^\circ(w) \le \mu, \ -\sigma_{\mathcal{C}}(-w) \ge 1, \ \mu > 0 \}$$

(5.4)
$$\overset{(i)}{=} \inf_{w,\mu} \{ \mu \mid \kappa^\circ(w) \le \mu, \ w \in \mathcal{C}', \ \mu > 0 \}$$

$$= \inf_{w,\mu} \{ \mu \mid \kappa^\circ(w) \le \mu, \ w \in \mathcal{C}' \}$$

$$= \inf_{w} \{ \kappa^\circ(w) \mid w \in \mathcal{C}' \} = v_g,$$

where equality (i) follows from the definition of $\mathcal{C}'$. This proves $v_f = 1/v_g$.

We now prove item (i). Assume that y^* solves (5.1). Then v_f is nonzero (by (5.2)) and finite, and so is $v_g = 1/v_f$. Then $y^* \in \operatorname{cone}\mathcal{C}'$ because $-\sigma_{\mathcal{C}}(-y^*) = \inf_{c \in \mathcal{C}}\langle c, y^* \rangle = v_f > 0$, and we see from (5.4) that 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 (5.4) that $y^*/v_g = v_f y^*$ solves (5.1). This completes the proof. $\square$

Fenchel duality theory allows us to use Theorem 5.1 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 5.2 (strong duality I). *Suppose that* $\operatorname{dom}\kappa^\circ \cap \mathcal{C}' \ne \emptyset$ *and* $\operatorname{ri}\operatorname{dom}\kappa \cap \operatorname{ri}\mathcal{C} \ne \emptyset$. *Then* $v_p v_g = 1$ *and the gauge dual* (D) *attains its optimal value.*

Proof. From $\operatorname{ri}\operatorname{dom}\kappa \cap \operatorname{ri}\mathcal{C} \ne \emptyset$ 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 5.1. $\square$

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

(5.5)
$$\underset{x}{\operatorname{minimize}} \quad \kappa^{\circ\circ}(x) \quad \text{subject to} \quad x \in \mathcal{C}'',$$

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 (5.5) 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 5.3. *Suppose that* κ *is closed. Then the optimal values of* (P) *and* (5.5) *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 (5.5) is equivalent to

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

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

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

COROLLARY 5.4 (strong duality II). *Suppose that* κ *is closed, and that* $\operatorname{ri}\operatorname{dom}\kappa \cap \operatorname{ri}\mathcal{C} \ne \emptyset$ *and* $\operatorname{ri}\operatorname{dom}\kappa^\circ \cap \operatorname{ri}\mathcal{C}' \ne \emptyset$. *Then* $v_p v_g = 1$ *and both values are attained.*

Proof. The conclusion follows from Corollary 5.2, Proposition 5.3, the fact that $\kappa = \kappa^{\circ\circ}$ for closed gauge functions, and the observation that $\operatorname{ri}\operatorname{dom}\kappa \cap \operatorname{ri}\mathcal{C} \ne \emptyset$ if and

only if $\operatorname{ri}\operatorname{dom}\kappa \cap \operatorname{ri}\mathcal{C}'' \neq \emptyset$, since $\operatorname{ri}\mathcal{C}'' = \bigcup_{\lambda>1}\lambda\operatorname{ri}\mathcal{C}$ (Rockafellar, 1970, p. 50) and $\operatorname{dom}\kappa$ is a cone. $\quad\square$

Before closing this section, we specialize Theorem 5.1 to study the relationship between the Lagrange (D_ℓ) and gauge (D_ρ) duals. Let v_l denote the optimal value of (D_ℓ). We use the fact that, for any y,

$$(5.6) \qquad -\sigma_{\mathcal{C}}(-y) = \inf_{c\in\mathcal{C}} \langle c, y\rangle \begin{cases} > 0 & \text{if } y \in \operatorname{cone}\mathcal{C}'\setminus\{0\}, \\ \leq 0 & \text{otherwise,} \end{cases}$$

which is directly verifiable using the definition of $\mathcal{C}'$.

COROLLARY 5.5. *Suppose that $\mathcal{C}$ is given by (4.1), where $0 \leq \sigma < \rho(b)$, assumption (4.3) 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. From (5.6), for any $y \in \operatorname{cone}\mathcal{C}'\setminus\{0\}$, we have $-\sigma_{\mathcal{C}}(-y) = \inf_{c\in\mathcal{C}}\langle c, y\rangle > 0$ and is hence finite. Note that $\inf_{c\in\mathcal{C}}\langle c, y\rangle = \inf_{c,r}\{\langle c, y\rangle \mid Ac + r = b, \ \rho(r) \leq \sigma\}$. Use this reformulation and proceed as in section 4.2 to obtain the dual function

$$\ell(u) = \inf_{c,\,\rho(r)\leq\sigma} \big\{\langle u, b\rangle - \langle u, r\rangle - \big(\langle A^*u, c\rangle - \langle c, y\rangle\big)\big\}$$

$$= \langle b, u\rangle - \sup_{\rho(r)\leq\sigma} \langle u, r\rangle - \sup_c \big\{\langle A^*u - y, c\rangle\big\}$$

$$= \langle b, u\rangle - \sigma\rho^\circ(u) - \delta_{A^*u=y}(u).$$

The dual problem to $\inf_{c\in\mathcal{C}}\langle c, y\rangle$ is given by maximizing ℓ over u. Because of assumption (4.3) and the finiteness of $-\sigma_{\mathcal{C}}(-y)$,

$$(5.7) \qquad \inf_{c\in\mathcal{C}} \langle c, y\rangle = \sup_{y=A^*u} \big\{\langle b, u\rangle - \sigma\rho^\circ(u)\big\},$$

and the supremum is attained, which is a consequence of Rockafellar (1970, Corollary 28.2.2 and Theorem 28.4). On the other hand, for any $y \notin \operatorname{cone}\mathcal{C}'\setminus\{0\}$, we have from weak duality and (5.6) that

$$(5.8) \qquad \sup_{y=A^*u} \big\{\langle b, u\rangle - \sigma\rho^\circ(u)\big\} \leq \inf_{c\in\mathcal{C}} \langle c, y\rangle \leq 0.$$

Since $\operatorname{dom}\kappa^\circ \cap \mathcal{C}' \neq \emptyset$, we can substitute (5.7) into (5.3) and obtain

$$0 < v_f = \sup\{\lambda \mid \kappa^\circ(y) \leq 1, \ -\sigma_{\mathcal{C}}(-y) \geq \lambda > 0\}$$

$$= \sup\{\langle b, u\rangle - \sigma\rho^\circ(u) \mid \kappa^\circ(A^*u) \leq 1, \ A^*u \in \operatorname{cone}\mathcal{C}'\setminus\{0\}\}$$

$$= \sup\{\langle b, u\rangle - \sigma\rho^\circ(u) \mid \kappa^\circ(A^*u) \leq 1\} = v_l,$$

where the last equality follows from (5.7), (5.8), and the positivity of v_f. This completes the first part of the proof. In particular, the Fenchel dual problem (5.1) has the same optimal value as the Lagrange dual problem (D_ℓ), and $y^* = A^*u^*$ solves (5.1) if and only if u^* solves (D_ℓ). Moreover, since assumption (4.3) holds, section 4.1 shows that (D) is equivalent to (D_ρ). The conclusion now follows from these and Theorem 5.1. $\quad\square$

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

COROLLARY 5.6. *Suppose that $\mathcal{C}$ and $\mathcal{C}_0$ are given by (4.1) and (4.2), where $0 \le \sigma < \rho(b)$. Suppose also that κ is closed,*

$$(5.9) \qquad \operatorname{ri} \operatorname{dom} \kappa \cap A^{-1} \operatorname{ri} \mathcal{C}_0 \ne \emptyset, \quad and \quad \operatorname{ri} \operatorname{dom} \kappa^\circ \cap A^* \operatorname{ri} \mathcal{C}_0' \ne \emptyset.$$

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

Proof. Since $A^{-1} \operatorname{ri} \mathcal{C}_0 \ne \emptyset$, A satisfies the assumption in (4.3). Then section 4.1 shows that (D) is equivalent to (D_ρ). Moreover, from Rockafellar (1970, Theorems 6.6 and 6.7), we see that $\operatorname{ri} \mathcal{C} = A^{-1} \operatorname{ri} \mathcal{C}_0$ and $\operatorname{ri} \mathcal{C}' = A^* \operatorname{ri} \mathcal{C}_0'$. The conclusion now follows from Corollary 5.4. ∎

This last result also holds if $\mathcal{C}_0$ were polyhedral; in that case, the assumptions (5.9) could be replaced with $\operatorname{ri} \operatorname{dom} \kappa \cap \mathcal{C} \ne \emptyset$ and $\operatorname{ri} \operatorname{dom} \kappa^\circ \cap \mathcal{C}' \ne \emptyset$.

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

$$(6.1) \qquad v(h, k) = \inf_x\ f(x, h, k),$$

with

$$(6.2) \qquad f(x, h, k) := \kappa(x) + \delta_{\operatorname{epi} \rho}(b + h - Ax, \sigma + k).$$

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:

$$f^*(z, y, \tau) = \sup_{x,h,k}\ \{\ \langle z, x \rangle + \langle y, h \rangle + \tau k - \kappa(x) - \delta_{\operatorname{epi} \rho}(b + h - Ax, \sigma + k)\ \}$$

$$= \sup_{x,w,\mu}\ \{\ \langle z + A^* y, x \rangle - \kappa(x) + \langle y, w \rangle + \tau \mu - \delta_{\operatorname{epi} \rho}(w, \mu)\ \} - \langle b, y \rangle - \tau \sigma$$

$$= \kappa^*(z + A^* y) + \delta^*_{\operatorname{epi} \rho}(y, \tau) - \langle b, y \rangle - \tau \sigma.$$

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

$$f^*(z, y, \tau) + \langle b, y \rangle + \tau \sigma = \delta_{\kappa^\circ(\cdot) \le 1}(z + A^* y) + \sigma_{\operatorname{epi} \rho}(y, \tau)$$

$$\stackrel{\text{(i)}}{=} \delta_{\kappa^\circ(\cdot) \le 1}(z + A^* y) + \delta_{(\operatorname{epi} \rho)^\circ}(y, \tau)$$

$$\stackrel{\text{(ii)}}{=} \delta_{\kappa^\circ(\cdot) \le 1}(z + A^* y) + \delta_{\operatorname{epi}(\rho^\circ)}(y, -\tau)$$

$$= \delta_{\kappa^\circ(\cdot) \le 1}(z + A^* y) + \delta_{\rho^\circ(\cdot) \le .}(y, -\tau),$$

where equality (i) follows from Proposition 2.6 and Proposition 2.3(i), and equality (ii) follows from Proposition 2.1(vi). Combining this with the definition of the value function $v(h, k)$,

$$v^*(y, \tau) = \sup_{h,k}\ \{\ \langle y, h \rangle + \tau k - v(h, k)\ \}$$

$$(6.3) \qquad = \sup_{x,h,k}\ \{\ \langle y, h \rangle + \tau k - f(x, h, k)\ \}$$

$$= f^*(0, y, \tau) = -\langle b, y \rangle - \sigma \tau + \delta_{\kappa^\circ(\cdot) \le 1}(A^* y) + \delta_{\rho^\circ(\cdot) \le .}(y, -\tau).$$

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

$$\partial v(0,0) = \operatorname*{argmax}_{y,\tau} \{ -f^*(0,y,\tau) \}$$

(6.4)
$$= \operatorname*{argmax}_{y,\tau} \{ \langle b, y \rangle + \sigma\tau \mid \kappa^\circ(A^*y) \leq 1, \ \rho^\circ(y) \leq -\tau \}$$

$$= \left\{ (y, -\rho^\circ(y)) \ \middle| \ y \in \operatorname*{argmax}_y \{ \langle b, y \rangle - \sigma\rho^\circ(y) \mid \kappa^\circ(A^*y) \leq 1 \} \right\},$$

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 5.5.

PROPOSITION 6.1. *For fixed (b, σ), define v as in (6.1) and f as in (6.2). Then*

$$\operatorname{dom} f(\cdot, 0, 0) \neq \emptyset \quad \Longleftrightarrow \quad 0 \in A \operatorname{dom} \kappa - [\rho(b - \cdot) \leq \sigma],$$

and hence

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

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

$$\partial v(0,0) = \left\{ (y, -\rho^\circ(y)) \ \middle| \ y \in \operatorname*{argmax}_y \{ \langle b, y \rangle - \sigma\rho^\circ(y) \mid \kappa^\circ(A^*y) \leq 1 \} \right\}$$

$$= \left\{ v(0,0) \cdot (y, -\rho^\circ(y)) \ \middle| \ y \in \operatorname*{argmin}_y \{ \kappa^\circ(A^*y) \mid \langle b, y \rangle - \sigma\rho^\circ(y) \geq 1 \} \right\}.$$

Proof. It is routine to verify the properties of the domain of $f(\cdot, 0, 0)$ and the interior of the domain of v. Suppose that $(0,0) \in \operatorname{int} \operatorname{dom} v$. Then the value function is continuous at $(0,0)$ and hence $\partial v(0,0) \neq \emptyset$. The first expression of $\partial v(0,0)$ follows directly from Rockafellar and Wets (1998, Theorem 11.39(b)) and the discussions preceding this proposition.

We next derive the second expression of $\partial v(0,0)$. Since $(0,0) \in \operatorname{int} \operatorname{dom} v$ implies $0 \in \operatorname{int}(A \operatorname{dom} \kappa - [\rho(b - \cdot) < \sigma])$, the linear map A satisfies assumption (4.3). Moreover, as another consequence of Rockafellar and Wets (1998, Theorem 11.39(a)), $(0,0) \in \operatorname{int} \operatorname{dom} v$ also implies that $v(0,0) = \sup_{y,\tau}\{-f^*(0,y,\tau)\}$, which is just the optimal value of the Lagrange dual problem (D_ℓ). Furthermore, $v(0,0)$ being finite and nonzero together with the definition of (D_ℓ) and (4.4) implies that $\operatorname{dom} \kappa^\circ \cap \mathcal{C}' \neq \emptyset$. The second expression of $\partial v(0,0)$ now follows from these three observations and Corollary 5.5. $\quad \square$

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 formulation (P_ρ) is to allow the gauge objective to be composed with a linear map, and for the addition of conic side constraints. The composite objective can be used to capture, for example, problems such as weighted basis pursuit (e.g., Candés, Wakin, and Boyd (2008); Friedlander et al. (2012)), or together with the conic constraint, problems such

as nonnegative total variation (Krishnan, Lin, and Yip, 2007). The following result generalizes the canonical primal-dual gauge pair (P_ρ) and (D_ρ).

PROPOSITION 7.1. *Let D be a linear map and let $\mathcal{K}$ be a convex cone. The following pair of problems constitute a primal-dual gauge pair:*

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

$$(7.1b) \qquad \underset{y,\,z}{\text{minimize}} \quad \kappa^\circ(z) \quad \text{subject to} \quad \langle y, b \rangle - \sigma\rho^\circ(y) \geq 1, \ D^*z - A^*y \in \mathcal{K}^*.$$

Proof. Reformulate (7.1a) as a gauge optimization problem by introducing additional variables, and lifting both the cone $\mathcal{K}$ and the epigraph epi ρ into the objective by means of their indicator functions: use the function $f(x, s, r, \tau) := \delta_{\mathcal{K}}(x) + \kappa(s) + \delta_{\text{epi}\,\rho}(r, \tau)$ 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 section 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:

$$\begin{aligned} f^\circ(u, z, y, \alpha) &= \max\left\{ \delta_{\mathcal{K}}^\circ(u), \ \kappa^\circ(z), \ \delta_{\text{epi}\,\rho}^\circ(y, \alpha) \right\} \\ &\overset{\text{(i)}}{=} \max\left\{ \delta_{\mathcal{K}^\circ}(u), \ \kappa^\circ(z), \ \delta_{(\text{epi}\,\rho)^\circ}(y, \alpha) \right\} \\ &\overset{\text{(ii)}}{=} \max\left\{ \delta_{\mathcal{K}^*}(-u), \ \kappa^\circ(z), \ \delta_{\text{epi}(\rho^\circ)}(y, -\alpha) \right\} \\ &\overset{\text{(iii)}}{=} \delta_{\mathcal{K}^*}(-u) + \kappa^\circ(z) + \delta_{\text{epi}(\rho^\circ)}(y, -\alpha), \end{aligned}$$

where (i) follows from Proposition 2.3(i), (ii) follows from Proposition 2.1(vi), and (iii) follows from the definition of indicator function. Moreover, use Corollary 4.2 to derive the antipolar of the linear constraint set

$$\mathcal{C} = \left\{ (x, s, r, \tau) \mid Dx = s, \ Ax + r = b, \ \tau = \sigma \right\},$$

which is given by

$$\mathcal{C}' = \left\{ (-D^*z + A^*y, z, y, \alpha) \mid \langle b, y \rangle + \sigma\alpha \geq 1 \right\}.$$

From the above discussion, we obtain the following gauge program:

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

Bringing the indicator functions down to the constraints leads to

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

further simplification by eliminating α yields the gauge dual problem (7.1b). □

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. Example 1.3 describes how a general conic optimization problem can be reformulated as an equivalent gauge problem; see (1.6).

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

$$(7.2) \qquad \underset{x}{\text{minimize}} \quad \langle c, x \rangle + \delta_{\mathcal{K}}(x) \quad \text{subject to} \quad \rho(b - Ax) \leq \sigma,$$

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

PROPOSITION 7.2. *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

$$(7.3) \qquad \kappa^{\circ}(u) = \inf \{ \alpha \geq 0 \mid \alpha c \in \mathcal{K}^* + u \},$$

with $\operatorname{dom} \kappa^{\circ} = \operatorname{span}\{c\} - \mathcal{K}^*$. 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. From Proposition 2.1, we have that

$$(7.4) \qquad \begin{aligned} \kappa^{\circ}(u) &= \sup \{ \langle u, x \rangle \mid \kappa(x) \leq 1 \} \\ &= \sup \{ \langle u, x \rangle \mid \langle c, x \rangle \leq 1 \text{ and } x \in \mathcal{K} \} \\ &= \inf \{ \alpha \geq 0 \mid \alpha c - u \in \mathcal{K}^* \}, \end{aligned}$$

where the strong (Lagrangian) duality relationship in the last equality stems from the following argument. First consider the case where $u \in \operatorname{dom} \kappa^{\circ}$. Because the maximization problem in (7.4) satisfies Slater's condition, equality follows from Rockafellar (1970, Corollary 28.2.2 and Theorem 28.4). Next, consider the case where $u \notin \operatorname{dom} \kappa^{\circ}$, where $\kappa^{\circ}(u) = \infty$. The last equality then follows from weak duality. For the domain, note that the minimization problem is feasible if and only if $u \in \operatorname{span}\{c\} - \mathcal{K}^*$; hence $\operatorname{dom} \kappa^{\circ} = \operatorname{span}\{c\} - \mathcal{K}^*$.

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

Remark 7.1. Note that even though the polar gauge in (7.3) is closed, it is not necessarily the case that it has a closed domain. For example, let $\mathcal{K}$ be the cone of PSD 2-by-2 matrices, and define

$$c = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad u_n = \begin{pmatrix} 0 & 1 \\ 1 & -\frac{1}{n} \end{pmatrix}$$

for each $n = 1, 2, \ldots$. Use the expression (7.3) to obtain that $\kappa^{\circ}(u_n) = n$. Hence, $u_n \in \operatorname{dom} \kappa^{\circ}$, but $\lim_{n \to \infty} u_n \notin \operatorname{dom} \kappa^{\circ}$.

The reader will find that the example is a more general result described by Ramana, Tunçel, and Wolkowicz (1997, Lemma 2.2), which shows that the cone of PSD matrices is *devious* (i.e., for every nontrivial proper face F of $\mathcal{K}$, span $F + \mathcal{K}$ is not closed). The concept of a devious cone seems to be intimately related to the closedness of the domain of polar gauges such as (7.3) because $\operatorname{span}\{c\} - \mathcal{K}^* = -(\operatorname{span} F + \mathcal{K}^*)$, where $F \subseteq \mathcal{K}^*$ is the smallest face of $\mathcal{K}^*$ that contains c; see (Tunçel and Wolkowicz, 2012, Proposition 3.2).

With that in mind, it is interesting to derive a representation for the closure of the domain of (7.3). It follows from Rockafellar (1970, Corollary 16.4.2) that $\operatorname{cl} \operatorname{dom} \kappa^{\circ} = \operatorname{cl} \left(\operatorname{span}\{c\} - \mathcal{K}^* \right) = \left(\{c\}^{\perp} \cap \operatorname{cl} \mathcal{K} \right)^{\circ}.$

7.2.1. Semidefinite conic gauge optimization. Here we give a concrete example of how to derive the gauge dual of a semidefinite conic gauge optimization problem. Consider the feasible semidefinite program

$$(7.5) \qquad \underset{X}{\text{minimize}} \quad \langle C, X \rangle \quad \text{subject to} \quad \mathcal{A}X = b, \ X \succeq 0,$$

where $C \succeq 0$, 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 7.2, with $\mathcal{K}$ equal to the (self-dual) PSD cone, gives the gauge polar

$$\kappa^\circ(U) = \inf \{ \, \alpha \geq 0 \mid \alpha C \succeq U \, \}.$$

The gauge dual (D_ρ) then specializes to

$$\underset{\alpha, y}{\text{minimize}} \quad \alpha \quad \text{subject to} \quad \alpha \geq 0, \ \langle b, y \rangle \geq 1, \ \alpha C \succeq \mathcal{A}^* y,$$

which is valid for all $C \succeq 0$. This dual can be simplified by noting that

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

where $\lambda_{\max}(U, C)$ is the largest generalized eigenvalue corresponding to the eigenvalue problem $Ux = \lambda Cx$ (which might be $+\infty$ as in the example given in Remark 7.1). Together with Corollary 4.2, which gives the antipolar of $\mathcal{C}$, and Theorem 5.1, which asserts that the optimal dual value is positive, the gauge dual problem can then be written as

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

The lifted formulation of phase retrieval (Candès, Strohmer, and Voroninski, 2012) is an example of the conic gauge optimization problem (7.5) with $C = I$. In that case, (7.5) is the problem of minimizing the trace of a PSD matrix that satisfies a set of linear constraints. The gauge dual problem above is simply minimizing the maximum eigenvalue of $\mathcal{A}^* y$ over 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, which may be useful for providing new avenues of exploration for modeling and algorithm development. 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, a 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 if 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 section 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." In this regard, if we define the value function as

$$\tilde{v}(b,\sigma) = \inf_x \left\{ \kappa(x) + \delta_{\mathrm{epi}\,\rho}(b - Ax, \sigma) \right\},$$

then $\tilde{v}$ is a gauge. It is conceivable that sensitivity analysis could be carried out based on studying its polar, given by

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

where $\mathcal{D} = \left\{ (y,\tau) \mid \kappa^\circ(A^*y) \leq 1, \rho^\circ(y) \leq -\tau \right\}$. This formula follows from Proposition 2.1(iv) and a computation of $\tilde{v}^*$ similar to the one leading to (6.3). 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. V. BURKE, AND M. P. FRIEDLANDER (2013), *Variational properties of value functions*, SIAM J. Optim., SIAM J. Optim. 23, pp. 1689–1717.

A. AUSLENDER AND M. TEBOULLE (2003), *Asymptotic Cones and Functions in Optimization and Variational Inequalities*, Springer Monogr. Math., Springer-Verlag, New York.

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

E. VAN DEN BERG (2009), *Convex Optimization for Generalized Sparse Recovery*, Ph.D. thesis, University of British Columbia, Vancouver, Canada.

E. VAN DEN BERG AND M. P. FRIEDLANDER (2011), *Sparse optimization with least-squares constraints*, SIAM J. Optim., 21, pp. 1201–1229.

A. BERMAN (1973), *Cones, Matrices and Mathematical Programming*, Lecture Notes in Econom. and Math. Systems 79, Springer-Verlag, Berlin, New York.

E. J. CANDÉS, M. B. WAKIN, AND S. P. BOYD (2008), *Enhancing sparsity by reweighted L1 minimization*, J. Fourier Anal. Appl., 14, pp. 877–905.

E. J. CANDÈS, T. STROHMER, AND V. VORONINSKI (2013), *Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming*, Comm. Pure Appl. Math., 66, pp. 1241–1274.

V. CHANDRASEKARAN, B. RECHT, P. A. PARRILO, AND A. S. WILLSKY (2012), *The convex geometry of linear inverse problems*, Found. Comput. Math., 12, pp. 805–849.

R. M. FREUND (1987), *Dual gauge programs, with applications to quadratic programming and the minimum-norm problem*, Math. Programming, 38, pp. 47–67.

M. P. FRIEDLANDER, H. MANSOUR, R. SAAB, AND O. YILMAZ (2012), *Recovering compressively sampled signals using partial support information*, IEEE Trans. Inform. Theory, 58, pp. 1122–1134.

M. X. GOEMANS AND D. P. WILLIAMSON (1995), *Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming*, J. Assoc. Comput. Mach., 42, pp. 1115–1145.

M. JAGGI (2013), *Revisiting Frank-Wolfe: Projection-free sparse convex optimization*, in Proceedings of the 30th International Conference on Machine Learning (ICML-13), pp. 427–435.

D. KRISHNAN, P. LIN, AND A. M. YIP (2007), *A primal-dual active-set method for non-negativity constrained total variation deblurring problems*, IEEE Trans. Image Process., 16, pp. 2766–2777.

L. LOVÀSZ (1983), *Submodular functions and convexity*, in Mathematical Programming: The State of the Art, A. Bachem, B. Korte, and M. Grötschel, eds., Springer, Berlin, pp. 235–257.

L. MCLINDEN (1978), *Symmetric duality for structured convex programs*, Trans. Amer. Math. Soc., 245, pp. 147–181.

M. V. RAMANA, L. TUNÇEL, AND H. WOLKOWICZ (1997), *Strong duality for semidefinite programming*, SIAM J. Optim., 7, pp. 641–662.

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

R. T. Rockafellar and R. J.-B. Wets (1998), *Variational Analysis*, Springer-Verlag, Berlin.

P. H. M. Ruys and H. N. Weddepohl (1979), *Economic Theory and Duality*, in Convex Analysis and Mathematical Economics, J. Kriens, ed., Springer, Berlin, New York, pp. 1–72.

T. Teuber, G. Steidl, and R. H. Chan (2013), *Minimization and parameter estimation for seminorm regularization models with* I-*divergence constraints*, Inverse Problems, 29, 035007.

L. Tunçel and H. Wolkowicz (2012), *Strong duality and minimal representations for cone optimization*, Comput. Optim. Appl., 53, pp. 619–648.