**Duality of optimization problems with gauge functions**

Shota Yamanaka a,b and Nobuo Yamashita a

aGraduate School of Informatics, Kyoto University, Kyoto, Japan; bFrontier Research Center, Toyota Motor Corporation, Toyota, Japan

**ABSTRACT**

Recently, Yamanaka and Yamashita proposed the so-called positively homogeneous optimization problem, which includes many important problems, such as the absolute-value and the gauge optimization problems. They presented a closed form of the dual formulation for the problem, and showed weak duality and the equivalence to the Lagrangian dual under some conditions. In this work, we focus on a special positively homogeneous optimization problem, whose objective function and constraints consist of some gauge and linear functions. We prove not only weak duality but also strong duality. We also study necessary and sufficient optimality conditions associated to the problem. Moreover, we give sufficient conditions under which we can recover a primal solution from a Karush-Kuhn-Tucker point of the dual formulation. Finally, we discuss how to extend the above results to general optimization problems by considering the so-called perspective functions.

**ARTICLE HISTORY**

Received 3 August 2020
Accepted 5 October 2022

**KEYWORDS**

Gauge optimization; duality theory; convex optimization; positively homogeneous functions

1. Introduction

The **gauge optimization** (GO) problem is described as follows [1–4]:

\[
\min_{x \in \mathcal{X}} \ g(x),
\]

where \(\mathcal{X} \subseteq \mathbb{R}^n\) is a closed convex set and \(g: \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}\) is a gauge function. Here, we say that \(g\) is a gauge function if \(g\) is convex, nonnegative, positively homogeneous and satisfies \(0 \in \text{dom} \ g\). Note that GO problems are convex. Freund [2] first introduced \((P_{GO})\), proposed a dual formulation called the gauge dual (which differs from the usual Lagrangian dual), and proved some duality results. He also showed that the class of gauge optimization problems includes the well-known linear programming with positive optimal value, \(p\)-norm optimization problems with \(p \in [1, \infty]\) and convex quadratic optimization problems [2].

Recently, Friedlander et al. [4] considered a specific form of GO problem in which \(\mathcal{X}\) is described as \(\mathcal{X} := \{x \in \mathbb{R}^n \mid h(b - Ax) \leq \sigma\}\), where \(h\) is a gauge
function, \( \sigma \) is a scalar, \( b \in \mathbb{R}^m \) and \( A \in \mathbb{R}^{m \times n} \). They gave a closed form of its gauge dual. Afterwards, Friedlander and Macêdo [3] applied this gauge duality to solve low-rank spectral optimization problems. Aravkin et al. [1] presented some theoretical results for the GO problem. In particular, they gave optimality conditions, and a way to recover a primal solution from the gauge dual. In that paper, they also extended their results to a more general convex optimization problem, where \( g \) and \( h \) were not necessarily gauge functions. In addition, they proposed the perspective duality, which is an extension of the gauge duality.

The gauge optimization problems in these previous works [1–4] do not involve linear terms in their objective functions. Therefore, these GO frameworks cannot directly handle linear conic optimization problems. More recently, Yamanaka and Yamashita [5] considered the following positively homogeneous optimization (PHO) problem:

\[
\begin{align*}
\min & \quad c^T x + d^T \Psi(x) \\
\text{s.t.} & \quad A x + B \Psi(x) = b, \\
& \quad H x + K \Psi(x) \leq p, \\
& \quad x \in \text{dom} \Psi,
\end{align*}
\]

where \( c \in \mathbb{R}^n, d \in \mathbb{R}^m, b \in \mathbb{R}^k, p \in \mathbb{R}^\ell, A \in \mathbb{R}^{k \times n}, B \in \mathbb{R}^{k \times m}, H \in \mathbb{R}^{\ell \times n} \) and \( K \in \mathbb{R}^{\ell \times m} \) are given constant vectors and matrices, \( \Psi : \mathbb{R}^n \to (\mathbb{R} \cup \{ \infty \})^m \) is defined by \( \Psi(\cdot) := (\psi_1(\cdot), \ldots, \psi_m(\cdot))^T \) where each function \( \psi_i : \mathbb{R}^{n_i} \to \mathbb{R} \cup \{ \infty \} \), \( \sum_{i=1}^m n_i = n \) is nonnegative, positively homogeneous and lower semi-continuous on \( \mathbb{R}^{n_i} \), and \( T \) denotes transpose. Here we assume \( m \leq n \). Moreover, \( \text{dom} \Psi \) denotes the effective domain of \( \Psi \), defined by \( \text{dom} \Psi := \{ x \in \mathbb{R}^n \mid \psi_i(x_i) < \infty, i = 1, \ldots, m \} \) where \( x_i \in \mathbb{R}^{n_i}, i = 1, \ldots, m \) are disjoint subvectors of \( x \). Problem (P_{PHO}) is not necessarily convex, and it includes (P_{GO}) with \( \mathcal{X} = \{ x \in \mathbb{R}^n \mid h(b - A x) \leq \sigma \} \) since gauge functions are nonnegative and positively homogeneous. Note that PHO can handle linear terms in its objective and constraint functions. Here, we explicitly include \( x \in \text{dom} \Psi \) in the constraints of (P_{PHO}). This is because we want to consider more general PHO problems than the ones used in the previous work [5], where \( \text{dom} \Psi = \mathbb{R}^n \) is assumed. For instance, an indicator function of a cone is nonnegative and positively homogeneous, then we can consider such function as \( \psi_i \). We will later show that the same results as in [5] can be obtained even when \( \text{dom} \Psi \neq \mathbb{R}^n \).

When \( n_i = 1 \) and \( \psi_i(x_i) = |x_i| \), (P_{PHO}) is reduced to the absolute value programming problem proposed by Mangasarian [6]. The other examples of PHO problems are \( p \)-order cone optimization problems [7, 8] with \( p \in (0, \infty) \), group Lasso-type problems [9, 10], and sum of norms optimization problems [11].

Yamanaka and Yamashita [5] proposed a closed-form dual formulation of the PHO, which they call the positively homogeneous dual, and showed that weak duality holds. They also investigated the relation between the positively homogeneous dual and the Lagrangian dual of (P_{PHO}), and proved that those problems are equivalent under some conditions. The result indicates that the Lagrangian...
dual of a PHO problem can be written in closed form even if it is nonconvex. Although the PHO problem has the above nice features, the theoretical analysis is still insufficient. In particular, the paper [5] does not discuss strong duality and primal recovery.

In this paper, we mainly study the following gauge optimization problem with possible linear functions:

\[
\begin{align*}
\min & \quad c^T x + d^T G(x) \\
\text{s.t.} & \quad Ax = b, \\
& \quad Hx + K G(x) \leq p, \\
& \quad x \in \text{dom } G,
\end{align*}
\]

where \( c, d, b, p, A, H, K \) are the same as in \((\text{P}_{\text{PHO}})\), \( G: \mathbb{R}^n \to (\mathbb{R} \cup \{\infty\})^m \) is defined by \( G(\cdot) := (g_1(\cdot), \ldots, g_m(\cdot))^T \) with \( g_i: \mathbb{R}^n \to \mathbb{R} \cup \{\infty\} \) as a gauge function for all \( i \), and each function \( g_i \) of \( G \) is lower semi-continuous on \( \mathbb{R}^n \). Note that there is no nonlinear term in the equality constraints, and problem \( (P) \) is convex when all elements of \( d \) and \( K \) are nonnegative. Problem \( (P) \) includes the convex GO problems considered in [1–4], and it is possible to explicitly handle linear terms. In this paper, we call \( (P) \) the gauge optimization problem when it is clear from the context.

In particular, we are interested in theoretical properties of problem \( (P) \) and its dual. We first define a dual problem of \( (P) \) as in [5], and then, give conditions under which weak and strong dualities hold for problem \( (P) \) and its dual. Moreover, we present necessary and sufficient optimality conditions for \( (P) \), that does not use differentials of \( g_i \) as in the Karush-Kuhn-Tucker (KKT) conditions. We further give sufficient conditions under which we can obtain a primal solution from a KKT point of the dual formulation. Finally, we show that the theoretical results for problem \( (P) \) can be extended to general optimization problems, by considering the so-called perspective functions.

The paper is organized as follows. In Section 2, we recall some important properties of \( (\text{P}_{\text{PHO}}) \) in [5]. We show that some of them hold even if \( \text{dom } \Psi \neq \mathbb{R}^n \). Section 3 presents the dual of problem \( (P) \), and gives some relations of \( (P) \) and its dual. In particular, we show weak and strong duality results, the optimality conditions for the problem, as well as the recovery of primal solutions by solving the dual problem. In Section 4, we discuss how to extend the obtained results to general optimization problems. Section 5 concludes the paper with final remarks and future works.

We use the following notations throughout the paper. We denote by \( \mathbb{R}_{++} \) the set of positive real numbers. Let \( x \in \mathbb{R}^n \) be an \( n \)-dimensional column vector, and \( A \in \mathbb{R}^{n \times m} \) be a matrix with dimension \( n \times m \). For two vectors \( x \) and \( y \), we denote the vector \((x^T, y^T)^T \) as \((x, y)^T \) for simplicity. For a vector \( x \in \mathbb{R}^n \), its \( i \)-th entry is denoted by \( x_i \). Moreover, if \( I \subseteq \{1, \ldots, n\} \), then \( x_I \) corresponds to the subvector of \( x \) with entries \( x_i, i \in I \). The \( n \)-dimensional vector of ones is given by \( e_n \), that is, \( e_n := (1, \ldots, 1)^T \in \mathbb{R}^n \). The identity matrix with dimension
is $E_n \in \mathbb{R}^{n \times n}$. For a matrix $A$, we write $A \succeq 0$ to denote $A$ is symmetric and positive semidefinite. The notation $\# J$ denotes the number of elements of a set $J$. A set $S$ is convex if $\alpha x + (1 - \alpha) y \in S$ holds for any $x, y \in S$ and $\alpha \in [0, 1]$. The convex hull of a set $S$ is denoted by $\text{co} S$. The effective domain of a function $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is defined by $\text{dom} f := \{x \mid f(x) < \infty\}$. A function $f$ is convex if $\text{dom} f$ is convex and the following inequality holds: $f(\alpha x + (1 - \alpha) y) \leq \alpha f(x) + (1 - \alpha) f(y)$ for any $x, y \in \mathbb{R}^n$ and $\alpha \in [0, 1]$. For a function $f$ and vectors $x$ and $y$, the subdifferential of $f(x, y)$ with respect to $x$ at $x_0$ is defined as

$$\partial_x f(x_0, y) := \{z \mid f(x, y) - f(x_0, y) \geq z^T (x - x_0) \text{ for all } x\}.$$ 

A function $f$ is lower semi-continuous if $f(x) \geq \liminf_{k \to \infty} f(x_k)$ for any sequence $\{x_k\} \subseteq \mathbb{R}^n$. We denote by $\| \cdot \|$ a norm function. Finally, $\delta_S : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is an indicator function of a set $S \subseteq \mathbb{R}^n$ defined by

$$\delta_S(x) := \begin{cases} 0 & \text{if } x \in S, \\ \infty & \text{otherwise}. \end{cases}$$

### 2. Positively homogeneous optimization problems and their duality

In this section, we recall positively homogeneous optimization problems and their properties in [5]. The positively homogeneous and vector positively homogeneous functions are defined respectively as follows.

**Definition 2.1 (Positively homogeneous functions):** A function $\psi : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is positively homogeneous if $\psi(\lambda x) = \lambda \psi(x)$ for all $x \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}^+$. 

**Definition 2.2 (Vector positively homogeneous functions):** A mapping $\Psi : \mathbb{R}^n \to (\mathbb{R} \cup \{\infty\})^m$ is a vector positively homogeneous function if it is defined as

$$\Psi(x) := \begin{bmatrix} \psi_1(x_{I_1}) \\ \vdots \\ \psi_m(x_{I_m}) \end{bmatrix}$$

with positively homogeneous functions $\psi_i : \mathbb{R}^{n_i} \to \mathbb{R} \cup \{\infty\}$, $i = 1, \ldots, m$, where $n = n_1 + \cdots + n_m$, $I_i \subseteq \{1, \ldots, n\}$ is a set of indices satisfying $I_i \cap I_j = \emptyset$ for all $i \neq j$, and $\# I_i = n_i$.

The polars $\psi^\circ$ and $\Psi^\circ$ of a positively homogeneous function $\psi$ and a vector positively homogeneous function $\Psi$ are defined as follows.

**Definition 2.3 (Polar of positively homogeneous functions):** Let $\psi : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ be a positively homogeneous function. Then, $\psi^\circ : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ defined by

$$\psi^\circ(y) := \sup\{x^T y \mid \psi(x) \leq 1\}$$

is called the polar of $\psi$. 


Definition 2.4 (Polar of vector positively homogeneous functions): Let \( \Psi : \mathbb{R}^n \to (\mathbb{R} \cup \{\infty\})^m \) be a vector positively homogeneous function. A function \( \Psi^\circ : \mathbb{R}^n \to (\mathbb{R} \cup \{\infty\})^m \) is the polar of a vector positively homogeneous function \( \Psi \) if \( \Psi^\circ \) is given as

\[
\Psi^\circ(y) = \begin{bmatrix}
\psi_1^\circ(y_{I_1}) \\
\vdots \\
\psi_m^\circ(y_{I_m})
\end{bmatrix}
\]

with the polar \( \psi_i^\circ : \mathbb{R}^{n_i} \to \mathbb{R} \cup \{\infty\} \) of positively homogeneous function \( \psi_i, i = 1, \ldots, m \).

Note that the paper [5] calls such polar functions dual functions. Note also that the polar of a positively homogeneous function is positively homogeneous and convex. Moreover, when \( \psi \) is a norm, \( \psi^\circ \) is the dual norm of \( \psi \).

Yamanaka and Yamashita [5] proposed the following dual of (P\( \text{PHO} \)):

\[
\begin{align*}
\max & \quad b^T u - p^T v \\
\text{s.t.} & \quad \Psi^\circ(A^T u - H^T v - c) + B^T u - K^T v \leq d, \\
& \quad v \geq 0,
\end{align*}
\]

where \((u, v) \in \mathbb{R}^k \times \mathbb{R}^\ell\). Note that if \((u, v)\) is feasible for (D\( \text{PHO} \)), then \(A^T u - H^T v - c \in \text{dom} \Psi^\circ\). Yamanaka and Yamashita [5] require the following assumption to show the weak duality of the PHO problems (P\( \text{PHO} \)) and (D\( \text{PHO} \)).

**Assumption 2.1:** Each positively homogeneous function \( \psi_i \) in \( \Psi \) is nonnegative, that is, \( \psi_i(x_{I_i}) \geq 0 \) for all \( x_{I_i} \in \mathbb{R}^{n_i} \).

Note that they assume another condition in [5, Assumption 2.1], but we do not use the condition and consider weaker assumption, which is Assumption 2.3, to prove the weak duality. We also note that they also need the condition that \( \text{dom} \Psi = \mathbb{R}^n \) for the weak duality, which is rather restrictive for practical use. In this paper, we will show the weak duality without the condition, that is, we assume \( \text{dom} \Psi \) may be different from \( \mathbb{R}^n \). More precisely, we assume the domain of each function \( \psi_i \) is cone.

**Assumption 2.2:** The domain of each function \( \psi_i \) in \( \Psi \) is cone, that is, \( cx \in \text{dom} \psi_i \) for every \( c \geq 0 \) and \( x \in \text{dom} \psi_i \).

To prove the weak duality, we first show the following lemma under Assumptions 2.1 and 2.2.
Lemma 2.1: Suppose that Assumptions 2.1 and 2.2 hold. Let \( \Psi \) and \( \Psi^\circ \) be a vector positively homogeneous function and its polar, respectively. Then, we have

\[
\begin{align*}
\Psi^\circ(y) & \geq 0, \\
\Psi(x)^T \Psi^\circ(y) & \geq x^T y
\end{align*}
\]

for all \( x \in \text{dom}\Psi \) and \( y \in \text{dom}\Psi^\circ \).

Proof: Since the first inequality can be shown as in the proof of [5, Proposition 2.1] by using Definitions 2.1, 2.3 and 2.4, we prove only the second inequality. Clearly, it is enough to show that \( \psi_i(x_{I_i}) \psi_i^\circ(y_{I_i}) \geq x_{I_i}^T y_{I_i} \). For simplicity, we denote \( \psi_i \) and \( x_{I_i} \) as \( \psi \) and \( x \), respectively.

If \( \psi(x) = 0 \), then we can show that \( x^T y \leq 0 \) for all \( y \in \text{dom} \psi^\circ \) as follows. Suppose to the contrary that there exists \( y \in \text{dom} \psi^\circ \) such that \( x^T y > 0 \), and hence \( tx^T y \to \infty \) as \( t \to \infty \). Moreover, since \( \psi(tx) = t\psi(x) = 0 \) for all \( t > 0 \), we have \( \psi^\circ(y) \geq \sup\{tx^T y \mid \psi(tx) \leq 1\} = \infty \), which contradicts the fact that \( y \in \text{dom} \psi^\circ \). Consequently, we obtain \( \psi(x) \psi^\circ(y) = 0 \geq x^T y \).

Next we consider the case where \( \psi(x) > 0 \). Let \( z = x/\psi(x) \). Since \( \psi \) is positively homogeneous, we obtain

\[
\psi(z) = \psi\left(\frac{x}{\psi(x)}\right) = \frac{1}{\psi(x)} \psi(x) = 1.
\]

Therefore, we have

\[
\psi^\circ(y) = \sup\{\xi^T y \mid \psi(\xi) \leq 1\} \geq z^T y = \frac{1}{\psi(x)} x^T y,
\]

which shows the second inequality. ■

For problems (P_{PHO}) and (D_{PHO}), the following weak duality holds.

Theorem 2.1 (Weak duality): Suppose that Assumptions 2.1 and 2.2 hold. Let \( x \in \mathbb{R}^n \) and \( (u, v) \in \mathbb{R}^k \times \mathbb{R}^\ell \) be feasible solutions of (P_{PHO}) and (D_{PHO}), respectively. Then, the following inequality holds:

\[
c^T x + d^T \Psi(x) \geq b^T u - p^T v.
\]

Proof: Using Lemma 2.1, we can show weak duality as in the proof of [5, Theorem 3.1]. ■

In the following, we show that the optimal values and solutions of problems (D_{PHO}) and the Lagrangian dual of (P_{PHO}) are the same. The Lagrangian
function $\mathcal{L} : \text{dom}\Psi \times \mathbb{R}^k \times \mathbb{R}^\ell \to \mathbb{R} \cup \{-\infty\}$ of $(P_{PHO})$ is defined as

$$
\mathcal{L}(x, u, v) := c^T x + d^T \Psi(x) + u^T (b - Ax - B\Psi(x)) + v^T (Hx + K\Psi(x) - p).
$$

Recall that $(P_{PHO})$ is equivalent to

$$
\inf_{x \in \text{dom}\Psi} \sup_{u \in \mathbb{R}^k, v \in \mathbb{R}^\ell} \mathcal{L}(x, u, v). 
$$

$(P_{PHO}^C)$

Then, the Lagrangian dual of problem $(P_{PHO}^C)$ is described as

$$
\sup_{u \in \mathbb{R}^k, v \in \mathbb{R}^\ell} \omega(u, v),
$$

$(D_{PHO}^C)$

where $\omega : \mathbb{R}^k \times \mathbb{R}^\ell \to \mathbb{R} \cup \{-\infty\}$ is defined by

$$
\omega(u, v) := \inf_{x \in \text{dom}\Psi} \mathcal{L}(x, u, v).
$$

Note that we explicitly require $x \in \text{dom}\Psi$ in the Lagrangian dual problem $(D_{PHO}^C)$.

Note that Yamanaka and Yamashita [5] further assumed a condition on positively homogeneous functions for the equivalence between $(D_{PHO})$ and $(D_{PHO}^C)$. The condition is that each component $\psi_i$ of $\Psi$ vanishes only at zero and $\text{dom}\psi_i = \mathbb{R}^n_i$. For example, a usual norm satisfies the condition, but neither an indicator function for a cone nor the function $\psi_i(x_{I_i}) = \max\{0, x_{I_i}\}$ satisfies it. Therefore, the condition is rather restrictive. Here, we suppose the following weaker assumption.

**Assumption 2.3:** For each $i$, one of the following conditions holds:

(a) $d_i \geq 0$, $B_{ji} = 0$ and $K_{ji} \geq 0$ for all $j$,
(b) $\text{dom}\psi_i = \mathbb{R}^{n_i}$ and there exists $\hat{x}_{I_i}$ such that $\psi_i(\hat{x}_{I_i}) \neq 0$.

Note that if problem $(P_{PHO})$ satisfies the first condition (a) of Assumption 2.3 for all $i$ and all $\psi_i$ are gauge functions, then it becomes a convex gauge optimization problem $(P)$.

We prove the following key lemma for the equivalence between $(D_{PHO})$ and $(D_{PHO}^C)$, which is an extension of [5, Lemma 4.1] to the case where $\text{dom}\psi_i \neq \mathbb{R}^{n_i}$.

**Lemma 2.2:** Let $\psi_i^*$ be the polar of positively homogeneous functions $\psi_i$ for $i = 1, \ldots, m$. Suppose that Assumptions 2.1, 2.2 and 2.3 hold. Assume also that $(\bar{u}, \bar{v})$ with $\bar{v} \geq 0$ is not a feasible solution of problem $(D_{PHO})$. Then $\omega(\bar{u}, \bar{v}) = -\infty$. 

**Proof:** Suppose that \((\bar{u}, \bar{v})\) with \(\bar{v} \geq 0\) is not a feasible solution of (D\(_{\text{PHO}}\)). Then, there exists an index \(j\) such that

\[
\psi^\circ_j(\alpha_{I_j}) > \beta_j, \quad (1)
\]

where \(\alpha := A^T \bar{u} - H^T \bar{v} - c \in \mathbb{R}^n\), and \(\beta := d - B^T \bar{u} + K^T \bar{v} \in \mathbb{R}^m\). Let \(\bar{x} := (0, \ldots, 0, \bar{x}_{I_j}, 0, \ldots, 0)\) with \(\bar{x}_{I_j} \in \text{dom} \psi_j \neq \emptyset\). Then we have \(\Psi(\bar{x}) = (0, \ldots, 0, \psi_j(\bar{x}_{I_j}), 0, \ldots, 0)\) and

\[
\mathcal{L}(\bar{x}, \bar{u}, \bar{v}) = -\alpha_j^T \bar{x}_{I_j} + \beta_j \psi_j(\bar{x}_{I_j}) + b^T \bar{u} + p^T \bar{v}. \quad (2)
\]

Now we consider three cases: \(\psi^\circ_j(\alpha_{I_j}) \in (0, \infty)\), \(\psi^\circ_j(\alpha_{I_j}) = \infty\), and \(\psi^\circ_j(\alpha_{I_j}) = 0\).

First we study the case where \(\psi^\circ_j(\alpha_{I_j}) \in (0, \infty)\). Recall that \(\psi^\circ_j(\alpha_{I_j})\) is defined as

\[
\psi^\circ_j(\alpha_{I_j}) = \sup_{x_{I_j}} \{x_{I_j}^T \alpha_{I_j} | \psi_j(x_{I_j}) \leq 1\}. \quad (3)
\]

Therefore, for all \(\varepsilon > 0\), there exists \(\bar{x}_{I_j}(\varepsilon)\) such that

\[
\psi^\circ_j(\alpha_{I_j}) - \varepsilon \leq \alpha_j^T \bar{x}_{I_j}(\varepsilon), \quad \psi_j(\bar{x}_{I_j}(\varepsilon)) \leq 1. \quad (4)
\]

Let \(\bar{\varepsilon}\) be a scalar such that \(\bar{\varepsilon} := \min\{\psi^\circ_j(\alpha_{I_j}) - \beta_j, \psi^\circ_j(\alpha_{I_j})\}/2 > 0\). Then \(\psi^\circ_j(\alpha_{I_j}) > \bar{\varepsilon} > 0\). Moreover, we show that there exists \(\tilde{x}_{I_j}\) such that

\[
\psi^\circ_j(\alpha_{I_j}) - \bar{\varepsilon} \leq \alpha_j^T \tilde{x}_{I_j}, \quad \psi_j(\tilde{x}_{I_j}) = 1. \quad (5)
\]

Since \(\psi^\circ_j(\alpha_{I_j}) > \bar{\varepsilon}\), the inequality (4) implies \(\alpha_j^T \bar{x}_{I_j}(\bar{\varepsilon}) > 0\), and hence \(\bar{x}_{I_j}(\bar{\varepsilon}) \neq 0\). If \(\psi_j(\bar{x}_{I_j}(\bar{\varepsilon})) \neq 0\), then we set \(\tilde{x}_{I_j} = \bar{x}_{I_j}(\bar{\varepsilon})/\psi_j(\bar{x}_{I_j}(\bar{\varepsilon}))\). This vector \(\tilde{x}_{I_j}\) satisfies conditions (5) as shown below:

\[
\psi^\circ_j(\alpha_{I_j}) - \bar{\varepsilon} \leq \alpha_j^T \bar{x}_{I_j}(\bar{\varepsilon}) \leq \alpha_j^T \frac{\tilde{x}_{I_j}(\bar{\varepsilon})}{\psi_j(\bar{x}_{I_j}(\bar{\varepsilon}))} = \alpha_j^T \tilde{x}_{I_j},
\]

\[
\psi_j(\tilde{x}_{I_j}) = \psi_j \left( \frac{\tilde{x}_{I_j}(\bar{\varepsilon})}{\psi_j(\tilde{x}_{I_j}(\bar{\varepsilon}))} \right) = \frac{1}{\psi_j(\tilde{x}_{I_j}(\bar{\varepsilon}))} \psi_j(\bar{x}_{I_j}(\bar{\varepsilon})) = 1,
\]

where the second inequality holds from Assumption 2.1 and (4). If \(\psi_j(\bar{x}_{I_j}(\bar{\varepsilon})) = 0\), then \(\psi_j(t \tilde{x}_{I_j}(\bar{\varepsilon})) = t \psi_j(\tilde{x}_{I_j}(\bar{\varepsilon})) = 0\) for all \(t > 0\) because \(\psi_j\) is positively homogeneous. From (3), we have \(\psi^\circ_j(\alpha_{I_j}) \geq t \tilde{x}_{I_j}(\bar{\varepsilon})^T \alpha_{I_j}\). Since \(\alpha_j^T \tilde{x}_{I_j}(\bar{\varepsilon}) > 0\), we obtain \(\psi^\circ_j(\alpha_{I_j}) \to \infty\) as \(t \to \infty\), which is a contradiction. Therefore, there exists \(\tilde{x}_{I_j}\) such that (5) holds.
We now denote \( \bar{t} = (0, \ldots, 0, t \bar{x}_j, 0, \ldots, 0) \) for \( t > 0 \). Then, we have from (2)

\[
\mathcal{L}(\bar{t}, \bar{u}, \bar{v}) = -t \alpha_j^T \bar{x}_j + \beta_j \psi_j(t \bar{x}_j) + b^T \bar{u} + p^T \bar{v} \\
\leq -t(\psi_j^\circ(\alpha_j) - \bar{\epsilon} - \beta_j \psi_j(\bar{x}_j)) + b^T \bar{u} + p^T \bar{v} \\
= -t(\psi_j^\circ(\alpha_j) - \bar{\epsilon} - \beta_j) + b^T \bar{u} + p^T \bar{v},
\]

where the second inequality and the third equality hold from (5). Since \( \bar{\epsilon} \leq (\psi_j^\circ(\alpha_j) - \beta_j)/2 \), we obtain

\[
\mathcal{L}(\bar{t}, \bar{u}, \bar{v}) \leq -t \left( \frac{\psi_j^\circ(\alpha_j) - \beta_j}{2} \right) + b^T \bar{u} + p^T \bar{v},
\]

which concludes \( \lim_{t \to \infty} \mathcal{L}(\bar{t}, \bar{u}, \bar{v}) = -\infty \).

Next we consider the case where \( \psi_j^\circ(\alpha_j) = \infty \). From (3), there exists a sequence \( \{\tilde{x}_j^k\} \subset \text{dom} \psi_j \) such that \( \psi_j(\tilde{x}_j^k) \leq 1 \) and \( (\tilde{x}_j^k)^T \alpha_j \to \infty \) as \( k \to \infty \). Let \( \tilde{x}_j = (0, \ldots, 0, \tilde{x}_j^k, 0, \ldots, 0) \). Then, it follows from (2) that

\[
\mathcal{L}(\tilde{x}_j, \bar{u}, \bar{v}) = -\alpha_j^T \tilde{x}_j + \beta_j \psi_j(\tilde{x}_j^k) + b^T \bar{u} + p^T \bar{v},
\]

and hence \( \lim_{k \to \infty} \mathcal{L}(\tilde{x}_j, \bar{u}, \bar{v}) = -\infty \).

We finally study the case where \( \psi_j^\circ(\alpha_j) = 0 \). Note that \( 0 > \beta_j \) from (1). When the first condition (a) of Assumption 2.3 holds, it then follows from \( \bar{v} \geq 0 \) that \( \beta_j = d_j - (B^T \bar{u})_j + (K^T \bar{v})_j \geq 0 \), which is a contradiction. Now, suppose that the second condition (b) of Assumption 2.3 holds. If \( \alpha_j \neq 0 \), we have \( \psi_j(\alpha_j) < +\infty \) because \( \alpha_j \in \text{dom} \psi_j = \mathbb{R}^n \). Then there exists \( \tilde{\epsilon} > 0 \) such that \( 1 \geq \psi_j(\tilde{\epsilon} \alpha_j) = \tilde{\epsilon} \psi_j(\alpha_j) \). Therefore we have

\[
\psi_j^\circ(\alpha_j) = \sup_{x_j} \{ x_j^T \alpha_j | \psi_j(x_j) \leq 1 \} \geq \tilde{\epsilon} \alpha_j^T \alpha_j > 0,
\]

which is a contradiction. Now we consider the case where \( \alpha_j = 0 \). From Assumption 2.3 (b), there exists \( \hat{x}_j \) such that \( \psi_j(\hat{x}_j) > 0 \). Let \( \hat{x}(t) = (0, \ldots, 0, t \hat{x}_j, 0, \ldots, 0) \) with \( t > 0 \). Then, it follows from (2) that

\[
\mathcal{L}(\hat{x}(t), \bar{u}, \bar{v}) = -\alpha_j^T \hat{x}_j(t) + \beta_j \psi_j(t \hat{x}_j) + b^T \bar{u} + p^T \bar{v} \\
= t \beta_j \psi_j(\hat{x}_j) + b^T \bar{u} + p^T \bar{v},
\]

and we conclude that \( \lim_{t \to \infty} \mathcal{L}(\hat{x}(t), \bar{u}, \bar{v}) = -\infty \).

Consequently, \( \omega(\bar{u}, \bar{v}) \) is unbounded from below.

The next theorem shows that problems \( (\text{DPHO}) \) and \( (\text{DPHO})^C \) are equivalent, which means that their optimal values and solutions of those problems are the same.
**Theorem 2.2:** Suppose that the Lagrangian dual problem \((\mathcal{D}^{\mathcal{L}}_{\text{PHO}})\) has a feasible solution. Suppose also that Assumptions 2.1–2.3 hold. Then, the optimal value and optimal solutions of problem \((\mathcal{D}^{\mathcal{L}}_{\text{PHO}})\) are the same as those of \((\mathcal{D}^{\mathcal{L}}_{\text{PHO}})\).

**Proof:** The result can be proved by using Lemma 2.2 as in the proof of [5, Theorem 4.1].

**Remark 2.1:** A referee suggested that Lemma 2.2 and Theorem 2.2 can be proved more simply by using the properties of a general Lagrangian duality, but here we describe the original proofs.

The next proposition is about the double dualization for the PHO problems which shows that the positively homogeneous dual of problem \((\mathcal{D}^{\text{PHO}})\) is similar to \((\mathcal{P}^{\text{PHO}})\).

**Proposition 2.1:** Suppose that problem \((\mathcal{D}^{\text{PHO}})\) is feasible. Then, the positively homogeneous dual of \((\mathcal{D}^{\text{PHO}})\) can be written as

\[
\begin{align*}
\min \quad & c^T x + d^T y \\
\text{s.t.} \quad & Ax + By = b, \quad \mathcal{P}^{\text{PHO}}
\end{align*}
\]

\[
Hx + Ky \leq p, \quad \Psi^\circ (x) \leq y,
\]

where \(\Psi^\circ\) denotes the polar of \(\Psi\), i.e. \(\Psi^\circ = (\Psi^\circ)^\circ\).

**Proof:** First, note that problem \((\mathcal{D}^{\text{PHO}})\) can be written as

\[
\begin{align*}
\min \quad & -b^T u + p^T v \\
\text{s.t.} \quad & \Psi^\circ (w) + B^T u - K^T v \leq d, \\
& w = A^T u - H^T v - c, \\
& -v \leq 0.
\end{align*}
\]

This problem is further reformulated as

\[
\begin{align*}
\min \quad & \hat{c}^T \theta \\
\text{s.t.} \quad & \hat{K} \hat{\Psi}^\circ (\theta) + \hat{H} \theta \leq \hat{p}, \\
& \hat{A} \theta = c,
\end{align*}
\]

where \(\theta = (u, v, w)^T \in \mathbb{R}^{k+\ell+n}\), \(\hat{c} = (-b, p, 0)^T \in \mathbb{R}^{k+\ell+n}\), \(\hat{p} = (d, 0)^T \in \mathbb{R}^{n+\ell}\), \(\hat{A} = (A^T , -H^T , -E_n)^T \in \mathbb{R}^{n+\ell}\), \(\hat{K} = \begin{bmatrix} 0 & 0 & E_n \\ 0 & 0 & 0 \end{bmatrix} \in \mathbb{R}^{(n+\ell) \times (k+\ell+n)}\), \(\hat{H} = \begin{bmatrix} B^T & -K^T & 0 \\ 0 & -E_\ell & 0 \end{bmatrix} \in \mathbb{R}^{(n+\ell) \times (k+\ell+n)}\),

and \(\hat{\Psi}^\circ\) is defined by \(\hat{\Psi}^\circ (\theta) := (\|u\|_2, \|v\|_2, \Psi^\circ (w))^T\). Note that \(\|u\|_2\) and \(\|v\|_2\) in \(\hat{\Psi}^\circ\) are dummy functions, and they do not affect the primal problem.
Moreover, the positively homogeneous dual of (6) can be described as

\[
\begin{align*}
\text{max} & \quad c^T x - \hat{p}^T y \\
\text{s.t.} & \quad \hat{\Psi}^{\circ\circ}(\hat{A}^T x - \hat{H}^T y - \hat{c}) - \hat{K}^T y \leq 0, \\
                 & \quad y \geq 0.
\end{align*}
\]

Let \( y = (y_1, y_2)^T \) with \( y_1 \in \mathbb{R}^n \) and \( y_2 \in \mathbb{R}^\ell \). Then, the above problem can be rewritten as

\[
\begin{align*}
\text{min} & \quad -c^T x + d^T y_1 \\
\text{s.t.} & \quad \|Ax - By_1 + b\|_2 \leq 0, \\
                 & \quad \|-Hx + Ky_1 + y_2 - p\|_2 \leq 0, \\
                 & \quad \hat{\Psi}^{\circ\circ}(-x) - y_1 \leq 0, \\
                 & \quad y \geq 0.
\end{align*}
\]

(7)

The first two inequality constraints are equivalent to

\[
\begin{align*}
-Ax + By_1 &= b, \\
-Hx + Ky_1 + y_2 &= p.
\end{align*}
\]

Since \( y_2 \geq 0 \) in (7), the second equality is further reduced to \(-Hx + Ky_1 \leq p\). Consequently, we can reformulate (7) as

\[
\begin{align*}
\text{min} & \quad -c^T x + d^T y_1 \\
\text{s.t.} & \quad -Ax + By_1 = b, \\
                 & \quad -Hx + Ky_1 \leq p, \\
                 & \quad \hat{\Psi}^{\circ\circ}(-x) \leq y_1,
\end{align*}
\]

which is precisely \((P'_{\text{PHO}})\) by denoting \(-x\) and \(y_1\) as \(x\) and \(y\), respectively. ■

In Lagrangian duality theory (see e.g. [12, p. 138]), it is well-known that the Lagrangian dual of \((D_{\text{PHO}}^C)\) is exactly the problem \((P_{\text{PHO}}^C)\). Therefore, we obtain the following result.

**Corollary 2.1:** Suppose that problem \((D_{\text{PHO}})\) is feasible. Suppose also that Assumptions 2.1–2.3 hold. Then, problem \((P'_{\text{PHO}})\) is equivalent to the original problem \((P_{\text{PHO}})\).

**Proof:** Recall that problem \((P_{\text{PHO}})\) is equivalent to \((P_{\text{PHO}}^C)\). Then, from Theorem 2.2, problem \((D_{\text{PHO}})\) is equivalent to \((D_{\text{PHO}}^C)\). Moreover, from Theorem 2.2, the positively homogeneous dual and the Lagrangian dual are equivalent under Assumptions 2.1 and 2.3. Therefore, the positively homogeneous dual of \((D_{\text{PHO}})\) is equivalent to the Lagrangian dual of \((D_{\text{PHO}}^C)\), which is \((P_{\text{PHO}}^C)\) and it is equivalent to \((P_{\text{PHO}})\). ■
3. Gauge optimization problems and their duality

In this section, we discuss the following gauge optimization problem:

$$\begin{align*}
\min & \quad c^T x + d^T G(x) \\
\text{s.t.} & \quad Ax = b, \\
& \quad Hx + KG(x) \leq p, \\
& \quad x \in \text{dom } G.
\end{align*}$$

(P)

We call $G$ a vector gauge function defined as $G := (g_1(\cdot), \ldots, g_m(\cdot))^T$ with $g_i : \mathbb{R}^{n_i} \to \mathbb{R} \cup \{+\infty\}$ as a gauge function for all $i$, and each function $g_i$ of $G$ is lower semi-continuous on $\mathbb{R}^{n_i}$.

Since (P) is a special case of (P PHO), the PHO dual of (P) is written as follows:

$$\begin{align*}
\max & \quad b^T u - p^T v \\
\text{s.t.} & \quad G^\circ (A^T u - H^T v - c) - K^T v \leq d, \\
& \quad v \geq 0,
\end{align*}$$

(D)

where $G^\circ$ is the polar function associated to $G$. Here, problem (D) is a convex optimization problem since each component $g_i^\circ$ of $G^\circ$ is convex.

The next proposition is a corollary of Lemma 2.1. Note that since a gauge function is positively homogeneous and nonnegative, Assumption 2.1 automatically holds.

**Proposition 3.1:** Let $G$ and $G^\circ$ be a vector gauge function and its polar, respectively. Suppose that Assumption 2.2 holds for each function $g_i$ of $G$. Then, we have

$$\begin{align*}
G^\circ(y) & \geq 0, \\
G(x)^T G^\circ(y) & \geq x^T y
\end{align*}$$

for any $x \in \text{dom } G$ and $y \in \text{dom } G^\circ$.

**Proof:** The proof follows from Lemma 2.1.

We have the weak duality theorem for problems (P) and (D), and the equivalence between (D) and the Lagrangian dual of (P) from Proposition 3.1 and Theorem 2.2. Throughout the paper, we denote the Lagrangian dual of (P) as $(D_L)$.

**Corollary 3.1 (Weak duality):** Suppose that Assumption 2.2 holds for each function $g_i$ of $G$. For problems (P) and (D), the following inequality holds:

$$c^T x + d^T G(x) \geq b^T u - p^T v$$

for all feasible points $x \in \mathbb{R}^n$ and $(u, v) \in \mathbb{R}^k \times \mathbb{R}^\ell$ of (P) and (D), respectively.
Proof: The proof directly follows from Proposition 3.1.

Corollary 3.2: Suppose that the Lagrangian dual problem (D_L) has a feasible solution. Suppose also that Assumptions 2.2 and 2.3 hold. Then, the optimal value and solutions of problem (D) are the same as (D_L).

Proof: The proof is a direct consequence of Theorem 2.2.

We now discuss the strong duality, necessary and sufficient optimality conditions, and the primal recovery for problem (P). To this end, we need (P) to be convex. Thus, from now on, we suppose the following assumption.

Assumption 3.1: All elements of d and K of problem (P) are nonnegative.

We now show that the dual of (D) becomes (P) under Assumption 3.1.

Corollary 3.3: Suppose that Assumptions 2.2 and 3.1 hold. Assume also that problem (D) is feasible. Then, the positively homogeneous dual of (D) is equivalent to (P).

Proof: Assumption 2.1 automatically holds because g_i is a gauge function. If Assumption 3.1 holds, then Assumption 2.3 holds for (P_{PHO}) with \Psi = G. Then we obtain the result from Corollary 2.1.

3.1. Strong duality

We focus on the strong duality between problems (P) and (D).

At first, we describe a theorem that we use to prove the strong duality. The theorem is shown in [12, Satz 5.19], and here we summarized only the parts that are relevant to the proof as follows.

Theorem 3.1 ([12, Satz 5.19]): Let f_p^* and f_d^* are the optimal values of problems (P_{PHO}^L) and (D_{PHO}^L), respectively. For problems (P_{PHO}) and (D_{PHO}), and the Lagrangian function L of (P_{PHO}), we assume the following conditions:

(A1) : \text{dom}\Psi is closed and convex,
(A2) : x \mapsto L(x, u, v) is for all (u, v) \in \mathbb{R}^k \times \mathbb{R}_+^\ell lower semi-continuous and convex on \text{dom}\Psi,
(A3) : (u, v) \mapsto -L(x, u, v) is for all x \in \text{dom}\Psi lower semi-continuous and convex on \mathbb{R}^k \times \mathbb{R}_+^\ell,
(A4) : There is x_0 \in \text{dom}\Psi such that L(x_0, u, v) \rightarrow -\infty for \| (u, v) \| \rightarrow +\infty, (u, v) \in \mathbb{R}^k \times \mathbb{R}_+^\ell.
We also assume \( f_d^* > -\infty \). Then problem \((D_{PHO}^L)\) has a solution and the strong duality holds, that is, \( f_d^* = f_p^* \).

We assume the following condition on functions \( g_i \) and prove strong duality for problems \((P)\) and \((D)\). As seen below, we require a certain constraint qualification for this purpose.

**Theorem 3.2 (Strong duality):** Suppose that Assumptions 2.2 and 3.1 hold. Suppose also that the Slater constraint qualification holds for \((P)\), that is, there exists a point \( x_0 \) satisfying \( x_0 \in \{ x \mid Ax = b, Hx + KG(x) < p, x \in \text{dom } G \} \). We also assume that \( \text{dom } G \) is closed and convex, and \((D)\) is bounded. Then, the strong duality holds for problems \((P)\) and \((D)\), i.e. if \((P)\) has an optimal solution \( x^* \), then \((D)\) also has an optimal solution \( (u^*, v^*) \) and the duality gap between \((P)\) and \((D)\) is zero, that is, \( c^T x^* + d^T G(x^*) = b^T u^* - p^T v^* \).

**Proof:** Item A1 of Theorem 3.1 and the condition \( f_d^* > -\infty \) hold from the assumption of the theorem. Item A2 of Theorem 3.1 clearly hold. Since the Lagrangian function of problem \((P)\) is described as

\[
\mathcal{L}(x, u, v) = c^T x + d^T G(x) + u^T (b - Ax) + v^T (Hx + KG(x) - p),
\]

item A3 of Theorem 3.1 is satisfied. We now denote \( x_0 \) as a point that satisfies Slater’s condition:

\[
x_0 \in \{ x \mid Ax = b, Hx + KG(x) < p, x \in \text{dom } G \}.
\]

Then, we have

\[
\mathcal{L}(x_0, u, v) = c^T x_0 + d^T G(x_0) + u^T (b - Ax_0) + v^T (Hx_0 + KG(x_0) - p) \to -\infty
\]

for \( v \to +\infty \), which implies that item A4 of Theorem 3.1 holds. Then, from Theorem 3.1, the optimal values of \((P)_{\mathcal{L}}\) and \((D)_{\mathcal{L}}\) are the same. Here, problems \((P)_{\mathcal{L}}\) and \((D)_{\mathcal{L}}\) are described as follows:

\[
\inf_{x \in \text{dom } G} \sup_{u \in \mathbb{R}_+, v \in \mathbb{R}_+^l} \mathcal{L}(x, u, v), \quad (P)_{\mathcal{L}}
\]

\[
\sup_{u \in \mathbb{R}_+, v \in \mathbb{R}_+^l} \inf_{x \in \text{dom } G} \mathcal{L}(x, u, v). \quad (D)_{\mathcal{L}}
\]

Note that problem \((P)_{\mathcal{L}}\) is equivalent to \((P)\). We also note that problem \((D)_{\mathcal{L}}\) is equivalent to \((D)\) from Corollary 3.2 because Assumption 3.1, which means Assumption 2.3, holds. Therefore, the optimal values of \((P)\) and \((D)\) are the same, which implies the strong duality holds between these problems. \(\square\)
Note that the constraint qualification is necessary for the strong duality. However, there exist a gauge optimization problem that holds the strong duality without the constraint qualification as seen below.

**Example 3.1:** Consider the following two dimensional gauge optimization problem:

$$\begin{align*}
\min & & e^T x \\
\text{s.t.} & & \|x\|_2 = 1. \\
\end{align*}$$

(P$_a$)

By considering the equality constraint as two inequality constraints, which indicate $\|x\|_2 \geq 1$ and $\|x\|_2 \leq 1$, we obtain the dual of (P$_a$) as follows:

$$\begin{align*}
\max & & -v_1 + v_2 \\
\text{s.t.} & & \sqrt{2} - v_1 + v_2 \leq 0, \\
& & v \geq 0, \\
\end{align*}$$

(D$_a$)

where $v = (v_1, v_2) \in \mathbb{R}^2$. Clearly, the Slater constraint qualification fails for problem (P$_a$) because there is no point that satisfies both $\|x\|_2 > 1$ and $\|x\|_2 < 1$. However, the optimal solutions of (P$_a$) and (D$_a$) are $x^* = (-\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2})$ and $v^* = (c, c - \sqrt{2}), c \geq 0$, respectively, and the optimal values of (P$_a$) and (D$_a$) are $-\sqrt{2}$.

### 3.2. Optimality conditions

The most well-known optimality conditions in the optimization literature are Karush-Kuhn-Tucker (KKT) conditions. These KKT conditions use gradients and/or subgradients of the functions involved in the problem. We now present alternative optimality conditions that do not require gradient information.

We first give sufficient optimality conditions for problems (P) and (D). Note that we do not assume the Slater constraint qualification and Assumption 3.1 here.

**Theorem 3.3 (Sufficient optimality conditions):** Points $x^*$ and $(u^*, v^*)$ are optimal for (P) and (D), respectively, if the following conditions hold:

(i) $Hx^* + K\mathcal{G}(x^*) \leq p$, $Ax^* = b$, $x^* \in \text{dom} \mathcal{G}$, (primal feasibility)
(ii) $\mathcal{G}^\circ(A^T u^* - H^T v^* - c) - K^T v^* \leq d$, $v^* \geq 0$, (dual feasibility)
(iii) $[d + K^T v^* - \mathcal{G}^\circ(A^T u^* - H^T v^* - c)]_i g_i(x^*_i) = 0, i = 1, \ldots, m$, (complementarity)
(iv) $[p - Hx^* - K\mathcal{G}(x^*)]_i v^*_i = 0, i = 1, \ldots, m, (complementarity)$
(v) $\mathcal{G}^\circ(A^T u^* - H^T v^* - c)^T \mathcal{G}(x^*) = (A^T u^* - H^T v^* - c)^T x^*$. (alignment)
Proof: From the complementarity conditions (iii) and (iv), we obtain

\[
0 = \left[ d + K^T v^* - G^\circ (A^T u^* - H^T v^* - c) \right]^T G(x^*) + \left[ p - Hx^* - KG(x^*) \right]^T v^*
\]

\[
= d^T G(x^*) - G^\circ (A^T u^* - H^T v^* - c)^T G(x^*) + p^T v^* - (Hx^*)^T v^*.
\]

It then follows from the alignment condition that we have

\[
d^T G(x^*) - G^\circ (A^T u^* - H^T v^* - c)^T G(x^*) + p^T v^* - (Hx^*)^T v^*
\]

\[
= c^T x^* + d^T G(x^*) - b^T u^* + p^T v^*,
\]

which indicates that the objective function values of the primal and the dual problems are the same for the feasible points \(x^*\) and \((u^*, v^*)\). From the weak duality theorem, \(x^*\) and \((u^*, v^*)\) are optimal for (P) and (D), respectively. □

Note that condition (v) in Theorem 3.3, called the alignment condition, is not standard, and seems to be strange at first glance. This is actually used in the previous work [1] about gauge duality, which is different from the duality considered here. Moreover, as it can be seen below, the alignment condition is one of the necessary conditions for optimality.

When the Slater constraint qualification for problem (P) and Assumption 3.1 hold, the sufficient optimality conditions in Theorem 3.3 become necessary.

**Theorem 3.4 (Necessary conditions for optimality):** Suppose that Assumptions 2.2 and 3.1 hold. Suppose also that the Slater constraint qualification holds for (P). Let \(x^*\) and \((u^*, v^*)\) be optimal solutions of (P) and (D), respectively. Then conditions (i)–(v) in Theorem 3.3 hold.

Proof: Since \(x^*\) and \((u^*, v^*)\) are optimal solutions of (P) and (D), respectively, the feasibility conditions (i) and (ii) clearly hold. Moreover, since strong duality holds for \(x^*\) and \((u^*, v^*)\) under the assumptions, we have

\[
0 = c^T x^* + d^T G(x^*) - b^T u^* + p^T v^*
\]

\[
= d^T G(x^*) - (A^T u^* - H^T v^* - c)^T x^* + p^T v^* - (Hx^*)^T v^*
\]

\[
\geq d^T G(x^*) - G^\circ (A^T u^* - H^T v^* - c)^T G(x^*) + p^T v^* - (Hx^*)^T v^*
\]

\[
= \left[ d + K^T v^* - G^\circ (A^T u^* - H^T v^* - c) \right]^T G(x^*) + \left[ p - Hx^* - KG(x^*) \right]^T v^*
\]

\[
\geq 0,
\]

where the second equality follows from the fact that \(Ax^* = b\), the third inequality follows from Proposition 3.1, and the last inequality follows from (i) and (ii).
Thus, the above inequalities hold with equalities, and hence we obtain conditions (iii), (iv) and (v).

We show an example of optimality conditions (i)-(v) for the Ridge-type problem [13].

**Example 3.2:** Consider the following Ridge-type optimization problem:

\[
\min_{x \in \mathbb{R}^n} \|Ax - b\|_2 + \sigma \|x\|_2, \quad (P_b)
\]

where \(A \in \mathbb{R}^{m \times n}\) and \(\sigma > 0\), and transform the problem into the standard gauge optimization form (P) as follows:

\[
\min_{x,y} \|y\|_2 + \sigma \|x\|_2 \quad \text{subject to} \quad Ax - y = b, \quad (P'_b)
\]

Then, the dual of \((P'_b)\) is described as

\[
\max \quad b^T u \\
\text{s.t.} \quad \|A^T u\|_2 \leq \sigma, \quad \|-u\|_2 \leq 1. \quad (D_b)
\]

Note that we remove the constraint \(v \geq 0\) from \((D_b)\) because it does not effect the optimal solutions and optimal value of \((D_b)\). The optimality conditions (i)-(v) in Theorem 3.3 for problems \((P'_b)\) and \((D_b)\) are described as follows:

1. \(Ax - y = b,\)
2. \(\|A^T u\|_2 \leq \sigma, \|-u\|_2 \leq 1,\)
3. \((\sigma - \|A^T u\|_2)\|x\|_2 = 0, \quad (1 - \|-u\|_2)\|y\|_2 = 0,\)
4. \(\|A^T u\|_2 \|x\|_2 + \|-u\|_2 \|y\|_2 = u^T Ax - u^T y.\)

From the above items (i), (iii) and (v), we obtain

\[
\sigma \|x\|_2 + \|Ax - b\|_2 = b^T u.
\]

Then, the optimality conditions for problems \((P'_b)\) and \((D_b)\) are

\[
\left\{ \begin{array}{l}
\|A^T u\|_2 \leq \sigma, \|-u\|_2 \leq 1, \\
\sigma \|x\|_2 + \|Ax - b\|_2 = b^T u.
\end{array} \right.
\]

Note that the above conditions are necessary and sufficient optimality conditions because Assumption 3.1 and the Slater constraint qualifications clearly hold for problems \((P'_b)\) and \((D_b)\). We also note that the left hand side of the above equality condition, which is an alignment condition, is the objective function of the original problem \((P_b)\). Therefore, the alignment condition in this example indicates that the strong duality holds for problems \((P_b)\) and \((D_b)\).
3.3. Primal recovery

Let us now discuss about the recovery of a primal optimal solution from a KKT point of the dual problem (D). For simplicity, we denote
\[ \Phi_1(u, v) := \mathcal{G}^\circ(A^T u - H^T v - c) \]
and
\[ \phi_i(u, v) := \mathcal{G}^\circ(A_i^T u - H_i^T v - c_i), i = 1, \ldots, m. \]
Then, the KKT conditions of (D) can be described as
\[ p + V^T \lambda - K \lambda - \mu = 0, \quad V \in \partial_v \Phi(u^*, v^*), \quad (8) \]
\[ -b + U^T \lambda = 0, \quad U \in \partial_u \Phi(u^*, v^*), \quad (9) \]
\[ d - \Phi(u^*, v^*) + K^T v^* \geq 0, \lambda \geq 0, \quad \lambda^T (d - \Phi(u^*, v^*) + K^T v^*) = 0, \quad (10) \]
\[ v^* \geq 0, \mu \geq 0, \quad v^T \mu = 0, \quad (11) \]
where \( \lambda \in \mathbb{R}^m \) and \( \mu \in \mathbb{R}^\ell \) are Lagrangian multipliers. Let \( A_i = A_{i_1} \) and \( H_i = H_{i_1} \) for all \( i = 1, \ldots, m \) in the subsequent discussion. Moreover, we divide matrices \( U \) and \( V \) as
\[ U = \begin{pmatrix} U_1 \\ \vdots \\ U_m \end{pmatrix}, \quad V = \begin{pmatrix} V_1 \\ \vdots \\ V_m \end{pmatrix}, \]
where \( U_i \in \mathbb{R}^{1 \times k} \) and \( V_i \in \mathbb{R}^{1 \times \ell} \) for all \( i = 1, \ldots, m \).

We now give the concrete formulae for the subdifferentials \( \partial_v \Phi \) and \( \partial_u \Phi \). First, for given \( u \in \mathbb{R}^k \) and \( v \in \mathbb{R}^\ell \), let us denote \( X_i(u, v) \) as the set of optimal solutions of the following problem:
\[ \sup_{x_{i_1}} u^T A_i x_{i_1} - v^T H_i x_{i_1} - c_i^T x_{i_1} \]
\[ \text{s.t.} \quad g_i(x_{i_1}) \leq 1. \quad (P_i) \]
Moreover, we assume the following condition to show key properties of \( X_i(u, v) \).

Assumption 3.2: For all \( i, g_i \) vanishes only at 0, that is, \( g_i(\bar{x}_{i_1}) = 0 \) if and only if \( \bar{x}_{i_1} = 0 \).

Lemma 3.1: Suppose that Assumption 3.2 holds. Then, \( X_i(u, v) \) is nonempty, convex and compact for all \( u \in \mathbb{R}^k \) and \( v \in \mathbb{R}^\ell \).

Proof: The feasible region of \( (P_i) \) is nonempty since \( g_i \) is a gauge function, and \( x_{i_1} = 0 \) is a feasible solution of problem \( (P_i) \). In addition, the feasible region is convex and closed because each function \( g_i \) is convex and closed from its lower semi-continuity. Moreover, Assumption 3.2 implies that the feasible region is bounded. To see this, let \( B_i := \{ z \in \mathbb{R}^{n_i} | \| z \| = 1 \} \) and \( \rho := \inf_{z \in B_i} g_i(z) \). Then \( \rho > 0 \) from Assumption 3.2. If \( \rho = +\infty \), that is, \( \text{dom } g_i = \emptyset \), then \( X_i(u, v) = \emptyset \) and this lemma holds. Now, suppose that \( \rho < \infty \). Then, the feasible region
is included in the compact set \( \overline{B}_i := \{ z \mid \|z\| \leq 1/\rho \} \) since for any \( s \not\in \overline{B}_i \) we have \( \|s\| > 1/\rho \) and
\[
g_i(s) = g_i(\|s\|s/\|s\|) = \|s\|g_i(s/\|s\|) > \frac{1}{\rho} \rho = 1,
\]
which shows that \( s \) is not a feasible solution of \((P_i)\). Consequently, the feasible region of \((P_i)\) is nonempty, convex and compact.

Since \((P_i)\) is a convex problem with a nonempty, compact and convex feasible region, the optimal solution set of \((P_i)\) is nonempty, convex and compact. \(\blacksquare\)

We now describe the concrete formulae for \( \partial_v \Phi \) and \( \partial_u \Phi \) by using \( X_i(u, v) \) as follows.

**Lemma 3.2:** Suppose that Assumption 3.2 holds for function \( G \). Then, we have
\[
\phi_i(u, v) = u^T A_i \tilde{x}_I - v^T H_i \tilde{x}_I - c_i^T \tilde{x}_I \quad \text{for all } \tilde{x}_I \in X_i(u, v),
\]
and
\[
\partial_u \phi_i(u, v) = \{ \tilde{x}_I^T A_i^T | \tilde{x}_I \in X_i(u, v) \} \tag{13}
\]

**Proof:** The first equation directly follows from the definitions of \( g_i^o \) and \( X_i(u, v) \). Since the set \( X_i(u, v) \) is nonempty, convex and compact from Lemma 3.1, we obtain
\[
\partial_u \phi_i(u, v) = \text{co} \{ \tilde{x}_I^T A_i^T | \tilde{x}_I \in X_i(u, v) \} = \{ \tilde{x}_I^T A_i^T | \tilde{x}_I \in X_i(u, v) \},
\]
\[
\partial_v \phi_i(u, v) = \text{co} \{ -\tilde{x}_I^T H_i^T | \tilde{x}_I \in X_i(u, v) \} = \{ -\tilde{x}_I^T H_i^T | \tilde{x}_I \in X_i(u, v) \},
\]
which are the desired formulae. \(\blacksquare\)

Finally, we present the main result of this subsection, which shows that it is possible to obtain a primal solution from a KKT point of problem \((D)\).

**Theorem 3.5 (Primal recovery):** Suppose that Assumptions 2.2, 3.1 and 3.2 hold for the function \( G \). Assume also that \((u^*, v^*, \lambda, \mu) \in \mathbb{R}^k \times \mathbb{R}^\ell \times \mathbb{R}^m \times \mathbb{R}^\ell \), \( V \in \partial_v \Phi(u^*, v^*) \) and \( U \in \partial_u \Phi(u^*, v^*) \) satisfy the KKT conditions (8)–(11) for the dual problem \((D)\). Then there exist \( \tilde{x}_I \in X_i(u^*, v^*) \) for all \( i = 1, \ldots, m \) such that \( U_i = (A_i \tilde{x}_I)^T \) and \( V_i = -(H_i \tilde{x}_I)^T \). Moreover, suppose that \( g_i(\tilde{x}_I) = 1 \) for \( i \) such that \( \lambda_i \neq 0 \). Let \( x^*_I = \lambda_i \tilde{x}_I \) for all \( i = 1, \ldots, m \). Then, \( x^* = (x^*_1, \ldots, x^*_m) \) is an optimal solution of \((P)\).
Proof: From the definitions of $\Phi$ and $G$, we have

$$
\Phi(u^*, v^*) = G(A^T u^* - H^T v^* - c) = \\
\begin{pmatrix}
g_1(A_1^T u^* - H_1^T v^* - c_1) \\
\vdots \\
g_m(A_m^T u^* - H_m^T v^* - c_m) 
\end{pmatrix}
$$

Moreover, since

$$
U \in \partial_u \Phi(u^*, v^*) \subseteq \\
\begin{pmatrix}
\partial_u \phi_1(u^*, v^*) \\
\vdots \\
\partial_u \phi_m(u^*, v^*)
\end{pmatrix}
$$

we have $U_i \in \partial_u \phi_i(u^*, v^*)$. In a similar way, we have $V_i \in \partial_v \phi_i(u^*, v^*)$. It then follows from (13) and (14) in Lemma 3.2 that, for all $i = 1, \ldots, m$, there exist $\tilde{x}_{I_i} \in X_i(u^*, v^*)$, such that $U_i = (A_i \tilde{x}_{I_i})^T$ and $V_i = -(H_i \tilde{x}_{I_i})^T$.

Now let $x_{I_i}^* = \lambda_i \tilde{x}_{I_i}$, $i = 1, \ldots, m$, and $x^* = (x_{I_1}^*, \ldots, x_{I_m}^*)^T$. We show that $x^*$ and $(u^*, v^*)$ satisfy the sufficient conditions (i)–(v) in Theorem 3.3. Note that the dual feasibility (ii) clearly holds. Moreover, since the assumption on $g_i(\tilde{x}_{I_i})$ implies

$$
g_i(x_{I_i}^*) = g_i(\lambda_i \tilde{x}_{I_i}) = \lambda_i g_i(\tilde{x}_{I_i}) = \lambda_i,
$$

we obtain

$$
G(x^*) = \lambda. \quad (15)
$$

We first show that the alignment condition (v) holds. From (12) in Lemma 3.2, we have

$$
g_i(A_i^T u^* - H_i^T v^* - c_i) = \phi_i(u^*, v^*) = (u^*)^T A_i \tilde{x}_{I_i} - (v^*)^T H_i \tilde{x}_{I_i} - c_{I_i}^T \tilde{x}_{I_i}.
$$

It then follows from (15) that

$$
g_i(A_i^T u^* - H_i^T v^* - c_i)^T g_i(x_{I_i}^*) = \lambda_i ((u^*)^T A_i \tilde{x}_{I_i} - (v^*)^T H_i \tilde{x}_{I_i} - c_{I_i}^T \tilde{x}_{I_i})
$$

$$
= (u^*)^T A_i x_{I_i}^* - (v^*)^T H_i x_{I_i}^* - c_{I_i}^T x_{I_i}^*
$$

$$
= (A_i^T u^* - H_i^T v^* - c_i)^T x_{I_i}^*,
$$

which shows that condition (v) holds.
Next we prove the primal feasibility (i). From the definition of $x^*$, we obtain

$$Ax^* = \sum_{i=1}^{m} \lambda_i A_i \bar{x}_i = \sum_{i=1}^{m} \lambda_i U_i^T = U^T \lambda = b,$$

where the second equality follows from (13) in Lemma 3.2 and the last equality is due to the KKT condition (9). Moreover, we have from (14) in Lemma 3.2 that

$$Hx^* = \sum_{i=1}^{m} \lambda_i H_i \bar{x}_i = -\sum_{i=1}^{m} \lambda_i V_i^T = -V^T \lambda. \quad (16)$$

It then follows from (15) that

$$Hx^* + KG(x^*) = -V^T \lambda + K \lambda = p - \mu \leq p,$$

where the equality and the inequality follow from the KKT conditions (8) and (11), respectively. Consequently, $x^*$ is a feasible solution of (P).

Finally, we show that the complementarity conditions (iii) and (iv) hold. First we consider condition (iii) as follows. If $\lambda_i = 0$, then $x_i^* = 0$ and $g_i(x_i^*) = 0$, and hence (iii) holds. If $\lambda_i \neq 0$, then $[d + K^T \nu^* - \mathcal{G}^*(A^T u^* - H^T \nu^* - c)]_i = 0$ from the KKT condition (10) and the definition of $\Phi$. Therefore, (iii) also holds.

Next we prove that condition (iv) is satisfied. If $\nu_i^* = 0$, then (iv) clearly holds. For this reason, we consider the case where $\nu_i^* \neq 0$. In such a case, $\mu_i = 0$ from the KKT condition (11), and hence $[p + V^T \lambda - K \lambda]_i = 0$ from the KKT condition (8). It then follows from (15) and (16) that

$$0 = \left[ p + V^T \lambda - K \lambda \right]_i = \left[ p - Hx^* - K \lambda \right]_i = \left[ p - Hx^* - KG(x^*) \right]_i.$$ 

Therefore, the complementarity condition (iv) holds.

From the previous discussion, we conclude that $x^*$ and $(u^*, \nu^*)$ satisfy all sufficient conditions for optimality, and hence $x^*$ is an optimal solution of (P). ■

Observe that the assumption that $g_i(\bar{x}_i) = 1$ for all $i$ such that $\lambda_i \neq 0$ seems to be rather restrictive. One sufficient condition for the assumption is that the effective domain of $g_i$ is $\mathbb{R}^{n_i}$ and $A_i^T u^* - H_i^T \nu^* - c_i \neq 0$ for all $i$. Under these conditions, the solution set $X_i(u^*, \nu^*)$ is included in the boundary of the feasible set of $(P_i)$, and thus $g_i(\bar{x}_i) = 1$ for all $\bar{x}_i \in X_i(u^*, \nu^*)$.

We now show an example of the primal recovery by using the Ridge-type optimization problem considered in Example 3.2.
Example 3.3: Consider the following positively homogeneous dual of the standard form of the Ridge-type optimization problem \((P'_b)\):

$$\begin{align*}
\max & \quad b^T u \\
\text{s.t.} & \quad \|A^T u\|_2 \leq \sigma \\
& \quad \|u\|_2 \leq 1.
\end{align*} \tag{D_c}$$

We assume that \(b \neq 0\). Note that \((D_c)\) has a solution \(u^*\) since the feasible set is nonempty and compact. Note also that the assumption \(b \neq 0\) implies \(u^* \neq 0\). Moreover Slater’s constraint qualification holds. We further note that the second constraint slightly changes comparing to \((D_b)\) because \(\| - u \|_2 = \| u \|_2\).

The Lagrangian function of \((D_c)\) is

$$L(u) = -b^T u + \lambda_1 (\|A^T u\|_2 - \sigma) + \lambda_2 (\|u\|_2 - 1),$$

where \(\lambda_1, \lambda_2 \in \mathbb{R}\) are the Lagrangian multipliers. When \(A^T u^* \neq 0\), there exist \((\lambda_1, \lambda_2)\) satisfying the following KKT conditions:

\begin{align*}
-b + \lambda_1 \frac{A A^T u^*}{\|A^T u^*\|_2} + \lambda_2 \frac{u^*}{\|u^*\|_2} &= 0, \quad (17) \\
\|A^T u^*\|_2 - \sigma &\leq 0, \quad (18) \\
\|u^*\|_2 - 1 &\leq 0, \quad (19) \\
\lambda_1 (\|A^T u^*\|_2 - \sigma) &= 0, \quad (20) \\
\lambda_2 (\|u^*\|_2 - 1) &= 0, \quad (21) \\
\lambda_1, \lambda_2 &\geq 0. \quad (22)
\end{align*}

Note that \(A^T u^* \neq 0\) when \(\text{rank}(A) = m\). However, if \(\text{rank}(A) < m\), the equation \(A^T u^* = 0\) may hold. In this case, we have to describe Equation (17) by using the subgradient. Then, Equation (17) becomes more complicated and we might fail to recover a primal solution.

Then, problems \((P_i), i = 1, 2\) are described as

$$\begin{align*}
\sup_x & \quad (u^*)^T A x \\
\text{s.t.} & \quad \|x\|_2 \leq 1, \quad (P_1)
\end{align*}$$

and

$$\begin{align*}
\sup_y & \quad -(u^*)^T y \\
\text{s.t.} & \quad \|y\|_2 \leq 1, \quad (P_2)
\end{align*}$$

respectively. Here the variables \(x\) and \(y\) in problems \((P_1)\) and \((P_2)\) are those of problem \((P'_b)\). Then, we obtain

$$X_1(u^*) = \left\{ \frac{A^T u^*}{\|A^T u^*\|_2} \right\}, \quad X_2(u^*) = \left\{ -\frac{u^*}{\|u^*\|_2} \right\},$$
and we observe that there exist $\bar{x} \in X_1(u^*)$ and $\bar{y} \in X_2(u^*)$ such that

$$U_1^T = A\bar{x} = \frac{AA^*u^*}{\|ATu^*\|_2}, \quad U_2^T = -E_m\bar{y} = \frac{u^*}{\|u^*\|_2}.$$ 

We also observe that

$$g_1(\bar{x}) = \|\bar{x}\|_2 = \begin{Vmatrix} \frac{ATu^*}{\|ATu^*\|_2} \end{Vmatrix}_2 = 1, \quad g_2(\bar{y}) = \|\bar{y}\|_2 = \begin{Vmatrix} -u^* \end{Vmatrix}_2 = 1.$$

Then, we consider four cases with respect to the Lagrangian multipliers $\lambda_1$ and $\lambda_2$. If $\lambda_1 = \lambda_2 = 0$, then we have $b = 0$ from Equation (17) which is a contradiction to the assumption here. If $\lambda_1 \neq 0$ and $\lambda_2 = 0$, then we have $\|ATu^*\|_2 = \sigma$ from (20) and

$$\lambda_1 AA^*u^* = \sigma b$$

from (17). By multiplying $(u^*)^T$, we obtain

$$\lambda_1 (u^*)^T AA^*u^* = \lambda_1 \|ATu^*\|_2^2 = \lambda_1 \sigma^2 = \sigma (u^*)^T b,$$

which results in

$$\lambda_1 = \frac{(u^*)^T b}{\sigma}.$$ 

Therefore, an optimal solution of problem $(P'_{b})$, that is $(x^*, y^*)$, is obtained by

$$(x^*, y^*) = (\lambda_1 \bar{x}, \lambda_2 \bar{y}) = \left( \frac{(u^*)^T b}{\sigma^2} A^*u^*, 0 \right).$$

If $\lambda_1 = 0$ and $\lambda_2 \neq 0$, then we have $\|u^*\|_2 = 1$ from (21) and $\lambda_2 u^* = b$ from (17). By multiplying $(u^*)^T$, we obtain $\lambda_2 = (u^*)^T b$. Therefore, the optimal solution is

$$(x^*, y^*) = (0, -(u^*)^T bu^*).$$

If $\lambda_1 \neq 0$ and $\lambda_2 \neq 0$, then we have $\|ATu^*\|_2 = \sigma$ and $\|u^*\|_2 = 1$ from (21) and (22). Then from (17), we obtain

$$-b + \frac{\lambda_1}{\sigma} AA^*u^* + \lambda_2 u^* = 0,$$

and by multiplying $(u^*)^T$ we have

$$-(u^*)^T b + \lambda_1 + \lambda_2 = 0.$$ 

Thus, we obtain

$$(x^*, y^*) = \left( \frac{\lambda_1}{\sigma} A^*u^*, (\lambda_1 \sigma - (u^*)^T b) \frac{u^*}{\|u^*\|_2} \right)$$

$$= \left( \frac{\lambda_1}{\sigma} A^*u^*, (\lambda_1 \sigma - (u^*)^T b) u^* \right).$$

Note that, for the original problem $(P_{b})$, if $\sigma$ is sufficiently large, then an optimal solution $x^*$ becomes zero. If $\sigma$ is small, then $x^* \neq 0$. The property can be
described by using the dual problem \((D_c)\) as follows. If \(\sigma\) is sufficiently large, then the first constraint of \((D_c)\): \(\|A^T u\|_2 \leq \sigma\) tend to be inactive, which indicates \(\lambda_1 = 0\), and thus \(x^* = 0\). On the other hand, if \(\sigma\) is sufficiently small, then the second constraint of \((D_c)\): \(\|u\|_2 \leq 1\) tend to be inactive, which indicate \(\lambda_1 \neq 0\) and \(\lambda_2 = 0\), and thus \(x^* \neq 0\).

4. Duality for general optimization problems

In this section we extend the previous results for gauge optimization to more general optimization problem:

\[
\begin{align*}
\min & \quad c^T x + d^T F(x) \\
\text{s.t.} & \quad A x = b, \\
& \quad H x + K F(x) \leq p, \\
\end{align*}
\]

(P\(_F\))

where \(F\) is a nonnegative vector convex function, that is, each component function \(f_i\) is an nonnegative convex function. Note that problem \((P_F)\) is convex if \(d \geq 0\) and \((K)_{ij} \geq 0\).

To this end, we first decompose general convex function of the problem, which is not necessarily nonnegative, into a linear and a nonnegative convex functions. Then, we consider the so-called perspective [1, 14] for the nonnegative convex function. The perspective function is a gauge one essentially equivalent to the original nonnegative convex function. Consequently, we reformulate the general convex function into a sum of linear function and a gauge one. The reformulation enables us to apply the results in the previous section for a general convex optimization problem.

4.1. Reformulation of a general convex function into sum of linear and gauge functions

Let us first observe that a convex function \(f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}\) can be written as a sum of a linear function and a nonnegative convex one. Let \(z \in \text{dom} f\) be a fixed vector, and let \(\eta \in \partial f(z) \neq \emptyset\). We can write

\[
f(x) = f(x) - f(z) - \eta^T (x - z) + f(z) + \eta^T (x - z).
\]

(23)

Note that \(f(x) - f(z) - \eta^T (x - z)\) is convex and nonnegative with respect to \(x\), because \(f\) satisfies the subgradient inequality [15, p. 214]: \(f(x) \geq f(z) + \eta^T (x - z)\). Moreover, the remaining term: \(f(z) + \eta^T (x - z)\) is linear with respect to \(x\). Thus, function \(f\) can be split into a nonnegative convex function and a linear one.

Next, we reformulate a nonnegative convex function into a gauge function through the so-called perspective of a nonnegative convex function [1, 14]. Recall that for any nonnegative convex function \(h : \mathbb{R}^n \to \mathbb{R}_+ \cup \{\infty\}\), its perspective
$h^p : \mathbb{R}^{n+1} \to \mathbb{R} \cup \{\infty\}$ is described as

$$h^p(x, \zeta) := \begin{cases} 
\zeta h(\zeta^{-1}x) & \text{if } \zeta > 0, \\
\delta_{\{0\}}(x) & \text{if } \zeta = 0, \\
\infty & \text{if } \zeta < 0,
\end{cases}$$

and its closure can be written by

$$h^\pi(x, \zeta) := \begin{cases} 
\zeta h(\zeta^{-1}x) & \text{if } \zeta > 0, \\
h^\infty(x) & \text{if } \zeta = 0, \\
\infty & \text{if } \zeta < 0,
\end{cases} \quad (24)$$

where $h^\infty$ is the recession function of $h$ [15, p. 66]. Note that if $h$ is a proper convex function, then $h^\pi$ is a positively homogeneous proper convex function [15, Theorem 8.5]. In addition, $h^\pi(0,0) = 0$ by definition, and hence $h^\pi$ becomes gauge. Therefore, $h$ is represented as the gauge function $h^\pi(x, \zeta)$ with $\zeta = 1$. Consequently, $f$ can be described as a sum of the linear function $f(z) + \eta^T(x - z)$ and a gauge function $h^\pi(x, 1)$, where $h(x) = f(x) - f(z) - \eta^T(x - z)$. We present an example of perspective and its polar.

**Example 4.1:** Let $f : \mathbb{R}^n \to \mathbb{R}$ be defined as $f(x) := \frac{1}{2}x^TAx$, where $A$ is an $n \times n$ symmetric positive definite matrix. Then, the perspective and its polar of the quadratic function $f$ are described as follows:

$$f^\pi(x, \zeta) = \begin{cases} 
\frac{1}{2\zeta}x^TAx & \text{if } \zeta > 0, \\
\delta_{\{0\}}(x) & \text{if } \zeta = 0, \\
\infty & \text{otherwise},
\end{cases}$$

$$f^\natural(y, \eta) = \begin{cases} 
-\frac{1}{2\eta}y^TA^{-1}y & \text{if } \eta < 0, \\
\delta_{\{0\}}(y) & \text{if } \eta = 0, \\
\infty & \text{otherwise}.
\end{cases}$$

**Proof:** The perspective $f^\pi$ directly follows from definition (24). Note that $A$ is positive definite, hence $f^\infty = \delta_{\{0\}}$ [15, p. 68]. The polar of $f^\pi$ is defined by

$$f^\natural(y, \eta) = \sup_{x, \zeta} \left\{ x^Ty + \zeta \eta \mid f^\pi(x, \zeta) \leq 1 \right\} . \quad (25)$$

We first consider the case where $\eta > 0$. Since $f^\pi(0, \zeta) = 0$ for $\zeta \geq 0$, $f^\natural(y, \eta) \geq \zeta \eta$ for $\zeta \geq 0$. Then $f^\natural(y, \eta) \to \infty$ as $\zeta \to \infty$. Next suppose that $\eta = 0$ and $y \neq 0$. Let $x(t) := ty$ with $t > 0$, and let $\zeta(t) := \frac{1}{2}x(t)^TAx(t)$. Since $A$ is positive definite, $\zeta(t) = \frac{1}{2}x(t)^TAx(t) > 0$. Then $f^\pi(x(t), \zeta(t)) = 1$ for all $t$. Consequently $f^\natural(y, 0) \geq x(t)^Ty + \zeta(t) \cdot 0 = t\|y\|^2$, and hence $f^\natural(y, 0) \to \infty$ as $t \to \infty$. 


Next, we study the case where \( y = 0 \) and \( \eta \leq 0 \). If \( (y, \eta) = (0, 0) \), then \( f^2(y, \eta) = 0 \). Note that \( f^2(x, \zeta) \leq 1 \) implies \( \zeta \geq 0 \), and \( f^2(0, 0) \leq 1 \). Therefore, when \( y = 0 \) and \( \eta < 0 \) we have \( f^2(y, \eta) = 0 \).

Finally, we investigate the case where \( y \neq 0 \) and \( \eta < 0 \). We now set
\[
x^* = -\frac{1}{\eta} A^{-1} y, \quad \zeta^* = \frac{1}{2\eta^2} y^T A^{-1} y,
\]
and
\[
\lambda^* = -\eta \frac{2(\zeta^*)^2}{(x^*)^T A x^*}.
\]

Since \( x^* \neq 0 \) and \( \zeta^* > 0 \), \( \lambda^* \) is well-defined and \( \lambda^* > 0 \). Moreover, we have from (26)
\[
\frac{1}{2} (x^*)^T A x^* = \frac{1}{2\eta^2} y^T A^{-1} y = \zeta^*.
\]
It then follows from (27) that
\[
\eta = -\frac{\lambda^*}{\zeta^*}.
\]

Then, Equations (26) and (28) give
\[
-y + \frac{\lambda^*}{\zeta^*} A x^* = 0.
\]

We note that the following conditions also hold:
\[
\frac{1}{2\zeta^*} x^* T A x^* - 1 \leq 0, \lambda^* \geq 0,
\]
\[
\lambda^* \left( \frac{1}{2\zeta^*} x^* T A x^* - 1 \right) = 0.
\]
Note also that \( f^\pi(x, \zeta) = \frac{1}{2\zeta^*} x^T A x \). Conditions (27), (29)–(31) are the KKT conditions of the convex optimization problem in the right-hand of (25). Therefore, the point \((x^*, \zeta^*)\) is its global optimal solution. Consequently, we obtain
\[
f^2(y, \eta) = (x^*)^T y + \zeta^* \eta = -\frac{1}{2\eta} y^T A^{-1} y,
\]
which completes the proof.

We now consider a vector function \( F: \mathbb{R}^n \to (\mathbb{R} \cup \{\infty\})^m \), which is defined by \( F(\cdot) := (f_1(\cdot), \ldots, f_m(\cdot)) \) with nonnegative convex functions \( f_i: \mathbb{R}^{n_i} \to \mathbb{R} \cup \{\infty\}, i = 1, \ldots, m \). We then define its perspective \( F^\pi: \mathbb{R}^{n+m} \to (\mathbb{R} \cup \{\infty\})^m \) as \( F^\pi(\cdot) := (f_1^\pi(\cdot), \ldots, f_m^\pi(\cdot)) \) with \( f_i^\pi: \mathbb{R}^{n_i+1} \to \mathbb{R} \cup \{\infty\} \). For simplicity, we denote \( F^\pi(x, \zeta) = (f_i^\pi(x_1, \zeta_1), \ldots, f_m^\pi(x_m, \zeta_m)) \) for any \( x \in \mathbb{R}^n \) and \( \zeta \in \mathbb{R}^m \). We also denote the polar of \( F^\pi \) as \( F^\circ(\cdot) := (F^\pi)^\circ(\cdot) = ((f_1^\pi)^\circ(\cdot), \ldots, (f_m^\pi)^\circ(\cdot)) \). Note that \( F^\pi(x, e_m) = (f_1^\pi(x_1, 1), \ldots, f_m^\pi(x_m, 1)) = F(x) \) by definition. We also observe that \( F^\pi \) is a vector gauge function if \( f_i \) is a nonnegative proper convex function for all \( i \).
4.2. Perspective dual problems

We now consider problem \((P_F)\). By using the perspective function of \(F\), we reformulate \((P_F)\) into a gauge optimization:

\[
\begin{align*}
\min & \quad \hat{c}^T z + d^T F^\pi (z) \\
\text{s.t.} & \quad \hat{A} z = \hat{b}, \\
& \quad \hat{H} z + K F^\pi (z) \leq p,
\end{align*}
\]

where \(F^\pi : \mathbb{R}^{n+m} \rightarrow \mathbb{R}^m\) is the perspective of \(F\), \(z = (x_I, \xi_1, \ldots, x_{I_m}, \zeta_1, \ldots, \zeta_m)^T \in \mathbb{R}^{n+m}\), \(\hat{c} = (c_I, 0, \ldots, c_{I_m}, 0)^T \in \mathbb{R}^{n+m}\), \(\hat{b} = (b, 1, \ldots, 1)^T \in \mathbb{R}^{2m}\), \(\hat{H} = [H_I, 0, \ldots, H_{I_m}, 0] \in \mathbb{R}^{\ell \times (n+m)}\) and

\[
\hat{A} = \begin{bmatrix}
A_{I_1} & 0 & \cdots & A_{I_m} & 0 \\
0 & 1 & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \cdots & 0 & 1
\end{bmatrix} \in \mathbb{R}^{2m \times (n+m)},
\]

where \(A_{I_i}\) is a submatrix of \(A\) with \(A_{j}, j \in I_i\) as its columns.

We obtain the PHO dual of \((P_\pi)\) as follows:

\[
\begin{align*}
\max & \quad b^T u - p^T v + e_m^T w \\
\text{s.t.} & \quad F^\natural \begin{pmatrix}
(A_{I_1})^T u - (H_{I_1})^T v - c_{I_1} \\
w_1 \\
\vdots \\
(A_{I_m})^T u - (H_{I_m})^T v - c_{I_m} \\
w_m
\end{pmatrix} - K^T v \leq d, \\
v \geq 0.
\end{align*}
\]

We call problem \((D_\pi)\) as the perspective dual of \((P_F)\).

Example 4.2: We now consider the following convex quadratic optimization problem as an example of \((P_F)\).

\[
\begin{align*}
\min & \quad \frac{1}{2} x^T A_0 x + b_0^T x \\
\text{s.t.} & \quad \frac{1}{2} x^T A_1 x + b_1^T x \leq c_1,
\end{align*}
\]

where \(A_0\) and \(A_1\) are symmetric and positive definite matrices. The problem can be rewritten as

\[
\begin{align*}
\min & \quad \frac{1}{2} y^T A_0 y + b_0^T y \\
\text{s.t.} & \quad \frac{1}{2} y^T A_1 y + b_1^T y \leq c_1, \\
x - y = 0.
\end{align*}
\]
Let $z := (x, y)^T$ and $F(z) := (f_0(x), f_1(y))^T = (\frac{1}{2}x^TA_0x, \frac{1}{2}y^TA_1y)^T$. Then the problem is described as follows:

$$\min \ (b_0^T, 0)z + (1, 0)F(z)$$

s.t. $$(0, b_1^T)z + (0, 1)F(z) \leq c_1,$$

$$(I, -I)z = 0.$$ 

Let $w := (x, \zeta_1, y, \zeta_2) \in \mathbb{R}^{2n+2}$ and $F^\pi(w) := (f_0^\pi(x, \zeta_1), f_1^\pi(y, \zeta_2))$. Then, a gauge optimization $(P_\pi)$ equivalent to $(P_{QP})$ is written as

$$\min \ (b_0^T, 0, 0, 0)w + (1, 0)F^\pi(w)$$

s.t. $$(0, 0, b_1^T, 0)w + (0, 1)F^\pi(w) \leq c_1,$$

$$\begin{bmatrix}
I & 0 & -I & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}w = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$ 

Let $F^\natural := (f_0^\natural, f_1^\natural)$ be the polar of $F^\pi$. Then the PHO dual of $(P_{QP})$ is given as

$$\max \ (0, 1, 1)u - c_1v$$

s.t. $F^\natural \left( \begin{bmatrix} I & 0 & 0 \\ 0 & 1 & 0 \\ -I & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}u - \begin{bmatrix} 0 \\ 0 \\ b_1 \end{bmatrix} \right) - \begin{bmatrix} 0 \\ 1 \end{bmatrix}v \leq \begin{bmatrix} 1 \\ 0 \end{bmatrix},$

$$v \geq 0.$$ 

Let $u = (u_1, u_2, u_3)^T \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}$. Then the dual problem can be further rewritten as

$$\max \ u_2 + u_3 - c_1v$$

s.t. $f_0^\natural(u_1 - b_0, u_2) \leq 1,$

$f_1^\natural(-u_1 - b_1v, u_3) \leq v,$

$v \geq 0.$ 

Recall that the functions $f_0^\natural$ and $f_1^\natural$ are described as in Example 4.1. It is easy to see that $(u, v)$ with $u_2 > 0$ or $u_3 > 0$ is not feasible for $(D_{QP})$.

The following lemma indicates the first two constraints in $(D_{QP})$ can be represented as semidefinite constraints.

**Lemma 4.1:** Let $f(x) = \frac{1}{2}x^TAx$, where $A$ is an $n \times n$ symmetric and positive definite matrix. Then,

$$f^\natural(y, \eta) \leq \gamma, \quad \gamma \geq 0$$

if and only if

$$\begin{bmatrix} Ay & y \\ y^T & -2\eta \end{bmatrix} \succeq 0.$$
Proof: First we suppose that (32) holds. The inequality $f^2(y, \eta) \leq \gamma$ implies $(y, \eta) = (0, 0)$ or $\eta < 0$ from the definition of $f^2$ in Example 4.1. If $(y, \eta) = (0, 0)$, then (33) holds since $A$ is positive definite and $\gamma \geq 0$. If $\eta < 0$, then (32) can be written as

$$-\frac{1}{2\eta}y^TA^{-1}y \leq \gamma, \quad \gamma \geq 0.$$ 

If $\gamma = 0$, then we have $y = 0$, and hence (33) holds. When $\gamma > 0$, we obtain

$$-2\eta - \frac{1}{\gamma}y^TA^{-1}y \geq 0, \quad \gamma \geq 0,$$

which results in (33) by using the Schur complement [16].

Next we assume that (33) holds. Then, we have $\eta \leq 0$ and $\gamma \geq 0$. If $\eta = 0$, then $y = 0$ from (33). It then follows from Example 4.1 that $f^2(y, \eta) = \delta_{\{0\}}(0) = 0 \leq \gamma$, and hence (32) holds. If $\gamma = 0$, then $y = 0$ once again. Then we obtain $f^2(y, \eta) = 0 = \gamma$, which indicates (32) holds. If $\eta < 0$ and $\gamma > 0$, then the Schur complement of (33) gives (34), which results in (32).

From Lemma 4.1, the perspective dual problem $(D_{\pi}^{QP})$ of problem $(P_{QP})$ is equivalent to the following semidefinite programming [17]:

$$\max \quad u_2 + u_3 - c_1 v$$

s.t.

$$\begin{bmatrix}
A_0 & u_1 - b_0 \\
(u_1 - b_0)^T & -2u_2
\end{bmatrix} \succeq 0,$$

$$\begin{bmatrix}
A_1v & u_1 + b_1 v \\
(u_1 + b_1 v)^T & -2u_3
\end{bmatrix} \succeq 0.$$  

In this section, we discussed the reformulation of a convex quadratic optimization problem into a gauge optimization problem and discuss the duality proposed in the previous section. On the other hand, it may be natural to apply the well-known Lagrangian duality directly to the convex optimization problem. We can surely write the Lagrangian dual of a convex problem in a closed-form when the problem has a special structure like quadratic programming. However, for the more complicated problems, we could fail to write their Lagrangian dual in a closed-form. Therefore, we believe that the results here can support such issues.

5. Conclusion

In this paper, we considered optimization problems with both gauge functions and linear functions in their objective and constraint functions. Using the positively homogeneous framework given in [5], we proved that weak and strong duality results hold for such gauge problems. We also discussed both necessary and sufficient optimality conditions associated to these problems, showing that it is possible to obtain a primal solution by solving the dual problem. We
also extended the results for gauge problems to general optimization problems. Important future works are to develop an efficient algorithm by using the theoretical results described here and to apply the results to real world problems, which includes location problems with a norm as a distance function and regularized regression problem such as Lasso, Ridge and their variants.

**Acknowledgments**

The authors are grateful to anonymous referees for their remarkably careful reading and helpful suggestions especially in the duality theory in Section 3. The authors are also grateful to Prof. Ellen. H. Fukuda for helpful and detailed comments and suggestions.

**Disclosure statement**

No potential conflict of interest was reported by the author(s).

**Funding**

This work was supported in part by a Grant-in-Aid for Scientific Research (C) [grant number 17K00032] from Japan Society for the Promotion of Science.

**ORCID**

Shota Yamanaka [http://orcid.org/0000-0002-9906-3786](http://orcid.org/0000-0002-9906-3786)

**References**

[1] Aravkin AY, Burke JV, Drusvyatskiy D, et al. Foundations of gauge and perspective duality. SIAM J Optim. 2018;28(3):2406–2434.
[2] Freund RM. Dual gauge programs, with applications to quadratic programming and the minimum-norm problem. Math Program. 1987;38(1):47–67.
[3] Friedlander MP, Macêdo I. Low-rank spectral optimization via gauge duality. SIAM J Sci Comput. 2016;38(3):A1616–A1638.
[4] Friedlander MP, Macêdo I, Pong TK. Gauge optimization and duality. SIAM J Optim. 2014;24(4):1999–2022.
[5] Yamanaka S, Yamashita N. Duality of nonconvex optimization with positively homogeneous functions. Comput Optim Appl. 2018;71(2):435–456.
[6] Mangasarian OL. Absolute value programming. Comput Optim Appl. 2007;36(1):43–53.
[7] Alizadeh F, Goldfarb D. Second-order cone programming. Math Program. 2003;95(1):3–51.
[8] Xue G, Ye Y. An efficient algorithm for minimizing a sum of $p$-norms. SIAM J Optim. 2000;10(2):551–579.
[9] Meier L, van de Geer S, Bühlmann P. The group lasso for logistic regression. J R Stat Soc Series B Stat Methodol. 2008;70(1):53–71.
[10] Yuan M, Lin Y. Model selection and estimation in regression with grouped variables. J R Stat Soc Series B Stat Methodol. 2006;68(1):49–67.
[11] Wolf GW. Facility location: concepts, models, algorithms and case studies. Heidelberg: Springer-Verlag Berlin; 2011.
[12] Göpfert A, Riedrich T, Tammer C. Angewandte funktionalanalyse: motivatio-
nen und methoden für mathematiker und wirtschaftswissenschaftler. Wiesbaden: Vieweg+Teubner Verlag; 2009.

[13] Hoerl AE, Kennard RW. Ridge regression: biased estimation for nonorthogonal prob-
lems. Technometrics. 1970;12(1):55–67.

[14] Aravkin AY, Burke JV, Friedlander MP. Variational properties of value functions. SIAM J Optim. 2013;23(3):1689–1717.

[15] Rockafellar RT. Convex analysis. Princeton: Princeton University Press; 1972.

[16] Boyd S, Vandenberghe L. Convex optimization. Cambridge: Cambridge University Press; 2004.

[17] Vandenberghe L, Boyd S. Semidefinite programming. SIAM Rev. 1996;38(1):49–95.