STUDIES ON
OPTIMIZATION PROBLEMS WITH
POSITIVELY HOMOGENEOUS FUNCTIONS
AND ASSOCIATED DUALITY RESULTS

SHOTA YAMANAKA

STUDIES ON OPTIMIZATION PROBLEMS WITH POSITIVELY HOMOGENEOUS FUNCTIONS AND ASSOCIATED DUALITY RESULTS

by

SHOTA YAMANAKA

Submitted in partial fullfillment of
the requirement for the degree of
DOCTOR OF INFORMATICS
(Applied Mathematics and Physics)

KYOTO UNIVERSITY
KYOTO 606-8501, JAPAN
SEPTEMBER 2021

Preface

Optimization problems involving the absolute values of variables and linear terms are called Absolute Value Optimization (AVO) problems. Since binary variables can be expressed by using absolute value of variables, any 0-1 integer optimization problem with linear terms can be reformulated into the AVO. Another application could be a continuous location problem on planer region using the ℓ_1-norm as a distance function. Although AVO problems have such applications in the real world, fewer results have been reported except one duality result and one specific application since AVO was proposed in 2007. It is partly because of the difficulty of solving AVO, which is NP-hard, and the limited ability of the absolute value of variables to express a wider range of applications.

Another interesting optimization problem is called gauge optimization (GO), which includes gauge functions in its objective function and constraints. Gauge functions generalize the absolute value function, GO problems have wider application than the AVO. The general framework of GO and its duality was investigated in the 1980's. Then, in recent years, some researches showed that the duality framework becomes concrete so that we can apply the results of GO to apply many problems. In spite of that series of results, the GO problem in the above research seems to have the limitation for applications. One of the reasons is that the GO problem in the previous works can take only one constraint into account and should include only one gauge function.

In this thesis, we analyze the theoretical properties of the generalization of AVO problems and develop a global optimization algorithm to solve AVO. In the theoretical part, we generalize the absolute value functions in the AVO problem to positively homogeneous functions. We call such problems as Positively Homogeneous Optimization (PHO) and investigate the duality, the optimality conditions, and the applications of the PHO. We also consider more general GO problems than the previous works by replacing positively homogeneous functions in PHO into gauge functions. Such general GO problems can directly handle linear terms and multiple constraints. The duality and the optimality conditions of the GO problem are also investigated in the same way as the PHO case. We also extend the results to general optimization problems by considering the so-called perspective functions. In the algorithmic part, we develop an algorithm to obtain a global solution of the AVO. To the best of the author's knowledge, there exist no algorithms to solve the AVO in general form except one algorithm proposed recently for a specific AVO.

The contributions of the thesis are three-fold. Firstly, we propose a dual formulation of the PHO problem that has a closed-form and some interesting properties. We also discuss the relation between the proposed duality and the Lagrangian one. The second contribution is to investigate the theoretical properties of the GO problems as a special case of PHO. The GO problems include multiple constraints and linear terms, which is different from the existing results. For the GO problems, we analyze the weak and strong duality, necessary and sufficient optimality conditions, and the extension to general optimization problems. Finally, we develop a global optimization algorithm for the AVO by using a branch-and-bound technique. The proposed algorithm involves the duality results of the AVO, the reformulation of the primal and dual AVO problems into a system of equations, and the method to solve equations including the absolute value of variables.

The author hopes that the results in this thesis will contribute to further research on optimization problems involving absolute values, norms, and gauge functions, both in terms of theory and algorithms.

Shota Yamanaka
June 2021

Acknowledgment

First of all, I would like to show my greatest appreciation to Professor Nobuo Yamashita of Kyoto University for accepting me as a doctoral student and supervising me. Since I was working full-time at a company while taking a doctoral course, it was often difficult for me to progress in my research. Through such times, he waited patiently and gave me a lot of thoughtful suggestions and advice. He also supported me with my thesis's detailed writing and structure, applying for grants, and preparing for presentations at international conferences.

I would like to express my gratitude to my undergraduate and master's supervisor, former Professor Masao Fukushima of Kyoto University. I am very grateful to him for helping me progress my undergraduate and master's research, the starting point of this thesis. In addition to the careful guidance, his enthusiasm and initiative for research, and his active interaction with other researchers greatly affected me.

I would also like to thank Toshiya Sanaka, my former supervisor at Mitsubishi Chemical Corporation. He readily agreed when I told him that I decided to go on to a doctoral course while working. I would like to express my gratitude to him and Mitsubishi Chemical Corporation for providing me with many opportunities to participate in international and domestic conferences, and enter the doctoral course.

I am deeply grateful to Associate Professor Ellen Hidemi Fukuda of Kyoto University. It is always a great help that she gives me advice for research and encourages me as a researcher and as my wife when I had problems in my research.

Finally, I would like to dedicate this thesis to my parents to appreciate their kind support and encouragement.

Contents

List of Figures

List of Tables

List of Notations

$\mathbb{R}$	the set of real numbers
$\mathbb{R}^n$	the set of real n-vectors
$\mathbb{R}^{m\times n}$	the set of real $m \times n$ matrices
$\mathbb{R}_+$, $\mathbb{R}_{++}$	the set of nonnegative and positive real numbers
I, J	the subset of $\{1, 2, \ldots, n\}$
$\#I$	the cardinality of set $\mathcal{I}$
x_i	the i-th entry of vector x
$(x, y)^T$	the column vector $(x^T, y^T)^T$
e	the vector of ones
E	the identity matrix
X^T	the transpose of matrix X
$\lvert x\rvert$	the absolute value of scalar x
$\\|x\\|_1$	the ℓ_1-norm of vector x
$\\|x\\|_2$	the ℓ_2-norm of vector x
$\\|x\\|_p$	the ℓ_p-norm of vector x
$\\|x\\|_\infty$	the supremum norm of vector x
f°	the polar of function f
$\mathrm{dom} f$	the effective domain of function f
$\partial_x f(x, y)$	the subdifferential of $f(x, y)$ with respect to x
δ_S	the indicator function of set S

Chapter 1

Introduction

In the real world, we have to optimize many problems that occur especially in business situations. Some of the problems are production planning, employee and/or production scheduling, and vehicle routing. Such problems have been studied for many years in the field of mathematical optimization, which is a branch of applied mathematics. In the mathematical optimization area, many researchers discuss modeling real-world problems appropriately, efficient algorithms to solve the model, and mathematical theories to support the modeling and the algorithm.

A mathematical optimization problem is in general described as follow:

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & x \in \mathcal{X}, \end{array} \tag{1.0.1}$$

where f is called the objective function, x is a vector of appropriate dimension, and $\mathcal{X}$ is a feasible region in some space. Optimization problem (1.0.1) are classified by the continuity, the differentiability and the structure of f, and functions describing $\mathcal{X}$. In the field of continuous optimization, for instance, problem (1.0.1) is called a quadratic optimization problem when f and the functions describing $\mathcal{X}$ are linear and quadratic functions.

In this thesis, we focus on optimization problems involving the absolute value or, more generally, the norm of variables. These problems are considered to occur in the real world and are solved in many situations. For example, location problems with the ℓ_1-norm as a distance function can be represented as optimization problems with linear terms and the absolute value of variables. Such problems are called Absolute Value Optimization (AVO), which is, in general, a nonconvex optimization problem. Another related problem to AVO is Gauge Optimization (GO), which contains the so-called gauge function in its objective function and constraints. The GO problem is convex, and it has interesting theoretical results and practical applications.

1.1 Absolute value optimization and its related problems

1.1.1 Absolute value equation

In recent years, the so-called *absolute value equations* (AVEs) and *absolute value optimization* (AVO) problems have been attracted much attention. The AVEs were introduced in 2004 by Rohn [82]. Basically, if $A, B \in \mathbb{R}^{m\times n}$ are given matrices, and $b \in \mathbb{R}^m$ is a given vector, one should find a vector x that satisfies the following equation:

$$Ax + B|x| = b, \tag{1.1.1}$$

where $|x|$ is a vector whose i-th entry is the absolute value of the i-th entry of x.

Linear Complementarity Problems

It is known that AVE (1.1.1) are equivalent to the *linear complementarity problems* (LCP) [40, 64, 77], which include many real-world applications.

Example 1.1 (Linear complementarity problem)**.** *The purpose of the LCP is to find $y, z \in \mathbb{R}^n$ such that*

$$y = Qz + q, \quad z \geq 0, \quad y \geq 0, \quad \text{and} \quad y^T z = 0,$$

which can be rewritten as

$$0 \leq z \perp Qz + q \geq 0, \tag{1.1.2}$$

where $Q \in \mathbb{R}^{n\times n}$ and $q \in \mathbb{R}^n$.

Mangasarian [61] showed how to reduce LCP (1.1.2) into AVE (1.1.1). We obtain the solution of LCP (1.1.2) by solving the following AVE:

$$(I + M)(I - M)^{-1}x - |x| = -((I + M)(I - M)^{-1} + I)q$$

and computing z by

$$z = (I - M)^{-1}(x + q).$$

Prokopyev [77] proved that AVE (1.1.1) can be reduced to LCP (1.1.2) when we denote $y = (r, s, t, \theta_1)^T \in \mathbb{R}^\ell$, $z = (p, w, \theta_2)^T \in \mathbb{R}^\ell$, $q = (0, -c, c, 0)^T \in \mathbb{R}^\ell$, and

$$Q = \begin{bmatrix} -E & 2E & 0 \\ A & B-A & 0 \\ -A & A-B & 0 \\ 0 & 0 & 0 \end{bmatrix} \in \mathbb{R}^{\ell \times \ell},$$

where $\ell = \max\{n+2m, 2n\}$, $\theta_1 \in \mathbb{R}^{\ell-n-2m}$, $\theta_2 \in \mathbb{R}^{\ell-2n}$ and the size of zero components in Q and q are set in order to $Q \in \mathbb{R}^{\ell\times\ell}$ and $q \in \mathbb{R}^\ell$.

Existing results

The main topic of theoretical research for the AVEs is about unique solvability [64, 82–85, 105, 112]. For a slightly specific AVE

$$Ax - |x| = b, \tag{1.1.3}$$

where $A \in \mathbb{R}^{n\times n}$, $b \in \mathbb{R}^n$, Mangasarian and Meyer [64] and Rohn et al. [85] showed that AVE (1.1.3) for any b has a unique solution when the smallest singular value of A is greater than 1. The unique solvability for the more general AVE

$$Ax - B|x| = b, \tag{1.1.4}$$

where $A, B \in \mathbb{R}^{n\times n}$, $b \in \mathbb{R}^n$, was also studied. Rohn et al. [85] and Rohn [84] proved that AVE (1.1.4) is uniquely solvable when the singular value of $|A^{-1}B|$ is less than one or when the minimum singular value of A exceeds the maximum singular value of $|B|$. Here, the absolute value of a matrix $A = (a_{ij})$ is defined by $|A| = (|a_{ij}|)$.

Since 2007, some methods for solving AVEs have been presented in the literature. For example, Rohn [83] considered an iterative algorithm using the sign of variables for the case that A and B are square matrices. For more general A and B, Mangasarian [61] provided a method involving successive linearization techniques. Other methods include a concave minimization approach, given by Mangasarian [60, 63], and Newton-type methods, proposed by Bello Cruz et al. [9], Caccetta et al. [16], Mangasarian [62], and Zhang and Wei [112]. Some generalizations of AVEs were also proposed. For example, Hu et al. [41] considered an AVE involving the absolute value of variables associated with the second-order cones. Miao et al. [66] investigated an AVE with the so-called circular cones. In both papers, quasi-Newton based algorithms were developed.

1.1.2 Absolute value optimization problems

As an extension of AVEs, Mangasarian [61] proposed in 2007 the AVO problems, which have the absolute value of variables in their objective and constraint functions. More precisely, the AVO problem is given by

$$\begin{aligned} \min \quad & c^T x + d^T |x| \\ \text{s.t.} \quad & Ax + B|x| = b, \\ & Hx + K|x| \geq p, \end{aligned} \tag{1.1.5}$$

where $c, d \in \mathbb{R}^n$, $b \in \mathbb{R}^m$, $p \in \mathbb{R}^\ell$, $A, B \in \mathbb{R}^{m\times n}$, $H, K \in \mathbb{R}^{\ell \times n}$.

It is clear that AVO (1.1.5) includes linear programming. However, it is, in general, hard to solve because AVO (1.1.5) is non-differentiable due to the absolute value functions and is not always convex depending on d, B and K. In particular, it is NP-hard to obtain even feasible solutions of the problem because solving AVEs was proved to be NP-hard.

Linear Programs with Linear Complementarity Constraints

Since AVE (1.1.1) and LCP (1.1.2) are equivalent, AVO (1.1.5) includes the so-called *Linear Programs with Linear Complementarity Constraints* (LPLCC). LPLCC is stated formally as follows [33, 39, 46, 109].

Example 1.2 (Linear programs with linear complementarity constraints)**.**

$$\begin{aligned} \min \quad & c^T x + d^T y \\ \text{s.t.} \quad & Ax + By \geq b, \\ & 0 \leq y \perp p + Hx + Ky \geq 0, \end{aligned} \tag{1.1.6}$$

where b, c, d and p are vectors, A, B, H and K are matrices of appropriate dimensions.

LPLCC (1.1.6) started to be investigated in the 1970's [42, 43, 47], and many theoretical results, algorithms, and applications are proposed [57]. In particular, it is proposed that the LPLCC can be applied to parameter calibration in machine learning [10, 54, 55]. Note that LPLCC (1.1.6) is a special case of mathematical program with equilibrium constraints (MPEC) [57]. MPEC also has many applications in various areas such as economics, engineering, and transportation [57]. However, MPEC is in general difficult to deal with since its feasible region is necessarily nonconvex and even disconnected.

0-1 integer optimization problems

We can rewrite an optimization problem with binary variables as an AVO because 0-1 integer variables can be represented by using the absolute value of variables. A 0-1 integer variable can be transformed into the absolute value form as follows:

$$x \in \{0,1\}^n \quad \Leftrightarrow \quad \left| x - \frac{1}{2} \right| = \frac{1}{2}. \tag{1.1.7}$$

Considering the above formulation, we give an example of a binary value optimization problem of facility location called the *Capacitated Facility Location Problem* (CFLP) [23, 101]. CFLP is one of the well-known location optimization problems in descrete form and it is NP-hard. CFLP has been studied since the 1960's [1, 53, 71], and even in recent years the variation of CFLP and algorithms are investigated [5, 103, 113].

CFLP minimizes the total cost of locating facilities and services between facilities and clients under the constraints about demands and the capacity of facilities.

Example 1.3 (Capacitated facility location problem [23, 101])**.**

$$\begin{aligned} \min \quad & \sum_{j \in \mathcal{J}} c_j x_j + \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}} h_{ij} y_{ij} & \\ \text{s.t.} \quad & \sum_{j \in \mathcal{J}} y_{ij} = b_i, & i \in \mathcal{I} \\ & \sum_{i \in \mathcal{I}} y_{ij} - u_j x_j \leq 0, & j \in \mathcal{J} \\ & x_j \in \{0,1\} & j \in \mathcal{J}, \end{aligned} \tag{1.1.8}$$

where $\mathcal{I}$ and $\mathcal{J}$ are the sets of clients and candidate locations, $c_j \in \mathbb{R}$ is a locating cost of facility at j, $h_{ij} \in \mathbb{R}$ is a service cost that a facility at j gives clients i, $b_i \in \mathbb{R}$ is a demand of client i, $u_j \in \mathbb{R}$ is a capacity of facility locating j, x_j represents whether facility is located at j or not and y_{ij} is the amount of service that client j recieves from facility at j.

The above problem minimizes the total cost of locating facilities and services under the first constraint about demands and the second constraint of capacity of facilities. Problem (1.1.8) can be written as AVO by using reformulation (1.1.7) as follows:

$$
\begin{aligned}
\min \quad & \sum_{j\in\mathcal{J}} c_j \left(z_j + \frac{1}{2}\right) + \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}} h_{ij} y_{ij} & \\
\text{s.t.} \quad & \sum_{j\in\mathcal{J}} y_{ij} = b_i, & i \in \mathcal{I} \\
& \sum_{i\in\mathcal{I}} y_{ij} - u_j \left(z_j + \frac{1}{2}\right) \le 0, & j \in \mathcal{J} \\
& |z_j| = \frac{1}{2}, &
\end{aligned}
\tag{1.1.9}
$$

where only continuous variables exist.

Continuous facility location problems

Since the absolute value of variables is one of the most popular measures for distance, facility location problems (FLPs) in continuous space could be another application of AVO (1.1.5). The classical studies of FLPs are focusing on minimizing the total distance or minimizing the maximum distances between new facilities and existing ones [74]. Mathematical models of location problems fall into four major categories, which are Analytic models, Continuous models, Network models and Discrete models [26]. Note that problem (1.1.8) is an example of Discrete models.

Here we introduce some instances of continuous location models in which the ℓ_1-norm or rectilinear distance functions are involved. In continuous models, we can consider many distance measure, such as the ℓ_1-norm, the ℓ_2-norm, the ℓ_p-norm, $p \ge 1$, a general/polyhedral gauge, the Hausdorff distance, and so on [35]. In the following, we first show minisum and minimax single facility location problems, then we give minisum and minimax multi-facility location problems which are an extension of single facility problems. Finally, maxisum and maximin multi-facility location problems are represented.

A single facility location problem is one of the simplest location problems. In the problem, we usually minimize an objective function using norms that represent the distances between new facility and existing ones. Some applications of the single facility location problem are locating a hospital in a metropolitan area, new classroom building on a college campus, and a new component in an electrical network [26]. In particular, the ℓ_1-norm as a distance function tends to be used in location problems because both the appropriateness as distance measure and the easiness to analyze the problem are compatible.

Location problems of a single facility

Location problems of a single facility in minisum and minimax forms are described as follows. In each example, n is the number of existing facilities, $w_i \in \mathbb{R}$ is the weight of the distance between x and facility i, and $p_i \in \mathbb{R}^2$ is the location of existing facility i.

Example 1.4 (Minisum location problem of a single facility [26])**.**

$$\min_x \quad \sum_{i=1}^{n} w_i\, e^T |x - p_i|, \tag{1.1.10}$$

The above problem can be easily transformed into the AVO by denoting $y_i := x - p_i$ as follows:

$$\begin{aligned} \min_x \quad & \sum_{i=1}^{n} w_i\, e^T |y_i| \\ \text{s.t.} \quad & x - y_i = p_i, \quad i = 1, \dots, n. \end{aligned} \tag{1.1.11}$$

Example 1.5 (Minimax location problem of a single facility [26])**.**

$$\min_x \quad \max_i \{\ w_i\, e^T |x - p_i| + h_i,\ 1 \le i \le n\ \}, \tag{1.1.12}$$

where $h_i \in \mathbb{R}$ is the cost for user i to prepare to go to the new facility.

By introducing a new variable z, problem (1.1.12) can be reformulated as

$$\begin{aligned} \min \quad & z \\ \text{s.t.} \quad & w_i\, e^T |x - p_i| + h_i \le z, \quad i = 1, \dots, n, \end{aligned} \tag{1.1.13}$$

which can be represented as the AVO by taking $y_i := x - p_i$ as follows:

$$\begin{aligned} \min \quad & z \\ \text{s.t.} \quad & w_i\, e^T |y_i| + h_i \le z, \quad i = 1, \dots, n, \\ & x - y_i = p_i, \qquad\qquad i = 1, \dots, n. \end{aligned} \tag{1.1.14}$$

Some of the algorithms for solving Minisum location problems (1.1.10) are the so-called median method, programmed mathematical method and contour line method. Minimax problems (1.1.12) are solved by transforming the problem to (1.1.13), introducing some inequalities, and using linear programming technique [26].

Location problems of multiple facilities

Problems of locating more than two facilities are the direct extension of a single FLP. Such problems are called multifacility location problems (MFLP), in which multiple new facilities are located in some areas where some facilities exist. The MFLP minimizes the sum of the cost proportional to the distances between new facilities and the one between new and existing facilities. The most popular MFLPs are Minisum and Minimax FLP describes as follows.

In the following examples, n and m are the numbers of new facilities and existing ones, respectively. We also denote $v_{jk} \in \mathbb{R}$ and $w_{ji} \in \mathbb{R}$ as weights for balancing the distances between new facilities and between new and existing ones.

Example 1.6 (Minisum location problem of multi-facilities [26])**.**

$$\min_x \quad \sum_{1 \le j < k \le n} v_{jk}\, e^T |x_j - x_k| + \sum_{j=1}^{n} \sum_{i=1}^{m} w_{ji}\, e^T |x_j - p_i|, \tag{1.1.15}$$

Example 1.7 (Minimax location problem of multi-facilities [26])**.**

$$\begin{aligned} \min_x \quad \max\{\ & w_{ji}\, e^T |x_j - p_i|, j = 1, \dots, n, i = 1, \dots, m; \\ & v_{jk}\, e^T |x_j - x_k|, 1 \le j < k \le n\ \}, \end{aligned} \tag{1.1.16}$$

Typical algorithms for solving minisum MFLP are linear programming approach [67, 79, 98, 99] and decomposition technique [15, 45]. Theoretical results about the optimality conditions have been reported in [20, 49, 76, 79].

Note that all minisum and minimax FLPs (1.1.10), (1.1.12), (1.1.15), (1.1.16) are convex optimization problems since the absolute value functions are convex. However, those problems become nonconvex when there exist constraints of some forbidden areas described by using the absolute value functions [13, 14, 52]. We consider such problems in Chapter 6 and show some results of numerical experiments.

In FLPs of a single facility or multiple facilities described above, we attempt to locate desirable facilities, which are preferable to be located nearby to inhabitants. For instance, hospitals, shopping stores and schools are categorized in such facilities. On the other hand, there is another class of facilities, which are undesirable ones in our community. Some of the examples are garbage dump sites, chemical plants and prisons. Although those facilities are necessary for our life, it is favorable for them to be located as far as a residential area. Such problems are modeled as maxisum and maximin FLPs [19, 22, 25, 80]. Although there are not many results about those

FLPs using the ℓ_1-norm in planer region, we can consider maxisum and maximin MFLP using the ℓ_1-norm analogious to problems (1.1.15) and (1.1.16). We give numerical experiments of problem (1.1.18) in Chapter 6.

Example 1.8 (Maxisum location problem of multi-facilities).

$$\max_x \quad \sum_{1\le j<k\le n} v_{jk}\, e^T|x_j - x_k| + \sum_{j=1}^{n}\sum_{i=1}^{m} w_{ji}\, e^T|x_j - p_i|, \tag{1.1.17}$$

Example 1.9 (Maximin location problem of multi-facilities).

$$\begin{aligned} \max_x \quad \min\{\, & w_{ji}\, e^T|x_j - p_i|, j = 1,\dots,n, i = 1,\dots,m; \\ & v_{jk}\, e^T|x_j - x_k|, 1 \le j < k \le n \,\}, \end{aligned} \tag{1.1.18}$$

Scheduling problem of automated vehicles

Qian et al. [78] showed a scheduling problem about automated vehicles at intersections without signals can be formulated as AVO as follows. For problem (1.1.19), they also gave an algorithm called the alternately iterative descent method and showed its efficiency through some case studies in which different traffic conditions are considered.

Example 1.10 (Scheduling problem of automated vehicles).

$$\begin{array}{ll} \min & e^T x \\ \text{s.t.} & |Ax + b| \ge c, \\ & |H_j x + k_j^i| \ge d_j, \quad \forall i \in \mathcal{I}_j, \quad j \in \mathcal{J} \\ & x \ge \alpha, \\ & x \ge \beta, \end{array} \tag{1.1.19}$$

where b, c, k_j^i, d_j, α, $\beta \in \mathbb{R}^n$, $A \in \mathbb{R}^{m\times n}$, and $H_j \in \mathbb{R}^{m\times n}$ are given vectors and matrices. The sets $\mathcal{J}$ and $\mathcal{I}_j$ denote lanes and vehicles at lane j, respectively.

Linear support vector machine

A linear support vector machine (SVM) can be an application of the AVO. SVM is a useful classification tool in the field of machine learning and data mining. Especially, for large-scale data, the computational cost of training and testing of linear SVM is low comparing to that of nonlinear SVM involving kernel methods. In SVM, the sum of the regularization term and loss function is minimized. In particular, some results about the so-called ℓ_1-norm SVM have been studied [12, 59, 110, 116], which is described as follows. Usually, problem (1.1.20) is transformed into linear programming and solved by using existing methods or software.

Example 1.11 (ℓ_1-norm linear support vector machine)**.**

$$\min_{w,b} \ \|w\|_1 + c\sum_{i=1}^{n}(\max\{0, 1 - y_i(w^T x_i + b)\}), \tag{1.1.20}$$

where $c > 0$ *is the regularization parameter,* $\{(x_i, y_i)\}_{i=1}^{n}, x_i \in \mathbb{R}^m, y_i \in \{-1, +1\}$ *is a set of given data. The first and second terms are called regularization term and loss function, respectively.*

Problems with capped ℓ_1-norm

In recent years, the so-called *Capped ℓ_1-norm* have been studied and got more attention for its robustness in the field of machine learning and data mining. Capped ℓ_1-norm is described by

$$\sum_{i=1}^{n} \min\{|x_i|, \alpha\}, \tag{1.1.21}$$

where $\alpha > 0$ is the parameter to control the domain that the ℓ_1-norm effects. In classification and regression, we solve the problems of minimizing sum of the loss function which describes the misfit between the actual data and model, and the regularization term which controls the fitting of the model to the data.

Many types of norms are used in loss functions and regularization terms. In particular, the ℓ_1-norm is more robust against outliers comparing to the ℓ_2-norm. However, the robustness is not enough when there exist heavy outliers [114, 115]. Capped ℓ_1-norm have been used to obtain more robustness in such situation. Also, capped ℓ_1-norm is a better approximation of the ℓ_0-norm. The effectiveness of the norm is shown in many papers of machine learning and pattern recognition [48, 56, 72, 96, 104, 114, 115].

Capped ℓ_1-norm (1.1.21) can be described by using only the absolute value functions as follows:

$$\sum_{i=1}^{n} \left(|x_i| + \alpha - \frac{1}{2}(|x_i + \alpha| + |x_i - \alpha|) \right). \tag{1.1.22}$$

Then the optimization problems with capped ℓ_1-norm can be transformed into the AVO. The following example is the so-called twin SVM [96] in which we find a pair of nonparallel hyperplanes $w_1 x + b_1 = 0$ and $w_2 x + b_2 = 0$, where $w_1, w_2, b_1, b_2 \in \mathbb{R}^n$.

Example 1.12 (Capped ℓ_1-norm Twin Support Vector Machines [96]).

$$\begin{array}{ll} \min\limits_{w_1,b_1} & \sum\limits_{i=1}^{m_1} \min(\|w_1 x_i + b_1\|_1, \alpha_1) + c_1 \sum\limits_{i=1}^{m_2} \min(\|\beta_{1,i}\|_1, \alpha_2) \\ \text{s.t.} & -Bw_1 - e_2 b_1 + \beta_1 \geq e_2, \end{array} \tag{1.1.23}$$

$$\begin{array}{ll} \min\limits_{w_2,b_2} & \sum\limits_{i=1}^{m_2} \min(\|w_2 x_i + b_2\|_1, \alpha_3) + c_2 \sum\limits_{i=1}^{m_1} \min(\|\beta_{1,i}\|_1, \alpha_4) \\ \text{s.t.} & Aw_2 + e_1 b_2 + \beta_2 \geq e_1, \end{array} \tag{1.1.24}$$

We have to solve the above two problems to obtain a pair of hyperplanes. Although it is not easy to solve because of the nonconvexity, Wang et al. [96] proposed an iterative algorithm through the so-called re-weighted trick [48, 72, 73] and reformulation of problems (1.1.23) and (1.1.24) into approximation ones.

Existing results and issues

To the best of the author's knowledge, the only theoretical result for AVO (1.1.5) is the duality proposed by Mangasarian [61]. The dual of (1.1.5) can be represented as

$$\begin{array}{ll} \max & b^T u + p^T v \\ \text{s.t.} & |A^T u + H^T v - c| + B^T u + K^T v \leq d, \\ & v \geq 0, \end{array} \tag{1.1.25}$$

and weak duality, in which

$$c^T x + d^T |x| \geq b^T u + p^T v$$

holds for feasible solutions x and (u, v) of problems (1.1.5) and (1.1.25), were proved.

An algorithm for the specific AVO (1.1.19) was investigated in [78]; however, to the best of the author's knowledge, there is no method to solve more general AVO (1.1.5) itself. When comparing to AVEs, the research associated with AVO problems is insufficient even if there are many applications described in this section. One of these reasons seems to be the difficulty of obtaining feasible solutions of the problems. In fact, their constraints include AVEs, which is known to be NP-hard [61].

1.2 Gauge Optimization problems

Another optimization problem related to AVO is *gauge optimization* (GO), which was introduced by Freund [30] and recently investigated by Friedlander et al. [32] and Aravkin et al. [6].

1.2.1 Gauge optimization problems in general form

GO problems basically consist of optimization problems involving the so-called *gauge* functions. The GO problem proposed in [6, 30–32] is described as follows:

$$\begin{array}{ll} \min & g(x) \\ \text{s.t.} & x \in \mathcal{X}, \end{array} \tag{1.2.1}$$

where $\mathcal{X} \subseteq \mathbb{R}^n$ is a closed convex set and $g\colon \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is a gauge function. Here, we say that g is a gauge function if g is convex, nonnegative, positively homogeneous (i.e., $g(\alpha x) = \alpha g(x)$, $\alpha > 0$) and satisfies $g(0) = 0$. For instance, the ℓ_p-norm for $p \in [1\,\infty]$ and max function are gauges. Note that GO problem (1.2.1) is convex because gauge functions are convex.

Freund [30] first introduced problem (1.2.1) and its gauge dual as follows:

$$\begin{array}{ll} \min & g^\circ(y) \\ \text{s.t.} & y \in \bar{\mathcal{X}}, \end{array} \tag{1.2.2}$$

where g° is the polar function of g, which is defined by

$$g^\circ(y) := \inf\{\mu \geq 0 \mid y^T x \leq \mu g(x)\}, \tag{1.2.3}$$

and $\bar{\mathcal{X}}$ is the so-called *anti-polar* set of $\mathcal{X}$ defined by

$$\bar{\mathcal{X}} := \{y \in \mathbb{R}^n \mid y^T x \geq 1 \text{ for all } x \in \mathbb{R}^n\}. \tag{1.2.4}$$

Note that the constraint of problem (1.2.1) can be nonlinear because the set $\mathcal{X}$ is just a closed convex set, not a convex polyhedron.

In [30], some examples of GO (1.2.1) and their dual are shown, and we introduce two of them in the following.

Example 1.13 (Convex quadratic programming [96])**.**

$$\begin{array}{ll} \min & \frac{1}{2} x^T Q x + q^T x \\ \text{s.t.} & Ax \geq b, \end{array} \tag{1.2.5}$$

where Q is a symmetric positive semidefinite matrix. Suppose that the matrix Q can be decomposed as $Q = M^T M$ for some matrix M. If M is nonsingular or if q lies in the row space of M, then q can be represented as $q = M^T s$ for some s. Then the above quadratic programming problem can be rewritten as

$$\begin{aligned} \min \quad & \frac{1}{2} x^T M^T M x + s^T M x \\ \text{s.t.} \quad & Ax \geq b, \end{aligned} \tag{1.2.6}$$

which is equivalent to the following GO problem:

$$\begin{aligned} \min \quad & \|Mx + s\|_2 \\ \text{s.t.} \quad & Ax \geq b, \end{aligned} \tag{1.2.7}$$

The gauge dual is also the following quadratic programming:

$$\begin{aligned} \min \quad & y^T A Q^{-1} A^T y \\ \text{s.t.} \quad & (b^T + q^T Q^{-1} A^T) y = 1, \\ & y \geq 0. \end{aligned} \tag{1.2.8}$$

Example 1.14 (Optimization problems with the ℓ_p-norm [96])**.**

$$\begin{aligned} \min_{x,y} \quad & \|x\|_p \\ \text{s.t.} \quad & Bx + Cy \geq d, \end{aligned} \tag{1.2.9}$$

where $1 \leq p \leq \infty$. The gauge dual of the above problem is written as

$$\begin{aligned} \min_{z} \quad & \|B^T z\|_q \\ \text{s.t.} \quad & C^T z = 0, \\ & d^T z = 1, \\ & z \geq 0, \end{aligned} \tag{1.2.10}$$

where q satisfies $1/p + 1/q = 1$.

Existing results and issues

Freund [30] investigated the gauge duality of problems (1.2.1) and (1.2.2). One of the results is the weak duality in the gauge framework, which indicates that

$$g(x)\, g(y) \geq 1 \tag{1.2.11}$$

holds for feasible solutions x and y of problems (1.2.1) and (1.2.2), respectively. He also showed the sufficient condition for optimality, which states x and y are optimal for problems (1.2.1) and (1.2.2) if $g(x)\,g(y) = 1$. Another result is that the gauge dual of (1.2.2) becomes the original problem (1.2.1).

Although problem (1.2.1) is slightly abstract to apply the gauge duality theory to more concrete forms of optimization problems, the paper [30] shows that $\mathcal{X}$ and $\bar{\mathcal{X}}$ can be written explicitly when $\mathcal{X} := \{x \mid Ax \geq b\}$. In that case, primal and dual gauge optimization pare is concretely represented as

$$\begin{array}{ll} \min & g(x) \\ \text{s.t.} & Ax \geq b, \end{array} \tag{1.2.12}$$

and

$$\begin{array}{ll} \max & b^T y \\ \text{s.t.} & g^\circ(M^T y) \leq 1, \\ & y \geq 0, \end{array} \tag{1.2.13}$$

respectively. Then the above duality theory can be applied to the well-known problems such as linear programming with positive optimal value, the ℓ_p-norm optimization problems with $p \in [1, \infty]$, and convex quadratic optimization problems. Considering applications to wider fields; however, it is desirable to handle nonlinear constraints in a concrete form analogous to the linear cases (1.2.12) and (1.2.13).

1.2.2 Gauge optimization problems in a specific form

More recently, Friedlander et al. [32] considered the following GO problem with a nonlinear constraint in a concrete form:

$$\begin{array}{ll} \min\limits_{x \in \mathcal{X}} & g(x) \\ \text{s.t.} & h(b - Ax) \leq \sigma, \end{array} \tag{1.2.14}$$

where $\mathcal{X}$ is a finite-dimensional Euclidean space, g and h are gauge functions, σ is a scalar, and b, A are, respectively, a vector and a matrix with appropriate dimensions. Although problem (1.2.14) is less general comparing to (1.2.1), it is sufficiently concrete to analyze theoretical properties and to apply the problem to more useful optimization problems. They also show the gauge dual of (1.2.14) as follows:

$$\begin{array}{ll} \min\limits_{y \in \mathcal{X}} & g^\circ(A^T y) \\ \text{s.t.} & y^T b - \sigma h^\circ(y) \geq 1. \end{array} \tag{1.2.15}$$

We show two examples of the above GO problems, which are finding minimum-length solutions and sparse optimization problem.

Example 1.15 (Norm and minimum length solutions [21, 32, 102]). *The primal problem that minimizes the norm of x under linear constraints and its gauge dual are described as follows:*

$$\begin{array}{ll} \min & \|x\| \\ \text{s.t.} & Ax = b, \end{array} \tag{1.2.16}$$

$$\begin{array}{ll} \min & \|A^T y\|_* \\ \text{s.t.} & b^T y \geq 1, \end{array} \tag{1.2.17}$$

where $\|\cdot\|_*$ denotes the dual norm.

Example 1.16 (Sparse optimization [32, 92, 93]). *The primal problem is solved in the field of sparse optimization, and its gauge dual is described by using the ℓ_∞-norm as follows:*

$$\begin{array}{ll} \min & \|x\|_1 \\ \text{s.t.} & \|Ax - b\|_2 \leq \sigma, \end{array} \tag{1.2.18}$$

$$\begin{array}{ll} \min & \|A^T y\|_\infty \\ \text{s.t.} & b^T y - \sigma \|y\|_2 \geq 1. \end{array} \tag{1.2.19}$$

The following example shows that a linear conic optimization problem can be seen as a GO problem.

Example 1.17 (Linear conic optimization [32, 75, 87]). *Consider the primal and dual pair of linear conic optimization*

$$\begin{array}{ll} \min & c^T x \\ \text{s.t.} & Ax = b, \ x \in \mathcal{K}, \end{array} \tag{1.2.20}$$

and

$$\begin{array}{ll} \max & b^T y \\ \text{s.t.} & c - A^T y \in \mathcal{K}^*. \end{array} \tag{1.2.21}$$

The above primal problem can be transformed into the GO problem by setting $\hat{c} = c - A^T \hat{y}$, where $\hat{y}$ is a feasible solution of the dual problem. Then the primal problem can be seen as the following GO problem:

$$\begin{array}{ll} \min & \hat{c}^T x + \delta_{\mathcal{K}}(x) \\ \text{s.t.} & Ax = b. \end{array} \tag{1.2.22}$$

Existing results and issues

In order to analyze the duality relationships, they first investigate a general GO problem (1.2.1) and its Fenchel dual. The Fenchel dual of (1.2.1) are described as

$$\begin{aligned} \max_{y} \quad & -(\delta_{\mathcal{X}})^*(-y) \\ \text{s.t.} \quad & g^\circ(y) \le 1. \end{aligned} \tag{1.2.23}$$

Denote v_p, v_{gd}, v_{fd} as the optimal values of problems (1.2.1), (1.2.2), and (1.2.23), respectively. Then we have

$$v_p \ge v_{fd} = \frac{1}{v_{gd}} > 0.$$

Moreover, if y^* is the solution of (1.2.23), then y^*/v_{fd} is the solution of (1.2.2). Conversely, if y^* solves (1.2.2), then $y^* v_{fd}$ solves (1.2.23).

For the more concrete GO (1.2.14), they showed its Lagrangian dual as follows:

$$\begin{aligned} \max_{y \in \mathcal{X}} \quad & b^T y - \sigma h^\circ(y) \\ \text{s.t.} \quad & g^\circ(A^T y) \le 1, \end{aligned} \tag{1.2.24}$$

and investigated the relations of solutions between problems (1.2.15) and (1.2.24) by using duality results about the Fenchel dual. Let v_{ld} denote the optimal value of (1.2.24). They prove that $v_{fd} = v_{ld}$ under some assumptions. In addition, if y^* is a solution of (1.2.24), then y^*/v_{ld} is the solution of (1.2.15). Also, if y^* solves (1.2.15) and $v_{gd} > 0$, then $y^* v_{ld}$ solves (1.2.24).

Aravkin et al. [6] presented other theoretical results for the GO problems (1.2.14) and (1.2.15). 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 by using the so-called *perspective* functions. In addition, they proposed the perspective duality, which is an extension of the gauge duality.

The applications of the gauge duality exist in developing algorithms by taking advantage of the gauge dual rather than the Lagrangian dual. The constraint of the gauge dual problem (1.2.15) is simpler compared with that of the Lagrangian dual (1.2.24). Therefore the computational cost of algorithms including projection can be low. Using this advantage, Friedlander and Macêdo [31] applied this gauge duality to solve low-rank spectral optimization problems. Also, Aravkin et al. [7] developed level-set methods for convex optimization problems.

The GO problems in these previous works [6, 30–32] do not involve linear terms in their objective functions. Therefore, these GO frameworks cannot directly handle such as linear conic optimization problems. Moreover, differently from AVO (1.1.5), GO problem (1.2.14) does not consider multiple constraints, but only one gauge constraint. It seems that these drawbacks make us unable to apply the GO framework to wider applications.

1.3 Motivations and contributions

The study on AVO is in its infancy, and, to the author's knowledge, there have been no works except for the above-mentioned duality results of Mangasarian [61] and the application of Qian et al. [78]. Moreover, there is no algorithm to solve AVO (1.1.5) in a general form. Although many theoretical and algorithmic results for AVEs exist, the research associated with AVO problems is not sufficient. One of the reasons seems that it is difficult to obtain even feasible solutions of the problem. In fact, their constraints include some AVEs, which are known to be NP-hard [61] to solve. Moreover, the ability of the expression of the absolute values seems not enough to model real world problems. On the other hand, the gauge function generalizes the absolute value function and includes a wider class of functions such as the ℓ_2-norm and the ℓ_p-norm, $p \geq 1$. However, the previous works related to the GO problems only involve one constraint that includes only one gauge function term.

In this thesis, we focus on investigating the theoretical property of the AVO problem and its generalizations and develop an algorithm for the AVO problem. From these aspects, the motivations and contributions of the thesis are itemized as follows.

(1) To generalize the AVO problem and analyze its theoretical results

We generalize the AVO problem by replacing the absolute value function with the positively homogeneous one. Examples of the positively homogeneous function include the absolute value function and the ℓ_p-norm function, where p is a positive real number. The problem is an extension of the AVO and is not necessarily convex. We call such a problem as positively homogeneous optimization and propose its dual formulation in a closed-form analogous to the AVO dual. Then we discuss the extension of the weak duality result investigated by Mangasarian for the AVO. We also study the relation between the dual formulation and the Lagrangian dual.

(2) To investigate the more general GO problem and its extension

The positively homogeneous function includes the gauge one. Then we can consider an optimization problem with multiple constraints involving gauge functions compared to the existing results which consider only one gauge constraint. Moreover, we extend the generalized GO problem to a general optimization problem by using the perspective framework.

(3) To develop an algorithm for the AVO problem

An algorithm for a specific AVO was proposed in [78]; however, to the best of the author's knowledge, there is no method to solve the general AVO (1.1.5). In this thesis, we develop a global optimization algorithm for AVO (1.1.5) by using the idea of a branch-and-bound procedure. We provide numerical experiments for some facility location problems and verify the effectiveness of the proposed algorithm.

1.4 Outline of the thesis

The thesis is organized as follows.

In Chapter 2, we introduce some preliminaries including notations, basic mathematical properties, and existing results that are necessary in the later Chapters.

In Chapter 3, we propose a dual formulation that, differently from the Lagrangian dual approach, has a closed-form and some interesting properties. In particular, we discuss the relation between the Lagrangian duality and the one proposed here and give some sufficient conditions under which these dual problems coincide. Finally, we show that some well-known problems, e.g., the sum of norms optimization and the group Lasso-type optimization problems, can be reformulated as positively homogeneous optimization problems.

In Chapter 4, 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 with 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.

In Chapter 5, we first propose an algorithm for the AVO, which is based on the branch-and-bound method. In the branching procedure, we generate two subproblems by restricting the sign of a component of the variable x in the primal AVO problem to be nonnegative or nonpositive. In the bounding procedure, we utilize the duality results in AVO to obtain a lower bound for each subproblem. Furthermore, to examine the effectiveness of the proposed algorithm, we apply it to solve facility location problems (FLPs). By using the ℓ_1-norm as a distance function, an FLP can naturally be formulated as an AVO. In particular, we can use the AVO formulation to deal with a nonconvex region in which facilities are located. We stress that such a problem is considerably difficult to solve compared with the conventional FLPs that assume the convexity of the region.

In Chapter 6, we conclude the thesis and mention some future issues.

Chapter 2

Preliminaries

In this chapter, we give mathematical notations and review the properties of theories and algorithms that will be appeared in this thesis.

2.1 Notations

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 n 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. We also denote by $\|\cdot\|$ the norm. For a function f and vectors x and y, we denote the subdifferential of $f(x, y)$ with respect to x as $\partial_x f(x, y)$. The effective domain of a function f is given by $\mathrm{dom} f$. The convex hull of a set S is denoted by $\mathrm{co}S$. Finally, $\delta_S \colon \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is an indicator function of a set $S \subseteq \mathbb{R}^n$ defined by

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

2.2 Convex analysis

We summarize the definitions related to sets and functions in convex analysis as follows.

Definition 2.1. *(Convex Sets) A set $S \subseteq \mathbb{R}^n$ is convex if*

$$\alpha x + (1-\alpha)y \in S$$

holds for any $x, y \in S$ and $\alpha \in [0,1]$.

Definition 2.2. *(Interior) A vector $x \in S \subseteq \mathbb{R}^n$ is an interior point of S if there exists an $\epsilon > 0$ such that*

$$\{y \mid \|y - x\|_2 \leq \epsilon\} \subseteq S.$$

The set of all interior points of S is called the interior of S and is denoted by $\mathrm{int}S$.

Definition 2.3. *(Open and closed set) A set $S \subseteq \mathbb{R}^n$ is open if $\mathrm{int}S = S$. A set $S \subseteq \mathbb{R}^n$ is closed if its complement set: $\mathbb{R}^n \setminus S = \{x \in \mathbb{R}^n \mid x \notin S\}$ is open.*

Definition 2.4. *(Closure and boundary) The closure of a set $S \subseteq \mathbb{R}^n$ is denoted $\mathrm{cl}S$ and is defined as $\mathrm{cl}S = \mathbb{R}^n \setminus \mathrm{int}(\mathbb{R}^n \setminus S)$. The boundary of a set $S \subseteq \mathbb{R}^n$ is denoted $\mathrm{bd}S$ and is defined as $\mathrm{bd}S = \mathrm{cl}S \setminus \mathrm{int}S$.*

Definition 2.5. *(Effective domain) The* effective domain *of a function $f \colon \mathbb{R}^n \to \mathbb{R} \cup +\infty$ is defined by $\mathrm{dom}f := \{x \mid f(x) < +\infty\}$.*

Definition 2.6. *(Convex function) A function $f \colon \mathbb{R}^n \to \mathbb{R}$ is convex if $\mathrm{dom}f$ is convex set 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 $\alpha \in [0,1]$.

Definition 2.7. *(Closed function) A function $f \colon \mathbb{R}^n \to \mathbb{R}$ is closed if the sublevel set*

$$\{x \in \mathrm{dom}f \mid f(x) \leq \alpha\}$$

for each $\alpha \in \mathbb{R}$ is closed.

2.2.1 Positively homogeneous function and its polar

Definition 2.8. *(Positively homogeneous functions) A function $\psi\colon \mathbb{R}^n \to \mathbb{R}$ is* positively homogeneous *if the following inequality holds:*

$$\psi(\lambda x) = \lambda\psi(x) \qquad \text{for all } x \in \mathbb{R}^n, \lambda \in \mathbb{R}_{++}.$$

Definition 2.9. *(Polar positively homogeneous functions) Let $\psi\colon \mathbb{R}^n \to \mathbb{R}$ be a positively homogeneous function. Then, $\psi^\circ\colon \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ defined by*

$$\psi^\circ(y) := \sup\{x^T y \mid \psi(x) \le 1\} \qquad \text{for all } y \in \mathbb{R}^n$$

is called the polar positively homogeneous *function of ψ.*

We observe that the ℓ_1-norm function is convex and positively homogeneous. Note that the ℓ_p-norm function, $0 < p < 1$, is positively homogeneous but nonconvex. The polar of both the ℓ_1-norm and the ℓ_p-norm, $0 < p < 1$, is the ℓ_∞-norm function.

2.2.2 Gauge function and its properties

Let $\mathcal{X} \subseteq \mathbb{R}^n$ be a closed convex set. If $g\colon \mathcal{X} \to \mathbb{R} \cup \{\infty\}$ is a gauge function, then g is convex, nonnegative, positively homogeneous, and satisfies $g(0) = 0$. It is clear that the gauge function includes the usual norm such as the ℓ_p-norm, $p \ge 1$.

All gauge functions are represented as a *Mincowski function* $\gamma_{\mathcal{C}}$ of some nonempty convex set $\mathcal{C}$ as follows:

$$g(x) = \gamma_{\mathcal{C}} := \inf\{\, \lambda \ge 0 \,|\, x \in \lambda\mathcal{C} \,\}. \tag{2.2.1}$$

The *polar of the gauge function* g is defined by

$$g^\circ(y) := \inf\{\, \mu > 0 \,|\, x^T y \le \mu g(x), \forall x \,\}$$

From this definition, we obtain the following Cauchy-Schwartz-like inequality

$$x^T y \le g(x) g^\circ(y), \quad \forall x, \forall y.$$

Note that if convex set $\mathcal{C}$ is a unit ball defined by $\mathcal{C} := \{x \mid \|x\|_2 \le 1\}$, then gauge function (2.2.1) becomes the ℓ_2-norm: $g(x) = \|x\|_2$. This is also the case when the unit ball is defined by any norm function. We also note that set $\mathcal{C}$ could not be symmetric like norms. In such cases, the gauge function calculate different value at two points located symmetric to the origin.

We itemize some important properties of the gauge function and its polar [32].

Proposition 2.1. *The gauge function* $g\colon \mathcal{X} \to \mathbb{R} \cup \{\infty\}$ *satisfies the following properties:*

(i) g° *is a closed gauge function*

(ii) $g^{\circ\circ} = \operatorname{cl} g$

(iii) $g^\circ(y) = \sup_x \{\, x^T y \mid g(x) \le 1 \,\}$ *for all* y

(iv) $\operatorname{dom} g^\circ = \mathcal{X}$ *if* g *is closed and* $g^{-1}(0) = \{0\}$

(v) $\operatorname{epi} g^\circ = \{\, (y, \lambda) \mid (y, -\lambda) \in (\mathrm{epi})^\circ \,\}$

Proposition 2.2. *Let* g_1 *and* g_2 *be gauges. Then* $g(x_1, x_2) := g_1(x_1) + g_2(x_2)$ *is a gauge, and its polar is represented as*

$$g^\circ(y_1, y_2) = \max\{g_1^\circ(y_1), g_2^\circ(y_2)\}.$$

2.3 Optimality conditions and duality

2.3.1 Lagrangian function and Lagrangian duality

We consider the following optimization problem in general form:

$$\begin{array}{rl} \min & f(x) \\ \text{s.t.} & g_i(x) \le 0, \quad i = 1, \dots, m, \\ & h_j(x) = 0, \quad j = 1, \dots, p, \end{array} \tag{P$_0$}$$

where $f\colon \mathbb{R}^n \to \mathbb{R}$, $g_i, \colon \mathbb{R}^n \to \mathbb{R}$ and $h_j\colon \mathbb{R}^n \to \mathbb{R}$. We define the domain of (P$_0$) as $\mathcal{D} = \mathrm{dom} f \cap \bigcap_{i=1}^m \mathrm{dom} g_i \cap \bigcap_{j=1}^p \mathrm{dom} h_j$ and assume that it is nonempty. Then the *Lagrangian* function associated with problem (P$_0$) is defined as

$$\mathcal{L}(x, u, v) := f(x) + \sum_{i=1}^m u_i g_i(x) + \sum_{j=1}^p v_j h_j(x),$$

where u_i, $i = 1, \dots, m$ and v_j, $j = 1, \dots, p$ are called the *Lagrangian* multipliers associated with the inequality and equality constraints, respectively. Note that $\mathrm{dom}\mathcal{L} = \mathcal{D} \times \mathbb{R}^m \times \mathbb{R}^p$.

We also define the *Lagrange dual function* $\omega\colon \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ as

$$\omega(u, v) := \inf_{x \in \mathcal{D}} \mathcal{L}(x, u, v),$$

which is the minimum value of the Lagrangian function with respect to x.

The Lagrangian dual problem associated with problem (P_0) is defined by

$$\begin{array}{rl} \sup & \omega(u,v) \\ \text{s.t.} & u \geq 0 \end{array} \tag{D_0}$$

Note that the objective function of problem (D_0) is always concave even if the primal problem (P_0) is not convex.

We denote f^* and ω^* are the optimal values of problems (P_0) and (D_0), respectively. Then the following inequality holds:

$$\omega^* \leq f^*$$

even if problem (P_0) is not convex. We refer to this property *weak duality.*

We call the difference of the two sides of the above inequality, which is $\Delta := f^* - \omega^*$, as the *optimal duality gap.* When $\Delta = 0$, we say that there exists no duality gap. Note that the duality gap Δ is always zero in the case of linear programming.

When $\Delta = 0$, i.e., the equality

$$\omega^* = f^*$$

holds, we say that *strong duality* holds for problem (P_0) and (D_0).

Strong duality does not hold for general optimization problems. However, if problem (P_0) is convex described as

$$\begin{array}{rl} \min & f(x) \\ \text{s.t.} & g_i(x) \leq 0, \quad i = 1, \ldots, m, \\ & Ax = b, \quad j = 1, \ldots, p, \end{array}$$

where f and g_i, $i = 1, \ldots, m$ are convex, strong duality holds under some conditions. Such conditions are called *constraint qualifications.*

One of the constraint qualifications is *Slater's condition*, which says that there exist an $x \in \mathcal{D}$ such that

$$g_i(x) < 0, \quad i = 1, \ldots, m, \qquad Ax = b.$$

2.3.2 Karush-Kuhn-Tucker conditions

For problem (P_0), we assume f, g_i, h_j are differentiable but not necessarily convex.

Let x^*, and (u^*, v^*) be optimal solutions with zero duality gap for problem (P_0) and (D_0), respectively. Then, the *Karush-Kuhn-Tucker* (KKT) conditions are described as follows:

$$\begin{array}{rcll} \nabla_x \mathcal{L}(x^*, u^*, v^*) = f(x^*) + \sum_{i=1}^m u_i^* g_i(x^*) + \sum_{j=1}^p v_j^* h_j(x^*) & = & 0, & \\ g_i(x^*) & \leq & 0, & i = 1, \ldots, m \\ u_i^* & \geq & 0, & i = 1, \ldots, m \\ g_i(x^*) u_i^* & = & 0, & i = 1, \ldots, m \\ h_j(x^*) & = & 0, & j = 1, \ldots, p \end{array}$$

The KKT conditions are the necessary optimality conditions for any optimization problem. Note that the KKT conditions are also sufficient optimality conditions if the primal problem is convex and satisfies the Slater's condition.

2.3.3 Gauge duality

Consider the following GO problem in general form:

$$\begin{array}{ll} \min & g(x) \\ \text{s.t.} & x \in \mathcal{X}, \end{array} \tag{2.3.1}$$

where $\mathcal{X} \subseteq \mathbb{R}^n$ is a closed convex set and $g \colon \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is a gauge function. The gauge dual of problem (2.3.1) is defined as follows:

$$\begin{array}{ll} \min & g^\circ(y) \\ \text{s.t.} & y \in \bar{\mathcal{X}}, \end{array} \tag{2.3.2}$$

where g° is the polar function of g and $\bar{\mathcal{X}}$ is the anti-polar set of $\mathcal{X}$ defined by

$$\bar{\mathcal{X}} := \{y \in \mathbb{R}^n \mid y^T x \geq 1 \text{ for all } x \in \mathbb{R}^n\}. \tag{2.3.3}$$

We also consider the following GO problem in a specific form:

$$\begin{array}{ll} \min\limits_{x \in \mathcal{X}} & g(x) \\ \text{s.t.} & h(b - Ax) \leq \sigma, \end{array} \tag{2.3.4}$$

where $\mathcal{X}$ is a finite-dimensional Euclidean space, g and h are gauge functions, σ is a scalar, and b and A are , respectively, a vector and a matrix with appropriate dimensions. The gauge dual of (2.3.4) are described as follows:

$$\begin{array}{ll} \min\limits_{y \in \mathcal{X}} & g^\circ(A^T y) \\ \text{s.t.} & y^T b - \sigma h^\circ(y) \geq 1. \end{array} \tag{2.3.5}$$

2.4 Global optimization algorithm

Global optimization algorithms focus on finding global optimal solutions of an optimization problem. Those algorithms are broadly classified into deterministic approaches and stochastic ones [27,28]. Some of the deterministic approach are branch-and-bound algorithms [2,3,68,101], cutting plane methods [91,101], difference of convex functions and reverse convex methods [90], and primal-dual methods [28,29,95]. The stochastic approach includes simulated annealing [50], genetic algorithms [37], for instance.

Here we provide the details of the branch-and-bound procedure, which we use to solve AVO (1.1.5). The branch-and-bound procedure is one of the popular and widely-used global optimization algorithms which provides exact global optimal solutions to NP-hard optimization problems. The method consists of branching and bounding procedures. In the branching procedure, we divide the feasible region of the original problem into some subregions to generate subproblems. On the other hand, in the bounding procedure, we check if a current subproblem can be discarded or not, by implementing some fathoming tests. The general branch-and-bound procedure for an optimization problem

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & x \in X \end{array}$$

are described as follows. Note that P refers to the above problem and $\hat{x}$ denotes an incumbent solution in the following description of the algorithm.

General Branch-and-Bound Algorithm

- **Step 0.** Set $\mathcal{A} := \{\mathrm{P}\}$ and initialize $\hat{x}$.
- **Step 1.** Choose a subproblem $\hat{\mathrm{P}}$ from the set $\mathcal{A}$ to explore.
  - **Step 1-a.** If $\hat{\mathrm{P}}$ is solved and a solution $\hat{x}'$ satisfying $f(\hat{x}') < f(\hat{x})$ found, set $\hat{x} = \hat{x}'$. Set $\mathcal{A} := \mathcal{A} \setminus \{\hat{\mathrm{P}}\}$ and go to Step 3.
  - **Step 1-b.** If P can not pruned, go to Step 2.
- **Step 2.** Generate subproblems $\mathrm{P}_1, \mathrm{P}_2, \ldots, \mathrm{P}_\mathrm{n}$ from P by dividing the feasible region X. Set $\mathcal{A} := \mathcal{A} \cup \{\bigcup_{i=1}^{n} \mathrm{P}_\mathrm{i}\} \setminus \{\hat{\mathrm{P}}\}$, and return to Step 1.
- **Step 3.** If $\mathcal{A} = \emptyset$, then terminate. The incumbent solution is an optimal solution of the original problem. Otherwise, return to Step 1.

Chapter 3

Duality of optimization problems with positively homogeneous functions

3.1 Introduction

Recently, the so-called *absolute value equations* (AVEs) and *absolute value optimization* (AVO) problems have been attracted much attention. The AVEs were introduced in 2004 by Rohn [82]. Basically, if $\tilde{A}, \tilde{B}$ are given matrices, and $\tilde{b}$ is a given vector, one should find a vector x that satisfies $\tilde{A}x + \tilde{B}|x| = \tilde{b}$, where $|x|$ is a vector whose i-th entry is the absolute value of the i-th entry of x. It is known that AVEs are equivalent to the *linear complementarity problems* (LCP) [40,64,77], which include many real-world applications. As an extension of AVEs, Mangasarian [61] proposed in 2007 the AVO problems, which have the absolute value of variables in their objective and constraint functions. More precisely, the AVO problem considered is given by

$$\begin{array}{ll} \min & \tilde{c}^T x + \tilde{d}^T |x| \\ \text{s.t.} & \tilde{A}x + \tilde{B}|x| = \tilde{b}, \\ & \tilde{H}x + \tilde{K}|x| \geq \tilde{p}, \end{array}$$

where $\tilde{A}, \tilde{B}, \tilde{H}, \tilde{K}$ are given matrices, and $\tilde{c}, \tilde{d}, \tilde{b}, \tilde{p}$ are vectors with appropriate dimensions. Since AVEs and LCP are equivalent, the AVO include the *mathematical programs with linear complementarity constraints* [57], which are one of the formulations of equilibrium problems. As another application of AVO, Yamanaka and

Fukushima [107] presented facility location problems and Qian et al. [78] showed a scheduling problem about automated vehicles.

Since 2007, some methods for solving AVEs have been presented in the literature. For example, Rohn [83] considered an iterative algorithm using the sign of variables for the case that $\tilde{A}$ and $\tilde{B}$ are square matrices. For more general $\tilde{A}$ and $\tilde{B}$, Mangasarian [61] provided a method involving successive linearization techniques. Other methods include a concave minimization approach, given by Mangasarian [60], and Newton-type methods, proposed by Caccetta et al. [16], Mangasarian [62], and Zhang and Wei [112]. Some generalizations of AVEs were also proposed. For example, Hu et al. [41] considered an AVE involving the absolute value of variables associated with the second-order cones. Miao et al. [66] investigated an AVE with the so-called circular cones. In both papers, quasi-Newton based algorithms were used.

As for AVO problems, Yamanaka and Fukushima [107] proposed to use a branch-and-bound technique. In the branching procedure, two subproblems are generated by fixing the sign of a variable as nonnegative or nonpositive. In the bounding procedure, the dual information are considered. However, to the best of our knowledge, there are no other methods that can find a global solution of AVO. When comparing to AVEs, the research associated with AVO problems is insufficient and one of these reasons is the difficulty for obtaining feasible solutions of the problems. In fact, their constraints include AVEs, which are known to be NP-hard [61].

Another optimization problem that is related to AVO was recently investigated by Friedlander et al. [32] and Aravkin et al. [6]. It is called *gauge optimization*, which basically consists of an optimization problem with the so-called gauge function. However, differently from AVO, this problem does not consider multiple constraints, but only one gauge constraint. In [6,32], the authors showed that the Lagrange dual of gauge optimization problems can be written in a closed-form by using the polar of the gauge functions.

In this chapter, similarly to [6,32], we introduce a generalized AVO problem, and show that it has a wider practical application compared to AVO problems. It is also more general than gauge optimization problems, because multiple constraints can be considered here. The generalization is done by replacing the absolute value functions with *positively homogeneous* functions. Therefore, the problem uses not only absolute value terms but also, for instance, p-norm functions with $p \in (0, \infty]$. This generalized problem is referred here as *positively homogeneous optimization* (PHO).

Here, we introduce the PHO dual problem and compare it with the Lagrange dual. We also show that the weak duality theorem holds, similarly to the AVO problems [61]. In addition, we investigate the relation between the positively homogeneous duality and the Lagrange duality, proving that these dual problems are equivalent under some conditions. In this case, the Lagrange dual of a positively homogeneous problem can be written in a closed-form. We point out that the gauge functions are special cases of positively homogeneous functions, which are not necessarily convex, differently from the gauge. Moreover, the proposed problems here have linear and positively homogeneous terms in their objective functions and constraints, which is different from the problem considered in [6, 32] that has only one gauge term. Here, we also give some applications for the positively homogeneous problems, which include p-order cone optimization, sum of norms optimization, and group Lasso-type optimization problems, and we show that their Lagrange dual can be written in a closed-form even without convexity assumptions.

This chapter is organized as follows. In Section 3.2.1, we give the definition of positively homogeneous functions as well as their dual, showing some of their properties. In Section 3.2.2, we define the PHO problems, and we prove that weak duality holds. In Section 3.2.3, the relation between the Lagrangian dual and the positively homogeneous dual is discussed. We give some applications for PHO problems in Section 3.3.

3.2 Positively homogeneous optimization problems and their duality

3.2.1 Positively homogeneous functions

In this section, we first introduce the definitions of positively homogeneous and vector positively homogeneous functions. Then, we define their dual, which will be used to describe the dual of PHO problems. Moreover, we show some properties associated with these functions.

Definition 3.1. *(Positively homogeneous functions) A function* $\psi\colon \mathbb{R}^n \to \mathbb{R}$ *is* positively homogeneous *if the following inequality holds:*

$$\psi(\lambda x) = \lambda \psi(x) \qquad \text{for all } x \in \mathbb{R}^n, \lambda \in \mathbb{R}_{++}.$$

Definition 3.2. *(Vector positively homogeneous functions) A mapping* $\Psi\colon \mathbb{R}^n \to \mathbb{R}^m$ *is a* vector positively homogeneous *function if the following property holds:*

$$\Psi(x) = \begin{bmatrix} \psi_1(x_{I_1}) \\ \vdots \\ \psi_m(x_{I_m}) \end{bmatrix} \qquad \text{for all } x \in \mathbb{R}^n,$$

where $\psi_i\colon \mathbb{R}^{n_i} \to \mathbb{R}$ *is a positively homogeneous function for all* $i = 1, \ldots, m$, $n = n_1 + \cdots + n_m$, $I_i \subseteq \{1, \ldots, n\}$ *is a set of indices satisfying*

$$I_i \cap I_j = \emptyset, \quad i \neq j, \qquad \text{and} \qquad \# I_i = n_i,$$

and $x_{I_i} \in \mathbb{R}^{n_i}$ *is a disjoint subvector of* x.

The above definition basically says that Ψ is vector positively homogeneous if its block components are all positively homogeneous. We now introduce the polar function of ψ, which can be seen as a generalization of the dual norm. Similarly, we also define the polar of vector positively homogeneous functions.

Definition 3.3. *(Polar positively homogeneous functions) Let* $\psi\colon \mathbb{R}^n \to \mathbb{R}$ *be a positively homogeneous function. Then,* $\psi^\circ\colon \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ *defined by*

$$\psi^\circ(y) := \sup\{x^T y \mid \psi(x) \leq 1\} \qquad \text{for all } y \in \mathbb{R}^n$$

is called the polar positively homogeneous *function of* ψ.

Note that ψ° is convex by definition. In fact, for all $y, z \in \mathbb{R}^n$ and $\alpha \in (0,1)$, we have

$$\begin{aligned} \psi^\circ(\alpha y + (1-\alpha) z) &= \sup\{x^T(\alpha y + (1-\alpha) z) \mid \psi(x) \leq 1\} \\ &\leq \alpha \sup\{x^T y \mid \psi(x) \leq 1\} + (1-\alpha) \sup\{x^T z \mid \psi(x) \leq 1\} \\ &= \alpha \psi^\circ(y) + (1-\alpha) \psi^\circ(z). \end{aligned}$$

Definition 3.4. *(Polar vector positively homogeneous functions) Let* $\Psi\colon \mathbb{R}^n \to \mathbb{R}^m$ *be a vector positively homogeneous function. A function* $\Psi^\circ\colon \mathbb{R}^n \to \mathbb{R}^m$ *is a* polar vector positively homogeneous *function associated with* Ψ *if the following property holds:*

$$\Psi^\circ(y) = \begin{bmatrix} \psi_1^\circ(y_{I_1}) \\ \vdots \\ \psi_m^\circ(y_{I_m}) \end{bmatrix}, \quad i = 1, \ldots, m, \qquad \text{for all } y \in \mathbb{R}^n$$

where $\psi_i^\circ\colon \mathbb{R}^{n_i} \to \mathbb{R}$ *is the polar of positively homogeneous function* ψ_i *for each* $i = 1, \ldots m$.

In this paper, we assume two conditions for positively homogeneous functions.

Assumption 3.1. *Let $\Psi \colon \mathbb{R}^n \to \mathbb{R}^m$ be a vector positively homogeneous function as in Definition 3.2. Then, for all $i = 1, \dots, m$, the positively homogeneous function ψ_i satisfies the following conditions:*

1. $\psi_i(x_{I_i}) \geq 0 \quad \text{for all } x_{I_i} \in \mathbb{R}^{n_i}$,
2. If $x_{I_i} \neq 0$, then $\psi_i(x_{I_i}) > 0$.

From the definition of positively homogeneous functions, we observe that $\psi_i(0) = 0$. In fact, if $x = 0$ then $0 = \psi(\lambda x) - \lambda\psi(x) = (1-\lambda)\psi_i(0)$ for all $\lambda \in \mathbb{R}_{++}$. Moreover, the second condition of the above assumption shows that zero is the only point that satisfies $\psi_i(x) = 0$. We also observe that if ψ_i is taken as the usual vector norm, then it satisfies these assumptions. Note that under the above assumption, the polar function ψ_i° always takes finite values.

We now show an important property satisfied by vector positively homogeneous functions and their polar.

Proposition 3.1. *Let Ψ and Ψ° be a vector positively homogeneous function and its polar, respectively. Suppose that Assumption 3.1 holds. Then, the following inequalities hold:*

$$\begin{aligned} \Psi^\circ(y) &\geq 0, \\ \Psi(x)^T \Psi^\circ(y) &\geq x^T y \end{aligned}$$

for any $x, y \in \mathbb{R}^n$.

Proof. For simplicity, we take an arbitrary index i and denote ψ_i and x_{I_i} as ψ and x, respectively. From Definition 3.1, we have $\psi(0) = 0$. Using this result and Definition 3.3, we obtain

$$\psi^\circ(y) = \sup\{x^T y \mid \psi(x) \leq 1\} \geq 0 \qquad \text{for all } y \in \mathbb{R}^n.$$

This shows that $\Psi^\circ(y) \geq 0$ for all $y \in \mathbb{R}^n$ from Definition 3.4.

If $x = 0$, then the second inequality of this proposition clearly holds. If $x \neq 0$, then $\psi(x) > 0$ from Assumption 3.1 and so

$$\psi\left(\frac{x}{\psi(x)}\right) = \frac{1}{\psi(x)}\psi(x) = 1$$

holds once again from Definition 3.1. Therefore, we obtain

$$\psi^\circ(y) \geq \left(\frac{x}{\psi(x)}\right)^T y \qquad \text{for all } y \in \mathbb{R}^n.$$

Then, for all $x, y \in \mathbb{R}^n$, we have

$$\psi(x)\psi^\circ(y) \geq x^T y,$$

which indicates that

$$\Psi(x)^T \Psi^\circ(y) = \sum_{i=1}^m \psi_{I_i}(x)\psi^\circ_{I_i}(y) \geq \sum_{i=1}^m x^T_{I_i} y_{I_i} = x^T y.$$

□

3.2.2 Positively homogeneous optimization problems

We consider the following *positively homogeneous optimization* (PHO) problem:

$$\begin{array}{ll} \min & c^T x + d^T \Psi(x) \\ \text{s.t.} & Ax + B\Psi(x) = b, \\ & Hx + K\Psi(x) \geq p, \end{array} \tag{$\mathrm{P_{PHO}}$}$$

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, and $\Psi\colon \mathbb{R}^n \to \mathbb{R}^m$ is a vector positively homogeneous function satisfying Assumption 3.1.

Now we give the Lagrangian dual of the problem ($\mathrm{P_{PHO}}$) as follows:

$$\sup_{\substack{u\\ v\geq 0}} \omega(u, v), \tag{$\mathrm{D^{\mathcal{L}}_{PHO}}$}$$

where $\omega\colon \mathbb{R}^k \times \mathbb{R}^\ell \to \mathbb{R}$ is given by

$$\omega(u, v) := \inf_x \mathcal{L}(x, u, v), \tag{3.2.1}$$

and $\mathcal{L}\colon \mathbb{R}^n \times \mathbb{R}^k \times \mathbb{R}^\ell \to \mathbb{R}$ is the Lagrangian function of ($\mathrm{P_{PHO}}$) defined by

$$\begin{aligned} \mathcal{L}(x, u, v) &:= c^T x + d^T \Psi(x) + u^T(b - Ax - B\Psi(x)) + v^T(p - Hx - K\Psi(x)) \\ &= b^T u + p^T v - (A^T u + H^T v - c)^T x + (d - B^T u - K^T v)^T \Psi(x), \end{aligned}$$

with $u \in \mathbb{R}^k$ and $v \in \mathbb{R}^\ell$ as the *Lagrange multipliers* associated with the equality and inequality constraints, respectively. Notice that it is difficult to write concretely the objective function of the problem ($\mathrm{D}^{\mathcal{L}}_{\mathrm{PHO}}$) because it is, in general, not convex with respect to x.

To obtain a closed-form dual problem, we consider a convex relaxation of the original problem ($\mathrm{P}_{\mathrm{PHO}}$) and its Lagrangian dual. For simplicity, we investigate the case where $\Psi(x) = |x| := (|x_1|, \ldots, |x_n|)^T$, and ($\mathrm{P}_{\mathrm{PHO}}$) has a linear objective function and only inequality constraints. More precisely, we analyze the following problem:

$$\begin{array}{ll} \min & c^T x \\ \text{s.t.} & Ax + B|x| \geq b. \end{array} \tag{$\mathrm{P_a}$}$$

If we set $x = x^+ - x^-$ and $|x| = x^+ + x^-$, where $x_i^+ = \max\{0, x_i\}$ and $x_i^- = \max\{0, -x_i\}$, then we can write ($\mathrm{P_a}$) as

$$\begin{array}{ll} \min & [c^T | -c^T] \begin{bmatrix} x^+ \\ x^- \end{bmatrix} \\ \text{s.t.} & [A| -A] \begin{bmatrix} x^+ \\ x^- \end{bmatrix} + [B|B] \begin{bmatrix} x^+ \\ x^- \end{bmatrix} \geq b, \end{array}$$

which is equivalent to the following problem:

$$\begin{array}{ll} \min & [c^T | -c^T] \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \\ \text{s.t.} & [A| -A] \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} + [B|B] \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \geq b, \\ & y_1, y_2 \geq 0, \\ & y_1^T y_2 = 0, \end{array}$$

where $y_1, y_2 \in \mathbb{R}^n$. Notice that the above problem is not convex due to the complementarity constraint $y_1^T y_2 = 0$. Therefore, we remove it from the problem and obtain the following relaxed one:

$$\begin{array}{ll} \min & [c^T | -c^T]\; y \\ \text{s.t.} & [A + B | -A + B]\; y \geq b, \\ & y \geq 0, \end{array}$$

where $y = (y_1, y_2)^T$. This problem is just a linear programming, then its Lagrangian dual can be written easily as

$$\begin{array}{rl} \max & b^T u \\ \text{s.t.} & \begin{bmatrix} A^T + B^T \\ -A^T + B^T \end{bmatrix} u \leq \begin{bmatrix} c \\ -c \end{bmatrix}, \\ & u \geq 0. \end{array}$$

Observing that the first constraint is equivalent to $|A^T u - c| + B^T u \leq 0$, we finally obtain the following closed-form dual problem:

$$\begin{array}{rl} \max & b^T u \\ \text{s.t.} & |A^T u - c| + B^T u \leq 0, \\ & u \geq 0. \end{array} \tag{$\mathrm{D_a}$}$$

In fact, the problem ($\mathrm{D_a}$) is the AVO dual of ($\mathrm{P_a}$) proposed by Mangasarian in [61], and the weak duality clearly holds in this case.

Let us return to the general problem ($\mathrm{P_{PHO}}$). Inspired by the above AVO dual problem ($\mathrm{D_a}$), we consider the following problem as the positively homogeneous dual problem:

$$\begin{array}{rl} \max & b^T u + p^T v \\ \text{s.t.} & \Psi^\circ(A^T u + H^T v - c) + B^T u + K^T v \leq d, \\ & v \geq 0, \end{array} \tag{$\mathrm{D_{PHO}}$}$$

where Ψ° is the polar vector positively homogeneous function associated with Ψ. Note that ($\mathrm{D_{PHO}}$) is a convex optimization problem since each component ψ_i° of Ψ° is a convex function.

The theorem below shows that the proposed dual problem ($\mathrm{D_{PHO}}$) is reasonable, in the sense that the weak duality holds between ($\mathrm{P_{PHO}}$) and ($\mathrm{D_{PHO}}$).

Theorem 3.1. *(Weak duality) For problems* ($\mathrm{P_{PHO}}$) *and* ($\mathrm{D_{PHO}}$)*, the following inequality holds:*

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

for all feasible points $x \in \mathbb{R}^n$ and $(u, v) \in \mathbb{R}^k \times \mathbb{R}^\ell$ of ($\mathrm{P_{PHO}}$) *and* ($\mathrm{D_{PHO}}$)*, respectively.*

Proof. Let $x \in \mathbb{R}^n$ and $(u, v) \in \mathbb{R}^k \times \mathbb{R}^\ell$ be feasible for ($\mathrm{P_{PHO}}$) and ($\mathrm{D_{PHO}}$), respectively. Then, we have

$$\begin{aligned} c^T x + d^T \Psi(x) &\geq c^T x + (\Psi^\circ(A^T u + H^T v - c) + B^T u + K^T v)^T \Psi(x) \\ &= c^T x + \Psi^\circ(A^T u + H^T v - c)^T \Psi(x) + u^T B \Psi(x) + v^T K \Psi(x), \end{aligned}$$

where the inequality holds from the first constraint of ($\mathrm{D_{PHO}}$) and the nonnegativity of Ψ. From the second inequality of Proposition 3.1, we also obtain:

$$\begin{aligned} c^T x + d^T \Psi(x) &\geq c^T x + (A^T u + H^T v - c)^T x + u^T B\Psi(x) + v^T K\Psi(x) \\ &= u^T(Ax + B\Psi(x)) + v^T(Hx + K\Psi(x)). \end{aligned}$$

Finally, the constraints of ($\mathrm{P_{PHO}}$) gives

$$c^T x + d^T \Psi(x) \geq b^T u + p^T v,$$

which completes the proof. □

The weak duality theorem itself is a powerful theoretical result, but it does not mention how large the duality gap between ($\mathrm{P_{PHO}}$) and ($\mathrm{D_{PHO}}$) is. And the duality gap can be large depending on problems, then the dual problem ($\mathrm{D_{PHO}}$) may be useless. Therefore, in the next section, we investigate the relation between the Lagrangian dual problem ($\mathrm{D^{\mathcal{L}}_{PHO}}$) and the one ($\mathrm{D_{PHO}}$) proposed here. As a result, surprisingly, we find that ($\mathrm{D^{\mathcal{L}}_{PHO}}$) and ($\mathrm{D_{PHO}}$) are equivalent.

3.2.3 The positively homogeneous duality and the Lagrangian duality

In this section, we consider the relation between the positively homogeneous duality and the more traditional Lagrangian duality of problem ($\mathrm{P_{PHO}}$), investigating conditions under which the Lagrangian dual problem ($\mathrm{D^{\mathcal{L}}_{PHO}}$) and the positively homogeneous dual problem ($\mathrm{D_{PHO}}$) are equivalent. Notice that the equivalence means the optimal values and solutions of ($\mathrm{D_{PHO}}$) and ($\mathrm{D^{\mathcal{L}}_{PHO}}$) are the same. Recalling (3.2.1), we first show a condition that makes $\omega(\bar{u}, \bar{v})$, the objective function of ($\mathrm{D^{\mathcal{L}}_{PHO}}$), unbounded from below for some $(\bar{u}, \bar{v})$.

Lemma 3.1. *Let ψ_i° be the dual of the positively homogeneous functions ψ_i for $i = 1, \ldots, m$. Suppose that Assumption 3.1 holds. Also, assume that $(\bar{u}, \bar{v})$ with $\bar{v} \geq 0$ is not a feasible solution of problem* ($\mathrm{D_{PHO}}$)*. Then, $\omega(\bar{u}, \bar{v})$ is unbounded from below.*

Proof. If $(\bar{u}, \bar{v})$ with $\bar{v} \geq 0$ is not feasible for problem ($\mathrm{D_{PHO}}$), then there exists an index i_0 satisfying

$$\psi_{i_0}^\circ(\alpha_{I_{i_0}}) > \beta_{i_0},$$

where $\alpha := A^T\bar{u} + H^T\bar{v} - c \in \mathbb{R}^n$, and $\beta := d - B^T\bar{u} - K^T\bar{v} \in \mathbb{R}^m$.

We now denote $\bar{\alpha}$ and $\bar{\alpha}(\lambda)$ as follows:

$$\begin{aligned}\bar{\alpha} &:= (\alpha_{I_1}, \alpha_{I_2}, \dots, \alpha_{I_{i_0}}, \dots, \alpha_{I_m}) \in \mathbb{R}^n,\\ \bar{\alpha}(\lambda) &:= (\alpha_{I_1}, \alpha_{I_2}, \dots, \lambda\hat{x}, \dots, \alpha_{I_m}) \in \mathbb{R}^n,\end{aligned}$$

where $\lambda \in \mathbb{R}_{++}$ and $\hat{x} \in \mathbb{R}^{n_{i_0}}$ is defined as the supreme point of the following problem:

$$\sup\{x^T\alpha_{I_{i_0}} \mid \psi_{i_0}(x) \le 1\}.$$

From the definition of $\hat{x}$, we obtain $\psi_{i_0}(\hat{x}) \le 1$. Then, from Definition 3.3, we have

$$\hat{x}^T\alpha_{I_{i_0}} = \psi^\circ_{i_0}(\alpha_{I_{i_0}}) \ge \psi_{i_0}(\hat{x})\psi^\circ_{i_0}(\alpha_{I_{i_0}}).$$

The above equality and the definition of the Lagrangian function give

$$\begin{aligned}\mathcal{L}(\bar{\alpha}(\lambda), \bar{u}, \bar{v}) &= b^T\bar{u} + p^T\bar{v} - \bar{\alpha}^T\bar{\alpha}(\lambda) + \beta^T\Psi(\bar{\alpha}(\lambda))\\ &= b^T\bar{u} + p^T\bar{v} - \sum_{i\neq i_0}\alpha_{I_i}^T\alpha_{I_i} - \lambda\hat{x}^T\alpha_{I_{i_0}}\\ &\quad + \sum_{i\neq i_0}\beta_i\psi_i(\alpha_{I_i}) + \beta_{i_0}\psi_{i_0}(\lambda\hat{x})\\ &= \gamma - \lambda\hat{x}^T\alpha_{I_{i_0}} + \beta_{i_0}\psi_{i_0}(\lambda\hat{x})\\ &\le \gamma - \lambda\psi_{i_0}(\hat{x})\psi^\circ_{i_0}(\alpha_{I_{i_0}}) + \beta_{i_0}\psi_{i_0}(\lambda\hat{x}),\end{aligned}$$

where $\gamma := b^T\bar{u} + p^T\bar{v} - \sum_{i\neq i_0}\alpha_{I_i}^T\alpha_{I_i} + \sum_{i\neq i_0}\beta_i\psi_i(\alpha_{I_i}) \in \mathbb{R}$ is constant with respect to λ. Moreover, Definition 3.1 shows that

$$\begin{aligned}\mathcal{L}(\bar{\alpha}(\lambda), \bar{u}, \bar{v}) &= \gamma - \lambda\psi_{i_0}(\hat{x})\psi^\circ_{i_0}(\alpha_{I_{i_0}}) + \lambda\beta_{i_0}\psi_{i_0}(\hat{x})\\ &= \gamma + \lambda\psi_{i_0}(\hat{x})(\beta_{i_0} - \psi^\circ_{i_0}(\alpha_{I_{i_0}}))\\ &\le \gamma + \lambda(\beta_{i_0} - \psi^\circ_{i_0}(\alpha_{I_{i_0}})).\end{aligned}$$

Therefore, $\mathcal{L}(\bar{\alpha}(\lambda), \bar{u}, \bar{v})$ converges to minus infinity when λ increases. Finally, if we set $x^k = \bar{\alpha}(\lambda^k)$ where $\lambda^k \to +\infty$ as $k \to +\infty$, then $\mathcal{L}(x^k, \bar{u}, \bar{v}) \to -\infty$, which shows that $\omega(\bar{u}, \bar{v})$ is unbounded from below. □

The above lemma indicates that a point $(\bar{u}, \bar{v})$ is a feasible solution of $(\mathrm{D_{PHO}})$ if $\omega(\bar{u}, \bar{v})$ with $\bar{v} \ge 0$ is bounded.

We now show that the positively homogeneous dual problem $(\mathrm{D_{PHO}})$ and the Lagrangian one $(\mathrm{D^{\mathcal{L}}_{PHO}})$ are equivalent under some conditions.

Lemma 3.2. *Suppose that Assumption 3.1 holds. Assume also that the positively homogeneous dual problem* $(\mathrm{D_{PHO}})$ *has a feasible solution* $(\bar{u}, \bar{v}) \in \mathbb{R}^k \times \mathbb{R}^\ell$*, and that there exists* $x^* \in \mathbb{R}^n$ *satisfying the following equality:*

$$(d - B^T\bar{u} - K^T\bar{v})^T\Psi(x^*) - (A^T\bar{u} + H^T\bar{v} - c)^T x^* = 0. \tag{3.2.2}$$

Then, the positively homogeneous dual problem $(\mathrm{D_{PHO}})$ *and the Lagrangian dual problem* $(\mathrm{D^{\mathcal{L}}_{PHO}})$ *have the same optimal value.*

Proof. From Lemma 3.1, the function ω is unbounded from below if there exists an index i_0 such that $\psi^\circ_{i_0}(\alpha_{I_{i_0}}) > \beta_{i_0}$, where $\alpha := A^T\bar{u} + H^T\bar{v} - c \in \mathbb{R}^n$, and $\beta := d - B^T\bar{u} - K^T\bar{v} \in \mathbb{R}^m$. Therefore, the problem $(\mathrm{D^{\mathcal{L}}_{PHO}})$ is equivalent to

$$\begin{array}{ll} \sup & \omega(u, v) \\ \text{s.t.} & \Psi^\circ(A^T u + H^T v - c) \leq d - B^T u - K^T v, \\ & v \geq 0. \end{array} \tag{$\hat{\mathrm{D}}^{\mathcal{L}}_{\mathrm{PHO}}$}$$

Let $(\bar{u}, \bar{v}) \in \mathbb{R}^k \times \mathbb{R}^\ell$ be the feasible solution of $(\hat{\mathrm{D}}^{\mathcal{L}}_{\mathrm{PHO}})$. From the definition of the Lagrangian function, we obtain:

$$\begin{aligned} \mathcal{L}(x, \bar{u}, \bar{v}) &= c^T x + d^T\Psi(x) + \bar{u}^T(b - Ax - B\Psi(x)) + \bar{v}^T(p - Hx - K\Psi(x)) \\ &= b^T\bar{u} + p^T\bar{v} - (A^T\bar{u} + H^T\bar{v} - c)^T x + (d - B^T\bar{u} - K^T\bar{v})^T\Psi(x). \end{aligned}$$

Then, taking $x^* \in \mathbb{R}^n$ that satisfies (3.2.2), we have

$$\mathcal{L}(x^*, \bar{u}, \bar{v}) = b^T\bar{u} + p^T\bar{v}.$$

Notice that x^* is the solution of the problem

$$\inf_x \mathcal{L}(x, \bar{u}, \bar{v}),$$

because $\mathcal{L}(x, \bar{u}, \bar{v}) \geq b^T\bar{u} + p^T\bar{v}$ holds from Proposition 3.1. Therefore, $(\hat{\mathrm{D}}^{\mathcal{L}}_{\mathrm{PHO}})$ can be described as follows:

$$\begin{array}{ll} \sup & b^T u + p^T v \\ \text{s.t.} & \Psi^\circ(A^T u + H^T v - c) \leq d - B^T u - K^T v, \\ & v \geq 0, \end{array}$$

which implies that the optimal value of the above problem is the same as that of the positively homogeneous dual problem $(\mathrm{D_{PHO}})$. $\square$

As a consequence of the above lemma, we obtain the following result.

Theorem 3.2. *Suppose that the Lagrangian dual problem* $(\mathrm{D}^{\mathcal{L}}_{\mathrm{PHO}})$ *has a feasible solution. Assume also that the vector positively homogeneous function* Ψ *satisfies Assumption 3.1. Then, the positively homogeneous dual problem* $(\mathrm{D}_{\mathrm{PHO}})$ *and the Lagrangian dual problem* $(\mathrm{D}^{\mathcal{L}}_{\mathrm{PHO}})$ *have the same optimal value and solutions.*

Proof. From Definition 3.1 and Assumption 3.1, we have $\Psi(0) = 0$. It means that equation (3.2.2) holds at $x^* = 0$. Thus, from Lemma 3.2, the problems $(\mathrm{D}_{\mathrm{PHO}})$ and $(\mathrm{D}^{\mathcal{L}}_{\mathrm{PHO}})$ have the same optimal value.

Moreover, we denote S_D and $S_{D_{\mathcal{L}}}$ as the sets of optimal solutions of problems $(\mathrm{D}_{\mathrm{PHO}})$ and $(\mathrm{D}^{\mathcal{L}}_{\mathrm{PHO}})$, respectively. Let us take $(u^*, v^*) \in S_D$. Then, it is clearly feasible for $(\mathrm{D}^{\mathcal{L}}_{\mathrm{PHO}})$. It follows from Theorem 3.2 that the optimal values of $(\mathrm{D}_{\mathrm{PHO}})$ and $(\mathrm{D}^{\mathcal{L}}_{\mathrm{PHO}})$ are the same, which is $b^T u^* + p^T v^*$, and so $(u^*, v^*) \in S_{D_{\mathcal{L}}}$. Conversely, let us take $(\bar{u}, \bar{v}) \in S_{D_{\mathcal{L}}}$. Then, the point $(\bar{u}, \bar{v})$ is feasible for $(\mathrm{D}^{\mathcal{L}}_{\mathrm{PHO}})$. Note that Lemma 3.1 indicates that if (u, v) is feasible for $(\mathrm{D}^{\mathcal{L}}_{\mathrm{PHO}})$ and the objective function value of $(\mathrm{D}^{\mathcal{L}}_{\mathrm{PHO}})$ at the point (u, v) is finite, then it is also feasible for $(\mathrm{D}_{\mathrm{PHO}})$. Thus, $(\bar{u}, \bar{v})$ is feasible for $(\mathrm{D}_{\mathrm{PHO}})$. Once again from Theorem 3.2, the optimal values of $(\mathrm{D}_{\mathrm{PHO}})$ and $(\mathrm{D}^{\mathcal{L}}_{\mathrm{PHO}})$ are the same, which means that $(\bar{u}, \bar{v}) \in S_D$. Consequently, we obtain $S_D = S_{D_{\mathcal{L}}}$. $\square$

The above theorem shows that the Lagrangian dual problem $(\mathrm{D}^{\mathcal{L}}_{\mathrm{PHO}})$ can be written in a closed-form when the function Ψ is positively homogeneous and satisfies Assumption 3.1. The paper [61] does not show that the same property holds for the AVO problem. We now give it as a direct consequence of Theorem 3.2.

Collolary 3.1. *If the dual of an AVO problem has a feasible solution, then its optimal value and optimal solutions are the same as those of the Lagrangian dual problem* $(\mathrm{D}^{\mathcal{L}}_{\mathrm{PHO}})$.

Proof. It holds from Theorem 3.2 and the fact that the absolute value function is positively homogeneous and satisfies Assumption 3.1. $\square$

Collolary 3.2. *If the optimal values of an AVO primal problem and its Lagrangian dual problem* $(\mathrm{D}^{\mathcal{L}}_{\mathrm{PHO}})$ *are the same, then the strong duality holds between the AVO primal and the AVO dual problem.*

Proof. It holds straightforward from Corollary 3.1. $\square$

From the results in Section 4, the positively homogeneous dual can be a tool for investigating the Lagrangian dual. Let us consider the following integer problem:

$$\begin{array}{rl} \min & \|Ax-b\|_2 \\ \text{s.t.} & |x| = e, \end{array} \tag{$\mathrm{P_b}$}$$

where A and b are given matrix and vector and e is a vector of ones. The positively homogeneous dual of $(\mathrm{P_b})$ can be written as follows:

$$\begin{array}{rl} \max & e^T u_1 + b^T u_2 \\ \text{s.t.} & \begin{bmatrix} I & -A^T \\ I & A^T \\ 0 & 0 \end{bmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{pmatrix} \|u_1\|_2 \\ \|u_2\|_2 \end{pmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}. \end{array} \tag{$\mathrm{D_b}$}$$

By taking the positively homogeneous dual of $(\mathrm{D_b})$ we obtain

$$\begin{array}{rl} \min & \|A(y-z)-b\|_2 \\ \text{s.t.} & y+z=e \\ & y, z \geq 0, \end{array} \tag{$\hat{\mathrm{P}}_{\mathrm{b}}$}$$

On the other hand, the equivalent reformulation of problem $(\mathrm{P_b})$ is as follows:

$$\begin{array}{rl} \min & \|A(x^+ - x^-)-b\|_2 \\ \text{s.t.} & x^+ + x^- = e, \end{array}$$

where $x_i^+ = \max\{0, x_i\}$ and $x_i^- = \max\{0, -x_i\}$. The above problem is further equivalent to

$$\begin{array}{rl} \min & \|A(y-z)-b\|_2 \\ \text{s.t.} & y+z=e \\ & y, z \geq 0 \\ & y^T z = 0. \end{array} \tag{$\mathrm{P'_b}$}$$

The above result shows that problem $(\hat{\mathrm{P}}_{\mathrm{b}})$ is obtained from problem $(\mathrm{P'_b})$ by removing the complementarity constraint $x^T y = 0$. It means that the relaxed problem $(\hat{\mathrm{P}}_{\mathrm{b}})$ can be obtained by using the positively homogeneous dual, which is actually the Lagrangian dual, and it is the limitation of using the Lagrangian dual. On the other hand, we see the potential of the Lagrange dual from the following one-dimensional nonconvex problems:

$$\begin{array}{rl} \min & x - 2|x| \\ \text{s.t.} & |x| \leq 1, \end{array} \tag{$\mathrm{P_{c1}}$}$$

$$\begin{array}{rl} \min & x + 2|x| \\ \text{s.t.} & |x| \geq 1. \end{array} \tag{P$_{\rm c2}$}$$

Problem ($\mathrm{P_{c1}}$) has a nonconvex objective function and its optimal value is -3 at $x^* = -1$, and problem ($\mathrm{P_{c2}}$) has a nonconvex constraint and its optimal value is 1 at $x^* = -1$. The positively homogeneous dual, which is equivalent to the Lagrangian dual, of each problem can be written as

$$\begin{array}{rl} \max & -v \\ \text{s.t.} & v \geq 3, \end{array} \tag{D$_{\rm c1}$}$$

$$\begin{array}{rl} \max & v \\ \text{s.t.} & v \geq 0 \\ & v \leq 1. \end{array} \tag{D$_{\rm c2}$}$$

Clearly, the optimal values of problems ($\mathrm{D_{c1}}$) and ($\mathrm{D_{c2}}$) are -3 and 1, respectively, and no duality gap exists. The above result indicates that we may obtain zero duality gap between a nonconvex problem and its Lagrangian dual. For this reason, we expect that the positively homogeneous dual supports the potential of the Lagrange dual for nonconvex problems.

3.3 Examples of positively homogeneous optimization problems

In this section, we present several applications that are formulated as PHO, and show their closed-form dual problems.

First, we observe that any p-norm function with $p \in [1, \infty)$ is positively homogeneous. So, if ψ is the p-norm, then ψ° becomes the q-norm, where $1/p + 1/q = 1$. Therefore, if ψ is taken as $\|\cdot\|_1, \|\cdot\|_2, \|\cdot\|_\infty$, then ψ° becomes $\|\cdot\|_\infty, \|\cdot\|_2, \|\cdot\|_1$, respectively. Moreover, in the case that $p \in (0, 1)$, the dual function ψ° is equal to $\|\cdot\|_\infty$ for all $p \in (0, 1)$, which is proved in Proposition 6 of Appendix A. From the result, we can consider any p-norm functions as ψ in PHO problems. And, even if such functions are nonconvex with $p \in (0, 1)$, the Lagrangian dual problem can be written in a closed-form from Theorem 3.2.

We now show some positively homogeneous problems using these p-norm functions. The first example is the so-called linear second-order cone optimization problem [4], which is one of the famous convex optimization problems.

Example 3.1. *Let $x = (x_1, x_2)^T \in \mathbb{R} \times \mathbb{R}^{n-1}$. Then, we consider the linear second-order cone optimization problem written by*

$$\begin{aligned} \min \quad & c^T x \\ \text{s.t.} \quad & Ax = b, \\ & x_1 - \|x_2\|_2 \geq 0, \end{aligned} \tag{P$_1$}$$

where $c \in \mathbb{R}^n$, $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. The above problem can be written in PHO form as

$$\begin{aligned} \min \quad & c^T x + 0^T \Psi(x) \\ \text{s.t.} \quad & Ax + 0\Psi(x) = b, \\ & Hx + K\Psi(x) \geq 0, \end{aligned}$$

with $H = (1, 0, \ldots, 0) \in \mathbb{R}^{1 \times n}, K = (0, -1) \in \mathbb{R}^{1 \times 2}$ and $\Psi \colon \mathbb{R}^n \to \mathbb{R}^2, \Psi(x) = (|x_1|, \|x_2\|_2)^T$. Then, recalling (D$_{\text{PHO}}$)*, its dual problem is given by*

$$\begin{aligned} \max \quad & b^T u \\ \text{s.t.} \quad & \Psi^\circ(A^T u + H^T v - c) + K^T v \leq 0, \\ & v \geq 0, \end{aligned}$$

where Ψ° is identical to Ψ in this case. Then, from the definition of Ψ, we have

$$\begin{aligned} \max \quad & b^T u \\ \text{s.t.} \quad & |(A^T u)_1 + v - c_1| \leq 0, \\ & \|(A^T u)_2 - c_2\|_2 \leq v, \\ & v \geq 0, \end{aligned}$$

with $(A^T u)_1$ as the first component of $A^T u$, $(A^T u)_2$ is the rest of it, and $c = (c_1, c_2)^T \in \mathbb{R} \times \mathbb{R}^{n-1}$. The first constraint of the above problem shows that

$$v = c_1 - (A^T u)_1,$$

and $v \geq 0$ automatically holds from the second constraint. Then, we obtain

$$\begin{aligned} \max \quad & b^T u \\ \text{s.t.} \quad & \|(A^T u)_2 - c_2\|_2 \leq c_1 - (A^T u)_1 \end{aligned}$$

as the dual problem of (P$_1$)*. In fact, the above problem is the standard dual of the linear second-order cone optimization problem [4].*

Although we use the 2-norm in the above example, any p-norm function with $p \in (0, \infty]$ can be considered. In this case, if $p \in [1, \infty]$, then the primal and dual problems are p-order cone and q-order cone optimization problems, respectively, where $1/p + 1/q = 1$ [106]. If $p \in (0, 1)$, then the dual is ∞-order cone optimization problem.

In the next example, we consider a gauge optimization problem, which is also a convex problem with multiple gauge functions in its objective and constraint functions. Here, we recall that f is a gauge function if and only if it is nonnegative, convex, positively homogeneous and satisfies $f(0) = 0$ [30]. For such a problem, we introduce its dual in PHO form.

Example 3.2. *Let $x \in \mathbb{R}^n$. We consider the following problem:*

$$\begin{aligned} \min \quad & \sum_{i=1}^{s} \alpha_i f_i(A_i x - a_i) \\ \text{s.t.} \quad & g_j(B_j x - b_j) \le \beta_j, \quad j = 1, \ldots, t, \end{aligned} \tag{P$_2$}$$

where $\alpha_i, \beta_j \in \mathbb{R}_+$, $A_i \in \mathbb{R}^{m_i \times n}$, $B_j \in \mathbb{R}^{k_j \times n}$, $a_i \in \mathbb{R}^{m_i}$ and $b_j \in \mathbb{R}^{k_j}$ are given for all $i = 1, \ldots, s$ and $j = 1, \ldots, t$, and $f_i \colon \mathbb{R}^{m_i} \to \mathbb{R}$ and $g_j \colon \mathbb{R}^{k_j} \to \mathbb{R}$ are gauge functions. Letting $y_i := A_i x - a_i$ and $z_j := B_j x - b_j$, (P$_2$) can be written as

$$\begin{aligned} \min \quad & \sum_{i=1}^{s} \alpha_i f_i(y_i) \\ \text{s.t.} \quad & g_j(z_j) \le \beta_j, \quad && j = 1, \ldots, t, \\ & A_i x - y_i = a_i, && i = 1, \ldots, s, \\ & B_j x - z_j = b_j, && j = 1, \ldots, t. \end{aligned}$$

The above problem does not have a gauge function defined for the variable x, so we introduce such a gauge function $x \mapsto \psi(x)$ and rewrite the problem in the following way:

$$\begin{aligned} \min \quad & 0 \times \psi(x) + \sum_{i=1}^{s} \alpha_i f_i(y_i) + 0 \times \sum_{j=1}^{t} g_j(z_j) \\ \text{s.t.} \quad & 0 \times \psi(x) \le 0, \\ & 0 \times f_i(y_i) \le 0, \quad && i = 1, \ldots, s, \\ & g_j(z_j) \le \beta_j, && j = 1, \ldots, t, \\ & A_i x - y_i = a_i, && i = 1, \ldots, s, \\ & B_j x - z_j = b_j, && j = 1, \ldots, t. \end{aligned}$$

Note that $\psi \colon \mathbb{R}^n \to \mathbb{R}$ is a dummy gauge function with x as its domain.

Let

$$\hat{x} := (x, y_1, \ldots, y_s, z_1, \ldots, z_t) \in \mathbb{R}^{n+\sum_{i=1}^{s} m_i + \sum_{j=1}^{t} k_j}$$

and

$$\Psi(\hat{x}) := (\psi(x), f_1(y_1), \ldots, f_s(y_s), g_1(z_1), \ldots, g_t(z_t))^T.$$

Then the above problem can be rewritten as

$$\begin{array}{ll} \min & d^T \Psi(\hat{x}) \\ \text{s.t.} & K\Psi(\hat{x}) \leq p, \\ & \hat{A}\hat{x} = \hat{b}, \end{array}$$

where $d = (0, \alpha_1, \ldots, \alpha_s, 0, \ldots, 0)^T \in \mathbb{R}^{1+s+t}$, $p = (0, \ldots, 0, \beta_1, \ldots, \beta_t)^T \in \mathbb{R}^{1+s+t}$,

$$K = \begin{bmatrix} 0 & 0 \\ 0 & E_t \end{bmatrix}, \hat{A} = \left[\begin{array}{c|ccc|ccc} A_1 & -E_{m_1} & & & & & \\ \vdots & & \ddots & & & 0 & \\ A_s & & & -E_{m_s} & & & \\ \hline B_1 & & & & -E_{k_1} & & \\ \vdots & & 0 & & & \ddots & \\ B_t & & & & & & -E_{k_t} \end{array}\right],$$

and

$$\hat{b} = \left[\begin{array}{c} a_1 \\ \vdots \\ a_s \\ \hline b_1 \\ \vdots \\ b_t \end{array}\right].$$

Moreover, its positively homogeneous dual problem is given by

$$\begin{array}{ll} \max & \hat{b}^T u - p^T v \\ \text{s.t.} & \Psi^{\circ}(\hat{A}^T u) - K^T v \leq d, \\ & v \geq 0. \end{array}$$

For simplification, let $u = (u_{11}, \ldots, u_{1s}, u_{21}, \ldots, u_{2t})^T$ *with* $u_{1i} \in \mathbb{R}^{m_i}, i = 1, \ldots, s$

and $u_{2j} \in \mathbb{R}^{k_j}, j = 1, \ldots, t$. *Then the above problem is rewritten as*

$$\begin{aligned} \max \quad & \sum_{i=1}^{s} a_i^T u_{1i} + \sum_{j=1}^{t} b_j^T u_{2j} - \sum_{\ell=1}^{t} \beta_\ell v_{1+s+\ell} \\ \text{s.t.} \quad & \sum_{i=1}^{s} A_i^T u_{1i} + \sum_{j=1}^{t} B_j^T u_{2j} = 0, \\ & f_i^\circ(-u_{1i}) \leq \alpha_i, \qquad i = 1, \ldots, s, \\ & g_j^\circ(-u_{2j}) \leq v_{1+s+j}, \quad j = 1, \ldots, t. \end{aligned} \tag{D$_2$}$$

Notice that the last constraint implies $v \geq 0$ *because* g_j° *is also a gauge function. Moreover,* (D$_2$) *does not include the polar function* ψ° *of the dummy gauge function* ψ.

Note that the objective and constraint functions f_i and g_j in the above example can be more general positively homogeneous functions. Thus, we can consider problems that have different positively homogeneous functions in their objective and constraint functions.

The next example is the group Lasso-type problems [65, 111], which is a special case of (P$_2$) and consist of unconstrained minimizations of the sum of certain norms. Such problems have many applications, in particular they appear in compressed sensing area [24, 88], where the sparsity of solutions are important. As an example, we consider a primal problem with p_1-norm and p_2-norm where $p_1, p_2 \in \mathbb{R}_+$, which are used in the regularization terms.

Example 3.3. *Let* $x \in \mathbb{R}^n$ *and* $p_1, p_2 \in \mathbb{R}_+$. *We consider the following problem:*

$$\min \quad \|Ax - b\|_2 + \lambda_1 \sum_{i=1}^{m'} \|x_{I_i}\|_{p_1} + \lambda_2 \sum_{i=m'+1}^{m} \|x_{I_i}\|_{p_2} \tag{P$_3$}$$

where $\lambda_1, \lambda_2 \in \mathbb{R}_+, b \in \mathbb{R}^m, A \in \mathbb{R}^{m \times n}$ *and* $0 < m' < m$.

Notice that the first term of the objective function of group Lasso-type problems are usually the square of 2*-norm functions. However, it is not positively homogeneous, so we removed the square and considered just the* 2*-norm functions.*

We obtain the above problem by setting, in (P$_2$), $s = m + 1$,

$$\alpha_i = \begin{cases} \lambda_1, & \text{if} \quad i = 1, \ldots, m', \\ \lambda_2, & \text{if} \quad i = m' + 1, \ldots, m, \\ 1, & \text{if} \quad i = m + 1, \end{cases}$$

$$A_i = \begin{cases} E_{I_i}, & \text{if} \quad i = 1, \dots, m, \\ A, & \text{if} \quad i = m+1, \end{cases}$$

where E_{I_i} *is a submatrix of* E_n *with* $E_j,\ j \in I_i$ *as its rows,*

$$a_i = \begin{cases} 0, & \text{if} \quad i = 1, \dots, m, \\ b, & \text{if} \quad i = m+1, \end{cases}$$

and

$$f_i(\cdot) = \begin{cases} \|\cdot\|_{p_1}, & \text{if} \quad i = 1, \dots, m', \\ \|\cdot\|_{p_2}, & \text{if} \quad i = m'+1, \dots, m, \\ \|\cdot\|_2, & \text{if} \quad i = m+1. \end{cases}$$

Then, recalling (P_2) *and* (D_2)*, the dual of* (P_3) *can be written as*

$$\begin{array}{ll} \max & b^T u_{1(m+1)} \\ \text{s.t.} & \displaystyle\sum_{i=1}^{m} E_{I_i}^T u_{1i} + A^T u_{1(m+1)} = 0, \\ & \|-u_{1i}\|_{q_1} \le \lambda_1, \quad i = 1, \dots, m', \\ & \|-u_{1i}\|_{q_2} \le \lambda_2, \quad i = m'+1, \dots, m, \\ & \|-u_{1(m+1)}\|_2 \le 1, \end{array}$$

where $q_i, i = 1, 2$ *are obtained by*

$$q_i = \begin{cases} \dfrac{p_i}{p_i - 1}, & \text{if} \quad p_i > 1, \\ \infty, & \text{if} \quad p_i \in (0, 1], \end{cases} \tag{3.3.1}$$

from Proposition 3.2 of Appendix A. Notice that the first equality constraint can be rewritten as

$$u_{1i} + (A^T)_{I_i} u_{1(m+1)} = 0, \quad i = 1, \dots, m.$$

Then, the above problem is described as

$$\begin{array}{ll} \max & b^T u \\ \text{s.t.} & \|(A^T)_{I_i} u\|_{q_1} \le \lambda_1, \quad i = 1, \dots, m', \\ & \|(A^T)_{I_i} u\|_{q_2} \le \lambda_2, \quad i = m'+1, \dots, m, \\ & \|-u\|_2 \le 1, \end{array}$$

where we denote $u_{1(m+1)}$ *as* u *for simplicity.*

The next example is also a Lasso-type problem. In this case, the objective function is a gauge, because the sum of gauge functions is also gauge. In order to obtain the dual of a gauge optimization problem, the polar of the objective function should be considered [6,32]. However, it may be difficult to obtain the polar of a sum of gauge functions. To overcome this drawback, we use here the PHO framework.

Example 3.4. *Let* $x \in \mathbb{R}^n$ *and* $p_1, p_2 \in \mathbb{R}_+$. *We consider the following problem:*

$$\begin{array}{ll} \min & \lambda_1 \|x\|_{p_1} + \lambda_2 \|x\|_{p_2} \\ \text{s.t.} & \|Ax - b\|_2 \leq \beta, \end{array} \tag{P$_4$}$$

where $\lambda_1, \lambda_2, \beta \in \mathbb{R}_+$, $A \in \mathbb{R}^{m \times n}$ *and* $b \in \mathbb{R}^m$. *The above problem can be obtained if we set, in* (P$_2$), $s = 2$, $t = 1$, $\alpha_1 = \lambda_1$, $\alpha_2 = \lambda_2$, $A_1 = A_2 = E_n$, $a_1 = a_2 = 0$, $B_1 = A$, $b_1 = b$, $f_1(\cdot) = \|\cdot\|_{p_1}$, $f_2(\cdot) = \|\cdot\|_{p_2}$, $g_1(\cdot) = \|\cdot\|_2$. *Then, recalling* (D$_2$), *the dual of* (P$_4$) *is written by*

$$\begin{array}{ll} \max & b^T u_{21} - \beta v_4 \\ \text{s.t.} & u_{11} + u_{12} + A^T u_{21} = 0, \\ & \| - u_{11}\|_{q1} \leq \lambda_1, \\ & \| - u_{12}\|_{q2} \leq \lambda_2, \\ & \| - u_{21}\| \leq v_4, \end{array}$$

which is finally rewritten as

$$\begin{array}{ll} \max & b^T u_2 - \beta v \\ \text{s.t.} & \|u_1 + A^T u_2\|_{q1} \leq \lambda_1, \\ & \| - u_1\|_{q2} \leq \lambda_2, \\ & \| - u_2\| \leq v, \end{array}$$

where we set u_{12}, u_{21} *and* v_4 *as* u_1, u_2 *and* v, *respectively, and* q_1 *and* q_2 *are defined in* (3.3.1).

In order to control the sparsity of the solutions of the above Lasso-type problems, we can use any combination of p-norm functions, with $p \in (0, \infty]$, as the regularization terms. Especially, it is reported that the p-norm functions with $p \in (0, 1)$ in (P$_3$) is useful because they give sparser solutions than 1-norm functions [17, 18, 69].

We now give another example: the sum of norms optimization problems, which are generally nonconvex. Such problems have applications, for example, in facility location, where locations of new facilities should be decided by analyzing the distance between the new and the existing facilities [100]. Moreover, the problem of the following example can be applied not only to the minimization of the distance but also maximization of it by taking the constant λ_i as $-\lambda_i$. Such a situation can be found, for instance, in locating obnoxious facilities in residential areas.

Example 3.5. *Let $x \in \mathbb{R}^n$. We consider the following problem:*

$$\begin{aligned} \min \quad & \sum_{i=1}^{s} \lambda_i f_i(A_i x - a_i), \\ \text{s.t.} \quad & Bx \leq b, \end{aligned} \tag{P$_5$}$$

where $\lambda_i \in \mathbb{R}$, $A_i \in \mathbb{R}^{m_i \times n}$, $B \in \mathbb{R}^{k \times n}$, $a_i \in \mathbb{R}^{m_i}$ and $b \in \mathbb{R}^k$ are given, and $f_i \colon \mathbb{R}^{m_i} \to \mathbb{R}$, $i = 1, \ldots, s$ are positively homogeneous functions. We now introduce its positively homogeneous dual by taking almost the same procedure as in Example 3.2. Let $y_i := A_i x - a_i$, then (P$_5$) *is equivalent to*

$$\begin{aligned} \min \quad & \sum_{i=1}^{s} \lambda_i f_i(y_i) \\ \text{s.t.} \quad & A_i x - y_i = a_i, \quad i = 1, \ldots s, \\ & Bx \leq b. \end{aligned}$$

By introducing additional constraints, we consider the following problem:

$$\begin{aligned} \min \quad & \sum_{i=1}^{s} \lambda_i f_i(y_i) \\ \text{s.t.} \quad & A_i x - y_i = a_i, \quad i = 1, \ldots s, \\ & Bx \leq b, \\ & c_i f_i(y_i) \leq d_i, \qquad i = 1, \ldots s, \end{aligned}$$

where c_i and d_i are strictly positive constants. Notice that the additional constraints ensure the boundedness of the each term of the objective function especially when λ_i is strictly negative. Without such constraints, (P$_5$) *can be unbounded depending on the linear constraint, and then its dual becomes infeasible. Note that the additional constraints do not change solutions, when we choose c_i and d_i so that the constraint $c_i f_i(y_i) \leq d_i$ will include reasonable solutions.*

Let $\hat{x} := (x, y_1, \ldots, y_s)^T \in \mathbb{R}^{n + \sum_{i=1}^{s} m_i}$ and

$$\Psi(\hat{x}) := (\psi(x), f_1(y_1), \ldots, f_s(y_s))^T \in \mathbb{R}^{1+s},$$

where $\psi(\cdot)$ is a dummy positively homogeneous function. Then the above problem can be described as

$$\begin{aligned} \min \quad & d^T \Psi(\hat{x}) \\ \text{s.t.} \quad & \hat{A}\hat{x} = \hat{a}, \\ & H\hat{x} + K\Psi(\hat{x}) \geq p, \end{aligned}$$

where $d = (0, \lambda_1, \ldots, \lambda_s)^T$, $\hat{a} = (a_1, \ldots, a_s)^T$, $p = (-b, -d_1, \ldots, -d_s)^T$,

$$\hat{A} = \begin{bmatrix} A_1 & -E_{m_1} & & 0 \\ \vdots & & \ddots & \\ A_s & 0 & & -E_{m_s} \end{bmatrix}, \; H = \begin{bmatrix} -B & 0 \\ 0 & 0 \end{bmatrix},$$

and

$$K = \left[\begin{array}{c|ccc} 0 & & 0 & \\ \hline & -c_1 & & \\ 0 & & \ddots & \\ & & & -c_s \end{array}\right].$$

Then, recalling the positively homogeneous dual $(\mathrm{D_{PHO}})$*, the dual of the above problem can be written as*

$$\begin{array}{rl} \max & \hat{a}^T u + p^T v \\ \text{s.t.} & \Psi^\circ(\hat{A}^T u + H^T v) \le d - K^T v, \\ & v \ge 0, \end{array}$$

which is rewritten by

$$\begin{array}{rl} \max & \displaystyle\sum_{i=1}^{s} a_i^T u_i - b^T v_1 - \sum_{i=1}^{s} d_i^T v_{i+1} \\ \text{s.t.} & \displaystyle\sum_{i=1}^{s} A_i^T u_i - B^T v_1 = 0, \\ & f_i^\circ(-u_i) \le \lambda_i + c_i, , \quad i = 1, \ldots s, \\ & v \ge 0, \end{array}$$

where $v = (v_1, \ldots, v_{s+1})^T$.

3.4 Conclusion

In this chapter, we proposed optimization problems with positively homogeneous functions, which we call positively homogeneous optimization problems. We also introduced their dual problems and showed the weak duality theorem between these problems. Moreover, we gave sufficient conditions for the equivalency between the proposed dual and the Lagrangian dual problems. Finally, we presented some examples of positively homogeneous problems to show their value in real-world applications. One natural future work will be to propose methods that obtain approximate solutions of positively homogeneous optimization problems. We believe the theoretical results described here are essential for that.

3.5 Appendix

The following proposition shows that the dual of the ℓ_p-norm function is the ℓ_∞-norm even when p is less than 1.

Proposition 3.2. *Suppose that $p \in (0, 1)$. Then, the dual of the p-norm function is equal to the ∞-norm.*

Proof. Let $y \in \mathbb{R}^n$ be an arbitrary vector. If $y = 0$, this proposition clearly holds. If $y \neq 0$, from Definition 3.3, we obtain

$$\begin{aligned} \|y\|_p^\circ &= \sup\{x^T y \mid \|x\|_p \leq 1\} \\ &\leq \sup\{|x^T y| \mid \|x\|_p \leq 1\} \\ &\leq \sup\left\{\sum_{i=1}^n |x_i||y_i| \mid \|x\|_p \leq 1\right\} \\ &\leq \max_j |y_j| \left(\sup\left\{\sum_{i=1}^n |x_i| \mid \|x\|_p \leq 1\right\}\right) \\ &= \max_j |y_j| \Big(\sup\{\|x\|_1 \mid \|x\|_p \leq 1\}\Big). \end{aligned}$$

Since $p \in (0, 1)$, we note that $\|x\|_1 \leq \|x\|_p$ holds [51]. Then, we have

$$\|y\|_p^\circ \leq \max_j |y_j| \Big(\sup\{\|x\|_p \mid \|x\|_p \leq 1\}\Big) = \max_j |y_j| = \|y\|_\infty.$$

Now, take an arbitrary $i_0 \in \text{argmax}_i |y_i|$, and define $\bar{x}_i$ as follows:

$$\bar{x}_i = \begin{cases} \text{sign}(y_i), & \text{if} \quad i = i_0, \\ 0, & \text{otherwise}, \end{cases}$$

where

$$\text{sign}(y_i) = \begin{cases} 1, & \text{if} \quad y_i > 0, \\ 0, & \text{if} \quad y_i = 0, \\ -1, & \text{if} \quad y_i < 0. \end{cases}$$

Then, $\|\bar{x}\|_p = 1$ and we have

$$\|y\|_p^\circ = \sup\{x^T y \mid \|x\|_p \leq 1\} \geq \bar{x}^T y = \max_i |y_i| = \|y\|_\infty,$$

which completes the proof. $\square$

Chapter 4

Duality of optimization problems with gauge functions

4.1 Introduction

The *gauge optimization* (GO) problem is in general described as follows [6, 30–32]:

$$\min_{x \in \mathcal{X}} \quad g(x), \tag{$\mathrm{P_{GO}}$}$$

where $\mathcal{X} \subseteq \mathbb{R}^n$ is a closed convex set and $g\colon \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is a *gauge* function. Here, we say that g is a gauge function if g is convex, nonnegative, positively homogeneous and satisfies $g(0) = 0$. Note that GO problems are convex because gauge functions are also convex. Freund [30] first introduced ($\mathrm{P_{GO}}$), 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, p-norm optimization problems with $p \in [1, \infty]$, and convex quadratic optimization problems [30].

Recently, Friedlander et al. [32] considered a specific form of the GO problem in which $\mathcal{X}$ is described as $\mathcal{X} := \{x \in \mathbb{R}^n \mid h(b - Ax) \le \sigma\}$, where h is a gauge function, σ 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 [31] applied this gauge duality to solve low-rank spectral optimization problems. Aravkin et al. [6] 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 [6, 30–32] 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 [108] considered the following *positively homogeneous optimization* (PHO) problem:

$$\begin{array}{ll} \min & c^T x + d^T \Psi(x) \\ \text{s.t.} & Ax + B\Psi(x) = b, \\ & Hx + K\Psi(x) \leq p, \\ & x \in \mathrm{dom}\Psi, \end{array} \tag{$\mathrm{P_{PHO^*}}$}$$

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, $\Psi\colon \mathbb{R}^n \to (\mathbb{R}\cup\{\infty\})^m$ is defined by $\Psi(\cdot) := (\psi_1(\cdot), \dots, \psi_m(\cdot))^T$ where each function $\psi_i\colon \mathbb{R}^{n_i} \to \mathbb{R}$, $\sum_{i=1}^m n_i = n$ is nonnegative and positively homogeneous, and T denotes transpose. Moreover, $\mathrm{dom}\Psi$ denotes the effective domain of Ψ, defined by $\mathrm{dom}\Psi := \{x \in \mathbb{R}^n \mid \psi_i(x_i) < \infty, i = 1, \dots, m\}$ where $x_i \in \mathbb{R}^{n_i}$ is a disjoint subvector of x. Problem ($\mathrm{P_{PHO^*}}$) is not necessarily convex, and it includes ($\mathrm{P_{GO}}$) with $\mathcal{X} = \{x \in \mathbb{R}^n \mid h(b - Ax) \leq \sigma\}$ 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 $x \in \mathrm{dom}\Psi$ in the constraints of ($\mathrm{P_{PHO^*}}$). This is because we want to consider more general PHO problems than the ones used in the previous work [108], where $\mathrm{dom}\Psi = \mathbb{R}^n$ is assumed. Then we can adopt the indicator function of some cones as ψ_i. We will later show that the same results as in [108] can be obtained even when $\mathrm{dom}\Psi \neq \mathbb{R}^n$.

When $n_i = 1$ and $\psi_i(x_i) = |x_i|$, ($\mathrm{P_{PHO^*}}$) is reduced to the absolute value optimization problem proposed by Mangasarian [61]. The other examples of PHO problems are p-order cone optimization problems [4, 106] with $p \in (0, \infty]$, group Lasso-type problems [65, 111], and sum of norms optimization problems [100].

Yamanaka and Yamashita [108] 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 ($\mathrm{P_{PHO^*}}$), 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 [108] does not discuss strong duality and primal recovery.

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

$$\begin{aligned} \min \quad & c^T x + d^T \mathcal{G}(x) \\ \text{s.t.} \quad & Ax = b, \\ & Hx + K\mathcal{G}(x) \le p, \\ & x \in \operatorname{dom} \mathcal{G}, \end{aligned} \tag{$\mathrm{P_{GO^*}}$}$$

where c, d, b, p, A, H, K are the same as in ($\mathrm{P_{PHO^*}}$), and $\mathcal{G}\colon \mathbb{R}^n \to (\mathbb{R} \cup \{\infty\})^m$ is defined by $\mathcal{G}(\cdot) := (g_1(\cdot), \dots, g_m(\cdot))^T$ with $g_i\colon \mathbb{R}^{n_i} \to \mathbb{R}$ as a gauge function for all i. Note that there is no nonlinear term in the equality constraints, and problem ($\mathrm{P_{GO^*}}$) is convex when all elements of d and K are nonnegative. Problem ($\mathrm{P_{GO^*}}$) includes the convex GO problems considered in [6, 30–32], and it is possible to explicitly handle linear terms. In this paper, we call ($\mathrm{P_{GO^*}}$) the gauge optimization problem when it is clear from the context.

In particular, we are interested in theoretical properties of problem ($\mathrm{P_{GO^*}}$) and its dual. We first define a dual problem of ($\mathrm{P_{GO^*}}$) as in [108], and then, give conditions under which weak and strong dualities hold for problem ($\mathrm{P_{GO^*}}$) and its dual. Moreover, we present necessary and sufficient optimality conditions for ($\mathrm{P_{GO^*}}$), 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 ($\mathrm{P_{GO^*}}$) can be extended to general convex optimization problems by considering the so-called perspective functions.

This chapter is organized as follows. In Section 4.2, we recall some important properties of ($\mathrm{P_{PHO^*}}$) in [108]. We show that some of them hold even if $\operatorname{dom}\Psi \neq \mathbb{R}^n$. Section 4.3 presents the dual of problem ($\mathrm{P_{GO^*}}$), and gives some relations of ($\mathrm{P_{GO^*}}$) 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 4.4, we discuss how to extend the obtained results to general convex optimization problems. Section 6 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 $I \subseteq \{1, \dots, n\}$, 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, $e_n := (1, \dots, 1)^T \in \mathbb{R}^n$. The identity matrix with dimension n 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. We also denote by $\| \cdot \|$ a norm function. For a function f and vectors x and y, we denote the subdifferential of $f(x, y)$ with respect to x as $\partial_x f(x, y)$. The effective domain of a function f is given by $\mathrm{dom} f$. The convex hull of a set S is denoted by $\mathrm{co} S$. Finally, $\delta_S \colon \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is an indicator function of a set $S \subseteq \mathbb{R}^n$ defined by

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

4.2 Positively homogeneous optimization problems and their duality

In this section, we recall positively homogeneous optimization problems and their properties in [108]. The positively homogeneous and vector positively homogeneous functions are defined respectively, as follows.

Definition 4.1. *(Positively homogeneous functions) A function $\psi \colon \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is positively homogeneous if $\psi(\lambda x) = \lambda \psi(x)$ for all $x \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}_{++}$.*

Definition 4.2. *(Vector positively homogeneous functions) A mapping $\Psi \colon \mathbb{R}^n \to (\mathbb{R} \cup \{\infty\})^m$ 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 $\psi_i \colon \mathbb{R}^{n_i} \to \mathbb{R} \cup \{\infty\}$, $i = 1, \dots, m$, where $n = n_1 + \cdots + n_m$, $I_i \subseteq \{1, \dots, n\}$ is a set of indices satisfying $I_i \cap I_j = \emptyset$ for all $i \neq j$, and $\#I_i = n_i$.

The polar of a positively homogeneous function ψ and similarly the polar of a vector positively homogeneous function Ψ are defined as follows. Note that the paper [108] calls such polar positively homogeneous functions dual functions.

Definition 4.3. *(Polar positively homogeneous functions) Let* $\psi\colon \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ *be a positively homogeneous function. Then,* $\psi^\circ\colon \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ *defined by*

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

is called the polar positively homogeneous function of ψ.

Note that a polar positively homogeneous function is positively homogeneous and convex. Moreover, when ψ is a norm, ψ° is the dual norm of ψ.

Definition 4.4. *(Polar vector positively homogeneous functions) Let* $\Psi\colon \mathbb{R}^n \to (\mathbb{R} \cup \{\infty\})^m$ *be a vector positively homogeneous function. A function* $\Psi^\circ\colon \mathbb{R}^n \to (\mathbb{R} \cup \{\infty\})^m$ *is the polar vector positively homogeneous function associated with* Ψ *if* Ψ° *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 $\psi_i^\circ\colon \mathbb{R}^{n_i} \to \mathbb{R} \cup \{\infty\}$ *of positively homogeneous function* ψ_i, $i = 1, \dots m$.

Yamanaka and Yamashita [108] assumed the nonnegativity of positively homogeneous functions for the weak duality of PHO problems as follows.

Assumption 4.1. *Each positively homogeneous function* ψ_i *in* Ψ *is nonnegative, that is,* $\psi_i(x_{I_i}) \geq 0$ *for all* $x_{I_i} \in \mathbb{R}^{n_i}$.

Note that we have to show the following lemma that corresponds to [108, Proposition 2.1] in which $\mathrm{dom}\Psi = \mathbb{R}^n$ is assumed, because we consider $\mathrm{dom}\Psi \neq \mathbb{R}^n$ in this chapter.

Lemma 4.1. *Let* Ψ *and* Ψ° *be a vector positively homogeneous function and its polar, respectively. Then, we have*

$$\Psi^\circ(y) \geq 0.$$

In addition, suppose that Assumption 4.1 holds. Then,

$$\Psi(x)^T \Psi^\circ(y) \geq x^T y$$

holds for all $x \in \mathrm{dom}\Psi$ *and* $y \in \mathrm{dom}\Psi^\circ$.

Proof. Since the first inequality has been shown in [108, Proposition 2.1] by using Definitions 4.1, 4.3 and 4.4, 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 ψ_i and x_{I_i} as ψ and x, respectively.

If $\psi(x) = 0$, then we can show that $x^T y \leq 0$ for all $y \in \operatorname{dom}\psi^\circ$ as follows. Suppose to the contrary that there exists $y \in \operatorname{dom}\psi^\circ$ 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 $y \in \operatorname{dom}\psi^\circ$. Consequently, we obtain $\psi(x)\psi^\circ(y) = 0 \geq x^T y$.

Next, we consider the case where $\psi(x) > 0$. Note that $x \in \operatorname{dom}\Psi$, and hence $\psi(x) < \infty$. Let $z = x/\psi(x)$. Since ψ 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. □

Yamanaka and Yamashita [108] proposed the following dual of $(\mathrm{P}_{\mathrm{PHO}^*})$:

$$\begin{array}{ll} \max & b^T u - p^T v \\ \text{s.t.} & \Psi^\circ(A^T u - H^T v - c) + B^T u - K^T v \leq d, \\ & v \geq 0, \end{array} \tag{$\mathrm{D}_{\mathrm{PHO}^*}$}$$

where $(u, v) \in \mathbb{R}^k \times \mathbb{R}^\ell$. Note that if (u, v) is feasible for $(\mathrm{D}_{\mathrm{PHO}^*})$, then $A^T u - H^T v - c \in \operatorname{dom}\Psi^\circ$. For problems $(\mathrm{P}_{\mathrm{PHO}^*})$ and $(\mathrm{D}_{\mathrm{PHO}^*})$, the following weak duality holds.

Theorem 4.1. *(Weak duality) Suppose that Assumption 4.1 holds. Let $x \in \mathbb{R}^n$ and $(u, v) \in \mathbb{R}^k \times \mathbb{R}^\ell$ be feasible solutions of $(\mathrm{P}_{\mathrm{PHO}^*})$ and $(\mathrm{D}_{\mathrm{PHO}^*})$, respectively. Then, the following inequality holds:*

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

Proof. Using Lemma 4.1, we can show the weak duality as in the proof of [108, Theorem 3.1]. □

In the following, we show that the optimal values and solutions of problems ($\mathrm{D}_{\mathrm{PHO}^*}$) and the Lagrangian dual of ($\mathrm{P}_{\mathrm{PHO}^*}$) are the same. We now define the Lagrangian function $\mathcal{L}\colon \mathrm{dom}\Psi \times \mathbb{R}^k \times \mathbb{R}^\ell \to \mathbb{R}$ of ($\mathrm{P}_{\mathrm{PHO}^*}$) by

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

and consider the following problem:

$$\inf_{x \in \mathrm{dom}\Psi} \sup_{u \in \mathbb{R}^k,\, v \in \mathbb{R}^\ell_+} \mathcal{L}(x,u,v). \tag{$\mathrm{P}^{\mathcal{L}}_{\mathrm{PHO}}$}$$

Note that problem ($\mathrm{P}^{\mathcal{L}}_{\mathrm{PHO}}$) is equivalent to the original problem ($\mathrm{P}_{\mathrm{PHO}^*}$) because the following relation holds:

$$\begin{aligned} &\inf_{x \in \mathrm{dom}\Psi} \sup_{u \in \mathbb{R}^k,\, v \in \mathbb{R}^\ell_+} \mathcal{L}(x,u,v) \\ &= \inf_{x \in \mathrm{dom}\Psi} \begin{cases} c^T x + d^T \Psi(x) & \text{if } Ax + B\Psi(x) = b \text{ and } Hx + K\Psi(x) \le p, \\ +\infty & \text{otherwise.} \end{cases} \end{aligned} \tag{4.2.1}$$

We also note that the Lagrangian dual of problem ($\mathrm{P}^{\mathcal{L}}_{\mathrm{PHO}}$) is described as

$$\sup_{u \in \mathbb{R}^k,\, v \in \mathbb{R}^\ell_+} \omega(u,v), \tag{$\mathrm{D}^{\mathcal{L}}_{\mathrm{PHO}}$}$$

where $\omega\colon \mathbb{R}^k \times \mathbb{R}^\ell \to \mathbb{R}$ is defined by

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

and we explicitly require $x \in \mathrm{dom}\Psi$ in the Lagrangian dual problem ($\mathrm{D}^{\mathcal{L}}_{\mathrm{PHO}}$).

Note that Yamanaka and Yamashita [108] further assumed a condition on positively homogeneous functions for the equivalence between ($\mathrm{D}_{\mathrm{PHO}^*}$) and ($\mathrm{D}^{\mathcal{L}}_{\mathrm{PHO}}$). The condition is that each component ψ_i of Ψ vanishes only at zero and $\mathrm{dom}\Psi = \mathbb{R}^n$. 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 4.2. *For each i, one of the following conditions hold:*

(a) $d_i \ge 0$, $B_{ji} = 0$ *and* $K_{ji} \ge 0$ *for all* j,

(b) $\mathrm{dom}\,\psi_i = \mathbb{R}^{n_i}$ *and there exists* $\hat{x}_{I_i}$ *such that* $\psi_i(\hat{x}_{I_i}) \neq 0$.

Note that if problem ($\mathrm{P}_{\mathrm{PHO}^*}$) satisfies the first condition (a) of Assumption 4.2 for all i and all ψ_i are gauge functions, then it becomes a convex gauge optimization problem ($\mathrm{P}_{\mathrm{GO}^*}$).

We prove the following key lemma for the equivalence between ($\mathrm{D}_{\mathrm{PHO}^*}$) and ($\mathrm{D}^{\mathcal{L}}_{\mathrm{PHO}}$). Note that it is an extension of [108, Lemma 4.1] to the case where $\mathrm{dom}\,\psi_i \neq \mathbb{R}^{n_i}$.

Lemma 4.2. *Let ψ_i° be the polar positively homogeneous functions of ψ_i for $i = 1, \dots, m$. Suppose that Assumptions 4.1 and 4.2 hold. Assume also that $(\bar{u}, \bar{v})$ with $\bar{v} \geq 0$ is not a feasible solution of problem* ($\mathrm{D}_{\mathrm{PHO}^*}$)*. 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 ($\mathrm{D}_{\mathrm{PHO}^*}$). Then, there exists an index j such that

$$\psi_j^\circ(\alpha_{I_j}) > \beta_j, \tag{4.2.2}$$

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 $\bar{x} := (0, \dots, 0, \bar{x}_{I_j}, 0, \dots, 0)$. Then we have $\Psi(\bar{x}) = (0, \dots, 0, \psi_j(\bar{x}_{I_j}), 0, \dots, 0)$ and

$$\mathcal{L}(\bar{x}, \bar{u}, \bar{v}) = -\alpha_{I_j}^T\bar{x}_{I_j} + \beta_j\psi_j(\bar{x}_{I_j}) + b^T\bar{u} + p^T\bar{v}. \tag{4.2.3}$$

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

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

$$\psi_j^\circ(\alpha_{I_j}) = \sup_{x_{I_j}}\{x_{I_j}^T\alpha_{I_j} \mid \psi_j(x_{I_j}) \leq 1\}. \tag{4.2.4}$$

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

$$\psi_j^\circ(\alpha_{I_j}) - \varepsilon \leq \alpha_{I_j}^T\bar{x}_{I_j}(\varepsilon), \quad \psi_j(\bar{x}_{I_j}(\varepsilon)) \leq 1. \tag{4.2.5}$$

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

$$\psi_j^\circ(\alpha_{I_j}) - \bar{\varepsilon} \leq \alpha_{I_j}^T\bar{x}_{I_j}, \quad \psi_j(\bar{x}_{I_j}) = 1. \tag{4.2.6}$$

Since $\psi_j^\circ(\alpha_{I_j}) > \bar{\varepsilon}$, the inequality (4.2.5) implies $\alpha_{I_j}^T\bar{x}_{I_j}(\bar{\varepsilon}) > 0$, and hence $\bar{x}_{I_j}(\bar{\varepsilon}) \neq 0$. If $\psi_j(\bar{x}_{I_j}(\bar{\varepsilon})) \neq 0$, then we set $\bar{x}_{I_j} = \bar{x}_{I_j}(\bar{\varepsilon})/\psi_j(\bar{x}_{I_j}(\bar{\varepsilon}))$. This vector $\bar{x}_{I_j}$ satisfies conditions (4.2.6) as shown below.

$$\psi_j^\circ(\alpha_{I_j}) - \bar{\varepsilon} \leq \alpha_{I_j}^T\bar{x}_{I_j}(\bar{\varepsilon}) \leq \alpha_{I_j}^T\frac{\bar{x}_{I_j}(\bar{\varepsilon})}{\psi_j(\bar{x}_{I_j}(\bar{\varepsilon}))} = \alpha_{I_j}^T\bar{x}_{I_j},$$

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

where the second inequality holds from Assumption 4.1 and (4.2.5). If $\psi_j(\bar{x}_{I_j}(\bar{\varepsilon})) = 0$, then $\psi_j(t\bar{x}_{I_j}(\bar{\varepsilon})) = t\psi_j(\bar{x}_{I_j}(\bar{\varepsilon})) = 0$ for all $t > 0$ because ψ_j is positively homogeneous. From (4.2.4), we have $\psi_j^\circ(\alpha_{I_j}) \geq t\bar{x}_{I_j}(\bar{\varepsilon})^T\alpha_{I_j}$. Since $\alpha_{I_j}^T\bar{x}_{I_j}(\bar{\varepsilon}) > 0$, we obtain $\psi_j^\circ(\alpha_{I_j}) \to \infty$ as $t \to \infty$, which is a contradiction. Therefore, there exists $\bar{x}_{I_j}$ such that (4.2.6) holds.

We now denote $\bar{t} = (0, \ldots, 0, t\bar{x}_{I_j}, 0, \ldots, 0)$ for $t > 0$. Then, we have from (4.2.3)

$$\begin{aligned}
\mathcal{L}(\bar{t}, \bar{u}, \bar{v}) &= -t\alpha_{I_j}^T\bar{x}_{I_j} + \beta_j\psi_j(t\bar{x}_{I_j}) + b^T\bar{u} + p^T\bar{v} \\
&\leq -t(\psi_j^\circ(\alpha_{I_j}) - \bar{\varepsilon} - \beta_j\psi_j(\bar{x}_{I_j})) + b^T\bar{u} + p^T\bar{v} \\
&= -t(\psi_j^\circ(\alpha_{I_j}) - \bar{\varepsilon} - \beta_j) + b^T\bar{u} + p^T\bar{v},
\end{aligned}$$

where the second inequality and the third equality hold from (4.2.6). Since $\bar{\varepsilon} \leq (\psi_j^\circ(\alpha_{I_j}) - \beta_j)/2$, we obtain

$$\begin{aligned}
\mathcal{L}(\bar{t}, \bar{u}, \bar{v}) &\leq -t(\psi_j^\circ(\alpha_{I_j}) - \bar{\varepsilon} - \beta_j) + b^T\bar{u} + p^T\bar{v} \\
&\leq -t\left(\frac{\psi_j^\circ(\alpha_{I_j}) - \beta_j}{2}\right) + b^T\bar{u} + p^T\bar{v},
\end{aligned}$$

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

Next, we consider the case where $\psi_j^\circ(\alpha_{I_j}) = \infty$. From (4.2.4), there exists a sequence $\{\bar{x}_{I_j}^k\} \subset \operatorname{dom}\psi_j$ such that $\psi_j(\bar{x}_{I_j}^k) \leq 1$ and $(\bar{x}_{I_j}^k)^T\alpha_{I_j} \to \infty$ as $k \to \infty$. Let $\bar{x}^k = (0, \ldots, 0, \bar{x}_{I_j}^k, 0, \ldots, 0)$. Then, it follows from (4.2.3) that

$$\mathcal{L}(\bar{x}^k, \bar{u}, \bar{v}) = -\alpha_{I_j}^T\bar{x}_{I_j}^k + \beta_j\psi_j(\bar{x}_{I_j}^k) + b^T\bar{u} + p^T\bar{v},$$

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

We finally study the case where $\psi_j^\circ(\alpha_{I_j}) = 0$. Note that $0 > \beta_j$ from (4.2.2). When the first condition (a) of Assumption 4.2 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 4.2 holds. If $\alpha_{I_j} \neq 0$, then there exists $\bar{\varepsilon} > 0$ such that $1 \geq \psi_j(\bar{\varepsilon}\alpha_{I_j}) = \bar{\varepsilon}\psi_j(\alpha_{I_j})$. Therefore we have

$$\psi_j^\circ(\alpha_{I_j}) = \sup_{x_{I_j}}\{x_{I_j}^T\alpha_{I_j} \mid \psi_j(x_{I_j}) \leq 1\} \geq \bar{\varepsilon}\alpha_{I_j}^T\alpha_{I_j} > 0,$$

which is a contradiction. Now we consider the case where $\alpha_{I_j} = 0$. From Assumption 4.2 (b), there exists $\hat{x}_{I_j}$ such that $\psi_j(\hat{x}_{I_j}) > 0$. Let $\hat{x}(t) = (0, \ldots, 0, t\hat{x}_{I_j}, 0, \ldots, 0)$

with $t > 0$. Then, it follows from (4.2.3) that

$$\begin{aligned}\mathcal{L}(\hat{x}(t), \bar{u}, \bar{v}) &= -\alpha_{I_j}^T \hat{x}_{I_j}(t) + \beta_j \psi_j(t\hat{x}_{I_j}) + b^T \bar{u} + p^T \bar{v} \\ &= t\beta_j \psi_j(\hat{x}_{I_j}) + b^T \bar{u} + p^T \bar{v},\end{aligned}$$

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 ($\mathrm{D}_{\mathrm{PHO}^*}$) and ($\mathrm{D}^{\mathcal{L}}_{\mathrm{PHO}}$) are equivalent, which means that their optimal values and solutions of those problems are the same.

Theorem 4.2. *Suppose that the Lagrangian dual problem* ($\mathrm{D}^{\mathcal{L}}_{\mathrm{PHO}}$) *has a feasible solution. Suppose also that Assumptions 4.1 and 4.2 hold. Then, the optimal value and optimal solutions of problem* ($\mathrm{D}_{\mathrm{PHO}^*}$) *are the same as those of* ($\mathrm{D}^{\mathcal{L}}_{\mathrm{PHO}}$).

Proof. The result can be proved by using Lemma 4.2 as in the proof of [108, Theorem 4.1]. □

The next proposition shows the positively homogeneous dual of problem ($\mathrm{D}_{\mathrm{PHO}^*}$), which is similar to ($\mathrm{P}_{\mathrm{PHO}^*}$).

Proposition 4.1. *Suppose that problem* ($\mathrm{D}_{\mathrm{PHO}^*}$) *is feasible. Then, the positively homogeneous dual of* ($\mathrm{D}_{\mathrm{PHO}^*}$) *can be written as*

$$\begin{aligned}\min\quad & c^T x + d^T y \\ \text{s.t.}\quad & Ax + By = b, \\ & Hx + Ky \le p, \\ & \Psi^{\circ\circ}(x) \le y,\end{aligned} \tag{$\mathrm{P}'_{\mathrm{PHO}^*}$}$$

where $\Psi^{\circ\circ}$ *denotes the polar of* Ψ°, *i.e.,* $\Psi^{\circ\circ} = (\Psi^\circ)^\circ$.

Proof. First, note that problem ($\mathrm{D}_{\mathrm{PHO}^*}$) can be written as

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

This problem is further reformulated as

$$\begin{aligned}\min\quad & \hat{c}^T \theta \\ \text{s.t.}\quad & \hat{K}\hat{\Psi}^\circ(\theta) + \hat{H}\theta \le \hat{p}, \\ & \hat{A}\theta = c,\end{aligned} \tag{4.2.7}$$

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

$$\hat{K} = \begin{bmatrix} 0 & 0 & E_n \\ 0 & 0 & 0 \end{bmatrix} \in \mathbb{R}^{(n+\ell)\times(k+\ell+n)}, \hat{H} = \begin{bmatrix} B^T & -K^T & 0 \\ 0 & -E_\ell & 0 \end{bmatrix} \in \mathbb{R}^{(n+\ell)\times(k+\ell+n)},$$

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 (4.2.7) can be described as

$$\begin{array}{rl} \max & c^T x - \hat{p}^T y \\ \text{s.t.} & \hat{\Psi}^{\circ\circ}(\hat{A}^T x - \hat{H}^T y - \hat{c}) - \hat{K}^T y \leq 0, \\ & y \geq 0. \end{array}$$

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

$$\begin{array}{rl} \min & -c^T x + d^T y_1 \\ \text{s.t.} & \|Ax - By_1 + b\|_2 \leq 0, \\ & \| - Hx + Ky_1 + y_2 - p\|_2 \leq 0, \\ & \Psi^{\circ\circ}(-x) - y_1 \leq 0, \\ & y \geq 0. \end{array} \tag{4.2.8}$$

The first two inequality constraints are equivalent to

$$\begin{aligned} -Ax + By_1 &= b, \\ -Hx + Ky_1 + y_2 &= p. \end{aligned}$$

Since $y_2 \geq 0$ in (4.2.8), the second equality is further reduced to $-Hx + Ky_1 \leq p$. Consequently, we can reformulate (4.2.8) as

$$\begin{array}{rl} \min & -c^T x + d^T y_1 \\ \text{s.t.} & -Ax + By_1 = b, \\ & -Hx + Ky_1 \leq p, \\ & \Psi^{\circ\circ}(-x) \leq y_1, \end{array}$$

which is precisely ($\mathrm{P}'_{\mathrm{PHO}^*}$) by denoting $-x$ and y_1 as x and y, respectively. □

In Lagrangian duality theory (see e.g. [34, p. 138]), it is well-known that the Lagrangian dual of ($\mathrm{D}^{\mathcal{L}}_{\mathrm{PHO}}$) is exactly the problem ($\mathrm{P}^{\mathcal{L}}_{\mathrm{PHO}}$). Therefore, we obtain the following result.

Collolary 4.1. *Suppose that problem* ($\mathrm{D_{PHO^*}}$) *is feasible. Suppose also that Assumptions 4.1 and 4.2 hold. Then, problem* ($\mathrm{P'_{PHO^*}}$) *is equivalent to the original problem* ($\mathrm{P_{PHO^*}}$).

Proof. From equality (4.2.1), problem ($\mathrm{P_{PHO^*}}$) is equivalent to problem ($\mathrm{P^{\mathcal{L}}_{PHO}}$). Then, from Theorem 4.2, problem ($\mathrm{D_{PHO^*}}$) is equivalent to ($\mathrm{D^{\mathcal{L}}_{PHO}}$). Moreover, from Theorem 4.2, the positively homogeneous dual and the Lagrangian dual are equivalent under Assumptions 4.1 and 4.2. Therefore, the positively homogeneous dual of ($\mathrm{D_{PHO^*}}$) is equivalent to the Lagrangian dual of ($\mathrm{D^{\mathcal{L}}_{PHO}}$), which is ($\mathrm{P^{\mathcal{L}}_{PHO}}$) and it is equivalent to ($\mathrm{P_{PHO^*}}$). □

4.3 Gauge optimization problems and their duality

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

$$\begin{array}{ll} \min & c^T x + d^T \mathcal{G}(x) \\ \text{s.t.} & Ax = b, \\ & Hx + K\mathcal{G}(x) \leq p, \\ & x \in \operatorname{dom} \mathcal{G}. \end{array} \tag{$\mathrm{P_{GO^*}}$}$$

We call $\mathcal{G}$ a vector gauge function defined as $\mathcal{G} := (g_1(\cdot), \ldots, g_m(\cdot))^T$ with $g_i \colon \mathbb{R}^{n_i} \to \mathbb{R} \cup \{\infty\}$ as a gauge function for all i. Since ($\mathrm{P_{GO^*}}$) is a special case of ($\mathrm{P_{PHO^*}}$), the PHO dual of ($\mathrm{P_{GO^*}}$) is written as follows:

$$\begin{array}{ll} \max & b^T u - p^T v \\ \text{s.t.} & \mathcal{G}^\circ(A^T u - H^T v - c) - K^T v \leq d, \\ & v \geq 0, \end{array} \tag{$\mathrm{D_{GO^*}}$}$$

where $\mathcal{G}^\circ$ is the polar function associated with $\mathcal{G}$. Here, problem ($\mathrm{D_{GO^*}}$) is a convex optimization problem since each component g_i° of $\mathcal{G}^\circ$ is convex.

The next proposition is a corollary of Lemma 4.1. Note that since gauge functions are nonnegative, Assumption 4.1 automatically holds.

Proposition 4.2. *Let $\mathcal{G}$ and $\mathcal{G}^\circ$ be a vector gauge function and its polar, respectively. Then, we have*

$$\begin{array}{rcl} \mathcal{G}^\circ(y) & \geq & 0, \\ \mathcal{G}(x)^T \mathcal{G}^\circ(y) & \geq & x^T y \end{array}$$

for any $x \in \operatorname{dom} \mathcal{G}$ *and* $y \in \operatorname{dom} \mathcal{G}^\circ$.

Proof. The proof follows from Lemma 4.1. □

We have the weak duality theorem for problems ($\mathrm{P}_{\mathrm{GO}^*}$) and ($\mathrm{D}_{\mathrm{GO}^*}$), and the equivalence between ($\mathrm{D}_{\mathrm{GO}^*}$) and the Lagrangian dual of ($\mathrm{P}_{\mathrm{GO}^*}$) from Proposition 4.2 and Theorem 4.2. Throughout the paper, we denote the Lagrangian dual of ($\mathrm{P}_{\mathrm{GO}^*}$) as ($\mathrm{D}^{\mathcal{L}}_{\mathrm{GO}^*}$).

Collolary 4.2. *(Weak duality) For problems* ($\mathrm{P}_{\mathrm{GO}^*}$) *and* ($\mathrm{D}_{\mathrm{GO}^*}$), *the following inequality holds:*

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

for all feasible points $x \in \mathbb{R}^n$ *and* $(u, v) \in \mathbb{R}^k \times \mathbb{R}^\ell$ *of* ($\mathrm{P}_{\mathrm{GO}^*}$) *and* ($\mathrm{D}_{\mathrm{GO}^*}$), *respectively.*

Proof. The proof directly follows from Proposition 4.2. □

Collolary 4.3. *Suppose that the Lagrangian dual problem* ($\mathrm{D}^{\mathcal{L}}_{\mathrm{GO}^*}$) *has a feasible solution. Suppose also that Assumption 4.2 holds. Then, the optimal value and solutions of problem* ($\mathrm{D}_{\mathrm{GO}^*}$) *are the same as* ($\mathrm{D}^{\mathcal{L}}_{\mathrm{GO}^*}$).

Proof. The proof is a direct consequence of Theorem 4.2. □

We now discuss the strong duality, necessary and sufficient optimality conditions, and the primal recovery for problem ($\mathrm{P}_{\mathrm{GO}^*}$). To this end, we need ($\mathrm{P}_{\mathrm{GO}^*}$) to be convex. Thus, from now on, we suppose the following assumption.

Assumption 4.3. *All elements of* d *and* K *of problem* ($\mathrm{P}_{\mathrm{GO}^*}$) *are nonnegative.*

Note that if Assumption 4.3 holds, then Assumption 4.2 holds for ($\mathrm{P}_{\mathrm{PHO}^*}$) with $\Psi = \mathcal{G}$. Moreover, we assume the following condition on each function g_i.

Assumption 4.4. *Each function* g_i *of* $\mathcal{G}$ *is lower semi-continuous on* $\mathbb{R}^{n_i}$.

We now show that the dual of ($\mathrm{D}_{\mathrm{GO}^*}$) becomes ($\mathrm{P}_{\mathrm{GO}^*}$) under Assumptions 4.3 and 4.4.

Collolary 4.4. *Suppose that Assumptions 4.3 and 4.4 hold. Assume also that problem* ($\mathrm{D}_{\mathrm{GO}^*}$) *is feasible. Then, the positively homogeneous dual of* ($\mathrm{D}_{\mathrm{GO}^*}$) *is equivalent to* ($\mathrm{P}_{\mathrm{GO}^*}$).

Proof. Since g_i is a gauge function for all i and satisfies Assumption 4.4, we have $\mathcal{G}^{\circ\circ} = \mathcal{G}$ by [81, Theorem 15.1]. Then, it follows from Proposition 4.1 that the dual of ($\mathrm{D_{GO^*}}$) becomes

$$\begin{array}{ll} \min & c^T x + d^T y \\ \text{s.t.} & Ax = b, \\ & Hx + Ky \le p, \\ & \mathcal{G}(x) \le y. \end{array} \qquad (\mathrm{P}'_{\mathrm{GO}^*})$$

We show that the optimal value of ($\mathrm{P_{GO^*}}$) is the same as that of ($\mathrm{P}'_{\mathrm{GO}^*}$). Let x^* be an optimal solution of ($\mathrm{P_{GO^*}}$). Then, $(\bar{x}, \bar{y}) := (x^*, \mathcal{G}(x^*))$ is feasible for ($\mathrm{P}'_{\mathrm{GO}^*}$), and hence $c^T x^* + d^T \mathcal{G}(x^*) \ge c^T \bar{x} + d^T \bar{y}$. This shows that the optimal value of ($\mathrm{P}'_{\mathrm{GO}^*}$) is less than or equal to that of ($\mathrm{P_{GO^*}}$).

Next, let $(\hat{x}, \hat{y})$ be an optimal solution of ($\mathrm{P}'_{\mathrm{GO}^*}$). From Assumptions 4.3 and the fact that $\mathcal{G}(\hat{x}) \le \hat{y}$, we have $c^T \hat{x} + d^T \mathcal{G}(\hat{x}) \le c^T \hat{x} + d^T \hat{y}$ and $H\hat{x} + K\mathcal{G}(\hat{x}) \le H\hat{x} + K\hat{y} \le p$. Therefore, $(\hat{x}, \mathcal{G}(\hat{x}))$ is also optimal for ($\mathrm{P}'_{\mathrm{GO}^*}$). Moreover, $\hat{x}$ is a feasible solution of ($\mathrm{P_{GO^*}}$) and $c^T \hat{x} + d^T \mathcal{G}(\hat{x}) \le c^T \hat{x} + d^T \hat{y}$. The result indicates that the optimal value of ($\mathrm{P_{GO^*}}$) is less than or equal to that of ($\mathrm{P}'_{\mathrm{GO}^*}$).

The above discussion shows that the optimal values of ($\mathrm{P_{GO^*}}$) and ($\mathrm{P}'_{\mathrm{GO}^*}$) are the same. Furthermore, if x^* is optimal for ($\mathrm{P_{GO^*}}$), then $(x^*, \mathcal{G}(x^*))$ is optimal for ($\mathrm{P}'_{\mathrm{GO}^*}$). Conversely, if $(\hat{x}, \hat{y})$ is an optimal solution of ($\mathrm{P}'_{\mathrm{GO}^*}$), then $\hat{x}$ is optimal for ($\mathrm{P_{GO^*}}$). □

4.3.1 Strong duality

We now focus on the strong duality between problems ($\mathrm{P_{GO^*}}$) and ($\mathrm{D_{GO^*}}$). As seen below, we require a certain constraint qualification for this purpose. Note that the Slater's constraint qualification of problem ($\mathrm{P_{GO^*}}$), which we use in the following theorem, indicates that there exists x_0 such that

$$x_0 \in \{x \mid Ax = b, Hx + K\mathcal{G}(x) < p, x \in \operatorname{dom} \mathcal{G}\}.$$

Theorem 4.3. *(Strong duality) Suppose that Assumption 4.3 holds. Suppose also that the Slater constraint qualification holds for* ($\mathrm{P_{GO^*}}$)*. Then, the strong duality holds for problems* ($\mathrm{P_{GO^*}}$) *and* ($\mathrm{D_{GO^*}}$)*, i.e., if* ($\mathrm{P_{GO^*}}$) *has an optimal solution x^*, then* ($\mathrm{D_{GO^*}}$) *also has an optimal solution (u^*, v^*), and the duality gap between* ($\mathrm{P_{GO^*}}$) *and* ($\mathrm{D_{GO^*}}$) *is zero, that is, $c^T x^* + d^T \mathcal{G}(x^*) = b^T u^* - p^T v^*$.*

Proof. Suppose that ($\mathrm{P_{GO^*}}$) has a solution. Since ($\mathrm{P_{GO^*}}$) is convex from Assumptions 4.3, and the Slater constraint qualification holds for ($\mathrm{P_{GO^*}}$), the strong duality holds between problems $(\mathrm{P}_{\mathrm{GO}^*})$ and $(\mathrm{D}^{\mathcal{L}}_{\mathrm{GO}^*})$. This means that $(\mathrm{D}^{\mathcal{L}}_{\mathrm{GO}^*})$ also has an optimal solution and the duality gap between $(\mathrm{P}_{\mathrm{GO}^*})$ and $(\mathrm{D}^{\mathcal{L}}_{\mathrm{GO}^*})$ is zero. It then follows from Corollary 4.3 that the optimal value of $(\mathrm{D}_{\mathrm{GO}^*})$ is the same as that of $(\mathrm{P}_{\mathrm{GO}^*})$. Moreover, since an optimal solution of $(\mathrm{D}^{\mathcal{L}}_{\mathrm{GO}^*})$ is that of $(\mathrm{D}_{\mathrm{GO}^*})$, $(\mathrm{D}_{\mathrm{GO}^*})$ has an optimal solution. □

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 4.1. *Consider the following one dimensional gauge optimization problem:*

$$\begin{array}{rl} \min & |x-1| \\ \text{s.t.} & |x| \leq 0. \end{array} \tag{P_a}$$

Then, the dual of (P_a) *is described as*

$$\begin{array}{rl} \max & u \\ \text{s.t.} & |u| \leq v \\ & |-u| \leq 1 \\ & v \geq 0, \end{array} \tag{D_a}$$

where $u, v \in \mathbb{R}$. *Clearly, the feasible region of* (P_a) *is* $\{0\}$ *and the Slater constraint qualification fails. However, the optimal solutions of* (P_a) *and* (D_a) *are* $x^* = 0$ *and* $(u^*, v^*) = (1, c), c \geq 1$, *respectively, and the optimal values of* (P_a) *and* (D_a) *are the same.*

4.3.2 Optimality conditions

The most well-known optimality conditions in the optimization field 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 $(\mathrm{P}_{\mathrm{GO}^*})$ and $(\mathrm{D}_{\mathrm{GO}^*})$. Note that we do not assume the Slater constraint qualification and Assumption 4.3 here.

Theorem 4.4. *(Sufficient optimality conditions) Points* x^* *and* (u^*, v^*) *are optimal for* $(\mathrm{P}_{\mathrm{GO}^*})$ *and* $(\mathrm{D}_{\mathrm{GO}^*})$, *respectively, if the following conditions hold:*

(i) $Hx^* + K\mathcal{G}(x^*) \leq p,\ Ax^* = b,\ x^* \in \operatorname{dom}\mathcal{G}$,

(ii) $\mathcal{G}^\circ(A^Tu^* - H^Tv^* - c) - K^Tv^* \leq d,\ v^* \geq 0$,

(iii) $\left[d + K^Tv^* - \mathcal{G}^\circ(A^Tu^* - H^Tv^* - c)\right]_i\ g_i(x^*_{I_i}) = 0,\ i = 1, \ldots, m$,

(iv) $[p - Hx^* - K\mathcal{G}(x^*)]_i\ v^*_i = 0,\ i = 1, \ldots, m$,

(v) $\mathcal{G}^\circ(A^Tu^* - H^Tv^* - c)^T\mathcal{G}(x^*) = (A^Tu^* - H^Tv^* - c)^Tx^*$,

where (i) *and* (ii) *describe the primal and dual feasibility, respectively, items* (iii) *and* (iv) *represent complementarity, and* (v) *is the so-called alignment condition.*

Proof. From the complementarity conditions (iii) and (iv), we obtain

$$\begin{aligned} 0 &= \left[d + K^Tv^* - \mathcal{G}^\circ(A^Tu^* - H^Tv^* - c)\right]^T\mathcal{G}(x^*) + [p - Hx^* - K\mathcal{G}(x^*)]^T v^* \\ &= d^T\mathcal{G}(x^*) - \mathcal{G}^\circ(A^Tu^* - H^Tv^* - c)^T\mathcal{G}(x^*) + p^Tv^* - (Hx^*)^Tv^*. \end{aligned}$$

It then follows from the alignment condition that we have

$$\begin{aligned} &d^T\mathcal{G}(x^*) - \mathcal{G}^\circ(A^Tu^* - H^Tv^* - c)^T\mathcal{G}(x^*) + p^Tv^* - (Hx^*)^Tv^* \\ &\quad = d^T\mathcal{G}(x^*) - (A^Tu^* - H^Tv^* - c)^Tx^* + p^Tv^* - (Hx^*)^Tv^* \\ &\quad = c^Tx^* + d^T\mathcal{G}(x^*) - b^Tu^* + p^Tv^*, \end{aligned}$$

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 $(\mathrm{P}_{\mathrm{GO}^*})$ and $(\mathrm{D}_{\mathrm{GO}^*})$, respectively. □

Note that condition (v) in Theorem 4.4, called the alignment condition, is not standard, and seems to be strange at first glance. This is actually used in the previous work [6] 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 $(\mathrm{P}_{\mathrm{GO}^*})$ and Assumption 4.3 hold, the sufficient optimality conditions in Theorem 4.4 become necessary.

Theorem 4.5. *(Necessary conditions for optimality) Suppose that Assumption 4.3 holds. Suppose also that the Slater constraint qualification holds for* $(\mathrm{P}_{\mathrm{GO}^*})$*. Let* x^* *and* (u^*, v^*) *be optimal solutions of* $(\mathrm{P}_{\mathrm{GO}^*})$ *and* $(\mathrm{D}_{\mathrm{GO}^*})$*, respectively. Then conditions (i)–(v) in Theorem 4.4 hold.*

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

$$\begin{aligned}
0 &= c^T x^* + d^T \mathcal{G}(x^*) - b^T u^* + p^T v^* \\
&= d^T \mathcal{G}(x^*) - (A^T u^* - H^T v^* - c)^T x^* + p^T v^* - (Hx^*)^T v^* \\
&\geq d^T \mathcal{G}(x^*) - \mathcal{G}^\circ(A^T u^* - H^T v^* - c)^T \mathcal{G}(x^*) + p^T v^* - (Hx^*)^T v^* \\
&= \left[d + K^T v^* - \mathcal{G}^\circ(A^T u^* - H^T v^* - c)\right]^T \mathcal{G}(x^*) + \left[p - Hx^* - K\mathcal{G}(x^*)\right]^T v^* \\
&\geq 0,
\end{aligned}$$

where the second equality follows from the fact that $Ax^* = b$, the third inequality follows from Proposition 4.2, 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 [36].

Example 4.2. *Consider the following Ridge-type optimization problem:*

$$\min_{x \in \mathbb{R}^n} \quad \|Ax - b\|_2 + \sigma \|x\|_2, \tag{P_b}$$

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

$$\begin{aligned}
\min_{x,y} \quad & \|y\|_2 + \sigma \|x\|_2 \\
& Ax - y = b,
\end{aligned} \tag{P'_b}$$

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

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

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 for problems* (P'_b) *and* (D_b) *are described as follows:*

(i) $Ax - y = b$,

(ii) $\|A^T u\|_2 \leq \sigma$, $\|-u\|_2 \leq 1$,

(iii) $(\sigma - \|A^T u\|_2)\|x\|_2 = 0$, $(1 - \|-u\|_2)\|y\|_2 = 0$,

(v) $\|A^T u\|_2\|x\|_2 + \|-u\|_2\|y\|_2 = u^T Ax - u^T y$.

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 $(\mathrm{P'}_b)$ *and* (D_b) *are*

$$\begin{cases} \|A^T u\|_2 \leq \sigma, \|-u\|_2 \leq 1, \\ \sigma\|x\|_2 + \|Ax - b\|_2 = b^T u. \end{cases}$$

Note that the above conditions are necessary and sufficient optimality conditions because Assumption 4.3 and the Slater constraint qualifications clearly hold for problems $(\mathrm{P'}_b)$ *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).

4.3.3 Primal recovery

Let us now discuss the recovery of a primal optimal solution from a KKT point of the dual problem $(\mathrm{D}_{\mathrm{GO}^*})$. For simplicity, we denote $\Phi(u, v) := \mathcal{G}^\circ(A^T u - H^T v - c)$ and $\phi_i(u, v) := g_i^\circ(A_{I_i}^T u - H_{I_i}^T v - c_{I_i}), i = 1, \dots, m$. Then, the KKT conditions of $(\mathrm{D}_{\mathrm{GO}^*})$ can be described as

$$\begin{aligned} p + V^T\lambda - K\lambda - \mu = 0, &\qquad V \in \partial_v \Phi(u^*, v^*), && (4.3.1)\\ -b + U^T\lambda = 0, &\qquad U \in \partial_u \Phi(u^*, v^*), && (4.3.2)\\ d - \Phi(u^*, v^*) - K^T v^* \geq 0, \lambda \geq 0, &\qquad \lambda^T(d - \Phi(u^*, v^*) - K^T v^*) = 0, && (4.3.3)\\ v^* \geq 0, \mu \geq 0, &\qquad v^{*T}\mu = 0, && (4.3.4) \end{aligned}$$

where $\lambda \in \mathbb{R}^m$ and $\mu \in \mathbb{R}^\ell$ are Lagrangian multipliers. Let $A_i = A_{I_i}$ and $H_i = H_{I_i}$ for all $i = 1, \dots, m$ in the subsequent discussion. Moreover, we divide the 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 $i = 1, \dots, m$.

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:

$$\begin{array}{ll} \sup\limits_{x_{I_i}} & u^T A_i x_{I_i} - v^T H_i x_{I_i} - c_{I_i}^T x_{I_i} \\ \text{s.t.} & g_i(x_{I_i}) \le 1. \end{array} \tag{P^i}$$

Moreover, we assume the following condition to show some key properties of the set $X_i(u, v)$.

Assumption 4.5. *For all i, g_i vanishes only at 0, that is, $g_i(\bar{x}_{I_i}) = 0$ if and only if $\bar{x}_{I_i} = 0$.*

Lemma 4.3. *Suppose that Assumptions 4.4 and 4.5 hold. Then, the set $X_i(u, v)$ 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 $x_{I_i} = 0$ 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 Assumption 4.4. Moreover, Assumption 4.5 implies that the feasible region is bounded. To see this, let $B_i := \{z \in \mathbb{R}^{n_i} \mid \|z\| = 1\}$ and $\rho := \inf_{z\in B_i} g_i(z)$. Then $\rho > 0$ from Assumption 4.5. If $\rho = +\infty$, that is, $\operatorname{dom} g_i = \{0\}$, then $X_i(u, v) = \{0\}$ and this lemma holds. Now, suppose that $\rho < \infty$. Then, the feasible region is included in the compact set $\bar{B}_i := \{z \mid \|z\| \le 1/\rho\}$ since for any $s \notin \bar{B}_i$ 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. □

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

Lemma 4.4. *Suppose that Assumptions 4.4 and 4.5 hold for function $\mathcal{G}$. Then, we have*

$$\phi_i(u, v) = u^T A_i \bar{x}_{I_i} - v^T H_i \bar{x}_{I_i} - c_{I_i}^T \bar{x}_{I_i} \quad \textit{for all } \bar{x}_{I_i} \in X_i(u, v), \tag{4.3.5}$$

$$\partial_u \phi_i(u,v) = \{\bar{x}_{I_i}^T A_i^T \mid \bar{x}_{I_i} \in X_i(u,v)\} \tag{4.3.6}$$

and

$$\partial_v \phi_i(u,v) = \{-\bar{x}_{I_i}^T H_i^T \mid \bar{x}_{I_i} \in X_i(u,v)\}. \tag{4.3.7}$$

Proof. The first equation directly follows from the definitions of g_i° and $X_i(u,v)$. Since the set $X_i(u,v)$ is nonempty, convex and compact from Lemma 4.3, we obtain

$$\begin{aligned}
\partial_u \phi_i(u,v) &= \mathrm{co}\{\bar{x}_{I_i}^T A_i^T \mid \bar{x}_{I_i} \in X_i(u,v)\} = \{\bar{x}_{I_i}^T A_i^T \mid \bar{x}_{I_i} \in X_i(u,v)\}, \\
\partial_v \phi_i(u,v) &= \mathrm{co}\{-\bar{x}_{I_i}^T H_i^T \mid \bar{x}_{I_i} \in X_i(u,v)\} = \{-\bar{x}_{I_i}^T H_i^T \mid \bar{x}_{I_i} \in X_i(u,v)\},
\end{aligned}$$

which are the desired formulae. □

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 ($\mathrm{D}_{\mathrm{GO}^*}$).

Theorem 4.6. *(Primal recovery) Suppose that Assumptions 4.3, 4.4 and 4.5 hold for the function $\mathcal{G}$. 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 (4.3.1)–(4.3.4) for the dual problem* ($\mathrm{D}_{\mathrm{GO}^*}$)*. Then there exist $\bar{x}_{I_i} \in X_i(u^*, v^*)$ for all $i = 1, \dots, m$ such that $U_i = (A_i \bar{x}_{I_i})^T$ and $V_i = -(H_i \bar{x}_{I_i})^T$. Moreover, suppose that $g_i(\bar{x}_{I_i}) = 1$ for i such that $\lambda_i \neq 0$. Let $x_{I_i}^* = \lambda_i \bar{x}_{I_i}$ for all $i = 1, \dots, m$. Then, $x^* = (x_{I_1}^*, \dots, x_{I_m}^*)^T$ is an optimal solution of* ($\mathrm{P}_{\mathrm{GO}^*}$)*.*

Proof. From the definitions of Φ and $\mathcal{G}^\circ$, we have

$$\begin{aligned}
\Phi(u^*, v^*) = \mathcal{G}^\circ(A^T u^* - H^T v^* - c) &= \begin{pmatrix} g_1^\circ(A_1^T u^* - H_1^T v^* - c_{I_1}) \\ \vdots \\ g_m^\circ(A_m^T u^* - H_m^T v^* - c_{I_m}) \end{pmatrix} \\
&= \begin{pmatrix} \phi_1(u^*, v^*) \\ \vdots \\ \phi_m(u^*, v^*) \end{pmatrix}.
\end{aligned}$$

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 (4.3.6) and (4.3.7) in Lemma 4.4 that, for all $i = 1, \dots, m$, there exist $\bar{x}_{I_i} \in X_i(u^*, v^*)$, such that $U_i = (A_i \bar{x}_{I_i})^T$ and $V_i = -(H_i \bar{x}_{I_i})^T$.

Now let $x^*_{I_i} = \lambda_i \bar{x}_{I_i}$, $i = 1, \ldots, m$, 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 4.4. Note that the dual feasibility (ii) clearly holds. Moreover, since the assumption on $g_i(\bar{x}_{I_i})$ implies

$$g_i(x^*_{I_i}) = g_i\left(\lambda_i \bar{x}_{I_i}\right) = \lambda_i g_i(\bar{x}_{I_i}) = \lambda_i,$$

we obtain

$$\mathcal{G}(x^*) = \lambda. \tag{4.3.8}$$

We first show that the alignment condition (v) holds. From (4.3.5) in Lemma 4.4, we have

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

It then follows from (4.3.8) that

$$\begin{aligned} g_i^\circ(A_i^T u^* - H_i^T v^* - c_{I_i})^T g_i(x^*_{I_i}) &= \lambda_i((u^*)^T A_i \bar{x}_{I_i} - (v^*)^T H_i \bar{x}_{I_i} - c_{I_i}^T \bar{x}_{I_i}) \\ &= (u^*)^T A_i x^*_{I_i} - (v^*)^T H_i x^*_{I_i} - c_{I_i}^T x^*_{I_i} \\ &= (A_i^T u^* - H_i^T v^* - c_{I_i})^T x^*_{I_i}, \end{aligned}$$

which shows that condition (v) holds.

Next we prove the primal feasibility (i). From the definition of x^*, we obtain

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

where the second equality follows from (4.3.6) in Lemma 4.4 and the last equality is due to the KKT condition (4.3.2). Moreover, we have from (4.3.7) in Lemma 4.4 that

$$Hx^* = \sum_{i=1}^m \lambda_i H_i \bar{x}_{I_i} = -\sum_{i=1}^m \lambda_i V_i^T = -V^T \lambda. \tag{4.3.9}$$

It then follows from (4.3.8) that

$$Hx^* + K\mathcal{G}(x^*) = -V^T \lambda + K\lambda = p - \mu \le p,$$

where the equality and the inequality follow from the KKT conditions (4.3.1) and (4.3.4), respectively. Consequently, x^* is a feasible solution of ($\mathrm{P}_{\mathrm{GO}^*}$).

Finally, we show that the complementarity conditions (iii) and (iv) hold. First we consider condition (iii) as follows. If $\lambda_i = 0$, then $x^*_{I_i} = 0$ and $g_i(x^*_{I_i}) = 0$, and hence (iii) holds. If $\lambda_i \neq 0$, then $\left[d + K^T v^* - \mathcal{G}^\circ(A^T u^* - H^T v^* - c)\right]_i = 0$ from the KKT condition (4.3.3) and the definition of Φ. Therefore, (iii) also holds.

Next we prove that condition (iv) is satisfied. If $v_i^* = 0$, then (iv) clearly holds. For this reason, we consider the case where $v_i^* \neq 0$. In such a case, $\mu_i = 0$ from the KKT condition (4.3.4), and hence $\left[p + V^T\lambda - K\lambda\right]_i = 0$ from the KKT condition (4.3.1). It then follows from (4.3.8) and (4.3.9) that

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

Therefore, the complementarity condition (iv) holds.

From the previous discussion, we conclude that x^* and (u^*, v^*) satisfy all sufficient conditions for optimality, and hence x^* is an optimal solution of $(\mathrm{P}_{\mathrm{GO}^*})$. □

Observe that the assumption that $g_i(\bar{x}_{I_i}) = 1$ 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 $A_i^T u^* - H_i^T v^* - c_{I_i} \neq 0$ for all i. Under these conditions, the solution set $X_i(u^*, v^*)$ is included in the boundary of the feasible set of (P^i), and thus $g_i(\bar{x}_{I_i}) = 1$ for all $\bar{x}_{I_i} \in X_i(u^*, v^*)$.

We now show an example of the primal recovery by using the Ridge-type optimization problem considered in Example 4.2.

Example 4.3. *Consider the following problem:*

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

where we assume $b \neq 0$. The problem is the positively homogeneous dual of the reformulated Ridge-type optimization problem $(\mathrm{P'}_b)$. Note that the second constraint slightly changes comparing to (D_b) because $\|-u\|_2 = \|u\|_2$. The Lagrangian function of (D_c) is

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

where $\lambda_1, \lambda_2 \in \mathbb{R}$ are the Lagrangian multipliers. Then, the KKT conditions are *obtained as follows:*

$$-b + \lambda_1 \frac{AA^T u}{\|A^T u\|_2} + \lambda_2 \frac{u}{\|u\|_2} = 0, \tag{4.3.10}$$
$$\|A^T u\|_2 - \sigma \leq 0, \tag{4.3.11}$$
$$\|u\|_2 - 1 \leq 0, \tag{4.3.12}$$
$$\lambda_1(\|A^T u\|_2 - \sigma) = 0, \tag{4.3.13}$$
$$\lambda_2(\|u\|_2 - 1) = 0, \tag{4.3.14}$$
$$\lambda_1, \lambda_2 \geq 0. \tag{4.3.15}$$

For simplicity, we denote that the point (u^*, v^*) *satisfying the above conditions as* (u, v)*. Note that we implicitly assume* $u \neq 0$ *and* $A^T u \neq 0$ *in equation* (4.3.10)*. We obtain* $u \neq 0$ *from the assumption that* $b \neq 0$*. If* $\mathrm{rank}(A) = m$*, we have* $A^T u \neq 0$ *for any* $u \neq 0$*. However, if* $\mathrm{rank}(A) < m$*, there exists* $\hat{u} \neq 0$ *such that* $A^T \hat{u} = 0$*. For the latter case, we have to describe equation* (4.3.10) *by using the subgradient. Then, equation* (4.3.10) *becomes more complicated and we might fail to recover a primal solution.*

Then, problems (P^i)*,* $i = 1, 2$ *are described as*

$$\begin{array}{ll} \sup\limits_{x} & u^T A x \\ \text{s.t.} & \|x\|_2 \leq 1, \end{array} \tag{P^1}$$

and

$$\begin{array}{ll} \sup\limits_{y} & -u^T y \\ \text{s.t.} & \|y\|_2 \leq 1, \end{array} \tag{P^2}$$

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*

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

We also observe that

$$g_1(\bar{x}) = \|\bar{x}\|_2 = \left\| \frac{A^T u}{\|A^T u\|_2} \right\|_2 = 1, \quad g_2(\bar{y}) = \|\bar{y}\|_2 = \left\| \frac{-u}{\|u\|_2} \right\|_2 = 1.$$

Then, we consider four cases with respect to the Lagrangian multipliers λ_1 *and* λ_2. *If* $\lambda_1 = \lambda_2 = 0$*, then we have* $b = 0$ *from equation* (4.3.10) *which is a contradiction to the assumption here. If* $\lambda_1 \neq 0$ *and* $\lambda_2 = 0$*, then we have* $\|A^T u\|_2 = \sigma$ *from* (4.3.13) *and*

$$\lambda_1 A A^T u = \sigma b$$

from (4.3.10). *By multiplying* u^T*, we obtain*

$$\lambda_1 u^T A A^T u = \lambda_1 \|A^T u\|_2^2 = \lambda_1 \sigma^2 = \sigma u^T b,$$

which results in

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

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

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

If $\lambda_1 = 0$ *and* $\lambda_2 \neq 0$*, then we have* $\|u\|_2 = 1$ *from* (4.3.14) *and* $\lambda_2 u = b$ *from* (4.3.10). *By multiplying* u^T*, we obtain* $\lambda_2 = u^T b$. *Therefore, the optimal solution is* $(x^*, y^*) = (0, -u^T b u)$.

If $\lambda_1 \neq 0$ *and* $\lambda_2 \neq 0$*, then we have* $\|A^T u\|_2 = \sigma$ *and* $\|u\|_2 = 1$ *from* (4.3.14) *and* (4.3.15). *Then from* (4.3.10)*, we obtain*

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

and by multiplying u^T *we have*

$$-u^T b + \lambda_1 \sigma + \lambda_2 = 0.$$

Thus, we obtain

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

Note that, for the original problem (P_b)*, if* σ *is sufficiently large, then an optimal solution* x^* *becomes zero. If* σ *is small, then* $x^* \neq 0$. *The property can be described by using the dual problem* (D_c) *as follows. If* σ *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* σ *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* $x^* \neq 0$.

4.4 Duality for general optimization problems

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

$$\begin{array}{ll} \min & c^T x + d^T F(x) \\ \text{s.t.} & Ax = b, \\ & Hx + KF(x) \le p, \end{array} \tag{P_F}$$

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

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* [6,8] 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 a linear function and a gauge one. The reformulation enables us to apply the results in the previous section for a general convex optimization problem.

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

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

$$f(x) = f(x) - f(z) - \eta^T(x - z) + f(z) + \eta^T(x - z). \tag{4.4.1}$$

Note that $f(x) - f(z) - \eta^T(x - z)$ is convex and nonnegative with respect to x, because f satisfies the subgradient inequality [81, p. 214]: $f(x) \ge f(z) + \eta^T(x - z)$. 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 using the so-called *perspective* of a nonnegative convex function. Recall that for any nonnegative convex function $h \colon \mathbb{R}^n \to \mathbb{R}_+ \cup \{\infty\}$, its perspective $h^p \colon \mathbb{R}^{n+1} \to \mathbb{R} \cup$

$\{\infty\}$ 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

$$h^\pi(x,\zeta) := \begin{cases} \zeta h(\zeta^{-1}x) & \text{if } \zeta > 0, \\ h^\infty(x) & \text{if } \zeta = 0, \\ \infty & \text{if } \zeta < 0, \end{cases} \tag{4.4.2}$$

where h^∞ is the *recession function* of h [81, p. 66]. Note that if h is a proper convex function, then h^π is a positively homogeneous proper convex function [81, Theorem 8.5]. In addition, $h^\pi(0,0) = 0$ by definition, and hence h^π 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 4.4. *Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be defined as $f(x) := \frac{1}{2}x^T Ax$, 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:*

$$\begin{aligned} f^\pi(x,\zeta) &= \begin{cases} \dfrac{1}{2\zeta}x^T Ax & \text{if } \zeta > 0, \\ \delta_{\{0\}}(x) & \text{if } \zeta = 0, \\ \infty & \text{otherwise}, \end{cases} \\ f^\natural(y,\eta) &= \begin{cases} -\dfrac{1}{2\eta}y^T A^{-1}y & \text{if } \eta < 0, \\ \delta_{\{0\}}(y) & \text{if } \eta = 0, \\ \infty & \text{otherwise}. \end{cases} \end{aligned}$$

Proof. Proof. The perspective f^π directly follows from definition (4.4.2). Note that A is positive definite, hence $f^\infty = \delta_{\{0\}}$ [81, p. 68]. The polar of f^π is defined by

$$f^\natural(y,\eta) = \sup_{x,\zeta} \left\{ x^T y + \zeta\eta \mid f^\pi(x,\zeta) \le 1 \right\}. \tag{4.4.3}$$

We first consider the case where $\eta > 0$. Since $f^\pi(0,\zeta) = 0$ for $\zeta \ge 0$, $f^\natural(y,\eta) \ge \zeta\eta$ for $\zeta \ge 0$. 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 $\zeta(t) := \frac{1}{2}x(t)^T Ax(t)$. Since A is positive definite, $\zeta(t) = \frac{1}{2}x(t)^T Ax(t) > 0$. Then $f^\pi(x(t), \zeta(t)) = 1$ for all t. Consequently $f^\natural(y, 0) \geq x(t)^T y + \zeta(t) \cdot 0 = t\|y\|^2$, 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 $f^\natural(y, \eta) = 0$. Note that $f^\pi(x, \zeta) \leq 1$ implies $\zeta \geq 0$, and $f^\pi(0, 0) \leq 1$. Therefore, when $y = 0$ and $\eta < 0$ we have $f^\natural(y, \eta) = 0$.

Finally, we investigate the case where $y \neq 0$ and $\eta < 0$. We now set

$$x^* = -\frac{1}{\eta}A^{-1}y, \quad \zeta^* = \frac{1}{2\eta^2}y^T A^{-1}y, \tag{4.4.4}$$

and

$$\lambda^* = -\eta\frac{2(\zeta^*)^2}{(x^*)^T Ax^*}. \tag{4.4.5}$$

Since $x^* \neq 0$ and $\zeta^* > 0$, λ^* is well-defined and $\lambda^* > 0$. Moreover, we have from (4.4.4)

$$\frac{1}{2}(x^*)^T Ax^* = \frac{1}{2\eta^2}y^T A^{-1}y = \zeta^*.$$

It then follows from (4.4.5) that

$$\eta = -\frac{\lambda^*}{\zeta^*}. \tag{4.4.6}$$

Then, equations (4.4.4) and (4.4.6) give

$$-y + \frac{\lambda^*}{\zeta^*}Ax^* = 0. \tag{4.4.7}$$

We note that the following conditions also hold:

$$\frac{1}{2\zeta^*}x^{*T}Ax^* - 1 \leq 0, \lambda^* \geq 0, \tag{4.4.8}$$

$$\lambda^*\left(\frac{1}{2\zeta^*}x^{*T}Ax^* - 1\right) = 0. \tag{4.4.9}$$

Note also that $f^\pi(x, \zeta) = \frac{1}{2\zeta}x^T Ax$. Conditions (4.4.5), (4.4.7), (4.4.8) and (4.4.9) are the KKT conditions of the convex optimization problem in the right-hand of (4.4.3). Therefore, the point (x^*, ζ^*) is its global optimal solution. Consequently, we obtain

$$f^\natural(y, \eta) = (x^*)^T y + \zeta^*\eta = -\frac{1}{2\eta}y^T A^{-1}y,$$

which completes the proof. $\square$

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

4.4.2 Perspective dual problems

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

$$\begin{array}{ll} \min & \hat{c}^T z + d^T F^\pi(z) \\ \text{s.t.} & \hat{A}z = \hat{b}, \\ & \hat{H}z + KF^\pi(z) \le p, \end{array} \tag{P$_\pi$}$$

where $F^\pi\colon \mathbb{R}^{n+m} \to \mathbb{R}^m$ is the perspective of F, $z = (x_{I_1},\zeta_1,\dots,x_{I_m},\zeta_m)^T \in \mathbb{R}^{n+m}$, $\hat{c} = (c_{I_1},0,\dots,c_{I_m},0)^T \in \mathbb{R}^{n+m}$, $\hat{b} = (b,1,\dots,1)^T \in \mathbb{R}^{2m}$, $\hat{H} = [H_{I_1},0,\dots,H_{I_m},0] \in \mathbb{R}^{\ell\times(n+m)}$ 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:

$$\begin{array}{ll} \max & b^T u - p^T v + e_m^T w \\ \text{s.t.} & F^\natural \begin{pmatrix} (A_{I_1})^T u - (H_{I_1})^T v - c_{I_1} \\ w_1 \\ \vdots \\ (A_{I_m})^T u - (H_{I_m})^T v - c_{I_m} \\ w_m \end{pmatrix} - K^T v \le d, \\ & v \ge 0. \end{array} \tag{D$_\pi$}$$

We call problem (D$_\pi$) as the perspective dual of (P$_F$).

Example 4.5. *We now consider the following convex quadratic optimization problem as an example of* (P_F)*.*

$$\begin{aligned} \min \quad & \frac{1}{2}x^T A_0 x + b_0^T x \\ \text{s.t.} \quad & \frac{1}{2}x^T A_1 x + b_1^T x \le c_1, \end{aligned} \tag{$\mathrm{P_{QP}}$}$$

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

$$\begin{aligned} \min \quad & \frac{1}{2}x^T A_0 x + b_0^T x \\ \text{s.t.} \quad & \frac{1}{2}y^T A_1 y + b_1^T y \le c_1, \\ & x - y = 0. \end{aligned}$$

Let $z := (x, y)^T$ and $F(z) := (f_0(x), f_1(y))^T = (\frac{1}{2}x^T A_0 x, \frac{1}{2}y^T A_1 y)^T$. Then the problem is described as follows:

$$\begin{aligned} \min \quad & (b_0^T, 0)z + (1, 0)F(z) \\ \text{s.t.} \quad & (0, b_1^T)z + (0, 1)F(z) \le c_1, \\ & (I, -I)z = 0. \end{aligned}$$

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_π) *equivalent to* $(\mathrm{P_{QP}})$ *is written as*

$$\begin{aligned} \min \quad & (b_0^T, 0, 0, 0)w + (1, 0)F^\pi(w) \\ \text{s.t.} \quad & (0, 0, b_1^T, 0)w + (0, 1)F^\pi(w) \le c_1, \\ & \begin{bmatrix} I & 0 & -I & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} w = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}. \end{aligned} \tag{$\mathrm{P_\pi^{QP}}$}$$

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

$$\begin{aligned} \max \quad & (0, 1, 1)u - c_1 v \\ \text{s.t.} \quad & F^\natural\left(\begin{bmatrix} I & 0 & 0 \\ 0 & 1 & 0 \\ -I & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} u - \begin{bmatrix} 0 \\ 0 \\ b_1 \\ 0 \end{bmatrix} v - \begin{bmatrix} b_0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \right) - \begin{bmatrix} 0 \\ 1 \end{bmatrix} v \le \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \\ & v \ge 0. \end{aligned}$$

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

$$\begin{array}{ll} \max & u_2 + u_3 - c_1 v \\ \text{s.t.} & f_0^\natural(u_1 - b_0, u_2) \le 1, \\ & f_1^\natural(-u_1 - b_1 v, u_3) \le v, \\ & v \ge 0. \end{array} \tag{$\mathrm{D}_\pi^{\mathrm{QP}}$}$$

Recall that the functions $f_0^\natural$ and $f_1^\natural$ are described as in Example 4.4. It is easy to see that (u, v) with $u_2 > 0$ or $u_3 > 0$ is not feasible for $(\mathrm{D}_\pi^{\mathrm{QP}})$.

The following lemma indicates the first two constraints in $(\mathrm{D}_\pi^{\mathrm{QP}})$ can be represented as semidefinite constraints.

Lemma 4.5. *Let $f(x) = \frac{1}{2}x^T Ax$, where A is an $n \times n$ symmetric and positive definite matrix. Then,*

$$f^\natural(y, \eta) \le \gamma, \quad \gamma \ge 0 \tag{4.4.10}$$

if and only if

$$\begin{bmatrix} A\gamma & y \\ y^T & -2\eta \end{bmatrix} \succeq 0. \tag{4.4.11}$$

Proof. Proof. First we suppose that (4.4.10) holds. The inequality $f^\natural(y, \eta) \le \gamma$ implies $(y, \eta) = (0, 0)$ or $\eta < 0$ from the definition of $f^\natural$ in Example 4.4. If $(y, \eta) = (0, 0)$, then (4.4.11) holds since A is positive definite and $\gamma \ge 0$. If $\eta < 0$, then (4.4.10) can be written as

$$-\frac{1}{2\eta} y^T A^{-1} y \le \gamma, \quad \gamma \ge 0.$$

If $\gamma = 0$, then we have $y = 0$, and hence (4.4.11) holds. When $\gamma > 0$, we obtain

$$-2\eta - \frac{1}{\gamma} y^T A^{-1} y \ge 0, \quad \gamma \ge 0, \tag{4.4.12}$$

which results in (4.4.11) by using the Schur complement [11].

Next, we assume that (4.4.11) holds. Then, we have $\eta \le 0$ and $\gamma \ge 0$. If $\eta = 0$, then $y = 0$ from (4.4.11). It then follows from Example 4.4 that $f^\natural(y, \eta) = \delta_{\{0\}}(0) = 0 \le \gamma$, and hence (4.4.10) holds. If $\gamma = 0$, then $y = 0$ once again. Then we obtain $f^\natural(y, \eta) = 0 = \gamma$, which indicates (4.4.10) holds. If $\eta < 0$ and $\gamma > 0$, then the Schur complement of (4.4.11) gives (4.4.12), which results in (4.4.10). □

From Lemma 4.5, the perspective dual problem ($\mathrm{D}_{\pi}^{\mathrm{QP}}$) of problem ($\mathrm{P}_{\mathrm{QP}}$) is equivalent to the following semidefinite programming [94]:

$$\begin{array}{ll} \max & u_2 + u_3 - c_1 v \\ \text{s.t.} & \begin{bmatrix} A_0 & u_1 - b_0 \\ (u_1 - b_0)^T & -2u_2 \end{bmatrix} \succeq 0, \\ & \begin{bmatrix} A_1 v & u_1 + b_1 v \\ (u_1 + b_1 v)^T & -2u_3 \end{bmatrix} \succeq 0. \end{array}$$

In this section, we discussed the reformulation of a 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 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.

4.5 Conclusion

In this chapter, we provided the details of optimization problems with both gauge functions and linear ones in their objective and constraint functions. Using the positively homogeneous framework given in [108], we proved that weak and strong duality results hold for such gauge problems. We also discussed both necessary and sufficient optimality conditions associated with 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.

Chapter 5

Branch-and-bound method for absolute value optimization problems

5.1 Introduction

In recent years, the absolute value equations (AVEs) [16, 41, 60, 62, 64, 77, 82, 83, 112] has attracted a growing attention. The absolute value optimization (AVO) is an extension of AVEs, which contains the absolute values of variables in its objective function and constraints. Formally, the AVO is stated as follows:

$$\begin{array}{ll} \min & c^T x + d^T |x| \\ \text{s.t.} & Ax + B|x| = b, \\ & Hx + K|x| \geq p, \end{array} \tag{$\mathrm{P_{AVO}}$}$$

where $c, d \in \mathbb{R}^n,\ b \in \mathbb{R}^m,\ p \in \mathbb{R}^\ell,\ A, B \in \mathbb{R}^{m \times n},\ H, K \in \mathbb{R}^{\ell \times n}$, and $|x|$ denotes the vector $|x| = (|x_1|, |x_2|, \ldots, |x_n|)^T \in \mathbb{R}^n$. Although this problem is a nonconvex optimization problem in general, Mangasarian [61] showed an interesting weak duality result and a sufficient optimality condition for the problem. In addition, the AVEs that appears in the constraints of the AVO is shown to be equivalent to a linear complementarity problem [64, 77]. This result indicates that the AVO is equivalent to a linear program with linear complementarity constraints, which is a special case of the mathematical program with equilibrium constraints (MPEC) [57]. MPEC has many applications in various areas such as economics, engineering, and transportation. However, MPEC is in general difficult to deal with, since its feasible region is necessarily nonconvex and even disconnected. The study on AVO is in its infancy and, to the authors' knowledge, there have been no works except for the above-mentioned duality results of Mangasarian [61].

In this paper, we first propose an algorithm for the AVO, which is based on the branch-and-bound method. In the branching procedure, we generate two subproblems by restricting the sign of a component of the variable x in ($\mathrm{P}_{\mathrm{AVO}}$) to be nonnegative or nonpositive. In the bounding procedure, we utilize the duality results in AVO to obtain a lower bound for each subproblem. Furthermore, to examine the effectiveness of the proposed algorithm, we apply it to solve facility location problems (FLPs). By using the ℓ_1 norm as a distance function, an FLP can naturally be formulated as an AVO. In particular, we can use the AVO formulation to deal with a nonconvex region in which facilities are located. We stress that such a problem is considerably difficult to solve compared with the conventional FLPs that assume the convexity of the region.

5.2 Absolute value optimization problems and their duality

The dual problem of AVO ($\mathrm{P}_{\mathrm{AVO}}$) is defined as follows [61]:

$$\begin{array}{ll} \max & b^T u + p^T v \\ \text{s.t.} & |A^T u + H^T v - c| + B^T u + K^T v \leq d, \\ & v \geq 0. \end{array} \tag{$\mathrm{D}_{\mathrm{AVO}}$}$$

Note that the inequality constraint can be represented as

$$\begin{aligned} & |A^T u + H^T v - c| \leq d - B^T u - K^T v \\ \iff \quad & -d + B^T u + K^T v \leq A^T u + H^T v - c \leq d - B^T u - K^T v \\ \iff \quad & \begin{cases} (-A + B)^T u + (-H + K)^T v \leq d - c, \\ (A + B)^T u + (H + K)^T v \leq d + c. \end{cases} \end{aligned}$$

Therefore, the dual problem ($\mathrm{D}_{\mathrm{AVO}}$) can be rewritten as follows:

$$\begin{array}{ll} \max & b^T u + p^T v \\ \text{s.t.} & (-A + B)^T u + (-H + K)^T v \leq d - c, \\ & (A + B)^T u + (H + K)^T v \leq d + c, \\ & v \geq 0. \end{array}$$

Notice that the dual problem ($\mathrm{D_{AVO}}$) is always a convex optimization problem, or more precisely, a linear program, although the primal problem ($\mathrm{P_{AVO}}$) is not convex in general. Moreover, a weak duality theorem and a sufficient optimality condition for AVO are shown in [61], which will be useful in our algorithm.

Theorem 5.1. [10] *If x and (u, v) are feasible solutions of* ($\mathrm{P_{AVO}}$) *and* ($\mathrm{D_{AVO}}$)*, respectively, then the following inequality holds:*

$$c^T x + d^T |x| \geq b^T u + p^T v.$$

This theorem says that we can get a lower bound of the optimal value of ($\mathrm{P_{AVO}}$) by solving the dual problem ($\mathrm{D_{AVO}}$). The next theorem gives a sufficient optimality condition for ($\mathrm{P_{AVO}}$).

Theorem 5.2. [10] *Let $\bar{x}$ be feasible in the primal AVO* ($\mathrm{P_{AVO}}$) *and $(\bar{u}, \bar{v})$ be feasible in the dual AVO* ($\mathrm{D_{AVO}}$) *with equal primal and dual objective values, that is,*

$$c^T \bar{x} + d^T |\bar{x}| = b^T \bar{u} + p^T \bar{v}.$$

Then $\bar{x}$ and $(\bar{u}, \bar{v})$ are optimal solutions of ($\mathrm{P_{AVO}}$) *and* ($\mathrm{D_{AVO}}$)*, respectively.*

5.3 Branch-and-bound method for absolute value optimization problem

In this section, we propose a branch-and-bound method for AVO. The branch-and-bound method is one of fundamental global optimization methods for nonconvex optimization problems [38] and combinatorial optimization problems [44]. The method consists of branching and bounding procedures. In the branching procedure, we divide the feasible region of the original problem into some subregions to generate subproblems. On the other hand, in the bounding procedure, we check if a current subproblem can be discarded or not, by implementing some fathoming tests. We now give the details of the branching and bounding procedures used in the proposed branch-and-bound method for solving AVOs.

A subproblem is constructed from ($\mathrm{P_{AVO}}$) by restricting some variables to be either nonpositive or nonnegative:

$$
\begin{array}{llll}
\mathrm{P}(\mathcal{I},\mathcal{J}) & \min & c^T x + d^T|x| & \\
 & \text{s.t.} & Ax + B|x| = b, & \\
 & & Hx + K|x| \geq p, & \\
 & & x_i \geq 0 \quad (i \in \mathcal{I}), & \\
 & & x_i \leq 0 \quad (i \in \mathcal{J}), &
\end{array}
$$

where $\mathcal{I}$ and $\mathcal{J}$ are subsets of $\{1, 2, \ldots, n\}$ such that $\mathcal{I} \cap \mathcal{J} = \emptyset$. Note that $(\mathrm{P}_{\mathrm{AVO}}) = \mathrm{P}(\emptyset, \emptyset)$. The branching procedure can conveniently be explained by using the enumeration tree, where each node corresponds to a subproblem. An example of the enumeration tree with $n = 2$ is shown in Fig. 5.1. At each node of the tree, branching means that we choose a variable x_i and restrict it to be nonnegative or nonpositive in the corresponding subproblem. The deepest nodes in the tree correspond to 2^n linear programs, which contain no absolute values of the variables. The branch-and-bound method maintains the set of subproblems that can be selected to apply a branching procedure. Such subproblems are said to be active, and the set of the current active subproblems is denoted by $\mathcal{A}$. For example, if we generate two subproblems $\mathrm{P}(\{1\}, \emptyset)$ and $\mathrm{P}(\emptyset, \{1\})$ at the root node $\mathrm{P}(\emptyset, \emptyset)$ in the enumeration tree of Fig. 5.1, then we have $\mathcal{A} = \{\mathrm{P}(\{1\}, \emptyset), \mathrm{P}(\emptyset, \{1\})\}$.

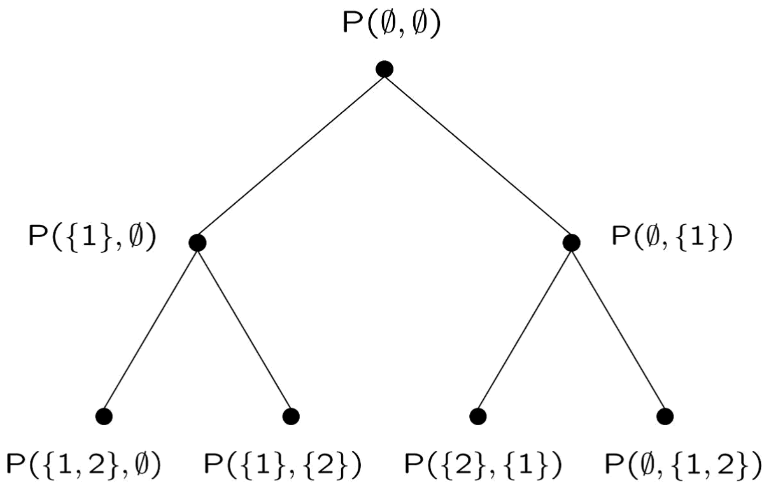

Fig. 5.1: Enumeration tree ($n = 2$)

In the bounding procedure, we consider the dual problem of $\mathrm{P}(\mathcal{I}, \mathcal{J})$ in order to get a lower bound of $\mathrm{P}(\mathcal{I}, \mathcal{J})$. For convenience, let $h_i := \sigma_i e_i \in \mathbb{R}^n$ for each

$i \in \mathcal{I} \cup \mathcal{J}$, where $e_i \in \mathbb{R}^n$ is the ith column of the $n \times n$ identity matrix, and $\sigma_i = 1$ if $i \in \mathcal{I}$ and $\sigma_i = -1$ if $i \in \mathcal{J}$. Then, the nonnegativity and nonpositivity constraints on variables x_i in $\mathrm{P}(\mathcal{I}, \mathcal{J})$ are represented as

$$h_i^T x \geq 0 \quad (i \in \mathcal{I} \cup \mathcal{J}).$$

Therefore, we can rewrite $\mathrm{P}(\mathcal{I}, \mathcal{J})$ as follows:

$$\begin{array}{lll} \mathrm{P}(\mathcal{I}, \mathcal{J}) & \min & c^T x + d^T |x| \\ & \text{s.t.} & Ax + B|x| = b, \\ & & \tilde{H}x + \tilde{K}|x| \geq \tilde{p}, \end{array}$$

where $\tilde{H} \in \mathbb{R}^{(\ell+|\mathcal{I}|+|\mathcal{J}|)\times n}$, $\tilde{K} \in \mathbb{R}^{(\ell+|\mathcal{I}|+|\mathcal{J}|)\times n}$, $\tilde{p} \in \mathbb{R}^{(\ell+|\mathcal{I}|+|\mathcal{J}|)}$ are defined by

$$\tilde{H} := \begin{bmatrix} H \\ \hline \vdots \\ h_i^T \\ \vdots \end{bmatrix}, \quad \tilde{K} := \begin{bmatrix} K \\ \hline \vdots \\ 0 \\ \vdots \end{bmatrix}, \quad \tilde{p} := \begin{bmatrix} p \\ \hline \vdots \\ 0 \\ \vdots \end{bmatrix}.$$

Moreover, the dual problem of $\mathrm{P}(\mathcal{I}, \mathcal{J})$ is written as

$$\begin{array}{lll} \mathrm{D}(\mathcal{I}, \mathcal{J}) & \max & b^T u + \tilde{p}^T v \\ & \text{s.t.} & |A^T u + \tilde{H}^T v - c| + B^T u + \tilde{K}^T v \leq d, \\ & & v \geq 0, \end{array}$$

which can further be rewritten as a linear program. Based on the result of solving the dual problem, the subproblem $\mathrm{P}(\mathcal{I}, \mathcal{J})$ can be fathomed if one of the following conditions holds:

(i) $\mathrm{D}(\mathcal{I}, \mathcal{J})$ is unbounded.

(ii) The optimal value of $\mathrm{D}(\mathcal{I}, \mathcal{J})$ is greater than the objective value of the incumbent solution, i.e., the best feasible solution of ($\mathrm{P}_{\mathrm{AVO}}$) found so far.

(iii) There is no duality gap between $\mathrm{P}(\mathcal{I}, \mathcal{J})$ and $\mathrm{D}(\mathcal{I}, \mathcal{J})$.

We now give more details about the bounding operations based on the above three conditions.

If the dual problem $\mathrm{D}(\mathcal{I}, \mathcal{J})$ is unbounded, then the primal problem $\mathrm{P}(\mathcal{I}, \mathcal{J})$ is infeasible from the weak duality theorem. In this case, any subproblem generated from the current subproblem by restricting the sign of some of its variables cannot be feasible. Hence we can discard the current subproblem $\mathrm{P}(\mathcal{I},\mathcal{J})$.

If the optimal value of $\mathrm{D}(\mathcal{I},\mathcal{J})$, which is a lower bound of the optimal value of $\mathrm{P}(\mathcal{I},\mathcal{J})$ by Theorem 5.1, is greater than the objective value of the incumbent solution, we have no chance to obtain an optimal solution of ($\mathrm{P}_{\mathrm{AVO}}$) by generating subproblems from $\mathrm{P}(\mathcal{I},\mathcal{J})$ further. Thus, we can discard the current subproblem.

If we find out that there is no duality gap between $\mathrm{P}(\mathcal{I},\mathcal{J})$ and $\mathrm{D}(\mathcal{I},\mathcal{J})$, then this means the subproblem $\mathrm{P}(\mathcal{I},\mathcal{J})$ is just solved. For this reason, we need not generate new subproblems from $\mathrm{P}(\mathcal{I},\mathcal{J})$ further, and we can discard the current subproblem. Moreover, if the optimal solution of $\mathrm{P}(\mathcal{I},\mathcal{J})$ is better than the incumbent solution, then we replace the incumbent solution by the optimal solution of $\mathrm{P}(\mathcal{I},\mathcal{J})$. We may check if there is no duality gap between $\mathrm{P}(\mathcal{I},\mathcal{J})$ and $\mathrm{D}(\mathcal{I},\mathcal{J})$ by solving the following system of absolute value equations and inequalities:

$$\begin{aligned} c^T x + d^T|x| &= f_D^*, \\ Ax + B|x| &= b, \\ \tilde{H}x + \tilde{K}|x| &\geq \tilde{p}, \end{aligned} \tag{S1}$$

where f_D^* is the optimal objective value of the dual problem $\mathrm{D}(\mathcal{I},\mathcal{J})$. If the system (S1) has a solution, then $\mathrm{P}(\mathcal{I},\mathcal{J})$ and $\mathrm{D}(\mathcal{I},\mathcal{J})$ have no duality gap. Moreover, by Theorem 5.2, it is an optimal solution of $\mathrm{P}(\mathcal{I},\mathcal{J})$.

We now formally state the algorithm.

Branch-and-Bound Algorithm for AVO:

- **Step 0.** Let $\mathcal{I} := \emptyset$ and $\mathcal{J} := \emptyset$. Find a feasible solution of problem $(\mathrm{P_{AVO}})=$ $\mathrm{P}(\emptyset, \emptyset)$. Let it be the incumbent solution and let f^* be the objective value at the incumbent solution. Set $\mathcal{A} := \{\mathrm{P}(\emptyset, \emptyset)\}$.

- **Step 1.** Choose a subproblem $\mathrm{P}(\mathcal{I}, \mathcal{J})$ from the set $\mathcal{A}$.

  - **Step 1-a.** If the dual problem $\mathrm{D}(\mathcal{I}, \mathcal{J})$ of $\mathrm{P}(\mathcal{I}, \mathcal{J})$ is infeasible, then go to Step 2. If $\mathrm{D}(\mathcal{I}, \mathcal{J})$ is unbounded, then fathom $\mathrm{P}(\mathcal{I}, \mathcal{J})$. Set $\mathcal{A} := \mathcal{A} \setminus \{\mathrm{P}(\mathcal{I}, \mathcal{J})\}$ and go to Step 3.
  - **Step 1-b.** Let f_D^* be the optimal objective value of the dual problem $\mathrm{D}(\mathcal{I}, \mathcal{J})$. If it satisfies $f_D^* > f^*$, then fathom $\mathrm{P}(\mathcal{I}, \mathcal{J})$. Set $\mathcal{A} := \mathcal{A} \setminus \{\mathrm{P}(\mathcal{I}, \mathcal{J})\}$ and go to Step 3.
  - **Step 1-c.** Solve the system (S1) of absolute value equations and inequalities. If we fail to get a solution of (S1), then go to Step 2. If we get a solution of (S1) and, in addition, the objective value at the solution, denoted $f_{(\mathcal{I},\mathcal{J})}$, satisfies $f_{(\mathcal{I},\mathcal{J})} \geq f^*$, then $\mathrm{P}(\mathcal{I}, \mathcal{J})$ is fathomed immediately. If $f_{(\mathcal{I},\mathcal{J})} < f^*$, then set $f^* := f_{(\mathcal{I},\mathcal{J})}$, update the incumbent solution, and fathom $\mathrm{P}(\mathcal{I}, \mathcal{J})$. Set $\mathcal{A} := \mathcal{A} \setminus \{\mathrm{P}(\mathcal{I}, \mathcal{J})\}$ and go to Step 3.

- **Step 2.** Choose a variable x_i such that $i \notin \mathcal{I} \cup \mathcal{J}$ as the branching variable, and generate two subproblems $\mathrm{P}(\mathcal{I} \cup \{i\}, \mathcal{J})$ and $\mathrm{P}(\mathcal{I}, \mathcal{J} \cup \{i\})$ from $\mathrm{P}(\mathcal{I}, \mathcal{J})$. Set $\mathcal{A} := \mathcal{A} \cup \{\mathrm{P}(\mathcal{I} \cup \{i\}, \mathcal{J}), \mathrm{P}(\mathcal{I}, \mathcal{J} \cup \{i\})\} \setminus \{\mathrm{P}(\mathcal{I}, \mathcal{J})\}$, and return to Step 1.

- **Step 3.** If $\mathcal{A} = \emptyset$, then terminate. The incumbent solution is an optimal solution of the original problem $(\mathrm{P_{AVO}})$. Otherwise, return to Step 1.

To get a feasible solution of $(\mathrm{P_{AVO}})$ in Step 0 and to solve (S1) in Step 1-c, we can use the successive linearization algorithm (SLA) for the system of absolute value equations and inequalities. This algorithm was first proposed by Mangasarian [64] to solve AVEs. We extend the algorithm so as to deal with a system that also contains absolute value inequalities (AVIs).

Here we describe the SLA for the system

$$\begin{cases} Ax + B|x| = b, \\ Hx + K|x| \geq p, \end{cases} \tag{S2}$$

which represents the constraints of ($\mathrm{P}_{\mathrm{AVO}}$). The algorithm can similarly be applied to solve the system (S1).

First, we give a result that relates the AVE-AVI system (S2) to the following concave minimization problem constructed from (S2):

$$\begin{array}{cr} \min\limits_{(x,t,s_1,s_2)\in\mathbb{R}^{n+n+m+\ell}} & \epsilon(-e^T|x| + e^T t) + e^T s_1 + e^T s_2 \\ \text{s.t.} & -s_1 \leq Ax + Bt - b \leq s_1, \\ & -Hx - Kt + p \leq s_2, \\ & 0 \leq s_2, \\ & -t \leq x \leq t, \end{array} \tag{5.3.1}$$

where $\epsilon > 0$ and e is the vector of ones.

Proposition 5.1. *If* (S2) *is solvable, then there exists some* $\bar{\epsilon} > 0$ *such that, for any* $\epsilon \in (0, \bar{\epsilon}]$, *any solution* $(\bar{x}, \bar{t}, \bar{s_1}, \bar{s_2})$ *of* (5.3.1) *satisfies*

$$\begin{aligned} |\bar{x}| &= \bar{t}, \\ A\bar{x} + B|\bar{x}| &= b, \\ H\bar{x} + K|\bar{x}| &\geq p. \end{aligned}$$

Proof. The proof is analogous to that of Proposition 3 in [64]. □

From this result, a solution of the AVE-AVI system (S2) may be obtained by solving the concave minimization problem (5.3.1) with a sufficiently small $\epsilon > 0$. We now give the SLA for the AVE-AVI system (S2), which is an extension of the SLA for AVEs [64]. Let $z = (x, t, s_1, s_2)^T$. Denote the feasible region of problem (5.3.1) by Z and its objective function by $\theta(z)$.

SLA for AVE-AVI:

- **Step 0.** Choose a starting point $z^0 \in Z$. Set $k := 0$.
- **Step 1.** Given z^k, find z^{k+1} such that

$$z^{k+1} \in \text{arg vertex} \min_{z \in Z} \; (\xi^k)^T(z - z^k),$$

where ξ^k is a subgradient of $\theta(z)$ at z^k, and arg vertex $\min_{z\in Z}(\xi^k)^T(z - z^k)$ is the set of vertex solutions of the linear program: $\min_{z\in Z}(\xi^k)^T(z - z^k)$.

- **Step 2.** If $(\xi^k)^T(z^{k+1} - z^k) = 0$, then stop. Otherwise, return to Step 1 with k increased by one.

In our numerical experiments, we compute a subgradient ξ^k of $\theta(z)$ at z^k as follows:

$$\xi^k = \begin{pmatrix} -\epsilon g^k \\ \epsilon e \\ e \\ e \end{pmatrix} \in \mathbb{R}^{n+n+m+\ell} \quad \text{with} \quad g_i^k = \begin{cases} 1 & (x_i^k > 0) \\ 0 & (x_i^k = 0) \\ -1 & (x_i^k < 0) \end{cases}, \quad i = 1, \cdots, n.$$

As is well-known, a concave minimization problem has at least one optimal solution at a vertex of the feasible region, provided a solution exists. Taking this fact into account, the SLA tries to find an optimal solution of (5.3.1) by solving a sequence of linear programs formed by linearizing the objective function of problem (5.3.1). The sequence generated by the SLA finitely converges to a point that satisfies a necessary optimality condition for the concave minimization problem [58, 64]. Notice that the solution obtained by this algorithm is not guaranteed to be a global optimal solution of (5.3.1). Nevertheless, we can easily check if the computed solution actually satisfies (S2) by direct substitution.

We now show the way to generate subproblems in Step 2 of the branch-and-bound algorithm. Recall that Step 2 is visited after either of the following two cases occurs.

Case 1. In Step 1-a, $\mathrm{D}(\mathcal{I}, \mathcal{J})$ is infeasible.

Case 2. In Step 1-c, (S2) cannot be solved.

If Case 1 occurs, then we generate two subproblems by choosing any variable x_i such that $i \notin \mathcal{I} \cup \mathcal{J}$ as the branching variable. In Case 2, we fail to have a solution of (S2), but a local optimal solution of problem (5.3.1) is obtained. In this case, we choose as the branching variable a variable x_i ($i \notin \mathcal{I} \cup \mathcal{J}$) such that $|x_i| \geq |x_j|$ for all $j \notin \mathcal{I} \cup \mathcal{J}$ at the obtained local optimal solution of (5.3.1).

In Step 1, a certain rule should be used to choose an active subproblem $\mathrm{P}(\mathcal{I}, \mathcal{J}) \in \mathcal{A}$. In the numerical experiments reported in the next section, we use the depth-first search, which generally chooses an active subproblem corresponding to the farthest node from the root node in the enumeration tree. In particular, when we return to Step 1 after generating two subproblems, we choose one of these subproblems. In this case, the choice depends on the above-mentioned two cases. If we generate two subproblems in Step 2 after Case 1 occurs, then we choose any of the two subproblems. In Case 2, as we mentioned above, we have a local optimal solution of (5.3.1) at hand. In this case, if the branching variable x_i in the local optimal solution takes a positive value, then we choose subproblem $\mathrm{P}(\mathcal{I} \cup \{i\}, \mathcal{J})$. Otherwise, we choose $\mathrm{P}(\mathcal{I}, \mathcal{J} \cup \{i\})$.

5.4 Numerical experiments

In this section, we consider facility location problems (FLPs) as an application of AVO, and show some numerical results with the proposed branch-and-bound algorithm applied to some examples of FLPs. All computations were carried out on an Intel® Core™ 2 Duo 3GHz machine with a MATLAB code. The CPLEX was used to solve linear programs in the SLA.

FLP is the problem of finding optimal locations of facilities in a given area, and it can be formulated as mathematical programs of different natures depending on the type of constraints and optimization criteria [26]. Generally speaking, there are two kinds of facilities from the residents' standpoint. The first category is a *desirable* facility such as schools, libraries, and fire stations. Such a facility should be located as closely as possible to the residents. The other category is an *undesirable* facility, which includes incineration plants, electric power stations, chemical factories, and so on. These facilities should be located far from the residential area. From the viewpoint of geography, there are three kinds of areas in which facilities can be located, i.e., continuous spaces, discrete spaces, and networks. Furthermore, the distance between two facilities or between a facility and a residential district can be measured by using various norms such as the Euclidean, the ℓ_1, and the ℓ_∞ norms.

In our numerical experiments, we consider two types of FLPs on a continuous space with the ℓ_1 distance, which can be reformulated as AVOs. Note that the ℓ_1 distance between two points x and y can be represented as $e^T|x-y|$.

5.5 Minimax Location Problem

A minimax multifacility location problem can be formulated as follows [26]:

$$\begin{array}{ll} \min & \max \ \{ \max\limits_{i\in I, j\in J} \alpha_{ij} e^T |x^i - P^j|, \max\limits_{i,k\in I, i\neq k} \beta_{ik} e^T |x^i - x^k| \ \} \\ \text{s.t.} & \qquad\qquad x^i \in X \quad (i \in I), \end{array} \tag{5.5.1}$$

where $x^i \in \mathbb{R}^2$ $(i \in I)$ and $P^j \in \mathbb{R}^2$ $(j \in J)$ denote the locations of the new and the existing facilities, respectively, I and J are finite index sets, α_{ij} and β_{ik} are positive weighting factors, and $X \subset \mathbb{R}^2$ is the region in which the facilities are located.

The problem is to minimize the maximum weighted distance between new and existing facilities, and between new facilities themselves. If each existing facility is regarded as a residential district, then this problem represents a mathematical model of locating desirable facilities, such as schools and fire stations, in a city. This kind of problems has been well-studied for the past decades. In particular, using the ℓ_1 norm as the distance function, Konforty and Tamir [52] studied the minimax single facility location problem with a forbidden region around each existing facility.

Problem (5.5.1) can be rewritten as the following problem by introducing a new variable $z \in \mathbb{R}$:

$$\begin{array}{ll} \min\limits_{x,z} & z \\ \text{s.t.} & z \geq \alpha_{ij} e^T |x^i - P^j| \quad (i \in I, j \in J), \\ & z \geq \beta_{ik} e^T |x^i - x^k| \quad (i, k \in I, i \neq k), \\ & x^i \in X \quad (i \in I). \end{array} \tag{5.5.2}$$

If X is a convex polyhedron, problem (5.5.2) is easy to solve because it reduces to a linear program. Here, we deal with the more general case where X is a nonconvex region.

We now give the details of the problem that we solve in numerical experiments. We define the region X as the set of points $x = (x_1, x_2)^T \in \mathbb{R}^2$ that satisfy the following inequalities:

$$\begin{aligned} &|x_1| + |x_2| + 0.2x_1 + 0.4x_2 \leq 10, \\ &|x_1| + 2|x_2 + 12| - 0.5x_1 \geq 12, \\ &|x_1 + 2| + 1.5|x_2 + 1| - 0.3x_1 - 0.5x_2 \geq 5. \end{aligned} \tag{5.5.3}$$

The region X is nonconvex, as shown in Fig. 5.2. Notice that since the region X is described by (5.5.3), problem (5.5.2) is an instance of AVO.

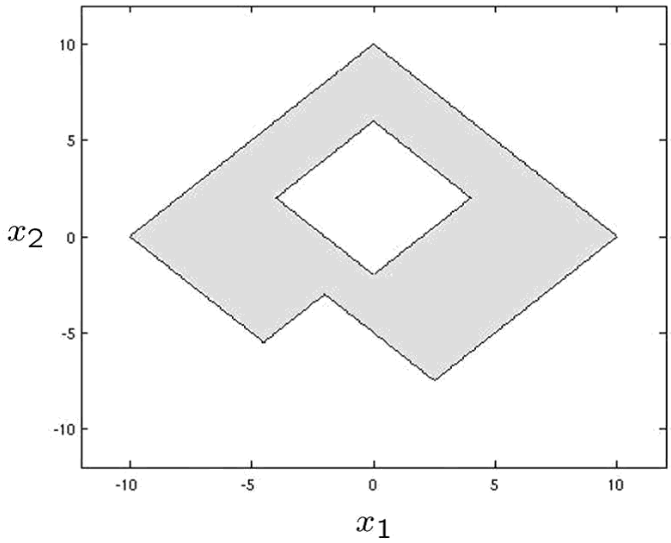

Fig. 5.2: Region X where the facilities are located.

In the numerical experiments, we let $I = \{1, 2\}$, $J = \{1, 2, 3\}$ and set the locations of the existing facilities as $P^1 = (-7, -5)$, $P^2 = (-2, 5)$, $P^3 = (7, -1)$. Moreover, we choose the positive weight $\beta_{12} = 1.0$, and use two data sets for the weights α_{ij} given as follows:

$$(\alpha_{11},\ \alpha_{12},\ \alpha_{13},\ \alpha_{21},\ \alpha_{22},\ \alpha_{23}) = (0.5,\ 1.0,\ 0.7,\ 0.5,\ 0.7,\ 1.0) \tag{5.5.4}$$

and

$$(\alpha_{11},\ \alpha_{12},\ \alpha_{13},\ \alpha_{21},\ \alpha_{22},\ \alpha_{23}) = (0.7,\ 1.0,\ 0.5,\ 0.5,\ 1.0,\ 0.7). \tag{5.5.5}$$

The problems with α_{ij}'s given by (5.5.4) and (5.5.5) are called Minimax-1 and Minimax-2, respectively. The branch-and-bound method was able to find solutions of Minimax-1 and Minimax-2, which are given by $x^1 = (0, 1.5)$, $x^2 = (0.98, -1.9)$ and $x^1 = (-3.42, 1.05)$, $x^2 = (-0.07, 1.55)$, respectively. The solutions are depicted in

Fig. 5.3 and Fig. 5.4. For each problem, the CPU time, the numbers of subproblems fathomed in Step 1-a, Step 1-b, Step 1-c, and the number of nodes explored are summarized in Table 1.

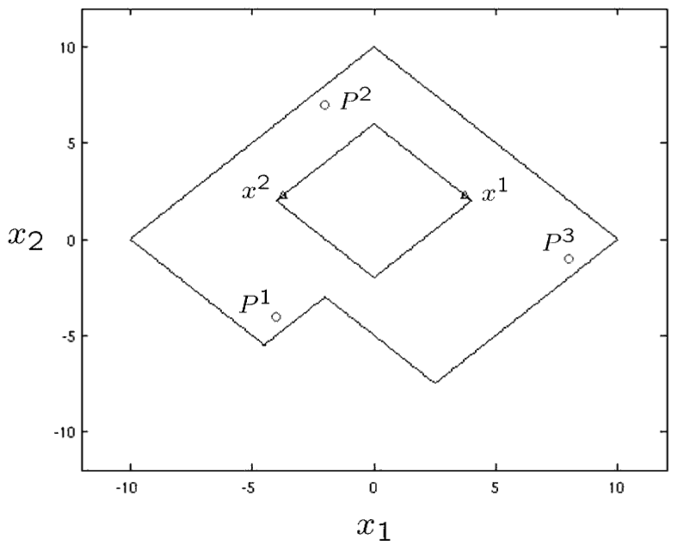

Fig. 5.3: Solution of Minimax-1

Table 5.1: Results for minimax location problems

	Time (sec)	Step 1-a	Step 1-b	Step 1-c	No. of nodes explored
Minimax-1	0.55	16	63	4	164
Minimax-2	2.7	230	96	30	710

5.6 Maximin Location Problem

A maximin multifacility location problem is generally formulated as follows [26]:

$$\begin{array}{lll} \max & \min \{ \min\limits_{i\in I, j\in J} \alpha_{ij} e^T |x^i - P^j|, \min\limits_{i,k\in I, i\neq k} \beta_{ik} e^T |x^i - x^k| \ \} \\ \text{s.t.} & x^i \in X \quad (i \in I), \end{array} \tag{5.6.1}$$

where x^i, P^j, α_{ij}, β_{ij} and X represent the same stuffs as in problem (5.5.1). Unlike the minimax location problem, this problem maximizes the minimum weighted distances between new and existing facilities, and between new facilities themselves.

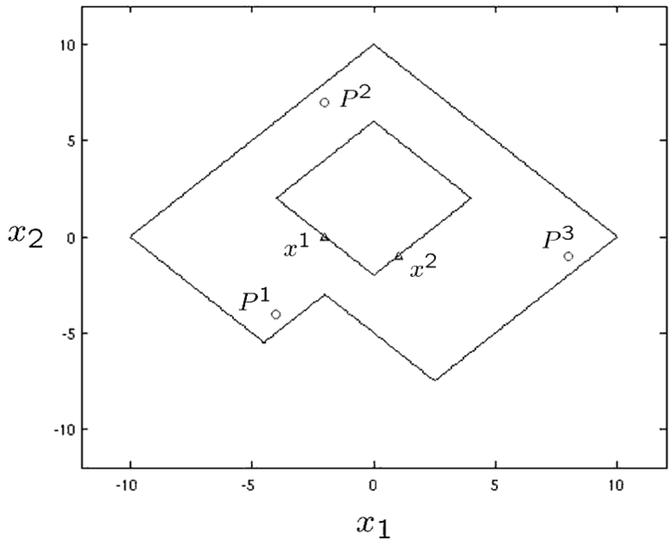

Fig. 5.4: Solution of Minimax-2

The maximin location problem will be useful in locating competing facilities such as convenience stores and gas stations.

Sayin [86] and Nadirler and Karasakal [70] reformulated a single facility maximin location problem on a convex region with the ℓ_1 distance as a mixed integer program. Tamir [89] proposed an algorithm for two-facility maximin location problems on a convex region with the ℓ_1 distance. In these approaches, the region for locating facilities is assumed to be convex. Here we solve multi-facility location problems on a nonconvex region.

Problem (5.6.1) can be rewritten as the following problem by introducing a new variable $z \in \mathbb{R}$ [89, 97]:

$$\begin{array}{ll} \max\limits_{x,z} & z \\ \text{s.t.} & z \leq \alpha_{ij} e^T |x^i - P^j| \quad (i \in I, j \in J), \\ & z \leq \beta_{ik} e^T |x^i - x^k| \quad (i, k \in I, i \neq k), \\ & x^i \in X \quad (i \in I). \end{array} \tag{5.6.2}$$

Notice that, unlike the inequality constraints in (5.5.2), those in this problem are nonconvex.

In the numerical experiments, we let the index sets of the new and the existing facilities be $I = \{1, 2\}$ and $J = \{1, 2, 3\}$, respectively. In addition, we set all the positive weights α_{ij} and β_{12} to be 1. The region X is the nonconvex region described by (5.5.3). Moreover, the locations of the existing facilities are given in the following two data sets:

$$P^1 = (-10, -1),\ P^2 = (-5, 2),\ P^3 = (2, 4) \tag{5.6.3}$$

and

$$P^1 = (-9, 1),\ P^2 = (-1, -3),\ P^3 = (6, -2). \tag{5.6.4}$$

The problems with the data sets (5.6.3) and (5.6.4) are called Maximin-1 and Maximin-2, respectively. By using the proposed branch-and-bound method, we obtained a solution $x^1 = (-2.42, -7.81)$, $x^2 = (6.27, -4.12)$ for Maximin-1 and a solution $x^1 = (-5.12, -9.84)$, $x^2 = (0, 6.96)$ for Maximin-2. Those solutions are shown in Fig. 5.5 and Fig. 5.6. For each problem, the CPU time, the numbers of subproblems fathomed in Step 1-a, Step 1-b, Step 1-c, and the number of nodes explored are shown in Table 2.

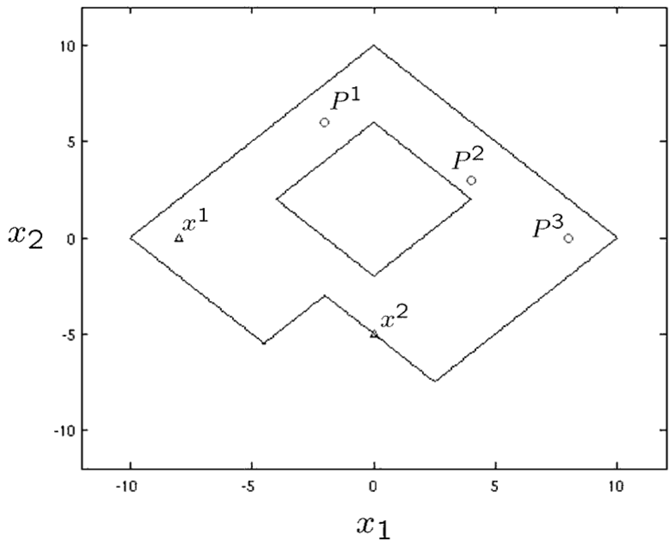

Fig. 5.5: Solution of Maximin-1

Table 5.2: Results for maximin location problems

	Time (sec)	Step 1-a	Step 1-b	Step 1-c	No. of nodes explored
Maximin-1	7.9	753	492	13	2514
Maximin-2	14.1	1478	686	18	4362

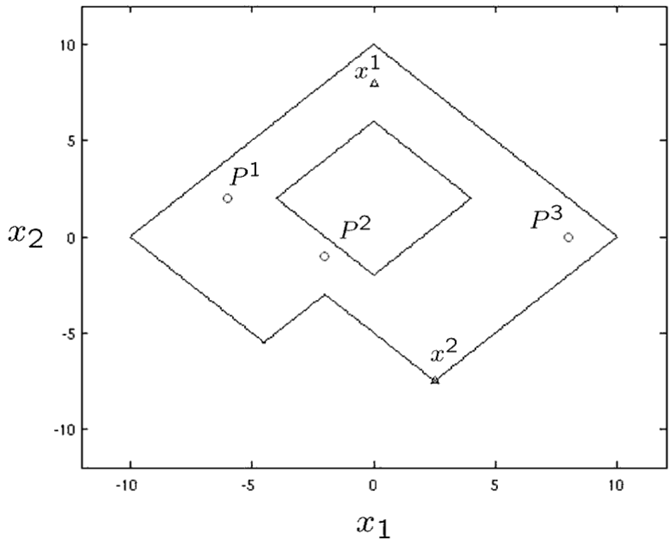

Fig. 5.6: Solution of Maximin-2

For each of the above examples Minimax-1, 2, and Maximin-1, 2, the AVO in the form of ($\mathrm{P_{AVO}}$) has 27 variables and 22 constraints. From the results shown in this section, we observe that a global optimal solution of each problem was found by exploring only a small number of nodes compared with the number of all possible nodes ($2^{27} - 1$) in the enumeration tree. Although problems (5.5.2) and (5.6.2) have the same number of variables and constraints, there is a significant difference in the CPU time between these two types of problems, as shown in Table 1 and Table 2. The reason for this phenomenon may be explained as follows. The minimax location problem (5.5.1) has a convex objective function, although the feasible region is nonconvex. On the other hand, the objective function of the maximin location problem (5.6.1) is neither convex nor concave. Such a problem is considered to be much more difficult to deal with in practice.

5.7 Conclusion

In this chapter, we developed an algorithm for the absolute value optimization, which is based on the branch-and-bound method. We have also carried out numerical experiments for nonconvex multi-facility location problems with the ℓ_1-norm, which can naturally be reformulated as absolute value optimization problems. The numerical results demonstrate the validity of the proposed algorithm.

Chapter 6

Conclusion

In this thesis, we first introduced an optimization problem that generalizes the so-called absolute value optimization problem and investigated the theoretical properties of the generalized problem. We also proposed a more general gauge optimization problem than the previous works and proved some important theoretical properties. Then, we develop a global optimization algorithm for the absolute value optimization by using a branch-and-bound method. The results in this thesis are summarized as follows.

- In Chapter 4, we proposed optimization problems with positively homogeneous functions, which we call positively homogeneous optimization problems. We also introduced their dual problems and showed the weak duality theorem between these problems. Moreover, we gave sufficient conditions for the equivalency between the proposed dual and the Lagrangian dual problems. Finally, we presented some examples of positively homogeneous problems to show their value in real-world applications. One natural future work will be to propose methods that obtain approximate solutions of positively homogeneous optimization problems. We believe the theoretical results described here are essential for that.

- Chapter 5 provided the details of optimization problems with both gauge functions and linear functions in their objective and constraint functions. Using the positively homogeneous framework given in [108], we proved that weak and strong duality results hold for such gauge problems. We also discussed both necessary and sufficient optimality conditions associated with 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. An important future work is to develop an efficient algorithm by using the theoretical results described here.

- In Chapter 6, we developed an algorithm for the absolute value optimization, which is based on the branch-and-bound method. We have also carried out numerical experiments for nonconvex multi-facility location problems with the ℓ_1-norm, which can naturally be reformulated as absolute value optimization problems. The numerical results demonstrate the validity of the proposed algorithm.

As we summarized above, we have contributed to the studies on absolute value optimization problems and their generalization. In the following, we list some future works for some unsolved issues.

- In Chapter 4, it is necessary to develop an algorithm for the PHO problem and investigating the effectiveness of the algorithm. In particular, optimization problems with the ℓ_p-norm, $0 < p < 1$ could be one appropriate application of the PHO. In sparse optimization, the ℓ_p-norm, $0 < p < 1$ gives sparser solutions than the ℓ_1-norm. However, it is usually avoided to use such ℓ_p-norm because of its nonconvexity. If an algorithm for the PHO is developed by using the theoretical results in this thesis, then sparse optimization problems with the ℓ_p-norm, $0 < p < 1$ can be solved efficiently.
- In Chapter 5, it is important to investigate the perspective transformation in more concrete form. We provided the perspective of a convex quadratic function and its polar as an example in this thesis. To make a list of the perspectives of popular convex functions in a concrete form should be necessary to apply the perspective framework for wider applications.
- In Chapter 6, to develop another global/local optimization algorithm could be the future works. Not only the duality results, considering the more reasonable problem than the general AVO by limiting the sign pattern of variables, smoothing of the absolute value function, and some heuristic methods might be useful.

Bibliography

[1] U. Akinc and B. M. Khumawala. An efficient branch and bound algorithm for the capacitated warehouse location problem. *Management Science*, 23(6):585–594, 1977.

[2] F. A. Al-Khayyal. Jointly constrained bilinear programs and related problems: An overview. *Computers & Mathematics with Applications*, 19(11):53–62, 1990.

[3] F. A. Al-Khayyal and J. E. Falk. Jointly constrained biconvex programming. *Mathematics of Operations Research*, 8(2):273–286, 1983.

[4] F. Alizadeh and D. Goldfarb. Second-order cone programming. *Mathematical Programming*, 95(1):3–51, 2003.

[5] H.-C. An, M. Singh, and O. Svensson. Lp-based algorithms for capacitated facility location. *SIAM Journal on Computing*, 46(1):272–306, 2017.

[6] A. Y. Aravkin, J. V. Burke, D. Drusvyatskiy, M. P. Friedlander, and K. MacPhee. Foundations of gauge and perspective duality. *SIAM Journal on Optimization*, 28(3):2406–2434, 2018.

[7] A. Y. Aravkin, J. V. Burke, D. Drusvyatskiy, M. P. Friedlander, and S. Roy. Level-set methods for convex optimization. *Mathematical Programming*, 174(1):359–390, 2019.

[8] A. Y. Aravkin, J. V. Burke, and M. P. Friedlander. Variational properties of value functions. *SIAM Journal on Optimization*, 23(3):1689–1717, 2013.

[9] J. Y. Bello Cruz, O. P. Ferreira, and L. F. Prudente. On the global convergence of the inexact semi-smooth newton method for absolute value equation. *Computational Optimization and Applications*, 65(1):93–108, 2016.

[10] K. P. Bennett, J. Hu, X. Ji, G. Kunapuli, and J.-S. Pang. Model selection via bilevel optimization. In *The 2006 IEEE International Joint Conference on Neural Network Proceedings*, pages 1922–1929. IEEE, 2006.

[11] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[12] P. S. Bradley and O. L. Mangasarian. Feature selection via concave minimization and support vector machines. In *ICML*, volume 98, pages 82–90. Citeseer, 1998.

[13] J. Brimberg and H. Juel. A minisum model with forbidden regions for locating a semi-desirable facility in the plane. *Location Science*, 6(1-4):109–120, 1998.

[14] J. Brimberg and G.O. Wesolowsky. The rectilinear distance minisum problem with minimum distance constraints. *Location Science*, 3(3):203–215, 1995.

[15] A. V. Cabot, R. L. Francis, and M. A. Stary. A network flow solution to a rectilinear distance facility location problem. *AIIE Transactions*, 2(2):132–141, 1970.

[16] L. Caccetta, B. Qu, and G. Zhou. A globally and quadratically convergent method for absolute value equations. *Computational Optimization and Applications*, 48(1):45–58, 2011.

[17] R. Chartrand. Exact reconstruction of sparse signals via nonconvex minimization. *IEEE Signal Processing Letters*, 14(10):707–710, 2007.

[18] R. Chartrand and W. Yin. Iteratively reweighted algorithms for compressive sensing. In *Acoustics, speech and signal processing, 2008. ICASSP 2008. IEEE international conference on*, pages 3869–3872. IEEE, 2008.

[19] B. Dasarathy and L. J. White. A maxmin location problem. *Operations Research*, 28(6):1385–1401, 1980.

[20] A. Dax. An efficient algorithm for solving the rectilinear multifacility location problem. *IMA journal of numerical analysis*, 6(3):343–355, 1986.

[21] D. L. Donoho and Y. Tsaig. Fast solution of ℓ_1-norm minimization problems when the solution may be sparse. *IEEE Transactions on Information Theory*, 54(11):4789–4812, 2008.

[22] Z. Drezner and G. O. Wesolowsky. Location of multiple obnoxious facilities. *Transportation Science*, 19(3):193–202, 1985.

[23] H. A. Eiselt and V. Marianov. *Foundations of location analysis*, volume 155. Springer Science & Business Media, 2011.

[24] Y. C. Eldar and M. Mishali. Robust recovery of signals from a structured union of subspaces. *IEEE Transactions on Information Theory*, 55(11):5302–5316, 2009.

[25] E. Erkut and S. Neuman. Analytical models for locating undesirable facilities. *European Journal of Operational Research*, 40(3):275–291, 1989.

[26] R. Z. Farahani and M. Hekmatfar. *Facility location: concepts, models, algorithms and case studies*. Springer, 2009.

[27] C. A. Floudas and C. E. Gounaris. A review of recent advances in global optimization. *Journal of Global Optimization*, 45(1):3, 2009.

[28] C. A. Floudas and V. Visweswaran. A global optimization algorithm (gop) for certain classes of nonconvex nlps—i. theory. *Computers & chemical engineering*, 14(12):1397–1417, 1990.

[29] C. A. Floudas and V. Visweswaran. Primal-relaxed dual global optimization approach. *Journal of Optimization Theory and Applications*, 78(2):187–225, 1993.

[30] R. M. Freund. Dual gauge programs, with applications to quadratic programming and the minimum-norm problem. *Mathematical Programming*, 38(1):47–67, 1987.

[31] M. P. Friedlander and I. Macêdo. Low-rank spectral optimization via gauge duality. *SIAM Journal on Scientific Computing*, 38(3):A1616–A1638, 2016.

[32] M. P. Friedlander, I. Macêdo, and T. K. Pong. Gauge optimization and duality. *SIAM Journal on Optimization*, 24(4):1999–2022, 2014.

[33] M. Fukushima and J.-S. Pang. Some feasibility issues in mathematical programs with equilibrium constraints. *SIAM Journal on Optimization*, 8(3):673–681, 1998.

[34] A. Göpfert, T. Riedrich, and C. Tammer. *Angewandte Funktionalanalysis: Motivationen und Methoden für Mathematiker und Wirtschaftswissenschaftler.* Springer, 2009.

[35] H. W. Hamacher and S. Nickel. Classification of location models. *Location Science*, 6(1-4):229–242, 1998.

[36] A. E. Hoerl and R. W. Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970.

[37] J. H. Holland. Genetic algorithms. *Scientific american*, 267(1):66–73, 1992.

[38] R. Horst, P. M. Pardalos, and N. Van Thoai. *Introduction to global optimization.* Springer Science & Business Media, 2000.

[39] J. Hu, J. E. Mitchell, J.-S. Pang, K. P. Bennett, and Gautam Kunapuli. On the global solution of linear programs with linear complementarity constraints. *SIAM Journal on Optimization*, 19(1):445–471, 2008.

[40] S. Hu and Z. Huang. A note on absolute value equations. *Optimization Letters*, 4(3):417–424, 2010.

[41] S. Hu, Z. Huang, and Q. Zhang. A generalized Newton method for absolute value equations associated with second order cones. *Journal of Computational and Applied Mathematics*, 235(5):1490–1501, 2011.

[42] T. Ibaraki. Complementarity programming. *Operations Research*, 19(6):1523–1528, 1971.

[43] T. Ibaraki. The use of cuts in complementary programming. *Operations Research*, 21(1):353–359, 1973.

[44] T. Ibaraki. Enumerative approaches to combinatorial optimization. *Annals of Operations Research*, 10-11, 1987.

[45] H. Idrissi, O. Lefebvre, and C. Michelot. Duality for constrained multifacility location problems with mixed norms and applications. *Annals of Operations Research*, 18(1):71–92, 1989.

[46] F. Jara-Moroni, J.-S. Pang, and A. Wächter. A study of the difference-of-convex approach for solving linear programs with complementarity constraints. *Mathematical Programming*, 169(1):221–254, 2018.

[47] R. G. Jeroslow. Cutting-planes for complementarity constraints. *SIAM Journal on Control and Optimization*, 16(1):56–62, 1978.

[48] W. Jiang, F. Nie, and H. Huang. Robust dictionary learning with capped l1-norm. In *Twenty-Fourth International Joint Conference on Artificial Intelligence*, 2015.

[49] H. Juel and R. F. Love. An efficient computational procedure for solving the multi-facility rectilinear facilities location problem. *Journal of the Operational Research Society*, 27(3):697–703, 1976.

[50] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *science*, 220(4598):671–680, 1983.

[51] U. S. Kirmaci, M. K. Bakula, M. E. Özdemir, and J. E. Pecaric. On some inequalities for $p-$norms. *Journal of Inequalities in Pure & Applied Mathematics*, 9(1):1–8, 2008.

[52] Y. Konforty and A. Tamir. The single facility location problem with minimum distance constraints. *Location Science*, 5(3):147–163, 1997.

[53] A. A. Kuehn and M. J. Hamburger. A heuristic program for locating warehouses. *Management science*, 9(4):643–666, 1963.

[54] G. Kunapuli, K Bennett, J. Hu, and J.-S. Pang. Bilevel model selection for support vector machines. In *CRM Proceedings and Lecture Notes*, volume 45, pages 129–158, 2008.

[55] G. Kunapuli, K. P. Bennett, J. Hu, and J.-S. Pang. Classification model selection via bilevel programming. *Optimization Methods & Software*, 23(4):475–489, 2008.

[56] Q. Lu, X. Li, Y. Dong, and D. Tao. Subspace clustering by capped l_1 norm. In *Chinese Conference on Pattern Recognition*, pages 663–674. Springer, 2016.

[57] Z.-Q. Luo, J.-S. Pang, and D. Ralph. *Mathematical Programs with Equilibrium Constraints*. Cambridge University Press, 1996.

[58] O. L. Mangasarian. Solution of general linear complementarity problems via nondifferentiable concave minimization. *Acta Mathematica Vietnamica*, 22(1):199–205, 1997.

[59] O. L. Mangasarian. Exact 1-norm support vector machines via unconstrained convex differentiable minimization. *Journal of Machine Learning Research*, 7(7), 2006.

[60] O. L. Mangasarian. Absolute value equation solution via concave minimization. *Optimization Letters*, 1(1):3–8, 2007.

[61] O. L. Mangasarian. Absolute value programming. *Computational Optimization and Applications*, 36(1):43–53, 2007.

[62] O. L. Mangasarian. A generalized Newton method for absolute value equations. *Optimization Letters*, 3(1):101–108, 2009.

[63] O. L. Mangasarian. A hybrid algorithm for solving the absolute value equation. *Optimization Letters*, 9(7):1469–1474, 2015.

[64] O. L. Mangasarian and R. R. Meyer. Absolute value equations. *Linear Algebra and Its Applications*, 419(2):359–367, 2006.

[65] L. Meier, S. van de Geer, and P. Bühlmann. The group Lasso for logistic regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(1):53–71, 2008.

[66] X. Miao, J. Yang, and S. Hu. A generalized Newton method for absolute value equations associated with circular cones. *Applied Mathematics and Computation*, 269:155–168, 2015.

[67] J. G. Morris. A linear programming solution to the generalized rectangular distance weber problem. *Naval Research Logistics Quarterly*, 22(1):155–164, 1975.

[68] D. R. Morrison, S. H. Jacobson, J. J. Sauppe, and E. C. Sewell. Branch-and-bound algorithms: A survey of recent advances in searching, branching, and pruning. *Discrete Optimization*, 19:79–102, 2016.

[69] N. Mourad and J. P. Reilly. Minimizing nonconvex functions for sparse vector reconstruction. *IEEE Transactions on Signal Processing*, 58(7):3485–3496, 2010.

[70] D. Nadirler and E. Karasakal. Mixed integer programming-based solution procedure for single-facility location with maximin of rectilinear distance. *Journal of the Operational Research Society*, 59(4):563–570, 2008.

[71] R. M. Nauss. An improved algorithm for the capacitated facility location problem. *Journal of the Operational Research Society*, 29(12):1195–1201, 1978.

[72] F. Nie, Z. Huo, and H. Huang. Joint capped norms minimization for robust matrix recovery. In *The 26th International Joint Conference on Artificial Intelligence (IJCAI 2017)*, 2017.

[73] F. Nie, J. Yuan, and H. Huang. Optimal mean robust principal component analysis. In *International conference on machine learning*, pages 1062–1070. PMLR, 2014.

[74] W. Ogryczak. Inequality measures and equitable approaches to location problems. *European Journal of Operational Research*, 122(2):374–391, 2000.

[75] G. Pataki. Strong duality in conic linear programming: facial reduction and extended duals. In *Computational and analytical mathematics*, pages 613–634. Springer, 2013.

[76] A. A. B. Pritsker and P. M. Ghare. Locating new facilities with respect to existing facilities. *AIIE Transactions*, 2(4):290–297, 1970.

[77] O. Prokopyev. On equivalent reformulations for absolute value equations. *Computational Optimization and Applications*, 44(3):363–372, 2009.

[78] B. Qian, H. Zhou, F. Lyu, J. Li, T. Ma, and F. Hou. Toward collision-free and efficient coordination for automated vehicles at unsignalized intersection. *IEEE Internet of Things Journal*, 6(6):10408–10420, 2019.

[79] M. R. Rao. On the direct search approach to the rectilinear facilities location problem. *AIIE Transactions*, 5(3):256–264, 1973.

[80] C. S. ReVelle and H. A. Eiselt. Location analysis: A synthesis and survey. *European journal of operational research*, 165(1):1–19, 2005.

[81] R. T. Rockafellar. *Convex analysis.* Princeton University Press, 1972.

[82] J. Rohn. A theorem of the alternatives for the equation $ax + b|x| = b$. *Linear and Multilinear Algebra*, 52(6):421–426, 2004.

[83] J. Rohn. An algorithm for solving the absolute value equation. *Electronic Journal of Linear Algebra*, 18(5):589–599, 2009.

[84] J. Rohn. On unique solvability of the absolute value equation. *Optimization Letters*, 3(4):603–606, 2009.

[85] J. Rohn, V. Hooshyarbakhsh, and R. Farhadsefat. An iterative method for solving absolute value equations and sufficient conditions for unique solvability. *Optimization Letters*, 8(1):35–44, 2014.

[86] S Sayin. A mixed integer programming formulation for the 1-maximin problem. *Journal of the operational research society*, 51(3):371–375, 2000.

[87] A. Shapiro. On duality theory of conic linear problems. In *Semi-infinite programming*, pages 135–165. Springer, 2001.

[88] M. Stojnic, F. Parvaresh, and B. Hassibi. On the reconstruction of block-sparse signals with an optimal number of measurements. *IEEE Transactions on Signal Processing*, 57(8):3075–3085, 2009.

[89] A. Tamir. Locating two obnoxious facilities using the weighted maximin criterion. *Operations Research Letters*, 34(1):97–105, 2006.

[90] H. Tuy. Global minimization of a difference of two convex functions. In *Nonlinear Analysis and Optimization*, pages 150–182. Springer, 1987.

[91] H. Tuy, T. V. Thieu, and N. Q. Thai. A conical algorithm for globally minimizing a concave function over a closed convex set. *Mathematics of operations research*, 10(3):498–514, 1985.

[92] E. Van Den Berg and M. P. Friedlander. Probing the pareto frontier for basis pursuit solutions. *SIAM Journal on Scientific Computing*, 31(2):890–912, 2009.

[93] E. Van den Berg and M. P. Friedlander. Sparse optimization with least-squares constraints. *SIAM Journal on Optimization*, 21(4):1201–1229, 2011.

[94] L. Vandenberghe and S. Boyd. Semidefinite programming. *SIAM review*, 38(1):49–95, 1996.

[95] V. Visweswaran and C. A. Floudas. New properties and computational improvement of the gop algorithm for problems with quadratic objective functions and constraints. *Journal of Global Optimization*, 3(4):439–462, 1993.

[96] C. Wang, Q. Ye, P. Luo, N. Ye, and L. Fu. Robust capped l1-norm twin support vector machine. *Neural Networks*, 114:47–59, 2019.

[97] S. B. Welch, S. Salhi, and Z. Drezner. The multifacility maximin planar location problem with facility interaction. *IMA journal of management mathematics*, 17(4):397–412, 2006.

[98] G. O. Wesolowsky and R. F. Love. Location of facilities with rectangular distances among point and area destinations. *Naval Research Logistics Quarterly*, 18(1):83–90, 1971.

[99] G. O. Wesolowsky and R. F. Love. The optimal location of new facilities using rectangular distances. *Operations Research*, 19(1):124–130, 1971.

[100] G. W. Wolf. *Facility Location: Concepts, Models, Algorithms and Case Studies*. Taylor & Francis, 2011.

[101] L. A. Wolsey and G. L. Nemhauser. *Integer and combinatorial optimization*, volume 55. John Wiley & Sons, 1999.

[102] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *IEEE transactions on pattern analysis and machine intelligence*, 31(2):210–227, 2008.

[103] L.-Y. Wu, X.-S. Zhang, and J.-L. Zhang. Capacitated facility location problem with general setup cost. *Computers & Operations Research*, 33(5):1226–1241, 2006.

[104] M.-J. Wu, J.-X. Liu, Y.-L. Gao, X.-Z. Kong, and C.-M. Feng. Feature selection and clustering via robust graph-laplacian pca based on capped l 1-norm. In *2017 IEEE international conference on Bioinformatics and Biomedicine (BIBM)*, pages 1741–1745. IEEE, 2017.

[105] S.-L. Wu and P. Guo. On the unique solvability of the absolute value equation. *Journal of Optimization Theory and Applications*, 169(2):705–712, 2016.

[106] G. Xue and Y. Ye. An efficient algorithm for minimizing a sum of $p-$norms. *SIAM Journal on Optimization*, 10(2):551–579, 2000.

[107] S. Yamanaka and M. Fukushima. A branch-and-bound method for absolute value programs. *Optimization*, 63(2):305–319, 2014.

[108] S. Yamanaka and N. Yamashita. Duality of nonconvex optimization with positively homogeneous functions. *Computational Optimization and Applications*, 71(2):435–456, 2018.

[109] B. Yu, J. E. Mitchell, and J.-S. Pang. Solving linear programs with complementarity constraints using branch-and-cut. *Mathematical Programming Computation*, 11(2):267–310, 2019.

[110] G.-X. Yuan, C.-H. Ho, and C.-J. Lin. Recent advances of large-scale linear classification. *Proceedings of the IEEE*, 100(9):2584–2603, 2012.

[111] M. Yuan and Y. Lin. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1):49–67, 2006.

[112] C. Zhang and Q. J. Wei. Global and finite convergence of a generalized Newton method for absolute value equations. *Journal of Optimization Theory and Applications*, 143(2):391–403, 2009.

[113] J. Zhang, B. Chen, and Y. Ye. A multiexchange local search algorithm for the capacitated facility location problem. *Mathematics of Operations Research*, 30(2):389–403, 2005.

[114] T. Zhang. Multi-stage convex relaxation for learning with sparse regularization. In *Advances in neural information processing systems*, pages 1929–1936, 2008.

[115] T. Zhang. Analysis of multi-stage convex relaxation for sparse regularization. *Journal of Machine Learning Research*, 11(3), 2010.

[116] J. Zhu, S. Rosset, R. Tibshirani, and T. J. Hastie. 1-norm support vector machines. *Neural Information Processing Systems*, 16, 2003.

Copyright notice