On the Principle of Lagrange in Optimization Theory and Its Application in Transportation and Location Problems

J. B. G. Frenk and Sonya Javadi

Abstract In mathematical optimization, the Lagrangian approach is a general method to find an optimal solution of a finite (infinite) dimensional constrained continuous optimization problem. This method has been introduced by the Italian mathematician Joseph-Louis Lagrange in 1755 in a series of letters to Euler. This approach became known under the name *The Principle of Lagrange* and was also applied much later to integer optimization problems. The basic idea behind this method is to replace a constrained optimization problem by a sequence of easier solvable optimization problems having fewer constraints and penalizing the deletion of some of the original constraints by replacing the original objective function. To select the best penalization, the so-called Lagrangian dual function needs to be optimized and a possible algorithm to do so is given by the so-called subgradient method. This method is discussed in detail at the end of this chapter. The Lagrangian approach led to the introduction of dual optimization problems and penalization methods in nonlinear programming and recently to the development of interior point methods and the identification of polynomial solvable classes of continuous optimization problems. Also it had its impact on how to construct algorithms to generate approximate solutions of integer optimization problems. In this chapter, we discuss in the first part the main ideas behind this approach for any type of finite dimensional optimization problem. In the remaining parts of this chapter we focus in more detail on how this approach is used in continuous optimization problems and show its full impact on the so-called K-convex continuous optimization problems. Also we consider its application within linear integer programming problems and show how it is used to solve these type of problems. To illustrate its application to the well-known integer programming problems, we consider in the final section

J. B. G. Frenk

Faculty of Engineering and Natural Science, Sabancı University, Istanbul, Turkey
e-mail: frenk@sabanciuniv.edu

S. Javadi (✉)

Department of Industrial Engineering, Doğuş University, Istanbul, Turkey
e-mail: sjavadi@dogus.edu.tr

its application to some classical vehicle routing and location models. As such this chapter should be regarded as an introduction to duality theory and the Lagrangian approach for less mathematically oriented readers proving at the same time most of the results using the simplest possible proofs.

1 Introduction

In this chapter we give a unified overview of the Lagrangian relaxation technique used in finite dimensional optimization problems with either continuous and/or discrete decision variables and show as an example of how to apply these techniques by solving some classical transportation and location models. The main purpose and goal of this chapter is to give an introduction to Lagrangian duality accessible for less mathematically oriented readers and explain its main ideas. At the same time we try to be rigorous proving the most well-known results about strong duality for the class of K-convex optimization problems by means of the simplest possible proofs. All the proofs related to strong duality will be given with one exception. We will only mention the most basic separation result in convex analysis separating a nonempty convex set from a point outside this set. In general, contrary to continuous K-convex optimization problems, no strong duality holds for integer linear programming problems. By means of elementary proofs, we show how to apply the Lagrangian relaxation technique to these problems. Another more abstract approach to Lagrangian duality given by the perturbation function approach (Frenk and Kassay, 2005; Rockafellar, 1970) requires the knowledge of conjugate and biconjugate functions well-known in convex analysis.

To show strong duality results for K-convex optimization problems, one uses in this approach the Fenchel–Moreau theorem. This theorem states that under some mild topological conditions a convex function equals its biconjugate function. Although this approach is more general and has the advantage of shortening proofs, applying directly the Fenchel–Moreau theorem the perturbation function approach hides the intuitive ideas that are shown more clearly in the Lagrangian relaxation approach. For this reason, we have chosen this approach to explain the main results in duality theory. This approach is closely related to the minmax approach used in a two-person noncooperative game with the decision maker playing against nature. In this framework strong duality is equivalent to the existence of a Nash equilibrium point (Frenk and Kassay, 2006, 2008).

In the first section we introduce the Lagrangian relaxation approach and its application to finite dimensional optimization problems. In the second section we apply this approach to (integer) linear programming problems and discuss the main relations between such an integer linear program and some of the most well-known relaxations used within this field. In the third section we consider the subgradient method to find the optimal Lagrangian multiplier in a Lagrangian dual problem.

Finally in the last section we consider some classical location and transportation models and show how the Lagrangian relaxation technique results in heuristics by solving these problems.

Although no new results are discussed, this chapter introduces the reader to the main ideas of the Lagrangian relaxation technique originally introduced by Joseph-Louis Lagrange in his treatise "Leçons sur le calcul des functions" (nouvelle edition) published in 1806 (Lagrange, 1806) and revealed in a series of letters to Euler between 1754 and 1756. Joseph-Louis Lagrange applied his approach to infinite dimensional optimization problems, and his approach eventually lead to the development of optimal control theory. To prove the claims of Lagrange, the theory of convex sets and functions was later developed. Without any difficulty, one can also extend the finite dimensional approach discussed in this chapter to infinite dimensional optimization problems using the well-known Hahn–Banach theorem in functional analysis (Choquet, 1976) or the related minmax approach (Frenk Kas and Kassay, 2007).

This chapter also tries to be complete in discussing in a hopefully transparent way some of the mathematical technicalities found in the books on convex analysis. As such the rationale behind the Lagrangian relaxation approach can be seen as one of the main ideas in optimization theory. It relates a difficult constrained optimization problem to an easier solvable constrained optimization problem with fewer constraints and another objective function. In this newly selected objective function the deleted constraints of the original problem are penalized using a linear penalty function.

Developments occurring 150 years later within the field of nonlinear and linear programming like the primal–dual relations in linear programming and the well-known Karush–Kuhn–Tucker conditions in nonlinear programming are special instances of this Lagrangian relaxation approach. Also the developments of penalty-based optimization methods in nonlinear programming (Fiacco and McCormick, 1969) can be seen as an easy extension of the ideas behind this relaxation approach. These penalty methods eventually led to the development of interior point polynomial algorithms in convex programming (see page 2 of Nesterov and Nemirovski (2001)). By these observations, it seems clear that the ideas of Lagrange play a fundamental role in the development of the theory of optimization known nowadays.

2   On the Principle of Lagrange

Before discussing the principle of Lagrange for general optimization problems with continuous and/or discrete decision variables, we first introduce some definitions. If the sets $A, B \subseteq \mathbb{R}^n$ are nonempty sets, then for every $\alpha, \beta \in \mathbb{R}$ we define the so-called Minkowski sum:

$$\alpha A + \beta B := \{\alpha \mathbf{x} + \beta \mathbf{y} : \mathbf{x} \in A, \mathbf{y} \in B\}. \tag{1}$$

Also we denote by $\mathbb{Z}_+^n$ ($\mathbb{R}_+^n$) the set of nonnegative integer valued (real valued) n-dimensional vectors and by $\mathbb{Z}_-^n := -\mathbb{Z}_+^n$ and $\mathbb{R}_-^n := -\mathbb{R}_+^n$.

Definition 2.1 A nonempty set $L \subseteq \mathbb{R}^n$ is called a *linear* space if $\alpha L + \beta L \subseteq L$ for every $\alpha, \beta \in \mathbb{R}$. A nonempty set $M \subseteq \mathbb{R}^n$ is called *affine* if $\alpha M + (1 - \alpha)M \subseteq M$ for every $\alpha \in \mathbb{R}$. A nonempty set $C \subseteq \mathbb{R}^n$ is called *convex* if $\alpha C + (1 - \alpha)C \subseteq C$ for every $0 < \alpha < 1$. A nonempty set $K \subseteq \mathbb{R}^n$ is called a *cone* if $\alpha K \subseteq K$ for every $\alpha > 0$ and it is called a *pointed cone* if K is a cone satisfying $K \cap (-K) = \{\mathbf{0}\}$. A nonempty set $A \subseteq \mathbb{R}^n$ is called *proper* if the set A is strictly contained in $\mathbb{R}^n$.

It is easy to verify that a pointed cone is proper and a cone K is convex if and only if $K + K \subseteq K$. To represent feasible sets within optimization theory, we need to introduce an ordering on a set.

Definition 2.2 A binary relation $\preceq$ on a nonempty set A is called a transitive ordering on the set A if for every $\mathbf{x}_i \in A$, $i = 1, 2, 3$ satisfying $\mathbf{x}_1 \preceq \mathbf{x}_2$ and $\mathbf{x}_2 \preceq \mathbf{x}_3$ it follows that $\mathbf{x}_1 \preceq \mathbf{x}_3$. A binary relation $\preceq$ on a nonempty set A is called a partial ordering on the set A if it is a transitive ordering and for every $\mathbf{x} \in A$ it satisfies $\mathbf{x} \preceq \mathbf{x}$ (reflexive property) and for every $\mathbf{x}_i \in A, i = 1, 2$, satisfying $\mathbf{x}_1 \preceq \mathbf{x}_2$ and $\mathbf{x}_2 \preceq \mathbf{x}_1$ it holds that $\mathbf{x}_1 = \mathbf{x}_2$ (antisymmetry property)

For any convex cone $K \subseteq \mathbb{R}^n$, we introduce the transitive ordering $\preceq_K$ on the set $\mathbb{R}^n$ (Boyd and Vandenberghe , 2004; Frenk and Kassay, 1999) given by

$$\mathbf{y} \preceq_K \mathbf{x} \Leftrightarrow \mathbf{x} - \mathbf{y} \in K. \tag{2}$$

It follows that the transitive ordering $\preceq_K$ on $\mathbb{R}^n$ for any convex pointed cone K is a partial ordering on $\mathbb{R}^n$. In most cases, unless otherwise specified, we will only consider orderings $\preceq_K$ on $\mathbb{R}^n$ with K a convex pointed cone. Consider now a nonempty set $X \subseteq \mathbb{R}^n$ and $K \subseteq \mathbb{R}^m$ a nonempty proper convex cone. Moreover, let $f : \mathbb{R}^n \to \mathbb{R}$ be some finite valued function and $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^m$ a finite valued vector function given by $\mathbf{g}(\mathbf{x}) := (g_1(\mathbf{x}), \ldots, g_m(\mathbf{x}))$ with $g_i : \mathbb{R}^n \to \mathbb{R}$ and introduce the minimization problem:

$$\upsilon(P) := \inf\{f(\mathbf{x}) : \mathbf{x} \in \mathcal{F}\} \tag{P}$$

with

$$\mathcal{F} := \{\mathbf{x} \in X : \mathbf{g}(\mathbf{x}) \preceq_K \mathbf{0}\} \tag{3}$$

the so-called feasible region. The optimization problem (P) is called a *primal problem*, and using the ordering $\preceq_K$ with K a proper convex cone yields the possibility to model a lot of different optimization problems. The value $\upsilon(P)$ is by definition equal to ∞ if the feasible region $\mathcal{F}$ of optimization problem (P) is empty. In this case optimization problem (P) is called *infeasible*. If the feasible region $\mathcal{F}$ is nonempty, then optimization problem (P) is called *feasible* and $-\infty \le \upsilon(P) < \infty$. It is not assumed that an *optimal feasible solution* exists and so we cannot replace

inf by min. Since

$$-\sup\{-f(\mathbf{x}) : \mathbf{x} \in \mathcal{F}\} = \inf\{f(\mathbf{x}) : \mathbf{x} \in \mathcal{F}\},$$

we also cover in this framework maximization problems. Important instances of optimization problem (P) are given by nonlinear optimization problems with feasible region $\mathcal{F}$ consisting of p inequality and $m-p$ equality constraints (Bazaraa, Sherali and Shetty, 1993; Nocedal and Wright, 2006). If this holds, the convex cone K is given by $K = \mathbb{R}_+^p \times \{\mathbf{0}\}, \mathbf{0} \in \mathbb{R}^{m-p}, 0 \le p \le m$. For $p = 0$, this reduces to a feasible region $\mathcal{F}$ only consisting of m equality constraints and we obtain $K = \{\mathbf{0}\}$. Special cases are linear programming problems (Bazaraa, Jarvis, and Sherali, 1990; Chvatal, 1983) and for X a mixed discrete–continuous set mixed integer linear programming problems (Nemhauser and Wolsey, 1988; Wolsey, 1998). Finally we also mention conic convex programming problems (Ben-Tal and Nemirovski, 2001) given by

$$\inf\{\mathbf{c}^\top \mathbf{x} : \mathbf{x} - \mathbf{b} \in L, \mathbf{x} \in K\} \tag{4}$$

with K a proper convex cone and L a linear space. In most cases it is difficult to find a feasible solution of optimization problem (P) and even more difficult (if it exists!) to find an optimal solution. A possible way to overcome this problem is to replace optimization problem (P) by an easier optimization problem selected in such a way that it resembles the original problem. Having this idea in mind we introduce the next definition.

Definition 2.3 An optimization problem given by

$$\upsilon(R) := \inf\{f_R(\mathbf{x}) : \mathbf{x} \in \mathcal{F}_R\} \tag{R}$$

is called a relaxation of the primal problem (P) if the feasible region $\mathcal{F}_R$ satisfies $\mathcal{F} \subseteq \mathcal{F}_R$ and $f_R(\mathbf{x}) \le f(\mathbf{x})$ for every $\mathbf{x}$ belonging to $\mathcal{F}$.

An immediate consequence of the definition of a relaxation is given by

$$-\infty \le \upsilon(R) \le \upsilon(P) \le \infty. \tag{5}$$

An easy way to obtain a relaxation is to delete some of the difficult constraints within the feasible set $\mathcal{F}$ given by relation (3). However, to achieve that the new constructed optimization problem resembles the original problem (P) it is in general not sufficient to delete constraints. A more refined way to achieve this goal is to delete constraints and *at the same time* incorporate within the objective function some penalty cost to compensate for the loss of these constraints and/or give some reward for still satisfying feasibility. To formalize this idea, we first introduce the standard inner product on $\mathbb{R}^n$ given by

$$\mathbf{x}^\top \mathbf{y} := \sum_{i=1}^n x_i y_i. \tag{6}$$

The most simplest way to achieve penalization of the loss of constraints is to consider a linear penalty function and introduce for some penalty parameter $\boldsymbol{\lambda} \in \mathbb{R}^m$ the problem

$$\inf\{f(\mathbf{x}) + \boldsymbol{\lambda}^\mathsf{T}\mathbf{g}(\mathbf{x}) : \mathbf{x} \in X\}. \tag{7}$$

Clearly it follows that $\mathcal{F} \subseteq X$ and to guarantee that

$$f(\mathbf{x}) + \boldsymbol{\lambda}^\mathsf{T}\mathbf{g}(\mathbf{x}) \leq f(\mathbf{x}) \tag{8}$$

for every $\mathbf{x}$ belonging to $\mathcal{F}$ it is sufficient and necessary to assume that $\boldsymbol{\lambda}^\mathsf{T}\mathbf{g}(\mathbf{x}) \leq 0$ for every $\mathbf{x}$ belonging to $\mathcal{F}$. Since we are dealing with minimization problems, a reward means that $\boldsymbol{\lambda}^\mathsf{T}\mathbf{g}(\mathbf{x}) \leq 0$ for every feasible $\mathbf{x}$. Due to $\mathbf{g}(\mathbf{x}) \preceq_K \mathbf{0}$ for every $\mathbf{x} \in \mathcal{F}$ and introducing the so-called dual cone K^* of a set $K \subseteq \mathbb{R}^m$ given by

$$K^* := \{\mathbf{y} \in \mathbb{R}^m : \mathbf{y}^\mathsf{T}\mathbf{x} \geq 0 \text{ for every } \mathbf{x} \in K\}, \tag{9}$$

this inequality certainly holds for every $\boldsymbol{\lambda}$ belonging to K^*. Also by the definition of a dual cone, the set K^* is a nonempty closed convex cone. A similar operation for the special case of a linear space $L \subseteq \mathbb{R}^m$ is to introduce its *orthogonal complement* $L^\perp = \{\mathbf{y} \in \mathbb{R}^m : \mathbf{y}^\mathsf{T}\mathbf{x} = 0 \text{ for every } \mathbf{x} \in L\}$. It is easy to show for a linear space L that

$$L^* = L^\perp. \tag{10}$$

At the moment, except for finiteness of the functions f and $\mathbf{g}$ on their domain we do not impose any other conditions on these functions and formally introduce the minimization problem:

$$\theta(\boldsymbol{\lambda}) := \inf\{L(\mathbf{x}, \boldsymbol{\lambda}) : \mathbf{x} \in X\} \tag{$D(\boldsymbol{\lambda})$}$$

with the function $L : X \times K^* \to [-\infty, \infty)$ the so-called Lagrangian *function* given by

$$L(\mathbf{x}, \boldsymbol{\lambda}) := f(\mathbf{x}) + \boldsymbol{\lambda}^\mathsf{T}\mathbf{g}(\mathbf{x}). \tag{11}$$

Since the set X is nonempty, it follows that $-\infty \leq \theta(\boldsymbol{\lambda}) < \infty$. The objective function $\theta : K^* \to [-\infty, \infty)$ is called the *Lagrangian dual function* and the vector $\boldsymbol{\lambda}$ the vector of *Lagrangian multipliers* or *dual variables*. To extend the definition of the function θ to $\mathbb{R}^m$, we set $\theta(\boldsymbol{\lambda}) = -\infty$ for every $\boldsymbol{\lambda}$ not belonging to K^*. By the construction of the Lagrangian dual function, the following result is easy to verify. Implicitly it is always assumed that solving optimization problem $(D(\boldsymbol{\lambda}))$ is much more easier for any $\boldsymbol{\lambda}$ than solving the original problem (P).

Lemma 2.1 *The function $\theta : \mathbb{R}^m \to [-\infty, \infty)$ is concave satisfying $\theta(\lambda) \leq \upsilon(P)$ for every $\lambda \in \mathbb{R}^m$.*

However, even for a primal problem having an optimal solution it might happen that the optimization problem $(D(\lambda))$ does not have an optimal solution for any $\lambda \in K^*$ and so $\theta(\lambda) = -\infty$ for every $\lambda \in \mathbb{R}^m$. This is shown by the following example. If this happens, the convex set $dom(\theta) = \{\lambda \in \mathbb{R}^m : \theta(\lambda) > -\infty\}$ is an empty set and solving optimization problem $(D(\lambda))$ for some given λ does not give any information about the value $\upsilon(P)$.

Example 2.1 Consider the function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = -x^2$ if $x < 0$ and $f(x) = x$ if $x \geq 0$ and consider the optimization problem:

$$\inf\{f(x) : 0 \leq x < \infty\} = \inf\{f(x) : x \preceq_K 0\}$$

with $K = \mathbb{R}_-$ and $X = \mathbb{R}$. This problem has optimal solution $x = 0$, and the dual cone K^* is given by $K^* = \mathbb{R}_-$. Constructing now the Lagrangian function, we obtain for $\lambda \leq 0$ the Lagrangian dual function:

$$\theta(\lambda) = \inf\{f(x) + \lambda x : x \in \mathbb{R}\}.$$

Since $\lim_{x \downarrow -\infty} f(x) + \lambda x = -\infty$ for every $\lambda \leq 0$, it follows that $\theta(\lambda) = -\infty$ for every $\lambda \leq 0$.

We now discuss how much the Lagrangian relaxation optimization problem $(D(\lambda))$ resembles the original problem (P). A similar result for an optimization problem consisting only of equality constraints or equivalently $K = \{0\}$ is discussed by Everett (1963).

Lemma 2.2 *If $\mathbf{x}_{opt}(\lambda)$ is an optimal solution of optimization problem:*

$$\theta(\lambda) = \inf\{f(\mathbf{x}) + \lambda^\top \mathbf{g}(\mathbf{x}) : \mathbf{x} \in X\}$$

for some $\lambda \in K^*$ with K a convex pointed cone, then $\mathbf{x}_{opt}(\lambda)$ is an optimal solution of the perturbed primal problem:

$$\inf\{f(\mathbf{x}) : \mathbf{g}(\mathbf{x}) \preceq_K \mathbf{g}(\mathbf{x}_{opt}(\lambda)), \mathbf{x} \in X\}.$$

Proof Let $\mathbf{x}_{opt}(\lambda)$ be an optimal solution of $\inf\{f(\mathbf{x}) + \lambda^\top \mathbf{g}(\mathbf{x}) : \mathbf{x} \in X\}$ for some $\lambda \in K^*$ and consider an arbitrary $\mathbf{x} \in X$ satisfying $\mathbf{g}(\mathbf{x}) \preceq_K \mathbf{g}(\mathbf{x}_{opt}(\lambda))$. We show that $\mathbf{x}_{opt}(\lambda)$ is feasible for the perturbed primal problem and $f(\mathbf{x}_{opt}(\lambda)) \leq f(\mathbf{x})$. Since the cone is pointed and hence $\mathbf{0} \in K$, the feasibility is obvious. To show $f(\mathbf{x}_{opt}(\lambda)) \leq f(\mathbf{x})$ for every $\mathbf{x} \in X$ and $\mathbf{g}(\mathbf{x}) \preceq_K \mathbf{g}(\mathbf{x}_{opt}(\lambda))$, we observe for $\lambda \in K^*$ that $\lambda^\top(\mathbf{g}(\mathbf{x}_{opt}(\lambda) - \mathbf{g}(\mathbf{x})) \geq 0$. This yields for $\lambda \in K^*$ and $\mathbf{x}_{opt}(\lambda)$ an

optimal solution of the minimization problem $(D(\lambda))$ that

$$
\begin{aligned}
f(\mathbf{x}_{opt}(\lambda)) &= L(\mathbf{x}_{opt}(\lambda), \lambda) - \lambda^{\mathsf{T}}\mathbf{g}(\mathbf{x}_{opt}(\lambda)) \leq L(\mathbf{x}, \lambda) - \lambda^{\mathsf{T}}\mathbf{g}(\mathbf{x}_{opt}(\lambda)) \\
&= f(\mathbf{x}) + \lambda^{\mathsf{T}}(\mathbf{g}(\mathbf{x}) - \mathbf{g}(\mathbf{x}_{opt}(\lambda))) \leq f(\mathbf{x})
\end{aligned}
$$

and we have verified the result. $\square$

To approximate the objective value $\upsilon(P)$ of a primal problem (P) as good as possible by means of the *relaxation approach*, it is therefore by Lemma 2.1 necessary (*maybe not sufficient!*) to maximize the dual function $\theta(\lambda)$ over the set K^*. This motivates the following definition.

Definition 2.4 For a primal problem (P), the optimization problem (D) given by

$$
\upsilon(D) = \sup\{\theta(\lambda) : \lambda \in K^*\} \tag{D}
$$

is called the Lagrangian dual of the primal problem (P).

An immediate consequence of Lemma 2.1 is given by the next result and this result is known as *weak duality*.

Lemma 2.3 *It follows that* $-\infty \leq \upsilon(D) \leq \upsilon(P) \leq \infty$.

In the next example we will use the principle of Lagrange to construct the Lagrangian dual problem of a canonical linear programming problem as proposed on page 237 in Bazaraa, Sherali and Shetty (1993). In different books canonical linear programming problems are defined differently, and one can derive similar results for these linear programs.

Example 2.2 Consider the canonical linear programming problem (Bazaraa, Sherali and Shetty, 1993):

$$
\upsilon(P) := \inf\{\mathbf{c}^{\mathsf{T}}\mathbf{x} : A\mathbf{x} \geq \mathbf{b}, \mathbf{x} \geq \mathbf{0}\}
$$

with A some $m \times n$ matrix, $\mathbf{c} \in \mathbb{R}^n$, and $\mathbf{b} \in \mathbb{R}^m$. It follows that

$$
\upsilon(P) = \inf\{\mathbf{c}^{\mathsf{T}}\mathbf{x} : \mathbf{g}(\mathbf{x}) \preceq_K \mathbf{0}, \mathbf{x} \geq \mathbf{0}\}
$$

with $\mathbf{g}(\mathbf{x}) = \mathbf{b} - A\mathbf{x}$ and $K = \mathbb{R}^m_+$. Hence by penalizing the constraints $\mathbf{b} - A\mathbf{x}$, we obtain for every $\lambda \in K^* = \mathbb{R}^m_+$ that the Lagrangian dual function $\theta : \mathbb{R}^m_+ \to [-\infty, \infty)$ equals

$$
\theta(\lambda) = \lambda^{\mathsf{T}}\mathbf{b} + \inf\{(\mathbf{c} - A^{\mathsf{T}}\lambda)^{\mathsf{T}}\mathbf{x} : \mathbf{x} \geq \mathbf{0}\} = \begin{cases} \lambda^{\mathsf{T}}\mathbf{b} & \text{if } \mathbf{c} - A^{\mathsf{T}}\lambda \geq \mathbf{0} \\ -\infty & \text{otherwise} \end{cases}
$$

This yields that the Lagrangian dual problem (D) is given by

$$\sup\{\theta(\boldsymbol{\lambda}) : \boldsymbol{\lambda} \in \mathbb{R}^m\} = \sup\{\boldsymbol{\lambda}^\mathsf{T}\mathbf{b} : A^\mathsf{T}\boldsymbol{\lambda} \leq \mathbf{c}, \boldsymbol{\lambda} \geq \mathbf{0}\},$$

and this coincides with the so-called dual linear programming of the above canonical linear programming problem mostly introduced without any explanation. To generalize this construction to conic convex problems, we first observe that for $X = K$

$$\inf\{(\mathbf{c} - \boldsymbol{\lambda})^\top\mathbf{x} : \mathbf{x} \in X\} = \begin{cases} 0 & \text{if } \mathbf{c} - \boldsymbol{\lambda} \in K^* \\ -\infty & \text{otherwise} \end{cases} \tag{12}$$

Introducing the vector function $\mathbf{g}(\mathbf{x}) = \mathbf{b} - \mathbf{x}$ and using that any linear space is a convex cone and $L^* = L^\top$, it follows that

$$\inf\{\mathbf{c}^\top\mathbf{x} : \mathbf{x} - \mathbf{b} \in L, \mathbf{x} \in K\} = \inf\{\mathbf{c}^\top\mathbf{x} : \mathbf{g}(\mathbf{x}) \preceq_L \mathbf{0}, \mathbf{x} \in K\}.$$

By a similar reasoning as in the first part of this example, using relation (12) we obtain that the Lagrangian dual of the above conic convex programming problem is given by $\sup\{\boldsymbol{\lambda}^\top\mathbf{b} : \mathbf{c} - \boldsymbol{\lambda} \in K^*, \boldsymbol{\lambda} \in L^\perp\}$.

In Example 2.2 it is shown that the well-known LP-dual of a linear programming problem (Bazaraa, Sherali and Shetty, 1993) mostly introduced without any explanation can be constructed using the principle of Lagrange. A direct consequence of Lemma 2.3 for standard linear programming problems is given by the observation that $\upsilon(P) = -\infty$ implies that the dual problem is infeasible. If in Lemma 2.3 we have $\upsilon(D) < \upsilon(P)$, a so-called duality gap exists. As shown by the following example, such a duality gap can even occur for linear programming problems (see also page 60 of Chvatal (1983)).

Example 2.3 Consider the linear programming problem:

$$\inf\{-x_1 - x_2 : x_1 - x_2 \geq 1, -x_1 + x_2 \geq 1, \mathbf{x} \in \mathbb{R}_+^2\}.$$

Clearly this optimization problem is infeasible and so $\upsilon(P) = \infty$. Penalizing the constraints $x_1 - x_2 - 1 \geq 0$ and $-x_1 + x_2 - 1 \geq 0$ using the nonpositive Lagrangian multipliers λ_1 and λ_2, we obtain that the Lagrangian dual function $\theta : R_-^2 \to [-\infty, \infty)$ is given by

$$\theta(\boldsymbol{\lambda}) = \inf\{x_1(\lambda_1 - \lambda_2 - 1) + x_2(\lambda_2 - \lambda_1 - 1) : \mathbf{x} \in \mathbb{R}_+^2\}.$$

Observe now for every $\boldsymbol{\lambda} \in \mathbb{R}_-^2$ that $\lambda_1 - \lambda_2 - 1 \geq 0 \Rightarrow \lambda_2 - \lambda_1 \leq -1$ and $\lambda_2 - \lambda_1 - 1 \geq 0 \Rightarrow \lambda_1 - \lambda_2 \leq -1$ and by this observation it follows that $\theta(\boldsymbol{\lambda}) = -\infty$ for every $\boldsymbol{\lambda} \in \mathbb{R}_-^2$ or equivalently $\upsilon(D) = -\infty$.

By Example 2.3, it is clear that one should be careful in applying the principle of Lagrange to a primal problem with an empty feasible region. Fortunately in a lot of practical problems one can decide beforehand that the feasible region $\mathcal{F}$ is nonempty and so we only consider primal problems satisfying $-\infty \leq \upsilon(P) < \infty$. By Lemma 2.3, we obtain that

$$\upsilon(P) = -\infty \Rightarrow \upsilon(P) = \upsilon(D) \tag{13}$$

and so in this case no duality gap exists and every $\lambda \in K^*$ is an optimal dual solution. By relation (13) and Example 2.3, we therefore only need to check for $\upsilon(P)$ finite whether a duality gap exists. For the class of feasible primal problems satisfying $\upsilon(P) > -\infty$, it is now important to decide under which conditions

$$\upsilon(D) = \upsilon(P). \tag{14}$$

An important issue in optimization theory is to determine the largest possible class satisfying the above equality. By identifying such a class, we are able in theory to solve an optimization problem belonging to this class using the dual approach or a combination of the dual and primal approach. The most well-known example of such an approach is the primal dual simplex method in linear programming (Chvatal, 1983). We return to this *identification problem* in the next section. Remember in Example 2.1 we constructed a primal optimization problem not having the property listed in relation (14). In this example an optimal solution of the primal optimization problem exists and $\upsilon(P) > \upsilon(D) = -\infty$ and the Lagrangian dual problem does not have any optimal solution. As shown in the next section, this never happens for linear programming problems or K-convex optimization problems satisfying a so-called regularity condition.

Despite the negative result as shown by Example 2.1, the Lagrangian relaxation approach is often used. This applies in particular to discrete optimization problems since for these problems mostly no strong duality holds. During the search for an upper bound on the objective value $\upsilon(P)$, a number of times the optimization problem $(D(\lambda))$ for different values $\lambda_i \in K^*, i = 1, \ldots, k$ is solved and we obtain the dual feasible optimal solutions $\mathbf{x}_{opt}(\lambda_i), 1 \leq i \leq k$. Sometimes the dual feasible optimal solution $\mathbf{x}_{opt}(\lambda_i)$ might also be primal feasible or equivalently $\mathbf{g}(\mathbf{x}_{opt}(\lambda_i)) \preceq_K \mathbf{0}$. However, in most cases this does not happen and so $\mathbf{x}_{opt}(\lambda_i)$ is not primal feasible. If $\mathbf{x}_{opt}(\lambda_i)$ is not primal feasible, it is sometimes possible to construct using the primal nonfeasible $\mathbf{x}_{opt}(\lambda_i), i = 1 \ldots, k$ a primal feasible $\mathbf{x}_i$ and we obtain a feasible candidate of our primal problem. The heuristic procedure that converts a primal nonfeasible $\mathbf{x}_{opt}(\lambda)$ into a primal feasible $\mathbf{x}$ is called a *Lagrangian heuristic*. On the other hand, it is also possible to generate for most problems a feasible solution by means of a so-called *primal heuristic*. Such a primal heuristic is mostly an ad hoc procedure particularly suited for the specific problem and it does not use dual information as done by a Lagrangian heuristic. We consider some examples of this approach in location and transportation problems. One might now wonder whether it is possible to say something about the quality of the generated

primal feasible solution by this Lagrangian heuristic with respect to the optimal objective value of the primal problem. This motivates the next definition.

Definition 2.5 For any $\epsilon > 0$, a vector $\mathbf{x} \in \mathbb{R}^n$ is called an ϵ-optimal solution of the optimization problem $\upsilon(P) = \inf\{f(\mathbf{x}) : \mathbf{x} \in \mathcal{F}\}$ if $f(\mathbf{x}) \leq \upsilon(P) + \epsilon$ and $\mathbf{x} \in \mathcal{F}$.

The next result discusses the quality of a primal feasible ϵ-optimal solution of the optimization problem $(D(\boldsymbol{\lambda}))$. The assumption of the next lemma automatically implies that $\upsilon(P)$ is finite.

Lemma 2.4 *If for some $\boldsymbol{\lambda} \in K^*$ there exists a primal feasible $\mathbf{x}_0$ with $\mathbf{x}_0$ an ϵ-optimal solution of optimization problem:*

$$\theta(\boldsymbol{\lambda}) = \inf\{f(\mathbf{x}) + \boldsymbol{\lambda}^\mathsf{T}\mathbf{g}(\mathbf{x}) : \mathbf{x} \in X\},$$

then

$$\upsilon(P) \leq f(\mathbf{x}_0) \leq \upsilon(P) + \epsilon - \boldsymbol{\lambda}^\mathsf{T}\mathbf{g}(\mathbf{x}_0) < \infty.$$

Proof Since $\mathbf{x}_0$ is primal feasible, we obtain by the weak duality result given by Lemma 2.3 that

$$\begin{aligned}
\upsilon(P) &\leq f(\mathbf{x}_0) = L(\mathbf{x}_0, \boldsymbol{\lambda}) - \boldsymbol{\lambda}^\mathsf{T}\mathbf{g}(\mathbf{x}_0) \\
&\leq \theta(\boldsymbol{\lambda}) + \epsilon - \boldsymbol{\lambda}^\mathsf{T}\mathbf{g}(\mathbf{x}_0) \leq \upsilon(D) + \epsilon - \boldsymbol{\lambda}^\mathsf{T}\mathbf{g}(\mathbf{x}_0) \\
&\leq \upsilon(P) + \epsilon - \boldsymbol{\lambda}^\mathsf{T}\mathbf{g}(\mathbf{x}_0)
\end{aligned}$$

and this shows the result. □

The next result is partly an easy implication of Lemma 2.4.

Lemma 2.5 *The following conditions are equivalent:*

1. The vector $\mathbf{x}_0$ is primal feasible ($\mathbf{g}(\mathbf{x}_0) \preceq_K \mathbf{0}$), $\boldsymbol{\lambda}_0$ is dual feasible ($\boldsymbol{\lambda}_0 \in K^*$), and $f(\mathbf{x}_0) = \theta(\boldsymbol{\lambda}_0)$.
2. The vector $\mathbf{x}_0$ is an optimal solution of optimization problem (P), the Lagrangian multiplier $\boldsymbol{\lambda}_0 \in K^*$ is an optimal solution of the Lagrangian dual problem (D), and $v(P) = v(D)$.
3. The primal feasible solution $\mathbf{x}_0$ is an optimal solution of $(D(\boldsymbol{\lambda}_0))$ with $\boldsymbol{\lambda}_0 \in K^*$ and $\theta(\boldsymbol{\lambda}_0)$ finite and $\boldsymbol{\lambda}_0^\mathsf{T}\mathbf{g}(\mathbf{x}_0) = 0$.

Proof To prove the implication $1 \to 3$, we observe

$$f(\mathbf{x}_0) = \theta(\boldsymbol{\lambda}_0) \leq f(\mathbf{x}_0) + \boldsymbol{\lambda}_0^\mathsf{T}\mathbf{g}(\mathbf{x}_0) \tag{15}$$

and this shows $\boldsymbol{\lambda}_0^\mathsf{T}\mathbf{g}(\mathbf{x}_0) \geq 0$. Since $\mathbf{g}(\mathbf{x}_0) \preceq_K \mathbf{0}$ and $\boldsymbol{\lambda}_0 \in K^*$, this yields $\boldsymbol{\lambda}_0^\mathsf{T}\mathbf{g}(\mathbf{x}_0) = 0$. By relation (15), we obtain $f(\mathbf{x}_0) + \boldsymbol{\lambda}_0^\mathsf{T}\mathbf{g}(\mathbf{x}_0) = \theta(\boldsymbol{\lambda}_0)$ and so $\mathbf{x}_0$ is an optimal solution of $\theta(\boldsymbol{\lambda}_0)$ showing part 3. The implication $3 \to 2$ is a direct consequence of Lemma 2.4, while the implication $2 \to 1$ is trivial.

The condition $\boldsymbol{\lambda}_0^\mathsf{T}\mathbf{g}(\mathbf{x}_0) = 0$ mentioned in part 3 of Lemma 2.5 is called the *complementary slackness condition*. Observe if $K = \{\mathbf{0}\}$ this complementary slackness condition is automatically satisfies for $\mathbf{x}_0$ primal feasible. Unfortunately in most integer linear programming problems, there exists a duality gap unless the matrix describing the feasible region has a special structure like total unimodularity (Nemhauser and Wolsey, 1988) and so in general condition 3 in Lemma 2.5 does not hold. Observe that we are interested in condition 3 since our procedure is to solve optimization problems $(D(\boldsymbol{\lambda}))$ for different values of $\boldsymbol{\lambda} \in K^*$. If during this procedure we find a vector $\mathbf{x}_0$ satisfying condition 3 of Lemma 2.5, then this vector $\mathbf{x}_0$ is an optimal solution of our problem (P). In Lemma 2.5 we assume that no duality gap exists and so it is important to know for which optimization problems such a result holds. This is discussed in the next section.

2.1 K-Convex Minimization Problems and Strong Duality

In this section we consider a class of optimization problems (P) for which under a so-called regularity condition on the feasible region no duality gap exists. To identify these classes of optimization problems, we introduce the following class of vector valued functions (Giannessi , 1984; Wolkowicz, 1981). Observe in this definition we do not assume that K is a convex pointed cone.

Definition 2.6 Let $X \subseteq \mathbb{R}^n$ be a nonempty convex set and $K \subseteq \mathbb{R}^m$ is a nonempty convex cone. The vector valued function $\mathbf{g} : X \to \mathbb{R}^m$ is called K-convex if and only if

$$\mathbf{g}(\alpha\mathbf{x}_1 + (1 - \alpha)\mathbf{x}_2) \preceq_K \alpha\mathbf{g}(\mathbf{x}_1) + (1 - \alpha)\mathbf{g}(\mathbf{x}_2) \tag{16}$$

for every $0 < \alpha < 1, \mathbf{x}_1, \mathbf{x}_2 \in X$ and $\preceq_K$ the transitive ordering introduced in relation (2).

If $K = \mathbb{R}^m_+$ (and so a convex pointed cone), then K-convexity reduces to the classical definition of convexity. Also it is easy to check that any affine function $\mathbf{g}(\mathbf{x}) = A\mathbf{x} + \mathbf{b}$ is K-convex for any nonempty convex pointed cone K. To relate the above definition of a K-convex function for $K \subseteq \mathbb{R}^m$ a convex pointed cone to the modern definition of a convex function (Rockafellar, 1970), we introduce for any convex pointed cone $K \subseteq \mathbb{R}^m$ and any vector function $\mathbf{g} : \mathbb{R}^n \to (\mathbb{R} \cup \{\infty\})^m$ the so-called epigraph $epi_K(\mathbf{g})$ given by

$$epi_K(\mathbf{g}) := \{(\mathbf{x}, \mathbf{r}) : \mathbf{g}(\mathbf{x}) \preceq_K \mathbf{r}\} \subseteq \mathbb{R}^{n+m}.$$

Since K is a convex pointed cone, we know that the ordering $\preceq_K$ is a partial ordering. It is shown in (Frenk and Kassay, 2007) for K a convex pointed cone and $\mathbf{h} : X \to \mathbb{R}^m$ some given vector function on a nonempty domain $X \subseteq \mathbb{R}^n$ that

the set $epi_K(\mathbf{h})$ is a convex set if and only if X is a convex set and $\mathbf{h} : X \to \mathbb{R}^m$ is K-convex and so for K a convex pointed cone the old definition of convexity coincides with the modern definition of convexity.

Definition 2.7 The optimization problem $\inf\{f(\mathbf{x}) : \mathbf{g}(\mathbf{x}) \preceq_K \mathbf{0}, \mathbf{x} \in X\}$ with K a proper convex cone and nonempty feasible region $\mathcal{F}$ is called a K-convex minimization problem if the set $X \subseteq \mathbb{R}^n$ is convex, the function $f : X \to \mathbb{R}$ is convex, and the function $\mathbf{g} : X \to \mathbb{R}^m$ is K-convex.

In the above definition we could also have assumed that K is a convex cone. However, using this alternative definition we only added the case $K = \mathbb{R}^n$, and since in this case $K^* = \{\mathbf{0}\}$, the Lagrangian dual function reduces to the original problem. At the same time, we could have imposed that the cone K is pointed (mostly assumed for K-convex optimization problems) but this condition only implies that the ordering $\preceq_K$ is a partial ordering and this property is not necessary for achieving strong duality for K-convex optimization problems. Since by relation (16) the function $\mathbf{x} \mapsto \boldsymbol{\lambda}^\top \mathbf{g}(\mathbf{x})$ is convex for $\boldsymbol{\lambda} \in K^*$ and any K-convex function $\mathbf{g} : X \to \mathbb{R}^m$, it follows for any K-convex minimization problem that the optimization problem $(D(\boldsymbol{\lambda}))$ is a convex minimization problem for every $\boldsymbol{\lambda} \in K^*$. To show that for the class of K-convex minimization problems satisfying some additional constraint qualification the strong duality property holds, we need the following definition.

Definition 2.8

1. If $S \subseteq \mathbb{R}^n$ is nonempty, then the set aff(S) is the smallest affine set containing S. This set is called the *affine hull* of S.
2. If $S \subseteq \mathbb{R}^n$ and $\mathbf{y} \in S$, then the point $\mathbf{y}$ is called a *relative interior point* of the set S if there exists some $\epsilon > 0$ such that the intersection of aff(S) and the set $\mathbf{y} + \epsilon\mathcal{B}$ with $\mathcal{B}$ the closed Euclidean ball given by

$$\mathcal{B} := \{\mathbf{x} \in \mathbb{R}^n : \| \mathbf{x} \| \leq 1\}$$

and $\| \, . \, \|$ the Euclidean norm is contained in S.
3. The *relative interior* ri(S) of a set $S \subseteq R^n$ is the collection of all relative interior points of the set S.

By the generalization of inequalities to cones, we needed the concept of K-convex functions. However, as for classical convex programming problems with equalities the property of K-convexity is stronger than is needed for the proof of the no-duality gap for K-convex optimization problems. What we need for this proof is the following implication of K-convex functions, which is weaker than K-convexity.

Lemma 2.6 *If $K \subseteq \mathbb{R}^m$ is a nonempty convex cone and the function $\mathbf{g} : X \to R^m$ is K-convex, then the set $\mathbf{g}(X) + ri(K)$ is nonempty and convex.*

Proof Since the set $K \subseteq \mathbb{R}^m$ is a nonempty convex cone, it is well-known (Frenk and Kassay, 2005; Rockafellar, 1970) that $ri(K)$ is also a nonempty convex cone and so the set $\mathbf{g}(X) + ri(K)$ is nonempty. To check that the set $\mathbf{g}(X) + ri(K)$ is

convex, we need to verify for any $\mathbf{x}_1, \mathbf{x}_2$ belonging to X and $0 < \alpha < 1$ that

$$\alpha(\mathbf{g}(\mathbf{x}_1) + ri(K)) + (1 - \alpha)(\mathbf{g}(\mathbf{x}_2) + ri(K)) \subseteq \mathbf{g}(X) + ri(K).$$

By the convexity of the set $ri(K)$, it follows that $\alpha ri(K) + (1 - a)ri(K) \subseteq ri(K)$ and this implies that

$$\alpha(\mathbf{g}(\mathbf{x}_1) + ri(K)) + (1 - \alpha)(\mathbf{g}(\mathbf{x}_2) + ri(K)) \subseteq \alpha \mathbf{g}(\mathbf{x}_1) + (1 - \alpha)\mathbf{g}(\mathbf{x}_2) + ri(K).$$

By the K-convexity of the function $\mathbf{g}$ and relation (2), we know that

$$\alpha \mathbf{g}(\mathbf{x}_1) + (1 - \alpha)\mathbf{g}(\mathbf{x}_2) \in \mathbf{g}(\alpha \mathbf{x}_1 + (1 - \alpha)\mathbf{x}_2) + K. \tag{17}$$

Since it is well-known that $K + ri(K) \subseteq ri(K)$ (Frenk and Kassay, 2005; Rockafellar, 1970) and the set X is convex, we obtain by relation (17) that

$$\begin{aligned}
\alpha(\mathbf{g}(\mathbf{x}_1) + ri(K)) + (1 - \alpha)(\mathbf{g}(\mathbf{x}_2) + ri(K)) &\subseteq \mathbf{g}(X) + K + ri(K) \\
&\subseteq \mathbf{g}(X) + ri(K)
\end{aligned}$$

and we have verified the desired result. $\qquad\square$

We now need a separation result to prove that under some regularity condition no duality gap exists for the class of K-convex minimization problems. However, we did not discuss what we mean by separation and so we list the following definition (Frenk and Kassay, 2005; Rockafellar, 1970).

Definition 2.9 If $C \subseteq \mathbb{R}^n$ is a nonempty set and $\mathbf{0} \in R^n$ does not belong to C, then the sets C and $\{\mathbf{0}\}$ are said to be properly separated if there exists some vector $\boldsymbol{\sigma}$ belonging to $\mathbb{R}^n$ satisfying

$$\inf\{\boldsymbol{\sigma}^\mathsf{T}\mathbf{x} : \mathbf{x} \in C\} \geq 0$$

and for some $\mathbf{x}$ belonging to C it follows that $\boldsymbol{\sigma}^\mathsf{T}\mathbf{x} > 0$. The vector $\boldsymbol{\sigma}$ is called the normal vector of the separating hyperplane between the sets C and $\{\mathbf{0}\}$.

By Definition 2.9, it follows that $\boldsymbol{\sigma}$ is not equal to the zero vector. The next theorem mentions an important separation result for an arbitrary proper convex set (Frenk and Kassay, 2005; Hiriart-Urruty and Lemaréchal, 2013; Rockafellar, 1970).

Theorem 2.1 *If $C \subseteq \mathbb{R}^n$ is a nonempty convex set and $\mathbf{0} \in \mathbb{R}^n$ does not belong to C, then the sets C and $\{\mathbf{0}\}$ can be properly separated. Moreover, the normal vector $\boldsymbol{\sigma}$ of the separating hyperplane can be chosen to belong to the affine hull $aff(C)$ of the set C.*

By the above separation result, it is possible to show that any K-convex minimization problem satisfying an additional constraint qualification has no duality gap.

Theorem 2.2 *If the K-convex minimization problem satisfies $-\infty < \upsilon(P) < \infty$ and the vector $\mathbf{0}$ belongs to $ri(\mathbf{g}(X) + K)$, then the Lagrangian dual (D) of the K-convex optimization problem given by $\sup\{\theta(\boldsymbol{\lambda}) : \boldsymbol{\lambda} \in K^*\}$ has an optimal solution and the duality gap equals zero.*

Proof Without loss of generality, we may assume that $\upsilon(P) = 0$. Since optimization problem (P) is a K-convex optimization problem, we obtain that the function $\mathbf{h} : X \to R^{m+1}$ given by

$$\mathbf{h}(\mathbf{x}) = (f(\mathbf{x}), \mathbf{g}(\mathbf{x}))$$

is K_e-convex with $K_e = [0, \infty) \times K$. This implies by Lemma 2.6 that the set $\mathbf{h}(X) + ri(K_e)$ is convex. If $\mathbf{0}$ belongs to the convex set $\mathbf{h}(X) + ri(K_e)$, then there exists some $\mathbf{x}_0 \in X$ satisfying $\mathbf{h}(\mathbf{x}_0) \in -ri(K_e)$. Since $ri(K_e) = (0, \infty) \times ri(K)$, we obtain $f(\mathbf{x}_0) < 0$ and $\mathbf{g}(\mathbf{x}_0) \preceq_K \mathbf{0}$. This contradicts our assumption $\upsilon(P) = 0$ and so $\mathbf{0} \notin \mathbf{h}(X) + ri(K_e)$. Applying now Theorem 2.1, there exists some nonzero vector $\boldsymbol{\sigma} = (\mu, \boldsymbol{\lambda}_0)$ satisfying for every $\mathbf{x} \in X$ and $\mathbf{k}_e = (r, \mathbf{k})$ with $r > 0$ and $\mathbf{k} \in ri(K)$

$$\boldsymbol{\sigma}^\top(\mathbf{h}(\mathbf{x}) + \mathbf{k}_e) = \mu(f(\mathbf{x}) + r) + \boldsymbol{\lambda}_0^\top(\mathbf{g}(\mathbf{x}) + \mathbf{k}) \geq 0 \qquad (18)$$

with a strict inequality for at least one $\mathbf{x}_0 \in X, r_0 > 0$ and $\mathbf{k}_0 \in ri(K)$. Also by Theorem 2.1, the component $\boldsymbol{\lambda}_0$ of this vector belongs to the set $aff(\mathbf{g}(X) + K)$. We now show that $\mu > 0$. Since $ri(K_e)$ is a convex cone, this implies for every $\alpha > 0$ and $\mathbf{k}_e = (r, \mathbf{k}) \in ri(K_e)$ that $\alpha\mathbf{k}_e \in ri(K_e)$ and so we obtain by relation (18) for $\mathbf{x} \in X$

$$\boldsymbol{\sigma}^\top(\mathbf{h}(\mathbf{x}) + \alpha\mathbf{k}_e) \geq 0. \qquad (19)$$

This shows by contradiction for $\mathbf{x} \in X$ fixed and letting α tend to infinity in relation (19) that $\boldsymbol{\sigma}^\top\mathbf{k}_e \geq 0$ for every $\mathbf{k}_e \in ri(K_e)$ and so $\boldsymbol{\sigma}$ belongs to $(ri(K_e))^* = K_e^* = [0, \infty) \times K^*$. Hence it follows that $\mu \geq 0$ and if we assume that $\mu = 0$ or $\boldsymbol{\sigma} = (0, \boldsymbol{\lambda}_0)$ it follows again by relation (18) that

$$\boldsymbol{\lambda}_0^\top(\mathbf{g}(\mathbf{x}) + \mathbf{k}) \geq 0 \qquad (20)$$

for every $\mathbf{x} \in X$ and $\mathbf{k} \in ri(K)$. Since by assumption $\mathbf{0}$ belongs to $ri(\mathbf{g}(X) + K)$ and $\boldsymbol{\lambda}_0 \in aff(\mathbf{g}(X) + K)$, one can find some $\epsilon > 0$ such that $-\epsilon^\top\boldsymbol{\lambda}_0 \in \mathbf{g}(X) + K$ and this implies by relation (20) that

$$-\epsilon\boldsymbol{\lambda}_0^\top\boldsymbol{\lambda}_0 = -\epsilon \parallel \boldsymbol{\lambda}_0 \parallel^2 \geq 0$$

or equivalently $\lambda_0 = \mathbf{0}$. Hence the normal vector σ equals the null vector and this contradicts Theorem 2.1. Therefore $\mu > 0$ and applying now relation (19), we obtain for any $\alpha > 0, r \in ri(K), \mathbf{x} \in X$ and $\mathbf{k} \in K$ that

$$(f(\mathbf{x}) + \alpha r) + \mu^{-1}\lambda_0^\intercal(\mathbf{g}(\mathbf{x}) + \mathbf{k}) \geq 0. \tag{21}$$

Since for $\mathbf{k} \in K$ it follows that $\alpha\mathbf{k} \in K$ for any $\alpha > 0$, we can replace $\mathbf{k}$ in relation (21) by $\alpha\mathbf{k}$ and letting α tend to zero we conclude from the same relation that for every $\mathbf{x} \in X$

$$f(\mathbf{x}) + \mu^{-1}\lambda_0^\intercal\mathbf{g}(\mathbf{x}) \geq 0$$

with $\mu^{-1}\lambda_0 \in K^*$. This shows $\theta(\mu^{-1}\lambda_0) \geq 0$ and applying the weak duality result yields $0 = \upsilon(D)$ and $\mu^{-1}\lambda_0$ is an optimal dual solution. □

The additional constraint qualification in Theorem 2.2 is called a *Slater-type condition* and as shown in the proof of Theorem 2.2 this condition is sufficient to guarantee that an optimal solution of the Lagrangian dual problem (D) of a general K-convex minimization problem exists. The constraint qualification is slightly stronger than the feasibility condition since

$$\mathcal{F} = \{\mathbf{x} \in X : \mathbf{g}(\mathbf{x}) \preceq_K \mathbf{0}\} \text{ nonempty} \iff \mathbf{0} \in \mathbf{g}(X) + K.$$

and it means that the feasible region has a nonempty relative interior. At the same time, it might happen that the primal problem does not have an optimal solution. Actually the *Slater-type condition* is a sufficient condition to guarantee the existence of an optimal solution of the dual problem and this condition guarantees that an optimal solution of the dual problem is contained in a compact subset of K^*. Due to the convexity-type assumptions on the bifunction $L : X \times K^*$ given by $L(\mathbf{x}, \lambda) = f(\mathbf{x}) + \lambda^\intercal\mathbf{g}(\mathbf{x})$ as listed in relation (11), the strong duality follows immediately as a corollary of Sion's minmax theorem already proven in 1928 by Neumann (1928) using Brouwer's fixed point theorem. This result was later rediscovered by Sion (1958) and proved using the Knaster–Kuratowski–Mazurkiewicz (KKM) lemma (George Xian-Zhi Yuan, 1999). This lemma is equivalent to Brouwer's fixed point theorem. This means that the Lagrangian relaxation approach can be seen as a particular instance of a noncooperative two-person game in which nature plays again the decision maker. The action set of nature is given by K^*, and the action set of the decision maker is given by $X \subseteq \mathbb{R}^n$. The strong duality result states now that the value of this particular two-person game exists, and Sion's minmax theorem shows under which sufficient conditions on the bifunction L this game has a value. In fact the strong duality result is an immediate consequence of the property that a convex set is connected and so it is a topological result. For more details on how to prove a generalization of Sion's minmax theorem using a more elementary approach by contradiction and Joos method and the obvious fact that a convex set is connected, the reader should consult (Frenk and Kassay, 2006). Connectedness of a set means

that any two points in such a set can be connected by a continuous curve contained within this set.

For linear programming problems given by $\inf\{\mathbf{c}^\top\mathbf{x} : A\mathbf{x} \leq \mathbf{b}, \mathbf{x} \in R^n\}$, one can verify by inspection of the last lemma in the simplex method (Chvatal, 1983; Saigal, 1997) used to solve the above linear programming problem that the Lagrangian dual of this linear programming problem has an optimal solution and no duality gap exists. *This means for linear programming problems that we only need to assume υ (P) is finite and so no Slater-type condition is needed.* Next to linear programming another class of important optimization problems is represented by the function $d_{cl(K)} : \mathbb{R}^n \to \mathbb{R}$ with

$$d_{cl(K)}(\mathbf{x}) := \inf\{\tfrac{1}{2}\|\mathbf{x} - \mathbf{y}\|^2 : \mathbf{y} \in cl(K)\}$$
$$= \inf\{\tfrac{1}{2}\|\mathbf{x} - \mathbf{y}\|^2 : -\mathbf{y} \preceq_{cl(K)} \mathbf{0}\}$$

and $K \subseteq \mathbb{R}^n$ a nonempty proper convex cone, $\|.\|$ denoting the Euclidean norm, and $cl(K)$ the closure of this convex cone. This optimization problem denotes the orthogonal projection of a vector on a given convex cone, and such a problem shows up in the subgradient method. It is easy to see that

$$d_{cl(K)}(\mathbf{x}) = 0 \Leftrightarrow \mathbf{x} \in cl(K) \tag{22}$$

and for any fixed $\mathbf{x}$ the above optimization problem is clearly a K-convex programming problem (even convex!) with $f(\mathbf{y}) = \tfrac{1}{2}\|\mathbf{x} - \mathbf{y}\|^2$ and $\mathbf{g}(\mathbf{y}) = -\mathbf{y}$ and $X = \mathbb{R}^n$. It follows trivially that $\mathbf{g}(\mathbb{R}^n) = \mathbb{R}^n$ and so the Slater-type condition reduces to $\mathbf{0} \in ri(\mathbb{R}^n + cl(K)) = ri(\mathbb{R}^n) = \mathbb{R}^n$, which is automatically satisfied. Since $(cl(K))^* = K^*$ and applying Theorem 2.2 we obtain that

$$d_{cl(K)}(\mathbf{x}) = \sup\{\theta(\boldsymbol{\lambda}) : \boldsymbol{\lambda} \in K^*\}$$

with

$$\theta(\boldsymbol{\lambda}) = \inf\left\{\frac{1}{2}\|\mathbf{x} - \mathbf{y}\|^2 - \boldsymbol{\lambda}^\top\mathbf{y} : \mathbf{y} \in \mathbb{R}^n\right\}.$$

Since the function $\mathbf{y} \to \tfrac{1}{2}\|\mathbf{x} - \mathbf{y}\|^2 - \boldsymbol{\lambda}^\top\mathbf{y}$ is strictly convex and differentiable on $\mathbb{R}^n$, an optimal solution for the optimization problem $(D(\boldsymbol{\lambda})$ is obtained by solving the necessary and sufficient first-order conditions. This implies with $\mathbf{y}_{opt}(\boldsymbol{\lambda})$ denoting the optimal feasible solution of the optimization problem $(D(\boldsymbol{\lambda}))$ that

$$-(\mathbf{x} - \mathbf{y}_{opt}(\boldsymbol{\lambda})) - \boldsymbol{\lambda} = \mathbf{0} \Leftrightarrow \mathbf{y}_{opt}(\boldsymbol{\lambda}) = \mathbf{x} + \boldsymbol{\lambda}$$

and so $\theta(\boldsymbol{\lambda}) = \frac{1}{2}\|\boldsymbol{\lambda}\|^2 - \boldsymbol{\lambda}^\top(\mathbf{x} + \boldsymbol{\lambda}) = -\frac{1}{2}\|\boldsymbol{\lambda}\|^2 - \boldsymbol{\lambda}^\top\mathbf{x}$. This finally shows

$$d_{cl(K)}(\mathbf{x}) = \max\left\{-\frac{1}{2}\|\boldsymbol{\lambda}\|^2 - \boldsymbol{\lambda}^\top\mathbf{x} : \boldsymbol{\lambda} \in K^*\right\}. \tag{23}$$

and using relation (23) the next so-called important bipolar theorem can be verified.

Theorem 2.3 *If $K \subseteq \mathbb{R}^n$ is a nonempty convex cone, then it follows that $cl(K) = K^{**}$ with $K^{**} := (K^*)^*$.*

Proof If $\mathbf{x}$ belongs to K, then by the definition of K^* it follows for every $\boldsymbol{\lambda} \in K^*$ that $\boldsymbol{\lambda}^\top\mathbf{x} \geq 0$ and this shows that $\mathbf{x}$ belongs to K^{**}. Hence $K \subseteq K^{**}$ and since K^{**} is closed we obtain $cl(K) \subseteq K^{**}$. To prove the reverse inclusion, let $\mathbf{x}$ belong to K^{**}. By the definition of K^{**}, it follows that $\mathbf{x}^\top\boldsymbol{\lambda} \geq 0$ for every $\boldsymbol{\lambda} \in K^*$ and this shows $-\frac{1}{2}\|\boldsymbol{\lambda}\|^2 - \boldsymbol{\lambda}^\top\mathbf{x} \leq -\frac{1}{2}\|\boldsymbol{\lambda}\|^2$ for every $\boldsymbol{\lambda} \in K^*$. Since $\mathbf{0}$ belongs to K^*, this implies by relation (23) that

$$0 \leq d_{cl(K)}(\mathbf{x}) \leq \max\left\{-\frac{1}{2}\|\boldsymbol{\lambda}\|^2 : \boldsymbol{\lambda} \in K^*\right\} = 0$$

and hence by relation (22) we obtain $\mathbf{x} \in cl(K)$. $\qquad\qquad\qquad\qquad\square$

A very special case of the bipolar theorem is the so-called lemma of Farkas known in linear programming, and this result can be used to show (Saigal, 1997) that any linear programming problem with a finite optimal objective function value satisfies the strong duality property. By the above bipolar theorem, it is easy to give a univariate characterization of K-convex functions in case K is a closed convex cone.

Lemma 2.7 *For a closed convex cone K, it follows that the function $\mathbf{x} \to \boldsymbol{\lambda}^\top\mathbf{h}(\mathbf{x})$ is convex on the convex set X for every $\boldsymbol{\lambda} \in K^*$ if and only if $\mathbf{h}$ is K-convex.*

An important implication of Theorem 2.2 and Lemma 2.7 is given by the following result. Observe a strictly convex function is a convex function with the inequalities in the definition replaced by strict inequalities for any strict convex combination of any two vectors.

Lemma 2.8 *If the objective function f in the K-convex optimization problem*

$$\inf\{f(\mathbf{x}) : \mathbf{g}(\mathbf{x}) \preceq_K \mathbf{0}, \mathbf{x} \in X\}$$

is strictly convex and $\mathbf{0} \in ri(\mathbf{g}(X) + K)$ and this optimization problem has an optimal solution, then the Lagrangian dual problem has an optimal solution $\boldsymbol{\lambda}_{opt}$ and the optimal solution $\mathbf{x}_{opt}(\boldsymbol{\lambda}_{opt})$ of the optimization problem $D(\boldsymbol{\lambda}_{opt})$ exists and is unique and coincides with the unique optimal solution of the primal problem.

Proof Since it is assumed that the primal K-convex optimization problem (P) has an optimal solution and the objective function f is strictly convex, the optimal solution of the problem (P) denoted by $\mathbf{x}_{opt}$ is unique. By Theorem 2.2, we know

that an optimal solution $\boldsymbol{\lambda}_{opt} \in K^*$ of the Lagrangian dual problem exists and it satisfies

$$f(\mathbf{x}_{opt}) = \upsilon(P) = \upsilon(D) = \theta(\boldsymbol{\lambda}_{opt}).$$

This shows that condition 1 of Lemma 2.5 is satisfied and again by Lemma 2.5 the primal feasible solution $\mathbf{x}_{opt}$ is an optimal solution of $D(\boldsymbol{\lambda}_0)$. Since by Lemma 2.7 we obtain that the function $\mathbf{x} \rightarrow \boldsymbol{\lambda}_{opt}^\mathsf{T} \mathbf{g}(\mathbf{x})$ is convex on X and hence the function $\mathbf{x} \rightarrow f(\mathbf{x}) + \boldsymbol{\lambda}^\mathsf{T} \mathbf{g}(\mathbf{x})$ is strictly convex, the optimization problem $D(\boldsymbol{\lambda}_{opt})$ has a unique optimal solution and since $\mathbf{x}_{opt}$ is such an optimal solution we obtain the desired result.

From a computational point of view, the following theorem for K-convex optimization problems is important and is a direct consequence of Lemma 2.5 and Theorem 2.2.

Theorem 2.4 *If the optimization problem (P) is a K-convex minimization problem and $\mathbf{0}$ belongs to $ri(\mathbf{g}(X) + K)$, then the vector $\mathbf{x}_0$ is an optimal solution of (P) if and only if for some optimal dual variable $\boldsymbol{\lambda}_0 \in K^*$ it follows that*

$$\mathbf{x}_0 = argmin\{f(\mathbf{x}) + \boldsymbol{\lambda}_0^\mathsf{T} \mathbf{g}(\mathbf{x}) : \mathbf{x} \in X\}, \boldsymbol{\lambda}_0^\mathsf{T} \mathbf{g}(\mathbf{x}_0) = 0 \ and \ -\mathbf{g}(\mathbf{x}_0) \in K.$$

Proof Since (P) is a K-convex optimization problem and $\mathbf{0} \in ri(\mathbf{g}(X) + K)$, it follows by Theorem 2.2 that there exists an optimal dual solution $\boldsymbol{\lambda}_0$ and $\upsilon(P) = \upsilon(D)$. Hence condition 2 of Lemma 2.5 is satisfied and this yields by $2 \rightarrow 3$ of Lemma 2.5 that $\mathbf{x}_0$ satisfies the above system of equations. The reverse implication is also a direct consequence of Lemma 2.5.

In the next result the so-called Karush–Kuhn–Tucker conditions for K-convex programs with differentiable functions are given and this result is an immediate consequence of Theorem 2.4. It also shows that the *Karush–Kuhn–Tucker vector* is an optimal solution of the Lagrangian dual problem.

Theorem 2.5 *If the optimization problem (P) is a K-convex optimization problem consisting of differentiable functions and the convex set X equals $\mathbb{R}^n$, then under the regularity condition $\mathbf{0} \in ri(\mathbf{g}(X) + K)$ the vector $\mathbf{x}_0$ is an optimal solution of (P) if and only if for some optimal dual variable $\boldsymbol{\lambda} \in K^*$ it follows that*

$$\nabla f(\mathbf{x}_0) + \sum_{i=1}^{n} \lambda_i \nabla g_i(\mathbf{x}_0) = \mathbf{0}, \boldsymbol{\lambda}^\mathsf{T} \mathbf{g}(\mathbf{x}_0) = 0 \ and \ -\mathbf{g}(\mathbf{x}_0) \in K.$$

Proof By Theorem 2.4, the vector $\mathbf{x}_0$ is an optimal solution of (P) if and only if for some optimal dual variable $\boldsymbol{\lambda} \in K^*$ we obtain

$$\mathbf{x}_0 = argmin\{f(\mathbf{x}) + \boldsymbol{\lambda}^\mathsf{T} \mathbf{g}(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^n\}, \boldsymbol{\lambda}^\mathsf{T} \mathbf{g}(\mathbf{x}_0) = 0, -\mathbf{g}(\mathbf{x}_0) \in K.$$

Since the optimization problem (P) is a differentiable K-convex optimization problem, it follows that the function $\mathbf{x} \to f(\mathbf{x}) + \boldsymbol{\lambda}^\mathsf{T}\mathbf{g}(\mathbf{x})$ is convex and differentiable on $\mathbb{R}^n$ and so the result follows by the sufficient and necessary first-order conditions for an unconstrained convex optimization problem.

If we consider a K-convex optimization problem satisfying Slater's constraint qualification and having an optimal solution and we identified an optimal dual solution, then by a similar proof as in the first part of Lemma 2.8 we can only show that the optimal solution set

$$\arg\min\{f(\mathbf{x}) + \boldsymbol{\lambda}^\mathsf{T}\mathbf{g}(\mathbf{x}) : \mathbf{x} \in X\}$$

contains a primal feasible optimal solution satisfying the complementary slackness condition. However, *the optimal solution set of the optimization problem ($D(\lambda)$) can contain more than one element and so it might be computationally difficult to identify this primal feasible element satisfying the complementary slackness condition.* The implication of Theorem 2.5 is that for differentiable K-convex minimization problems satisfying the Slater-type condition the Karush–Kuhn–Tucker vector $\boldsymbol{\lambda}$ completely coincide with the optimal dual variables and so the Karush–Kuhn–Tucker conditions for K-convex programs are actually duality results. These conditions are also known in linear programming as *primal–dual relations*. We show in the next section the relation between Lagrangian relaxation and linear programming relaxation applied to an integer linear programming problem.

2.2 On Integer Linear Programming and Lagrangian and LP Relaxations

In this section we consider for D some $p \times n$ matrix and A some $m \times n$ matrix the integer linear programming problem:

$$\inf\{\mathbf{c}^\top\mathbf{x} : \mathbf{x} \in \mathcal{F}_{INT}\} \qquad\qquad (INT)$$

with

$$\mathcal{F}_{INT} := \{\mathbf{x} \in \mathbb{Z}_+^n : D\mathbf{x} \le \mathbf{d}, A\mathbf{x} = \mathbf{b}\}$$

a nonempty feasible region and apply the Lagrangian relaxation approach to this problem. In the above representation it is assumed that the optimization problem $\inf\{\mathbf{h}^\top\mathbf{x} : A\mathbf{x} = \mathbf{b}, \mathbf{x} \in \mathbb{Z}_+^n\}$ is relatively easy to solve for any $\mathbf{h} \in \mathbb{R}^n$ by means of a polynomially bounded algorithm, while the original problem (INT) is extremely difficult to solve. Due to this assumption, we penalize in optimization problem (INT) the constraint $D\mathbf{x} - \mathbf{d} \le \mathbf{0}$ and so the Lagrangian dual problem

(sometimes called the *partial dual* problem) is given by

$$\sup\{\theta(\boldsymbol{\lambda}) : \boldsymbol{\lambda} \geq \mathbf{0}\} \tag{D}$$

with

$$\theta(\boldsymbol{\lambda}) = -\boldsymbol{\lambda}^\top \mathbf{d} + \inf\{(\mathbf{c} + D^\top \boldsymbol{\lambda})^\top \mathbf{x} : A\mathbf{x} = \mathbf{b}, \mathbf{x} \in \mathbb{Z}_+^n\}. \tag{24}$$

Also, a relaxation of optimization problem (INT) is given by the optimization problem:

$$\inf\{\mathbf{c}^\top \mathbf{x} : D\mathbf{x} \leq \mathbf{d}, \mathbf{x} \in conv(X)\} \tag{LPR}$$

with $conv(X)$ denoting the smallest convex set containing $X = \{\mathbf{x} \in \mathbb{Z}_+^n : A\mathbf{x} = \mathbf{b}\}$. The set conv($X$) (also called the *convex hull* of the set X) can be completely characterized by linear inequalities and so the optimization problem (LPR) is actually a linear programming problem. Unfortunately it is not possible to generate the linear inequality description of the set conv(X) and so optimization problem (LPR) cannot be solved by a linear programming solver. Remember we only used a relaxation of the original problem if this relaxation can be easily solved and for problem (LPR) this is certainly not true. A relaxation of problem (INT) that can be solved by a linear programming solver is given by the standard LP relaxation:

$$\inf\{\mathbf{c}^\top \mathbf{x} : D\mathbf{x} \leq \mathbf{d}, A\mathbf{x} = \mathbf{b}, \mathbf{x} \geq 0\} \tag{SLPR}$$

of optimization problem (INT). In the next result we relate the optimal objective function values of the above optimization problems.

Lemma 2.9 *If $\mathcal{F}_{INT}$ is nonempty, then*

$$\upsilon(SLPR) \leq \upsilon(D) = \upsilon(LPR) \leq \upsilon(INT).$$

Proof The inequalities $\upsilon(SLPR) \leq \upsilon(LPR) \leq \upsilon(INT)$ are obvious and so we only need to verify that $\upsilon(D) = \upsilon(LPR)$. As observed, the optimization problem (LPR) is a linear programming problem and if $\upsilon(LPR)$ is finite, we obtain by strong duality and the observation after Theorem 2.2 that $\upsilon(LPR) = \sup_{\boldsymbol{\lambda} \geq \mathbf{0}}\{\theta(\boldsymbol{\lambda})\}$ with

$$\theta(\boldsymbol{\lambda}) = -\boldsymbol{\lambda}^\mathsf{T}\mathbf{d} + \inf\{(\mathbf{c} + D^\top \boldsymbol{\lambda})^\top \mathbf{x} : \mathbf{x} \in conv(X)\}.$$

Since the function $\mathbf{x} \to (\mathbf{c} + D^\top \boldsymbol{\lambda})^\top \mathbf{x}$ is linear on $conv(X)$, it follows that

$$\inf_{\mathbf{x} \in conv(X)}\{(\mathbf{c} + D^\mathsf{T}\boldsymbol{\lambda})^\mathsf{T}\mathbf{x}\} = \inf_{\mathbf{x} \in X}\{(\mathbf{c} + D^\mathsf{T}\boldsymbol{\lambda})^\mathsf{T}\mathbf{x}\}$$
$$= \inf\{(\mathbf{c} + D^\mathsf{T}\boldsymbol{\lambda})^\mathsf{T}\mathbf{x} : A\mathbf{x} = \mathbf{b}, \mathbf{x} \in \mathbb{Z}_+^n\}$$

and this shows by relation (24) the desired result.   □

In Lemma 2.9 it might happen that the optimal objective function value $\upsilon(D)$ is strictly above $\upsilon(SLPR)$ and in this case we obtain a stronger lower bound on $\upsilon(INT)$. Hence it might be computationally more efficient to compute $\upsilon(D)$ and use this approach in a branch and bound procedure. In general it is computationally easier to compute $\upsilon(SLPR)$ instead of $\upsilon(D)$ but due to the lower bound difference computing $\upsilon(D)$ might be more efficient in branching in a classical branch and bound procedure. However, under some conditions the values $\upsilon(SLPR)$ and $\upsilon(D)$ are equal, and this happens if the optimization problem satisfies the so-called integrability property. Observe in this case all the lower bounds discussed in Lemma 2.9 are the same.

Definition 2.10 The optimization problem $\upsilon(R) = \inf\{\mathbf{h}^\top\mathbf{x} : A\mathbf{x} = \mathbf{b}, \mathbf{x} \in \mathbb{Z}_+^n\}$ satisfies the so-called integrability property if $\upsilon(R) = \inf\{\mathbf{h}^\top\mathbf{x} : A\mathbf{x} = b, \mathbf{x} \geq \mathbf{0}\}$ for every vector $\mathbf{h}$.

A sufficient condition for the integrability property is that the matrix A is totally unimodular (Schrijver (1998)), and these matrices are extremely important in polyhedral combinatorics. For sufficient conditions to check whether a matrix is totally unimodular, consult Schrijver (1998)). These matrices appear a lot in network flow problems. By a similar argument as used in Lemma 2.9, one can now show the following result.

Lemma 2.10 *If $\mathcal{F}_{INT}$ is nonempty and the optimization problem* $\inf\{\mathbf{h}^\top\mathbf{x} : A\mathbf{x} = \mathbf{b}, \mathbf{x} \in \mathbb{Z}_+^n\}$ *satisfies the integrability property, then* $\upsilon(SLPR) = \upsilon(D) = \upsilon(LPR)$.

Proof By penalizing the constraints $\mathbf{D}x \leq \mathbf{d}$ in problem $(SLPR)$ and using the integrability property, we obtain by the strong duality result that $\upsilon(SLPR) = \sup_{\boldsymbol{\lambda} \geq \mathbf{0}}\{\theta(\boldsymbol{\lambda})\}$ with

$$\theta(\boldsymbol{\lambda}) = -\boldsymbol{\lambda}^\top\mathbf{d} + \inf\{(\mathbf{c} + D^\top\boldsymbol{\lambda})^\top\mathbf{x} : A\mathbf{x} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}\}$$
$$= -\boldsymbol{\lambda}^\top\mathbf{d} + \inf\{(\mathbf{c} + D^\top\boldsymbol{\lambda})^\top\mathbf{x} : A\mathbf{x} = \mathbf{b}, \mathbf{x} \in \mathbb{Z}_+^n\}$$

and the desired result follows.   □

By the above observation, it follows that the Lagrangian dual has the same optimal objective function value as the standard linear programming relaxation of (INT) and so it seems useless to solve the Lagrangian dual. However, from a computational point of view it is sometimes more efficient to solve the Lagrangian dual than to use a standard linear programming package to solve the standard linear programming relaxation of (INT). Hence to solve the Lagrangian dual

or the standard linear programming relaxation depends on the problem under consideration. This concludes our discussion of the principle of Lagrange for finite dimensional optimization problems. Until now we only dealt in general with the principle of Lagrange and we did not discuss how to optimize the Lagrangian dual function. This is the topic of the next section.

2.3   On Subgradient Optimization and the Dual Problem

Until now we only dealt with the general structure of a Lagrangian dual problem and we did not discuss how to solve the Lagrangian dual problem $\sup\{\theta(\boldsymbol{\lambda}) : \boldsymbol{\lambda} \in K^*\}$ with K^* a closed convex cone and the Lagrangian dual function $\theta : K^* \to [-\infty, \infty)$ given by

$$\theta(\boldsymbol{\lambda}) = \inf\{f(\mathbf{x}) + \boldsymbol{\lambda}^{\mathsf{T}}\mathbf{g}(\mathbf{x}) : \mathbf{x} \in X\}. \tag{25}$$

Since the set $X \subseteq R^n$ is nonempty, it follows $-\infty \le \theta(\boldsymbol{\lambda}) < \infty$. Moreover, by Lemma 2.1 the function $-\theta$ is convex and

$$-\sup\{\theta(\boldsymbol{\lambda}) : \boldsymbol{\lambda} \in K^*\} = \inf\{-\theta(\boldsymbol{\lambda}) : \boldsymbol{\lambda} \in K^*\}. \tag{$-D$}$$

The optimization problem $(-D)$ has the same optimal solution as optimization problem (D) if an optimal solution exists and without loss of generality we always assume that the set $\mathrm{dom}(-\theta) := \{\theta \in K^* : \theta(\boldsymbol{\lambda}) > -\infty\}$ is nonempty. In most practical cases the nonemptyness of the set $\mathrm{dom}(-\theta)$ can be easily verified (look at location problems!). It is now easy to check that

$$-\infty \le \upsilon(-D) < \infty \Leftrightarrow \mathrm{dom}(-\theta) \text{ is nonempty.}$$

Observe that the function $-\theta : K^* \to (-\infty, \infty]$ is in general not differentiable. This can be easily checked by means of a picture if we assume for simplicity that the set X consists of two elements. Also the optimization problem $(-D)$ is an *unconstrained convex programming problem* in case $K = \{\mathbf{0}\}$ and a *constrained convex programming problem* otherwise. Hence in principle we can apply to this problem the subgradient optimization procedure (Hiriart-Urruty and Lemaréchal, 2013) if it is possible to compute a ϵ-*subgradient* with $\epsilon \ge 0$ of the function $-\theta : K^* \to (-\infty, \infty]$ at every point $\boldsymbol{\lambda} \in K^*$. Observe that the set of ϵ-subgradients of the function $-\theta$ at the point $\boldsymbol{\lambda} \in K^0$ is called the ϵ-*subgradient set* and this set is denoted by $\partial_\epsilon(-\theta)(\boldsymbol{\lambda})$. In case ϵ equals 0 a 0-subgradient can be seen as a generalization of a gradient for a differentiable function.

Definition 2.11 A vector $\lambda_0^* \in \mathbb{R}^n$ is called a ϵ-subgradient for some $\epsilon \geq 0$ of the function $f : K^* \to (-\infty, \infty]$ at the point $\lambda_0 \in K^*$ if for every $\lambda \in K^*$ it follows that

$$f(\lambda) \geq f(\lambda_0) + (\lambda - \lambda_0)^\mathsf{T} \lambda_0^* - \epsilon.$$

It is called a subgradient if $\epsilon = 0$.

In general it is difficult to *compute* a ϵ-subgradient for an arbitrary finite valued convex function. However, for the function $-\theta$ a ϵ-subgradient for this function at a given point $\lambda \in K^*$ is easy to compute by solving optimization problem $(D(\lambda))$. This is shown in the next lemma and can be easily proved.

Lemma 2.11 *If for some $\lambda_0 \in K^*$ the optimization problem:*

$$\theta(\lambda_0) := \inf\{f(\mathbf{x}) + \lambda_0^\mathsf{T} \mathbf{g}(\mathbf{x}) : \mathbf{x} \in X\}$$

has an optimal solution $\mathbf{x}_{opt}(\lambda_0)$, then it follows that $-\mathbf{g}(\mathbf{x}_{opt}(\lambda_0))$ is a subgradient of the function $-\theta$ at λ_0. Moreover, if the feasible solution $\mathbf{x}_\epsilon(\lambda_0)$ is an ϵ-optimal solution of the optimization problem $(D(\lambda_0))$, then the vector $-\mathbf{g}(\mathbf{x}_\epsilon(\lambda_0))$ is a ϵ-subgradient of the function $-\theta$ at λ_0.

Observe that we introduce ϵ-subgradients since in general one cannot perform exact calculations on a computer and so in most cases we actually are calculating a ϵ-optimal solution. Solving the optimization problem $(D(\lambda))$ with $\lambda = \lambda_0$ and obtaining the optimal solution $\mathbf{x}(\lambda_0)$ yield both the value $-\theta(\lambda_0)$ and the affine function $\lambda \to \theta(\lambda_0) - (\lambda - \lambda_0)^\mathsf{T} \mathbf{g}(\mathbf{x}_{opt}(\lambda_0))$. By Lemma 2.11, the above affine function serves as a lower bound on the convex function $-\theta$ on K^*.

We now discuss in detail the subgradient method applied to the optimization problem:

$$\inf\{f(\lambda) : \lambda \in D\} \qquad\qquad (C)$$

with $D \subseteq \mathbb{R}^n$ a closed convex set and $f : \mathbb{R}^n \to (-\infty, \infty]$ with $f(\lambda) < \infty$ for some $\lambda \in D$. By our assumption, it follows that $-\infty \leq \upsilon(C) < \infty$. Moreover, for this problem we assume that for every $\lambda \in D$ a ϵ-subgradient of f at λ can be computed and by Lemma 2.11 this condition is clearly satisfied by our optimization problem $\inf\{-\theta(\lambda) : \lambda \in K^*\}$ selecting an optimal Lagrangian multiplier. To introduce the subgradient method, we observe the following. If λ_{opt} is the unknown optimal solution of optimization problem (C) and we apply an algorithm to find this optimal solution, this algorithm generates a sequence $\{\lambda_n : n \in N\} \subseteq D$. It is *desirable* that at each next step of this algorithm the newly generated point λ_{n+1} is closer to λ_{opt}, then the presently generated point λ_n. This means that the inequality:

$$\| \lambda_{n+1} - \lambda_{opt} \|^2 < \| \lambda_n - \lambda_{opt} \|^2 \qquad\qquad (26)$$

with

$$\| \lambda \|^2 := \sum\nolimits_{i=1}^{n} \lambda_i^2$$

the squared Euclidean norm should hold for our algorithm. To achieve this, suppose that we are after n steps of our iteration procedure at the vector λ_n and some oracle supplies us with a search direction $\mathbf{d}_n$. If the new point λ_{n+1} is given by

$$\lambda_{n+1} := \lambda_n - t_n \mathbf{d}_n$$

for some $t_n > 0$, then it follows that

$$\| \lambda_{n+1} - \lambda_{opt} \|^2 = \| \lambda_n - \lambda_{opt} \|^2 + t_n^2 \| \lambda_{opt} \|^2 + 2 t_n \mathbf{d}_n^\mathsf{T} (\lambda_{\text{opt}} - \lambda_n). \qquad (27)$$

To guarantee that relation (26) holds, it is *necessary* (maybe not sufficient!) to choose the direction $\mathbf{d}_n$ in such a way that

$$\mathbf{d}_n^\mathsf{T} (\lambda_{opt} - \lambda_n) \leq 0. \qquad (28)$$

If $\mathbf{d}_n$ is an ϵ-subgradient of the function f at λ_n for some $\epsilon \geq 0$ (remember for $\epsilon = 0$ the ϵ-subgradient is actually a subgradient!), this implies

$$\mathbf{d}_n^\mathsf{T} (\lambda_{opt} - \lambda_n) \leq f(\lambda_{opt}) - f(\lambda_n) + \epsilon \leq \epsilon \qquad (29)$$

and since it is *necessary* that relation (28) holds it follows by relation (29) that a hopefully good choice of the direction $\mathbf{d}_n$ is selecting a ϵ-subgradient of f at λ_n. By this observation, we obtain the usual *subgradient iteration scheme* given by the following algorithm.

Algorithm 2.1 Subgradient optimization scheme:

$$\lambda_{n+1} := \lambda_n - t_n \lambda_n^* , t_n > 0 \text{ and } \lambda_n^* \text{ a } \epsilon\text{-subgradient of } f \text{ at } \lambda_n .$$

However, using the above formula might create problems for constrained optimization problems. It can happen that the point λ_n belongs to D and λ_{n+1} does not belong to D and so the point λ_{n+1} becomes infeasible. A way to solve this is to use the orthogonal projection of λ_{n+1} onto the closed convex set D. This means that the *modified subgradient iteration scheme* has the form:

Algorithm 2.2 Modified subgradient iteration scheme:

$$\lambda_{n+1} = P_D(\lambda_n - t_n \lambda_n^*), \, t_n > 0 \text{ and } \lambda_n^* \text{ a } \epsilon\text{-subgradient of } f \text{ at } \lambda_n \qquad (30)$$

with

$$P_D(\mathbf{x}_0) := \arg\min\{\|\mathbf{x}_0 - \boldsymbol{\lambda}\|^2 : \boldsymbol{\lambda} \in D\}, \mathbf{x}_0 \text{ fixed vector}$$

and $\arg\min\{Q\}$ denoting the optimal solution of optimization problem (Q).

Since the function $\boldsymbol{\lambda} \to \|\mathbf{x}_0 - \boldsymbol{\lambda}\|^2$ for $\mathbf{x}_0$ fixed is strictly convex and D is a closed nonempty convex set, there exists a *unique optimal solution* and so the vector $P_D(\mathbf{x}_0)$ is well-defined or equivalently $\mathbf{x} \to P_D(\mathbf{x})$ represents a function. Moreover, we mention without proof (check first-order conditions of the above constrained optimization problem!) that the function $\mathbf{x} \to P_D(\mathbf{x})$ is a so-called contraction and this means that

$$\| P_D(\mathbf{x}_0) - P_D(\mathbf{x}_1) \| \leq \| \mathbf{x}_0 - \mathbf{x}_1 \| \tag{31}$$

for every $\mathbf{x}_0$ and $\mathbf{x}_1$ belonging to $\mathbb{R}^n$. For our primal optimization problem, it almost always follows that $K = \mathbb{R}^m_+ \times \mathbb{R}^{n-m}$ and so $K^* = \mathbb{R}^m_+ \times \{\mathbf{0}\}$. This means that we can solve the projection by hand (make a picture!). It is now possible to show the following fundamental inequality (Correa and Lemaréchal , 1993).

Lemma 2.12 *If $\boldsymbol{\lambda} \in D$ and $\boldsymbol{\lambda}_n^*$ is an ϵ_n-subgradient of the function f at the present iteration point $\boldsymbol{\lambda}_n$ for some $\epsilon_n \geq 0$ and $\boldsymbol{\lambda}_{n+1} = P_D(\boldsymbol{\lambda}_n - t_n \boldsymbol{\lambda}_n^*)$, then it follows that*

$$\| \boldsymbol{\lambda}_{n+1} - \boldsymbol{\lambda} \|^2 \leq \| \boldsymbol{\lambda}_n - \boldsymbol{\lambda} \|^2 + t_n^2 \| \boldsymbol{\lambda}_n^* \|^2 + 2t_n(f(\boldsymbol{\lambda}) - f(\boldsymbol{\lambda}_n) + \epsilon_n).$$

Proof Since obviously $P_D(\boldsymbol{\lambda}) = \boldsymbol{\lambda}$ for $\boldsymbol{\lambda} \in D$, it follows by relation (30) that

$$\| \boldsymbol{\lambda}_{n+1} - \boldsymbol{\lambda} \|^2 = \| P_D(\boldsymbol{\lambda}_n - t_n \boldsymbol{\lambda}_n^*) - P_D(\boldsymbol{\lambda}) \|^2$$

and this implies by relation (31) that

$$\| \boldsymbol{\lambda}_{n+1} - \boldsymbol{\lambda} \|^2 \leq \|\boldsymbol{\lambda}_n - \boldsymbol{\lambda} \|^2 + t_n^2 \| \boldsymbol{\lambda}_n^* \|^2 + 2t_n(\boldsymbol{\lambda} - \boldsymbol{\lambda}_n)^\mathsf{T}\boldsymbol{\lambda}_n^*. \tag{32}$$

Since $\boldsymbol{\lambda}_n^*$ is an ϵ_n-subgradient of the function f at $\boldsymbol{\lambda}_n$, we obtain that $f(\boldsymbol{\lambda}) - f(\boldsymbol{\lambda}_n) \geq (\boldsymbol{\lambda} - \boldsymbol{\lambda}_n)^\mathsf{T}\boldsymbol{\lambda}_n^* - \epsilon_n$ and this implies by relation (32) that

$$\| \boldsymbol{\lambda}_{n+1} - \boldsymbol{\lambda} \|^2 \leq \| \boldsymbol{\lambda}_n - \boldsymbol{\lambda} \|^2 + t_n^2 \| \boldsymbol{\lambda}_n^* \|^2 + 2t_n(f(\boldsymbol{\lambda}) - f(\boldsymbol{\lambda}_n) + \epsilon_n)$$

showing the desired inequality.   □

By the above inequality, the convergence of our subgradient optimization scheme is easily proved and this is shown by the following result.

Theorem 2.6 *If $\lim_{n\uparrow\infty} \epsilon_n = 0$ and the positive sequence $\{t_n : n \in \mathbb{N}\}$ satisfies*

$$\sum_{n=1}^{\infty} t_n = \infty \text{ and } \lim_{n\uparrow\infty} t_n \| \boldsymbol{\lambda}_n^* \|^2 = 0,$$

then it follows that

$$\lim_{n\uparrow\infty} m_n = \upsilon(C) \geq -\infty$$

with

$$m_n := \min\{f(\lambda_k) : k \leq n\}.$$

Proof Clearly the sequence $m_n, n \in N$ is decreasing and so it has a limit c. If the limit c equals $-\infty$, the result is proved and so we assume that the limit c is finite. If $c > \upsilon(C)$, one can find by the definition of an infimum some $\delta > 0$ and $\lambda_e \in D$ satisfying

$$\upsilon(C) \leq f(\lambda_e) \leq c - \delta \leq f(\lambda_n) - \delta \qquad (33)$$

for every $n \in \mathbb{N}$. Hence by Lemma 2.12 with $\lambda = \lambda_e$, we obtain that

$$\| \lambda_{n+1} - \lambda_e \|^2 \leq \| \lambda_n - \lambda_e \|^2 + t_n^2 \| \lambda_n^* \|^2 + 2t_n(f(\lambda_e) - f(\lambda_n) + \epsilon_n).$$

This implies by relation (33) that

$$\| \lambda_{n+1} - \lambda_e \|^2 \leq \| \lambda_n - \lambda_e \|^2 + t_n(t_n \| \lambda_n^* \|^2 + 2(\epsilon_n - \delta)).$$

Since we assume that

$$\lim_{n\uparrow\infty} t_n \| \lambda_n^* \|^2 = 0 \text{ and } \lim_{n\uparrow\infty} \epsilon_n = 0,$$

one can find some $n_0 \in N$ such that for every $n \geq n_0$ it follows that

$$t_n \| \lambda_n^* \|^2 + 2(\epsilon_n - \delta) \leq -\delta$$

and this yields for every $n \geq n_0$ that

$$\| \lambda_{n+1} - \lambda_e \|^2 \leq \| \lambda_n - \lambda_e \|^2 - \delta t_n. \qquad (34)$$

Hence we obtain by relation (34) that

$$0 \leq \| \lambda_m - \lambda_e \|^2 \leq \| \lambda_{n_0} - \lambda_e \|^2 - \delta \sum_{k=n_0}^{m-1} t_k$$

for every $m \geq n_0$ and this implies by the assumption $\sum_{k=n_0}^{\infty} t_k = \infty$ that

$$0 \leq \lim_{m\uparrow\infty} \| \lambda_m - \lambda_e \|^2 = -\infty.$$

Hence we obtain a contradiction and so it must follow that $c = \upsilon(C)$. $\square$

Since it can happen that an optimal solution of optimization problem (C) does not exist, we cannot show that the sequence λ_n generated by the subgradient method converges to the optimal solution of (C). In general it also does not hold that the sequence $f(\lambda_n)$ is decreasing. Although we have shown under fairly general conditions that the sequence m_n converges to $v(C)$, it is very difficult to give a *proper stopping rule* for the subgradient method. Also in case we apply the subgradient method to optimization problem ($-D$), the selection of the step sizes $t_n, n \in N$ is a matter of trial and error (Nemhauser and Wolsey, 1988) and so devising a proper scheme is an art in itself. For more details on these practical issues, the reader is referred to Beasley et al. (1993). We finally consider one theoretical case in which one knows in advance that the current iteration point is indeed optimal. This result is presented in the next lemma.

Lemma 2.13 *If at the $(n + 1)th$ step it follows that*

$$\lambda_{n+1} := P_D(\lambda_n - t_n\lambda_n^*)$$

equals λ_n with λ_n^* a ϵ_n-subgradient of the function f at λ_n, then the vector λ_n is an ϵ_n-optimal solution of optimization problem (C).

Proof We only give a proof of this result in case $\lambda_n - t_n\lambda_n^*$ belongs to D. If this holds, it follows since $t_n > 0$ that

$$\lambda_{n+1} := P_D(\lambda_n - t_n\lambda_n^*) = \lambda_n - t_n\lambda_n^*$$

and so $\lambda_n^* = \mathbf{0}$. Since λ_n^* is an ϵ_n-subgradient, this implies that

$$f(\lambda) \geq f(\lambda_n) - \epsilon_n$$

for every $\lambda \in D$ and the desired result follows.   □

For more theoretical results regarding the subgradient method, the reader is referred to Shor (2012) and Correa and Lemaréchal (1993). In the next section we apply the above results to derive Lagrangian heuristics for location and transportation problems.

3   Application of the Principle of Lagrange to Problems in Transportation and Location Theory

In this section we consider two classical problems in location and transportation theory and apply to these problems the Lagrangian relaxation approach. In the first subsection we discuss the well-known set covering problem and in the second subsection several versions of the capacitated facility location problem.

3.1 The Set Covering Problem

In this subsection we formulate the set covering problem. Observe that this problem shows up in a lot of different applications within computer and management science. To give a particular example in transportation, consider a company that needs to serve every day m different customers given by the set $\mathcal{M} = \{1, \ldots, m\}$. To serve these customers, the company selects from a given set $\mathcal{N} = \{1, \ldots, n\}$ of known routes a subset of these routes covering all present day customers. Each route $j \in \mathcal{N}$, has a known cost $c_j > 0$ and so the company wants to select the subset that serves all customers on a given day and has minimal cost. Selecting a route $j \in \mathcal{N}$ for today is a yes\no decision and so we introduce for each route $j \in \mathcal{N}$ the decision variable

$$x_j = \begin{cases} 1 & \text{if route } j \text{ is selected} \\ 0 & \text{otherwise} \end{cases}.$$

To model that every customer who needs to be served today is at least covered by one selected route, we construct the $m \times n$ matrix $A = (a_{ij})$, given by

$$a_{ij} = \begin{cases} 1 & \text{if customer } i \text{ is on route } j \\ 0 & \text{otherwise} \end{cases}.$$

Clearly the value $\sum_{j \in \mathcal{N}} a_{ij} x_j, i \in \mathcal{M}$ yields the number of times customer $i \in \mathcal{M}$ is on the subset of selected routes and we need to solve the optimization problem:

$$\inf \left\{ \sum_{j \in \mathcal{N}} c_j x_j : \sum_{j \in \mathcal{N}} a_{ij} x_j \geq 1, i \in \mathcal{M}, x_j \in \mathbb{B}, j \in \mathcal{N} \right\} \qquad \text{(SP)}$$

with $\mathbb{B} = \{0, 1\}$. Hence the set covering problem (SP) is given by

$$\inf \{ \sum_{j \in \mathcal{N}} c_j x_j : \mathbf{g}(\mathbf{x}) \preceq_K \mathbf{0}, \mathbf{x} \in \mathbb{B}^n \}$$

with $\mathbf{g}(\mathbf{x}) = (g_1(\mathbf{x}), \ldots, g_m(\mathbf{x}))$, $g_i(\mathbf{x}) = 1 - \sum_{j \in \mathcal{N}} a_{ij} x_j$ and $K = \mathbb{R}_+^m$. The feasible set contains only a finite number of elements bounded above by 2^n and so the set covering problem has an optimal solution. It is well-known that the set covering problem is $\mathcal{NP}$-hard (Garey and Johnson, 1979) and so by present day belief it seems very unlikely that there exists a polynomially bounded algorithm to solve this problem. To analyze this problem by means of the Lagrangian relaxation approach, we penalize the constraints $\sum_{j \in \mathcal{N}} a_{ij} x_j \geq 1, i \in \mathcal{M}$ and construct the Lagrangian dual function $\theta : \mathbb{R}_+^m \to \mathbb{R}$ given by

$$\theta(\boldsymbol{\lambda}) = \min \left\{ \sum_{j \in \mathcal{N}} c_j x_j + \sum_{i \in \mathcal{M}} \lambda_i (1 - \sum_{j \in \mathcal{N}} a_{ij} x_j) : \mathbf{x} \in \mathbb{B}^n \right\}.$$

After some easy calculations, it follows that

$$\theta(\boldsymbol{\lambda}) = \sum_{i \in \mathcal{M}} \lambda_i + \sum_{j \in \mathcal{N}} \min \left\{ c_j - \sum_{i \in \mathcal{M}} a_{ij} \lambda_i, 0 \right\} \tag{35}$$

and the optimal solution $\mathbf{x}(\boldsymbol{\lambda}) = (x_1(\boldsymbol{\lambda}), \ldots, x_n(\boldsymbol{\lambda}))$ is given by

$$x_j(\boldsymbol{\lambda}) = 1_{S_j}(\boldsymbol{\lambda}) \tag{36}$$

with $S_j = \{\boldsymbol{\lambda} \in \mathbb{R}_+^m : c_j \leq \sum_{i \in \mathcal{M}} a_{ij} \lambda_i\}$, $j \in \mathcal{N}$. We now apply the subgradient method to solve the Lagrangian dual problem

$$\upsilon(D) = \min_{\lambda \geq \mathbf{0}} \left\{ \sum_{i \in \mathcal{M}} \lambda_i + \sum_{j \in \mathcal{N}} \min\{c_j - \sum_{i \in \mathcal{M}} a_{ij} \lambda_i, 0\} \right\}$$

and obtain as an optimal dual solution a vector $\boldsymbol{\lambda}_{opt}$ and construct the solution $\mathbf{x}(\boldsymbol{\lambda}_{opt})$ given in relation (36). By Lemma 2.9, this dual problem satisfies

$$\upsilon(D) = \min \left\{ \sum_{j \in \mathcal{N}} c_j x_j : \sum_{j \in \mathcal{N}} a_{ij} x_j \geq 1, i \in \mathcal{M}, 0 \leq x_j \leq 1, j \in \mathcal{N} \right\}.$$

If the solution $\mathbf{x}(\boldsymbol{\lambda}_{opt})$ is primal feasible and it satisfies the complementary slackness conditions

$$\lambda_i (1 - \sum_{j \in \mathcal{N}} a_{ij} x_j(\boldsymbol{\lambda}_{opt})) = 0, i \in \mathcal{M},$$

then it follows by Lemma 2.5 that this solution is an optimal primal solution. If the solution $\mathbf{x}(\boldsymbol{\lambda}_{opt})$ is only primal feasible, we have found a feasible solution for our problem and so $\theta(\boldsymbol{\lambda}_{opt}) \leq \upsilon(SP) \leq \sum_{j \in \mathcal{N}} c_j x_j(\boldsymbol{\lambda}_{opt})$. If the solution is not primal feasible, then we use the following Lagrangian heuristic to convert the solution $\mathbf{x}(\boldsymbol{\lambda}_{opt})$ into a primal feasible solution. Consider the set of routes selected by our solution $\mathbf{x}(\boldsymbol{\lambda}_{opt})$ given by

$$R(\boldsymbol{\lambda}_{opt}) = \{1 \leq j \leq n : x_j(\boldsymbol{\lambda}_{opt}) = 1\}.$$

Starting with the initial set $R(\boldsymbol{\lambda}_{opt})$, we check for each customer $i \in \mathcal{M}$ whether this customer is visited by one of the routes in the presently selected set. If not, we select the cheapest possible route covering this customer and add this route to this presently selected set of routes. After evaluating all the customers, we have constructed a set of routes that covers all customers. Some customers might be assigned to different routes and we delete them from all the routes except one. Observe that this heuristic is applied in a branch and bound procedure and at each branch we solve a special case of the set covering problem with some routes already selected. This concludes our discussion of the set covering problem. In the next section we consider the (un)capacitated facility location problem.

3.2   Fixed Charged Location Models on Discrete Spaces

To introduce the so-called fixed charged location problems in discrete location, we denote as before the set of customers by $\mathcal{M} = \{1, \ldots, m\}$ and the set of possible sites of facilities by $\mathcal{N} = \{1, \ldots, n\}$. We denote by $f_j, j \in \mathcal{N}$ the fixed setup cost of site j, while c_{ij} denotes the transportation costs of supplying the demand of customer $i \in \mathcal{M}$ by site $j \in \mathcal{N}$. The *uncapacitated facility location problem* tries to determine a subset of the set $\mathcal{N}$ of possible sites in such a way that the total transportation and setup costs are minimized. In this problem the demand of all customers should be satisfied. Moreover, each open facility has unlimited capacity. A more difficult and related problem is given by the capacitated version of this problem. In this problem each site j has a fixed capacity q_j. For the capacitated facility location problem, there actually exist two versions. In the first version the total demand of each customer can only be supplied by one facility and this problem is called the *single source capacitated facility location* problem, while in the other version the total demand of each customer can be supplied by more than one facility. This problem is called the *multiple source capacitated facility location* problem. To formulate an optimization problem for the uncapacitated version, we introduce the binary decision vector $\mathbf{y} = (y_j)_{j \in \mathcal{N}}$ given by

$$y_j = 1 \Leftrightarrow \text{facility } j \text{ is opened at site } j.$$

Since in the uncapacitated facility location problem the capacity of each facility is unlimited, we assign each customer to the "cheapest" open facility and this implies that the model can be represented by the optimization problem:

$$\inf \left\{ \sum_{i \in \mathcal{M}} \min\{c_{ij} : y_j = 1\} + \sum_{j \in \mathcal{N}} f_j y_j : \mathbf{y} \in \mathbb{B}^n \right\}..$$

However, in this optimization problem the objective function is not linear and to linearize the objective function we introduce the decision variables $\mathbf{x} = (x_{ij})_{i \in \mathcal{M}, j \in \mathcal{N}}$ given by

$$x_{ij} = \text{fraction of demand of customer } i \in \mathcal{M} \text{ supplied by site } j \in \mathcal{N}.$$

Using these decision variables, the objective function is given by

$$\sum_{i \in \mathcal{M}} \sum_{j \in \mathcal{N}} c_{ij} x_{ij} + \sum_{j \in \mathcal{N}} f_j y_j.$$

Since all demand should be delivered, it is clear that these decision variables need to satisfy the so-called *assignment constraints* $\sum_{j \in \mathcal{N}} x_{ij} = 1$ for every $i \in \mathcal{M}$ and so the feasible region $\mathcal{F}_{UFL}$ of the uncapacitated facility location problem is given by

$$\mathcal{F}_{UFL} = \{(\mathbf{x}, \mathbf{y}) : 0 \leq x_{ij} \leq y_j, \sum_{j \in \mathcal{N}} x_{ij} = 1, y_j \in \mathbb{B}, i \in \mathcal{M}, j \in \mathcal{N}\}. \tag{37}$$

Hence the optimization problem for the uncapacitated facility location problem is given by

$$\inf \left\{ \sum_{i \in \mathcal{M}} \sum_{j \in \mathcal{N}} c_{ij} x_{ij} + \sum_{j \in \mathcal{N}} f_j y_j : (\mathbf{x}, \mathbf{y}) \in \mathcal{F}_{UFL} \right\}. \tag{UFL}$$

If we consider the multiple-source capacitated facility location problem, we need to add the capacity constraints $\sum_{i \in \mathcal{M}} d_i x_{ij} \leq q_j y_j$ for every $j \in \mathcal{N}$ with d_i denoting the total demand of customer $i \in \mathcal{M}$. Hence the feasible region $\mathcal{F}_{MSCFL}$ of the multiple-source capacitated facility location problem is given by

$$\mathcal{F}_{MSCFL} = \mathcal{F}_{UFL} \cap \left\{ (\mathbf{x}, \mathbf{y}) : \sum_{i \in \mathcal{M}} d_i x_{ij} \leq q_j y_j, j \in \mathcal{N} \right\}$$

and we need to solve the optimization problem:

$$\inf \left\{ \sum_{i \in \mathcal{M}} \sum_{j \in \mathcal{N}} c_{ij} x_{ij} + \sum_{j \in \mathcal{N}} f_j y_j : (\mathbf{x}, \mathbf{y}) \in \mathcal{F}_{MSCFL} \right\}. \tag{$MSCFL$}$$

If we consider the single source capacitated facility location problem, each customer can only be supplied by exactly one open facility and this implies that the decision variables $x_{ij} \in \mathbb{B}$. Hence the feasible region $\mathcal{F}_{SSCFL}$ is given by

$$\mathcal{F}_{SSCFL} = \mathcal{F}_{UFL} \cap \left\{ (\mathbf{x}, \mathbf{y}) : \sum_{i \in \mathcal{M}} d_i x_{ij} \leq q_j y_j, x_{ij} \in \mathbb{B}, i \in \mathcal{N}, j \in \mathcal{M} \right\}$$

and the optimization problem is given by

$$\inf \left\{ \sum_{i \in \mathcal{M}} \sum_{j \in \mathcal{N}} c_{ij} x_{ij} + \sum_{j \in \mathcal{N}} f_j y_j : (\mathbf{x}, \mathbf{y}) \in \mathcal{F}_{SSCFL} \right\}. \tag{$SSCFL$}$$

For an extensive overview of transportation-location models on discrete spaces, the reader should consult Daskin (2011) or Francis, McGinnis and White (1992). It is well-known that the uncapacitated facility location problem (UFL) is $\mathcal{NP}$-hard (Garey and Johnson, 1979) and as already observed represented by

$$\inf \left\{ \sum_{i \in \mathcal{M}} \sum_{j \in \mathcal{N}} c_{ij} x_{ij} + \sum_{j \in \mathcal{N}} f_j y_j : \mathbf{g}(\mathbf{x}) \preceq_K \mathbf{0}, 0 \leq x_{ij} \leq y_j, \right.$$

$$\left. i \in \mathcal{N}, j \in \mathcal{M}, \mathbf{y} \in \mathbb{B}^n \right\}$$

with

$$\mathbf{g}(\mathbf{x}) = (g_1(\mathbf{x}), \ldots, g_m(\mathbf{x})), \, g_i(\mathbf{x}) = \sum_{j \in \mathcal{N}} x_{ij} - 1$$

and $K = \{\mathbf{0}\} \subset \mathbb{R}^m$. To construct the Lagrangian dual function of the uncapacitated facility location model, we first observe for $f_j = 0$ that we always open a facility at site j and so without loss of generality we may assume that $f_j > 0$. In the uncapacitated facility location model we penalize the assignment constraints $\sum_{j \in \mathcal{N}} x_{ij} - 1 = 0$ for every $i \in \mathcal{M}$. Since $K^* = \{\mathbf{0}\}^* = \mathbb{R}^m$, the Lagrangian dual function $\theta : \mathbb{R}^m \to [-\infty, \infty)$ is defined on $\mathbb{R}^m$ and it is given by

$$\theta(\boldsymbol{\lambda}) = -\sum_{i \in \mathcal{M}} \lambda_i + \sum_{j \in \mathcal{N}} \inf \left\{ \sum_{i \in \mathcal{M}} (c_{ij} + \lambda_i) x_{ij} + f_j y_j : (\mathbf{x}, \mathbf{y}) \in X_j \right\} \tag{38}$$

with $X_j := \{(\mathbf{x}, \mathbf{y}) : 0 \leq x_{ij} \leq y_j, i \in \mathcal{M}, y_j \in \mathbb{B}\}$, $j \in \mathcal{N}$. To solve the minimization problem:

$$\inf \left\{ \sum_{i \in \mathcal{M}} (c_{ij} + \lambda_i) x_{ij} + f_j y_j : (\mathbf{x}, \mathbf{y}) \in X_j \right\} \tag{H_j}$$

for a given $j \in \mathcal{N}$, we take x_{ij} as large (small) as possible in case $c_{ij} + \lambda_i$ is negative (positive). This implies that an optimal solution $(\mathbf{x}(\boldsymbol{\lambda}), \mathbf{y}(\boldsymbol{\lambda}))$ of optimization problem (H_j) is given by

$$x_{ij}(\boldsymbol{\lambda}) = \begin{cases} y_j(\boldsymbol{\lambda}) & \text{if } c_{ij} + \lambda_i < 0 \\ 0 & \text{if } c_{ij} + \lambda_i \geq 0 \end{cases}. \tag{39}$$

By relation (39), it follows that the decision variables x_{ij}, $i \in \mathcal{M}$, $j \in \mathcal{N}$ can be eliminated and we obtain

$$\inf \left\{ \sum_{i \in \mathcal{M}} (c_{ij} + \lambda_i) x_{ij} + f_j y_j : (\mathbf{x}, \mathbf{y}) \in X_j \right\} = \inf\{(f_j + b_j(\boldsymbol{\lambda})) y_j : y_j \in \mathbb{B}\} \tag{40}$$

with $b_j : \mathbb{R}^m \to (-\infty, \infty)$ given by

$$b_j(\boldsymbol{\lambda}) := \sum_{i \in \mathcal{M}} \min\{c_{ij} + \lambda_i, 0\}. \tag{41}$$

To determine the optimal objective value of problem (H_j), we finally observe that

$$\inf\{(f_j + b_j(\boldsymbol{\lambda})) y_j : y_j \in \mathbb{B}\} = \min\{f_j + b_j(\boldsymbol{\lambda}), 0\} \tag{42}$$

with optimal solution $\mathbf{y}(\lambda) = (y_j(\lambda))_{j \in \mathcal{N}}$ given by

$$
y_j(\lambda) = \begin{cases} 1 & \text{if } f_j + b_j(\lambda) \leq 0 \\ 0 & \text{if } f_j + b_j(\lambda) > 0 \end{cases}. \tag{43}
$$

Hence the Lagrangian dual function $\theta : \mathbb{R}^m \to \mathbb{R}$ of the uncapacitated facility location problem is given by

$$
\theta(\lambda) = -\sum_{i \in \mathcal{N}} \lambda_i + \sum_{j \in \mathcal{M}} \min\{f_j + b_j(\lambda), 0\}
$$

and its Lagrangian dual by

$$
\sup\{\theta(\lambda) : \lambda \in \mathbb{R}^n\}.
$$

Applying the subgradient method to $\inf\{-\theta(\lambda) : \lambda \in \mathbb{R}^m\}$, we determine the optimal Lagrangian multiplier λ_{opt}. In case the optimal solution $(\mathbf{x}(\lambda_{opt}), \mathbf{y}(\lambda_{opt}))$ given by relations (39) and (43) is also primal feasible, we obtain since $K = \{\mathbf{0}\}$ that the complementary slackness conditions are automatically satisfied. This implies by Lemma 2.5 that this solution is an optimal solution of the uncapacitated facility location problem. However, in most cases this solution is not primal feasible and this means that either no facilities are opened or there exists a customer whose demand is not satisfied. The following Lagrangian heuristic now transforms the solution $(\mathbf{x}(\lambda_{opt}), \mathbf{y}(\lambda_{opt}))$ into a primal feasible solution.

Algorithm 3.1 Lagrangian heuristic:

1. Consider the possibly empty set $S_1(\lambda_{opt})$ of open facilities given by

$$
S_1(\lambda_{opt}) := \{j \in \mathcal{N} : y_j(\lambda_{opt}) = 1\}
$$

   and go to step 2.
2. If this set is empty, open a facility and assign all customers to this facility. If this set is nonempty, consider all customers "overassigned" or not assigned and make them nonassigned. Assign now these nonassigned customers to the cheapest facility belonging to the set $S_1(\lambda_{opt})$.

As for the set covering problem, we need to solve in a branch and bound procedure at each branch a special instance of the uncapacitated facility location problem with some facilities already opened and or some customers already assigned to some open facilities. A more efficient and special purpose procedure to solve the uncapacitated facility location problem starts with the observation that for positive setup costs we may replace the decision variables $y_j \in \mathbb{B}$ in problem (UFL) by $y_j \in \mathbb{Z}_+$ without changing the set of optimal solutions and the optimal objective

value. This means that we replace the feasible region $\mathcal{F}_{UFL}$ by the bigger set $\mathcal{F}_{RUFL}$ given by

$$\mathcal{F}_{RUFL} := \{(\mathbf{x}, \mathbf{y}) : 0 \leq x_{ij} \leq y_j, \sum\nolimits_{j \in \mathcal{N}} x_{ij} = 1, y_j \in \mathbb{Z}_+\},$$

and so we obtain the relaxation:

$$\inf \left\{ \sum\nolimits_{i \in \mathcal{M}} \sum\nolimits_{j \in \mathcal{N}} c_{ij} x_{ij} + \sum\nolimits_{j \in \mathcal{M}} f_j y_j : (\mathbf{x}, \mathbf{y}) \in \mathcal{F}_{RUFL} \right\}. \qquad (RUFL)$$

As already observed, it follows that $\upsilon(UFL) = \upsilon(RUFL)$ and the optimal solutions of both optimization are the same. Penalizing now the assignment constraints in the above optimization problem ($RUFL$), it follows by the same procedure as used for the uncapacitated facility location problem that the Lagrangian dual problem is given by

$$\upsilon(D) = \sup \left\{ -\sum\nolimits_{i \in \mathcal{M}} \lambda_i : f_j + b_j(\boldsymbol{\lambda}) \geq 0, j \in \mathcal{M}, \boldsymbol{\lambda} \in \mathbb{R}^m \right\}.$$

This Lagrangian dual problem is also called the *condensed dual*. There exists now a special purpose heuristic called DUALOC (Erlenkotter, 1978) to solve the above dual problem and this is surprisingly effective. Finally we look at the multisource capacitated version of the facility location problem. Again penalizing the assignment constraints, we obtain by the same approach as for the uncapacitated version that the Lagrangian dual function $\theta : \mathbb{R}^m \to \mathbb{R}$ is given by

$$\theta(\boldsymbol{\lambda}) = -\sum\nolimits_{i \in \mathcal{N}} \lambda_i + \sum\nolimits_{j \in \mathcal{M}} \min\{0, f_j + h_j(\boldsymbol{\lambda})\}$$

with

$$\begin{aligned} h_j(\boldsymbol{\lambda}) &= \inf \left\{ \sum\nolimits_{i \in \mathcal{M}} (c_{ij} + \lambda_j) x_{ij} : 0 \leq x_{ij} \leq 1, \sum\nolimits_{i \in \mathcal{M}} d_i x_{ij} \leq q_j, j \in \mathcal{N} \right\} \\ &= -\sup \{ \sum\nolimits_{i \in \mathcal{M}} (-c_{ij} - \lambda_j) x_{ij} : 0 \leq x_{ij} \leq 1, \sum\nolimits_{i \in \mathcal{M}} d_i x_{ij} \leq q_j, j \in \mathcal{N} \}. \end{aligned}$$

The last problem is a LP-relaxation of the knapsack problem. Without loss of generality, we may now assume that $-c_{ij} - \lambda_j > 0$ for every $i \in \mathcal{M}$ (if not set $x_{ij} = 0$ and reduce problem!). To solve this problem, it is well-known that we need to order the ratios $\frac{-c_{ij} - \lambda_j}{d_i}$ in decreasing order given by

$$\frac{-c_{\pi(1)j} - \lambda_j}{d_{\pi(1)}} \geq \frac{-c_{\pi(2)j} - \lambda_j}{d_{\pi(2)}} \geq \dots \geq \frac{-c_{\pi(m)j} - \lambda_j}{d_{\pi(m)}}$$

for some permutation π and start assigning the demand of customers $\pi(1), \pi(2), \ldots, \pi(n)$ to location j until the capacity of this location is satisfied. The last customer assigned might only get a part of his demand delivered from that facility j. Again we can easily construct as for the uncapacitated version a

Lagrangian heuristic to convert the optimal solution of $\theta(\lambda)$ to a primal feasible solution. Observe that for the single source capacitated facility location problem we obtain the classical knapsack problem. Although this problem is $\mathcal{NP}$-hard, there are fast heuristics solving this problem approximately (Martello, 1990). This ends our discussion of fixed charged location problems on discrete spaces.

4 Conclusion

In this study, the Lagrangian relaxation approach and the main ideas behind this approach are discussed in detail for finite dimensional optimization problems with continuous and/or discrete decision variables. Contrary to many courses taught in graduate programs distinguishing between continuous and discrete optimization problems, we start in this chapter with a generic description of an optimization problem incorporating both type of problems and apply to that optimization problem the Lagrangian relaxation approach. By introducing the main ideas behind this approach, we identify the so-called Lagrangian dual problem and a class of continuous optimization problems for which the optimal objective value of the primal problem equals the optimal objective value of the Lagrangian dual problem. At the same time, we show that many important results much later developed in the field of Operational Research are special instances of the Lagrangian relaxation approach. The most important ones are the primal–dual relations in linear programming and the dual linear program and the Karush–Kuhn–Tucker conditions in nonlinear programming. As such this chapter does not contain any new results but tries to give an easy introduction to duality theory and its use in finite dimensional optimization problems for less mathematically oriented readers using the simplest possible proofs. Since many models in Operational Research can be formulated as optimization problems and the Lagrangian relaxation technique tries to find solutions of these models, a basic understanding of this technique is of importance to more application oriented researchers in this field. Also it is important to understand that seemingly different algorithms for solving certain problems are actually based on the same ideas. As an example (although not discussed in this chapter), we mention the Dantzig–Wolfe decomposition technique in linear programming and the dual version of it called Benders method. To illustrate its use to some specific examples, we showed in the last part of this chapter how these techniques can be applied in generating approximative solutions for the classical set covering and fixed charged facility location models.

References

Bazaraa, M. S., Jarvis, J. J., & Sherali, H. D. (1990). *Linear programming and network flows*. New York: John Wiley and Sons.

Bazaraa, M. S., Sherali, H. D., & Shetty, C. M. (1993). *Nonlinear programming: theory and algorithms*. New York: John Wiley and Sons.

Beasley, J., Dowsland, K., Glover, F., Laguna, M., Peterson, C., Reeves, C. R., & Soderberg, B. (1993). *Modern heuristic techniques for Combinatorial Problems*. New York: John Wiley and Sons.

Ben-Tal, A., & Nemirovski, A. (2001), *Lectures on modern convex optimization: analysis, algorithms and engineering applications*. Philadelphia: SIAM.

Boyd, S., & Vandenberghe, L. (2004). *Convex optimization*. Cambridge: Cambridge University Press.

Choquet, G. (1976). *Lectures on analysis volume II: representation theory*. London: W.A. Benjamin.

Chvatal, V. (1983). *Linear programming*. New York: Macmillan Company.

Correa, R., & Lemaréchal, C. (1993). Convergence of some algorithms for convex minimization. *Mathematical Programming, 62*(1), 261–275.

Daskin, M. S. (2011). *Network and discrete location: models, algorithms, and applications*. New York: John Wiley and Sons.

Erlenkotter, D. (1978). A dual-based procedure for uncapacitated facility location. *Operations Research, 26*(6), 992–1009.

Everett III, H. (1963). Generalized Lagrange multiplier method for solving problems of optimum allocation of resources. *Operations Research, 11*(3), 399–417.

Fiacco, A. V., & McCormick, G. P. (1969). *Nonlinear programming: sequential unconstrained minimization techniques*. New York: Wiley.

Francis, R. L., McGinnis, L. F., & White, J. A. (1992). *Facility Layout and Location: an analytical approach*. Pearson College Division.

Frenk, J. B. G., & Kassay, G. (1999). On classes of generalized convex functions, Gordan–Farkas type theorems, and Lagrangian duality. *Journal of Optimization Theory and Applications, 102*(2), 315–343.

Frenk, J. B. G., & Kassay, G. (2005). Introduction to convex and quasiconvex analysis. In *Handbook of generalized convexity and generalized monotonicity* (pp. 3–87). New York: Springer.

Frenk, J. B. G., & Kassay, G. (2006). The level set method of Joo and its use in minimax theory. *Mathematical Programming, 105*(1), 145–155.

Frenk, J. B. G., & Kassay, G. (2007). Lagrangian duality and cone convexlike functions. *Journal of Optimization Theory and Applications, 134*(2), 207–222.

Frenk, J. B. G., & Kassay, G. (2008). On noncooperative games, minimax theorems and equilibrium problems. In A. Chinchuluum, P. M. Pardalos, A. Migdalas & L. Pitsoulis (Eds.) *On Noncooperative Games, Minimax Theorems and Equilibrium Problems.* (pp. 211–250). New York: Springer.

Frenk, J. B. G., Kas, P., & Kassay, G. (2007). On linear programming duality and necessary and sufficient conditions in minimax theory. *Journal of Optimization Theory and Applications, 132*, 423–439.

Garey, M. R., & Johnson, D. S. (1979). *Computers and intractability, a guide to the theory of NP completeness*. San Francisco: W.F Freeman.

Giannessi, F. (1984). Theorems of the alternative and optimality conditions. *Journal of Optimization Theory and Applications, 42*(3), 331–365.

Hiriart-Urruty, J. B., & Lemaréchal, C. (2013). *Convex analysis and minimization algorithms I: fundamentals*. New York: Springer Verlag.

Lagrange, J. L. (1806). *Leçons sur le calcul des fonctions*. Paris: Courcier.

Martello, S. (1990). *Knapsack problems: algorithms and computer implementations*. New York: John Wiley and Sons.

Nemhauser, G. L., & Wolsey, L. A. (1988). *Integer and combinatorial optimization*. New York: John Wiley and Sons.

Neumann, J. V. (1928). Zur theorie der gesellschaftsspiele. *Mathematische Annalen, 100*(1), 295–320.

Nesterov, Y., & Nenirovski, A. (2001). *Interior point polynomial algorithms in convex programming*. Philadelphia: SİAM.

Nocedal, J., & Wright, S. (2006). *Numerical optimization*. New York: Springer Verlag.

Rockafellar, R. (1970). *Convex analysis*. Princeton: Princeton University Press.

Saigal, R. (1997). Linear programming: a modern integrated analysis. *Journal of the Operational Research Society, 48*(7), 762–762.

Schrijver, A. (1998). *Theory of Linear and Integer Programming*. New York: John Wiley and Sons.

Shor, N. Z. (2012). *Minimization methods for non-differentiable functions*. New York: Springer Verlag.

Sion. (1958). On general minimax theorems. *Pacific Journal of Mathematics* (8), 171–178.

Wolkowicz, H. (1981). Some applications of optimization in matrix theory. *Linear Algebra and Its Applications, 40*, 101–118.

Wolsey, L. A. (1998). *Integer programming*. New York: John Wiley and Sons.

Yuan, G. X. (1999). *KKM theory and applications to nonlinear analysis*. New York: Marcel Dekker.