New insights into conjugate duality

Von der Fakultät für Mathematik der Technischen Universität Chemnitz
genehmigte

D i s s e r t a t i o n

zur Erlangung des akademischen Grades
Doctor rerum naturalium
(Dr. rer. nat.)

vorgelegt
von

Dipl. - Math. Sorin - Mihai Grad
geboren am 02.03.1979 in Satu Mare (Rumänien)

eingereicht am 26.01.06

Gutachter: Prof. Dr. rer. nat. habil. Gert Wanka
Prof. Dr. Juan Enrique Martínez - Legaz
Prof. Dr. rer. nat. habil. Winfried Schirotzek

Tag der Verteidigung: 13.07.2006

Bibliographical description

Sorin - Mihai Grad

New insights into conjugate duality

Dissertation, **116** pages, Chemnitz University of Technology, Faculty of Mathematics, 2006

Report

With this thesis we bring some new results and improve some existing ones in conjugate duality and some of the areas it is applied in.

First we recall the way Lagrange, Fenchel and Fenchel - Lagrange dual problems to a given primal optimization problem can be obtained via perturbations and we present some connections between them. For the Fenchel - Lagrange dual problem we prove strong duality under more general conditions than known so far, while for the Fenchel duality we show that the convexity assumptions on the functions involved can be weakened without altering the conclusion. In order to prove the latter we prove also that some formulae concerning conjugate functions given so far only for convex functions hold also for almost convex, respectively nearly convex functions.

After proving that the generalized geometric dual problem can be obtained via perturbations, we show that the geometric duality is a special case of the Fenchel - Lagrange duality and the strong duality can be obtained under weaker conditions than stated in the existing literature. For various problems treated in the literature via geometric duality we show that Fenchel - Lagrange duality is easier to apply, bringing moreover strong duality and optimality conditions under weaker assumptions.

The results presented so far are applied also in convex composite optimization and entropy optimization. For the composed convex cone - constrained optimization problem we give strong duality and the related optimality conditions, then we apply these when showing that the formula of the conjugate of the precomposition with a proper convex K - increasing function of a K - convex function on some non - empty convex set $X \subseteq \mathbb{R}^n$, where K is a non - empty closed convex cone in $\mathbb{R}^k$, holds under weaker conditions than known so far. Another field were we apply these results is vector optimization, where we provide a general duality framework based on a more general scalarization that includes as special cases and improves some previous results in the literature. Concerning entropy optimization, we treat first via duality a problem having an entropy - like objective function, from which arise as special cases some problems found in the literature on entropy optimization. Finally, an application of entropy optimization into text classification is presented.

Keywords

perturbation functions; conjugate duality; conjugate functions; nearly convex sets and functions; almost convex functions; Fenchel - Lagrange duality; composed convex optimization problems; cone constraint qualification; geometric programming; entropy optimization; weak and strong duality; optimality conditions; duality in multiobjective convex optimization; efficient solutions; properly efficient solutions

Contents

Chapter 1

Introduction

Duality has played an important role in optimization and its applications especially during the last half of century. Several duality approaches were introduced in the literature, we mention here the classical Lagrange duality, Fenchel duality and geometric duality alongside the recent Fenchel - Lagrange duality, all of them being studied and used in the present thesis. Conjugate functions are of great importance for the latter three types.

Within this work we gathered our new results regarding the weakening of the sufficient conditions given until now in the literature that assure strong duality for the Fenchel, Fenchel - Lagrange and, respectively, geometric dual problems of some primal convex optimization problem with geometric and inequality constraints, showing moreover that the latter dual is actually a special case of the second one. Then we give duality statements also for composed convex optimization problems, with applications in multiobjective duality, and for optimization problems having entropy - like objective functions, generalizing some results in entropy optimization.

1.1 Duality: about and applications

Recognized as a basic tool in optimization, duality consists in attaching a dual problem to a given primal problem. Usually the dual problem has only geometric and/or linear constraints, but this is not a general rule. Among the advantages of introducing a dual to a given problem we mention also the lower bound assured by weak duality for the objective value of the primal problem and the easy derivation of necessary and sufficient optimality conditions. Of major interest is to give sufficient conditions that assure the so - called strong duality, i.e. the coincidence of the optimal objective values of the two problems, primal and dual, and the existence of an optimal solution to the dual problem.

Although there are several types of duality considered in the literature (for instance Weir - Mond duality and Wolfe duality), we restricted our interest to the following: Lagrange duality, Fenchel duality, geometric duality and Fenchel - Lagrange duality. The first of them is the oldest and perhaps the most used in the literature and consists in attaching the so - called Lagrangian to a primal minimization problem where the variable satisfies some geometric and (cone -) inequality constraints. The Lagrangian is constructed using the so - called Lagrange multiplicators that take values in the dual of the cone that appears in the constraints. Fenchel duality attaches to a problem consisting in the minimization of the sum of two functions a dual maximization problem containing in the objective function the conjugates of these functions. The conjugate functions started being intensively used in optimization since Rockafellar's book [72]. A combination of these two duality approaches has been recently brought into light by Boţ and Wanka (cf. [8,92]). They called the new dual problem they introduced the Fenchel - Lagrange dual problem and it contains both conjugate functions and Lagrange multiplicators. Moreover, this combined duality approach has as particular case another famous and widely - used duality concept, namely geometric programming duality. Geometric programming includes also posynomial programming and signomial programming. Geometric programming is due mostly to Peterson (see, for example, [71]) and it is still used despite being applicable only to some special classes of problems.

To assure strong duality there were taken into consideration various conditions, the most famous being the one due to Slater in Lagrange duality and the one involving relative interiors of the domains of the functions involved in Fenchel duality. There is a continuous challenge to give more and more general sufficient conditions for strong duality. An important prerequisite for strong duality in all the mentioned duality concepts is that the functions and sets involved are convex. Another direction which brought some interesting results regarding the weakening of the assumptions that deliver strong duality has been opened by the generalized convexity concepts.

Depending on the functions involved in the primal optimization problems we can distinguish different non - disjoint types of optimization problems. For instance there are differentiable optimization, linear optimization, discrete optimization, combinatorial optimization, complex optimization, DC optimization, entropy optimization and so on. When a composition of functions appears in a problem it is usually classified as a composite optimization problem. To the class of the composite optimization problems belong also many problems of the already mentioned types where the objective or constraint functions can be written as compositions of functions.

Applications of the duality can be detected in both theoretical and practical areas. Even if mentioning only a few fields where duality is successfully present one could not avoid multiobjective optimization, variational inequalities, theorems of the alternative, algorithms, maximal monotone operators from the first category, respectively economy and finance, data mining and support vector machines, image recognition and reconstruction, location and transports, and many others.

1.2 A description of the contents

In this section we present the way this thesis is organized, underlining the most important results contained within. The name of the present chapter fully represents its contents. After an overview on duality and its applications and this detailed presentation of the work we recall some notions and definitions needed later. Let us mention that all along this thesis we work in finite dimensional real spaces.

The second chapter deals with conjugate duality, introducing more general assumptions that assure strong duality for some primal - dual pairs of problems than known so far in the literature. It begins with a short presentation of the way dual problems are obtained via perturbations. Given a primal optimization problem consisting in minimizing a proper convex function subject to geometric and cone - inequality constraints, for suitable choices of the perturbation function one obtains the three dual problems we are mostly interested in within this work, namely Lagrange, Fenchel and Fenchel - Lagrange. The relations between these three duals are also recalled. Then we give the most general condition known so far that assures strong duality between the primal problem and its Fenchel - Lagrange dual (cf. [9]). We prove that, in the special case when the primal problem is the ordinary convex programming problem this new condition becomes the weakest constraint qualification known so far that guarantees strong duality in that situation (see also [32,72]). The last part of the chapter deals with the classical Fenchel duality. We show that it holds under weaker requirements for the functions involved (cf. [11,14]), i.e. when they are considered only almost convex (according to the definition introduced by Frenk and Kassay in [36]), respectively nearly convex (in the sense due to Aleman in [1]). In order to prove this we give also some other new results regarding these kinds of functions and their conjugates. A small application of these new results in game theory is presented, too.

The aim of the third chapter is to show that geometric programming duality is a particular case of the recently introduced Fenchel - Lagrange duality. First we show that the generalized geometric dual problem (cf. [71]) can be obtained also via perturbations (cf. [13]). Then we determine the Fenchel - Lagrange dual problem of the primal problem used in geometric programming (cf. [48]) and it turns out to be exactly the geometric dual problem known in the literature (cf. [15]). Specializing also the conditions that guarantee strong duality one can notice that the requirements we consider are more general than the ones usually considered in geometric programming, as the functions and some sets are not asked to be also lower semicontinuous, respectively closed, like in the existing literature. We have collected some applications of the geometric programming from the literature, including the classical posynomial programming, and we show for each of these problems that they do not have to be artificially transformed in order to fulfill the needs of geometric programming, as they may be easier treated by means of Fenchel - Lagrange duality. Moreover, when studying them via the latter, the strong duality statements and optimality conditions for all of these problems arise under weaker conditions than considered in the original papers.

In the fourth part of our work we give duality statements for some classes of problems, extending some results in the second chapter. First we deal with the so - called composed convex optimization problem that consists in the minimization of the composition of a proper convex K - increasing function with a function K - convex on the set where it is defined, subject to geometric and cone inequality constraints, where K is a closed convex cone. Strong duality and optimality conditions for such problems are proven under a weak constraint qualification (cf. [9]). The unconstrained case delivers us the tools to rediscover the formula of the conjugate of the precomposition with a proper K - increasing convex function of a K - convex function on the set where it is defined, which is shown to remain valid under weaker assumptions than known until now. Another application of the duality for the composed convex optimization problem is in multiobjective optimization where we present a new duality framework arising from a more general scalarization method than the usual linear one widely used (cf. [10]). The linear, maximum and norm scalarizations, usually used in the literature, turn out to be particular instances of the scalarization we consider. New duality results based on the Fenchel - Lagrange scalar dual problem are delivered.

The second section of this chapter deals with problems having entropy - like objective functions. Starting from the classical Kullback - Leibler entropy measure $\sum_{i=1}^{n} x_i \ln(x_i/y_i)$, which is applied in various fields such as pattern and image recognition, transportation and location problems, linguistics, etc., we have constructed the problem of minimizing a sum of functions of the type $\sum_{i=1}^{k} f_i(x) \ln(f_i(x)/g_i(x))$ subject to some geometric and inequality constraints. One may notice that this objective function contains as special cases the three most important entropy measures, namely the ones due to Shannon (cf. [83]), Kullback and Leibler (cf. [59]) and, respectively, Burg (cf. [24]). After giving strong duality and optimality conditions for such a problem we show that some problems in the literature on entropy optimization are rediscovered as special cases (cf. [12]). An application in text classification is provided, which contains also an algorithm that generalizes an older one due to Darroch and Ratcliff (cf. [16]).

An index of the notations and a comprehensive list of references close this thesis.

1.3 Preliminaries: definitions and results

We define now some notions that appear often within the present work, mentioning moreover some basic results called later. As usual, $\mathbb{R}^n$ denotes the n - *dimensional real space* for any positive integer n, $\overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$ the set of *extended reals*, $\mathbb{R}_+ = [0, +\infty)$ contains the non - negative reals, $\mathbb{Q}$ is the set of *real rationals* and $\mathbb{N}$ is the set of *positive integers*. Throughout this thesis all the vectors are considered as column vectors and an upper index ^{T} transposes a column vector to a row one and viceversa. The *inner product* of two vectors $x = (x_1, \ldots, x_n)^T$ and $y = (y_1, \ldots, y_n)^T$ in the n - dimensional real space is denoted by $x^T y = \sum_{i=1}^n x_i y_i$. Denote by "$\leqq$" the partial ordering introduced on any finite dimensional real space by the corresponding positive orthant considered as a cone. Because there are in the literature several different definitions for it, let us mention that by a *cone* in $\mathbb{R}^n$ we understand (cf. [44]) a set $C \subseteq \mathbb{R}^n$ with the property that whenever $x \in C$ and $\lambda > 0$ it follows $\lambda x \in C$. We use also the notation "$\geqq$" in the sense that $x \geqq y$ if and only if $y \leqq x$.

We extend the addition and the multiplication from $\mathbb{R}$ onto $\overline{\mathbb{R}}$ according to the following rules

$$\begin{aligned}
&a + (+\infty) = +\infty \ \forall a \in (-\infty, +\infty], \ a + (-\infty) = -\infty \ \forall a \in [-\infty, +\infty), \\
&a(+\infty) = +\infty \text{ and } a(-\infty) = -\infty \ \forall a \in (0, +\infty], \\
&a(+\infty) = -\infty \text{ and } a(-\infty) = +\infty \ \forall a \in [-\infty, 0), \\
&0(+\infty) = 0(-\infty) = 0.
\end{aligned}$$

Let us mention moreover that $(+\infty) + (-\infty)$ and $(-\infty) + (+\infty)$ are not defined.

Given some non - empty subset X of $\mathbb{R}^n$ we denote its *closure* by $\mathrm{cl}(X)$, its *interior* by $\mathrm{int}(X)$, its *border* by $\mathrm{bd}(X)$ and its *affine hull* by $\mathrm{aff}(X)$, while to write its *relative interior* we use the prefix ri. We need also the well - known *indicator* function of X, namely $\delta_X : \mathbb{R}^n \to \overline{\mathbb{R}}$ which is defined by

$$\delta_X(x) = \begin{cases} 0, & \text{if } x \in X, \\ +\infty, & \text{if } x \notin X, \end{cases}$$

and the *support* function of X,

$$\sigma_X : \mathbb{R}^n \to \overline{\mathbb{R}}, \ \sigma_X(y) = \sup_{x \in X} y^T x.$$

Notice that $\sigma_X = \delta_X^*$. When C is a non - empty cone in $\mathbb{R}^n$, its *dual* cone is given as

$$C^* = \left\{ x^* \in \mathbb{R}^n : {x^*}^T x \geq 0 \ \forall x \in C \right\}.$$

For a function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ we have the *effective domain* $\mathrm{dom}(f) = \{x \in \mathbb{R}^n : f(x) < +\infty\}$ and the *epigraph* $\mathrm{epi}(f) = \{(x, r) \in \mathbb{R}^n \times \mathbb{R} : f(x) \leq r\}$. The function f is said to be *proper* if one has concomitantly $\mathrm{dom}(f) \neq \emptyset$ and $f(x) > -\infty \ \forall x \in \mathbb{R}^n$. We also reserve the notation $\bar{f}$ for the *lower semicontinuous envelope* of f.

Take the function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$. It is called *convex* if for any $x, y \in \mathbb{R}^n$ and any $\lambda \in [0, 1]$ one has

$$f(\lambda x + (1 - \lambda) y) \leq \lambda f(x) + (1 - \lambda) f(y),$$

whenever the sum in the right - hand side is defined. When X is a non - empty convex subset of $\mathbb{R}^n$ the function $g : X \to \mathbb{R}$ is called *convex on* X if for all $x, y \in X$

and all $\lambda \in [0,1]$ one has

$$g(\lambda x + (1-\lambda)y) \leq \lambda g(x) + (1-\lambda)g(y).$$

Let us moreover notice that if $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ takes the value $g(x)$ at any $x \in X$, being equal to $+\infty$ otherwise, f is convex if and only if X is convex and g is convex on the set X. We call $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ *concave* if $-f$ is convex and $g : X \to \mathbb{R}$ *concave on* X if $-g$ is convex on X. We have $\text{epi}(\bar{f}) = \text{cl}(\text{epi}(f))$. For $x \in \mathbb{R}^n$ such that $f(x) \in \mathbb{R}$ we define the *subdifferential* of f at x by $\partial f(x) = \{x^* \in X^* : f(y) - f(x) \geq \langle x^*, y - x\rangle \ \forall y \in \mathbb{R}^n\}$.

When C is a non - empty closed convex cone in $\mathbb{R}^n$, a function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is called C - *increasing* if for $x, y \in \mathbb{R}^n$ fulfilling $x - y \in C$, follows $f(x) \geq f(y)$. If, moreover, whenever $x \neq y$ we have $f(x) > f(y)$, the function is called C - *strongly increasing*. Consider a non - empty convex set $X \subseteq \mathbb{R}^n$ and a non - empty closed convex cone K in $\mathbb{R}^k$. When a vector function $F : X \to \mathbb{R}^k$ fulfills the property

$$\forall x, y \in X \ \forall \lambda \in [0,1] \Rightarrow \lambda F(x) + (1-\lambda)F(y) - F\big(\lambda x + (1-\lambda)y\big) \in K,$$

it is called K - *convex on* X. In order to deal with conjugate duality we need first to introduce the conjugate functions. For $\emptyset \neq X \subseteq \mathbb{R}^n$ and a function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ we have the so - called *conjugate* function of f *regarding the set* X defined as

$$f_X^* : \mathbb{R}^n \to \overline{\mathbb{R}}, \ f_X^*(x^*) = \sup_{x \in X} \Big\{ {x^*}^T x - f(x) \Big\}.$$

When $X = \mathbb{R}^n$ or $\text{dom}(f) \subseteq X$ the conjugate regarding the set X turns out to be the classical (Legendre - Fenchel) conjugate function of f, denoted by f^*. This notion is extended also for functions defined on X as follows. Let $g : X \to \mathbb{R}^n$. Its conjugate regarding the set X is

$$g_X^* : \mathbb{R}^n \to \overline{\mathbb{R}}, \ g_X^*(x^*) = \sup_{x \in X} \Big\{ {x^*}^T x - g(x) \Big\}.$$

Concerning the conjugate functions we have the following inequality known as the Fenchel - Young inequality

$$f_X^*(x^*) + f(x) \geq {x^*}^T x \ \forall x \in X \ \forall x^* \in \mathbb{R}^n.$$

If $A : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation, then by $A^* : \mathbb{R}^m \to \mathbb{R}^n$ we denote its *adjoint* defined by $(Ax)^T y^* = x^T (A^* y^*) \ \forall x \in \mathbb{R}^n \ \forall y^* \in \mathbb{R}^m$. Let us also note that everywhere within this work we write min (max) instead of inf (sup) when the infimum (supremum) is attained and for an optimization problem (P) we denote its optimal objective value by $v(P)$.

Acknowledgements

This thesis would not have been possible without the continuous and careful assistance and advice of Prof. Dr. Gert Wanka, my supervisor, and Dr. Radu Ioan Boţ, my colleague and long - time friend. I want to thank them for proposing me this topic and for the continuous and careful supervision of my research work. I am also grateful to Karl - und Ruth - Mayer - Stiftung for the financial support between October 2002 and September 2003 and to the Faculty of Mathematics at the Chemnitz University of Technology for providing me a good research environment. Finally, I would like to thank my girlfriend Carmen and my family for love, patience and understanding.

Chapter 2

Conjugate duality in scalar optimization

Within this chapter we improve some general results in conjugate duality which generalize earlier statements given in the literature. Consider an optimization problem that consists in the minimization of a function subject to geometric and cone - inequality constraints. First we sketchily present the way perturbation theory is applied in duality and we recall the particular choices of the perturbation function that deliver us the Lagrange, Fenchel and, respectively, Fenchel - Lagrange dual problems to the considered primal. As the first two of them are widely - known and used, while the last one is very recent, being a combination of the other two, we focus on it. The second section has as its climax the introduction of a new constraint qualification that guarantees strong duality between the primal problem and its Fenchel - Lagrange dual. Under this new constraint qualification we deliver also necessary and sufficient optimality conditions. We prove that this new constraint qualification becomes, when the primal problem is the so - called ordinary convex program (cf. [72]), the weakest constraint qualification that assures strong duality in this special case. The final section of this chapter brings into attention new results involving some generalizations of the convexity. Some important results concerning conjugate functions known so far to be valid only for convex functions are proven also when the functions involved are almost convex or nearly convex. Finally we show that even the classical duality theorem due to Fenchel is valid when the functions involved are only almost convex, respectively nearly convex.

2.1 Dual problems obtained by perturbations and relations between them

2.1.1 Motivation

Various duality approaches and frameworks have been considered in literature, and in each case the challenge was to bring the weakest possible condition whose fulfilment guaranteed the annulment of the duality gap that normally exists between the optimal objective value of the primal and of the dual, respectively. One of the methods successfully used to introduce new dual problems was the one using perturbations. Wanka and Boţ (cf. [92]) have shown that the well - known dual problems usually named Lagrange dual and, respectively, Fenchel dual can be obtained by appropriately perturbing the given primal problem. Moreover, choosing a perturbation that combines the ones used to obtain the two mentioned dual problems, they have obtained a new dual problem to a given primal one, namely the

Fenchel - Lagrange dual problem, consisting in the minimization of a function with both geometric and inequality constraints. As within this thesis we give results concerning all the three dual problems mentioned above, this introductory section is necessary for better understanding the connections between these dual problems.

2.1.2 Problem formulation and dual problems obtained by perturbations

Take X a non - empty subset of $\mathbb{R}^n$ and C a non - empty closed convex cone in $\mathbb{R}^m$. Consider the functions $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ and $g = (g_1, \ldots, g_m)^T : X \to \mathbb{R}^m$. The primal optimization problem we consider in this section is

$$(P) \qquad \inf_{\substack{x \in X, \\ g(x) \in -C}} f(x).$$

Definition 2.1 *Any $x \in X$ fulfilling $g(x) \in -C$ is said to be a feasible element to the problem (P). Moreover, x is called an optimal solution for (P) if the optimal objective value of the problem is attained at x, i.e. $f(x) = v(P)$.*

The set $\mathcal{A} = \{x \in X : g(x) \in -C\}$ is called the *feasible set* of the problem (P).

To be sure that the problem makes sense, we assume from the very beginning the feasible set nonempty and moreover that it and $\mathrm{dom}(f)$ have at least a point in common. Thus $v(P) < +\infty$. Let us also remark that the function g could be defined on any subset of $\mathbb{R}^n$ containing X without affecting our future results.

Using an approach based on the theory of conjugate functions described by Ekeland and Temam in [31], we construct different dual problems to the primal problem (P). In order to do it, let us first consider the general unconstrained optimization problem

$$(PG) \qquad \inf_{x \in \mathbb{R}^n} F(x),$$

with $F : \mathbb{R}^n \to \overline{\mathbb{R}}$. Let us consider the perturbation function $\Phi : \mathbb{R}^n \times \mathbb{R}^m \to \overline{\mathbb{R}}$ which has the property that $\Phi(x, 0) = F(x)$ for each $x \in \mathbb{R}^n$. Here, $\mathbb{R}^m$ is the space of the perturbation variables. For each $p \in \mathbb{R}^m$ we obtain a new optimization problem

$$(PG_p) \qquad \inf_{x \in \mathbb{R}^n} \Phi(x, p).$$

For any $p \in \mathbb{R}^m$, the problem (PG_p) is a perturbed problem attached to (PG). In order to introduce a dual problem to (PG), we calculate the conjugate of Φ which is the function $\Phi^* : \mathbb{R}^n \times \mathbb{R}^m \to \overline{\mathbb{R}}$,

$$\Phi^*(x^*, p^*) = \sup_{\substack{x \in \mathbb{R}^n, \\ p \in \mathbb{R}^m}} \left\{ (x^*, p^*)^T (x, p) - \Phi(x, p) \right\} = \sup_{\substack{x \in \mathbb{R}^n, \\ p \in \mathbb{R}^m}} \left\{ {x^*}^T x + {p^*}^T p - \Phi(x, p) \right\}.$$

Now we can define the following optimization problem

$$(DG) \qquad \sup_{p^* \in \mathbb{R}^m} \{-\Phi^*(0, p^*)\}.$$

The problem (DG) is called the dual problem to (PG) and its optimal objective value is denoted by $v(DG)$.

This approach presents an important feature: between the primal and the dual problem weak duality always holds. The following statement proves this fact.

Theorem 2.1 *(weak duality) (cf. [31]) The following chain of inequalities is always valid*

$$-\infty \leq v(DG) \leq v(PG) \leq +\infty.$$

Our next aim is to show how this approach can be applied to the constrained optimization problem (P). Therefore, take

$$F(x) = \begin{cases} f(x), & \text{if } x \in X \text{ and } g(x) \in -C, \\ +\infty, & \text{otherwise.} \end{cases}$$

It is easy to notice that the primal problem (P) is equivalent to

$$(PG) \qquad \inf_{x \in \mathbb{R}^n} F(x),$$

and, since the perturbation function $\Phi : \mathbb{R}^n \times \mathbb{R}^m \to \overline{\mathbb{R}}$ satisfies $\Phi(x, 0) = F(x)$ for each $x \in \mathbb{R}^n$, this must fulfill

$$\Phi(x, 0) = \begin{cases} f(x), & \text{if } x \in X \text{ and } g(x) \in -C, \\ +\infty, & \text{otherwise.} \end{cases} \tag{2. 1}$$

For special choices of the perturbation function Φ we obtain different dual problems to (P) as shown in the following.

First let the perturbation function be (cf. [8, 92])

$$\Phi_L : \mathbb{R}^n \times \mathbb{R}^m \to \overline{\mathbb{R}},\ \Phi_L(x, q) = \begin{cases} f(x), & \text{if } x \in X \text{ and } g(x) - q \in -C, \\ +\infty, & \text{otherwise,} \end{cases}$$

with the perturbation variable $q \in \mathbb{R}^m$. The formula of the conjugate of this function, $\Phi_L^* : \mathbb{R}^n \times \mathbb{R}^m \to \overline{\mathbb{R}}$ is

$$\Phi_L^*(x^*, q^*) = \begin{cases} \sup\limits_{x \in X} \left\{ x^{*T}x + q^{*T}g(x) - f(x) \right\}, & \text{if } q^* \in -C^*, \\ +\infty, & \text{otherwise.} \end{cases}$$

The dual obtained by the perturbation function Φ_L to the problem (P) is

$$(D^L) \qquad \sup_{q^* \in \mathbb{R}^m} \{-\Phi_L^*(0, q^*)\},$$

which has actually the following formulation

$$(D^L) \qquad \sup_{q^* \in C^*} \inf_{x \in X} \left[f(x) + q^{*T}g(x) \right].$$

One can immediately notice that the problem (D^L) is actually the well - known Lagrange dual problem to (P).

Take now the perturbation function

$$\Phi_F : \mathbb{R}^n \times \mathbb{R}^n \to \overline{\mathbb{R}},\ \Phi_F(x, p) = \begin{cases} f(x + p), & \text{if } x \in X \text{ and } g(x) \in -C, \\ +\infty, & \text{otherwise,} \end{cases}$$

with the perturbation variable $p \in \mathbb{R}^n$. Its conjugate function is

$$\Phi_F^*(x^*, p^*) = f^*(p^*) - \inf_{x \in \mathcal{A}} \left\{ (p^* - x^*)^T x \right\} = f^*(p^*) + \sigma_{\mathcal{A}}(x^* - p^*),$$

and the corresponding dual problem to (P),

$$(D^F) \qquad \sup_{p^* \in \mathbb{R}^n} \{-\Phi_F^*(0, p^*)\},$$

turns out to be

$$(D^F) \qquad \sup_{p^* \in \mathbb{R}^n} \{-f^*(p^*) - \sigma_{\mathcal{A}}(-p^*)\}.$$

It is easy to remark that the primal problem may be rewritten as

$$(P) \qquad \inf_{x \in \mathbb{R}} [f(x) + \delta_{\mathcal{A}}(x)],$$

and the latter dual problem is actually

$$(D^F) \qquad \sup_{p^* \in \mathbb{R}^n} \big\{ - f^*(p^*) - \delta^*_{\mathcal{A}}(-p^*)\big\},$$

which is the Fenchel dual problem to (P) (cf. [72]).

Thus we have obtained via the presented perturbation theory two classical dual problems to the primal problem (P) by appropriately choosing the perturbation function. The natural question of what happens when one combines the two perturbation functions considered above has been answered by Boţ and Wanka in a series of recent papers beginning with [92]. They took the perturbation function

$$\Phi_{FL} : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m \to \overline{\mathbb{R}},$$

$$\Phi_{FL}(x, p, q) = \begin{cases} f(x+p), & \text{if } x \in X \text{ and } g(x) - q \in -C, \\ +\infty, & \text{otherwise,} \end{cases}$$

with the perturbation variables $p \in \mathbb{R}^n$ and $q \in \mathbb{R}^m$. Φ_{FL} satisfies the condition (2. 1) required in order to be a perturbation function to the problem (P) and its conjugate is

$$\Phi^*_{FL}(x^*, p^*, q^*) = \begin{cases} f^*(p^*) + \sup\limits_{x \in X} \Big\{ (x^* - p^*)^T x + {q^*}^T g(x) \Big\}, & \text{if } q^* \in -C^*, \\ +\infty, & \text{otherwise.} \end{cases}$$

The dual problem that arises in this case is

$$(D^{FL}) \qquad \sup_{\substack{p^* \in \mathbb{R}^n, \\ q^* \in \mathbb{R}^k}} \big\{ - \Phi^*_{FL}(0, p^*, q^*)\big\},$$

which is actually

$$(D^{FL}) \qquad \sup_{\substack{p^* \in \mathbb{R}^n, \\ q^* \in C^*}} \Big\{ - f^*(p^*) - ({q^*}^T g)^*_X(-p^*)\Big\},$$

where by $q^T g$ we denote the function defined on X whose value at any $x \in X$ is equal to $\sum_{j=1}^m q_j g_j(x)$, with $q = (q_1, \ldots, q_m)^T$. Because of the way it was constructed this problem has been called (cf. [92], see also [9] and [15]) the Fenchel - Lagrange dual problem to (P). As we shall see later it can be obtained also by considering the Lagrange dual problem to (P) and then the Fenchel dual to the inner minimization problem.

Between the optimal objective values of the three dual problems attached to (P) we take into consideration and the primal problem itself there hold the following relations (see also Theorem 2.1, [8, 92])

$$v(D^{FL}) \leq \begin{matrix} v(D^L) \\ v(D^F) \end{matrix} \leq v(P).$$

Between $v(D^L)$ and $v(D^p)$ no general order can be given, see [8] or [92] for examples where $v(D^L) > v(D^p)$ and, respectively, $v(D^L) < v(D^p)$. Sufficient conditions to assure the fulfillment of the inequalities above as equalities were already given in the literature, see, for instance, [8] or [92]. We are interested to close the gap between $v(D^{FL})$ and $v(P)$, i.e. to give sufficient conditions to guarantee the simultaneous satisfaction of the inequalities above as equalities and moreover the existence of an optimal solution to the dual problem. The next section deals with this problem and its connections in the literature.

2.2 Strong duality and optimality conditions for Fenchel - Lagrange duality

2.2.1 Motivation

Introduced by Wanka and Boţ (cf. [92]), the Fenchel - Lagrange dual problem is a combination of the well - known Lagrange and Fenchel dual problems. Although new, it has proven to have some important applications, from multiobjective optimization (cf. [8, 18, 19, 89–91, 93]) to theorems of the alternative and Farkas type results (cf. [22]). Moreover, we show in the third chapter of the present thesis that the well - known and widely used geometric programming duality is a special case of the Fenchel - Lagrange duality, so all its applications can be taken over, too. For a given primal optimization problem consisting in the minimization of a function with both geometric and inequality constraints, the initial constraint qualification considered in order to achieve strong duality between the primal and the new dual problem was based on the well - known condition due to Slater. Then these results have been refined in [15] and generalized in [9]. In the latter paper the inequality constraints are considered over some non - empty closed convex cone and the new constraint qualification may work also when the cone has empty interior, while the classical constraint qualification due to Slater fails in such a situation, as it asks the cone to have a non - empty interior.

2.2.2 Duality and optimality conditions

To the given primal optimization problem (P) we have introduced three dual problems, two of them being widely used and known, while the third, their combination, has been recently introduced. In order to give the strong duality statement for the pair of problems (P) - (D^{FL}) we need to introduce a constraint qualification, inspired by the one used in [37]. Further, unless otherwise specified, consider moreover X a non - empty convex set and that f is a proper convex function and g a C - convex function on X. First we give an equivalent formulation of the constraint qualification considered in [37], which is in this case

$$(CQ_{FK}) \qquad\qquad 0 \in \mathrm{ri}(g(X \cap \mathrm{dom}(f)) + C).$$

Lemma 2.1 *Let $U \subseteq X$ a non - empty convex set. Then*

$$\mathrm{ri}(g(U) + C) = g(\mathrm{ri}(U)) + \mathrm{ri}(C).$$

Proof. Consider the set

$$M = \{(x, y) : \ x \in U, \ y \in \mathbb{R}^m, \ y - g(x) \in C\},$$

which is easily provable to be convex. For each $x \in U$ consider now the set $M_x = \{y \in \mathbb{R}^m : \ (x, y) \in M\}$. When $x \notin U$ it is obvious that $M_x = \emptyset$, while in the complementary case we have $y \in M_x \Leftrightarrow y - g(x) \in C \Leftrightarrow y \in g(x) + C$, so we conclude

$$M_x = \begin{cases} g(x) + C, & \text{if } x \in U, \\ \emptyset, & \text{if } x \notin U. \end{cases}$$

Therefore M_x is also convex for any $x \in U$. Let us see now how we can characterize the relative interior of the set M. According to Theorem 6.8 in [72] we have $(x, y) \in \operatorname{ri}(M)$ if and only if $x \in \operatorname{ri}(U)$ and $y \in \operatorname{ri}(M_x)$. On the other hand, for any $x \in \operatorname{ri}(U)$, $y \in \operatorname{ri}(M_x)$ means actually $y \in \operatorname{ri}(g(x) + C) = g(x) + \operatorname{ri}(C)$, so we can write further

$$\operatorname{ri}(M) = \{(x, y) : \ x \in \operatorname{ri}(U), \ y - g(x) \in \operatorname{ri}(C)\}.$$

Consider now the linear transformation $A : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ defined by $A(x, y) = y$. Let us prove that $A(M) = g(U) + C$. Take first an element $y \in A(M)$. It follows that there is an $x \in U$ such that $y - g(x) \in C$, which yields $y \in g(U) + C$. Reversely, for any $y \in g(U) + C$ there is an $x \in U$ such that $y \in g(x) + C$, so $y - g(x) \in C$. This means $(x, y) \in M$, followed by $y \in A(M)$.

Finally, by Theorem 6.6 in [72] we get

$$\operatorname{ri}(g(U) + C) = \operatorname{ri}(A(M)) = A(\operatorname{ri}(M)) = g(\operatorname{ri}(U)) + \operatorname{ri}(C). \qquad \square$$

According to this lemma, the constraint qualification (CQ_{FK}) is equivalent to

$$(CQ_R) \qquad \exists x' \in \operatorname{ri}(X \cap \operatorname{dom}(f)) \text{ such that } g(x') \in -\operatorname{ri}(C).$$

This condition is sufficient to assure duality between (P) and (D^L), but in order to close the gap between (P) and (D^{FL}) we introduce the following constraint qualification

$$(CQ) \qquad \exists x' \in \operatorname{ri}(\operatorname{dom}(f)) \cap \operatorname{ri}(X) : g(x') \in -\operatorname{ri}(C).$$

Remark 2.1 Notice, using eventually Theorem 6.5 in [72], that the validity of (CQ) guarantees the satisfaction of (CQ_R). We are now ready to formulate the strong duality statement.

Theorem 2.2 *(strong duality) (see [9]) Consider the constraint qualification (CQ) fulfilled. Then there is strong duality between the problem (P) and its dual (D^{FL}), i.e. $v(P) = v(D^{FL})$ and the latter has an optimal solution if $v(P) > -\infty$.*

Proof. First we deal with the Lagrange dual problem to (P), which is

$$(D^L) \qquad \sup_{q^* \in C^*} \inf_{x \in X} \Big[f(x) + {q^*}^T g(x) \Big].$$

According to [37], the convexity assumptions introduced above and the fulfillment of the condition (CQ_{FK}) (or (CQ_R), due to Lemma 2.1) are sufficient to assure the coincidence of $v(P)$ and $v(D^L)$, moreover guaranteeing the existence of an optimal solution $\overline{q^*}$ to (D^L) when $v(P) > -\infty$. As (CQ) implies (CQ_{FK}) (see Lemma 2.1 and Remark 2.1), $v(P)$ and $v(D^L)$ coincide, the latter having moreover a solution when $v(P) > -\infty$.

Now let us write the Fenchel dual problem to the inner infimum in (D^L). For any $q^* \in C^*$, ${q^*}^T g$ is a real-valued function convex on X, so in order to apply rigourously Fenchel's duality theorem (cf. [72]) we have to consider its convex extension to $\mathbb{R}^n$, $\widetilde{{q^*}^T g}$, which takes the value $+\infty$ outside X. As $\operatorname{dom}(\widetilde{{q^*}^T g}) = X$ and $\operatorname{ri}(\operatorname{dom}(f)) \cap \operatorname{ri}(X) \neq \emptyset$, we have by Fenchel's duality theorem (cf. Theorem 31.1 in [72])

$$\inf_{x \in X} \Big[f(x) + {q^*}^T g(x) \Big] = \inf_{x \in \mathbb{R}^n} \Big[f(x) + \widetilde{{q^*}^T g}(x) \Big] = \sup_{p^* \in \mathbb{R}^n} \Big\{ - f^*(p^*) - \widetilde{{q^*}^T g}^{\,*}(-p^*) \Big\}.$$

It is not difficult to notice that $\widetilde{{q^*}^T g}^{\,*}(-p^*) = ({q^*}^T g)^*_X(-p^*)$, so it is straightforward that

$$v(P) = \sup_{q^* \in C^*} \inf_{x \in X} \Big[f(x) + {q^*}^T g(x)\Big] = \sup_{\substack{q^* \in C^*, \\ p^* \in \mathbb{R}^n}} \Big\{-f^*(p^*) - ({q^*}^T g)^*_X(-p^*)\Big\} = v(D^{FL}).$$

In case $v(P)$ is finite, because of the existence of an optimal solutions for the Lagrange dual, we get

$$v(P) = \sup_{q^* \in C^*} \inf_{x \in X} \Big[f(x) + {q^*}^T g(x)\Big] = \inf_{x \in X} \Big[f(x) + \overline{q^*}^T g(x)\Big].$$

Further, by Fenchel's duality theorem (cf. [72]),

$$v(P) = \inf_{x \in X} \Big[f(x) + \overline{q^*}^T g(x)\Big] = \max_{p^* \in \mathbb{R}^n} \Big\{ - f^*(p^*) - (\overline{q^*}^T g)^*_X(-p^*)\Big\},$$

the latter being attained at some $\overline{p^*} \in \mathbb{R}^n$. This means exactly that $(\overline{p^*}, \overline{q^*})$ is an optimal solution to (D^{FL}). $\square$

Now let us deliver necessary and sufficient optimality conditions regarding the convex optimization problems (P) and (D^{FL}).

Theorem 2.3 *(optimality conditions)*
(a) *If the constraint qualification (CQ) is fulfilled and the primal problem (P) has an optimal solution $\bar{x}$, then the dual problem has an optimal solution $(\overline{p^*}, \overline{q^*})$ and the following optimality conditions are satisfied*

(i) $f^*(\overline{p^*}) + f(\bar{x}) = \overline{p^*}^T \bar{x}$,

(ii) $(\overline{q^*}^T g)^*_X(-\overline{p^*}) + \overline{q^*}^T g(\bar{x}) = -\overline{p^*}^T \bar{x}$,

(iii) $\overline{q^*}^T g(\bar{x}) = 0$.

(b) *If $\bar{x}$ is a feasible point to the primal problem (P) and $(\overline{p^*}, \overline{q^*})$ is feasible to the dual problem (D^{FL}) fulfilling the optimality conditions $(i) - (iii)$, then there is strong duality between (P) and (D^{FL}) and the mentioned feasible points turn out to be optimal solutions to the corresponding problems.*

Proof. (a) Theorem 2.2 guarantees strong duality between (P) and (D^{FL}), so the dual problem has an optimal solution, say $(\overline{p^*}, \overline{q^*})$. The equality of the optimal objective values of (P) and (D^{FL}) implies

$$f(\bar{x}) + f^*(\overline{p^*}) + (\overline{q^*}^T g)^*(-\overline{p^*}) = 0. \tag{2. 2}$$

The Fenchel - Young inequality states

$$f^*(\overline{p^*}) + f(\bar{x}) \geq \overline{p^*}^T \bar{x}$$

and

$$\big(\overline{q^*}^T g\big)^*_X(-\overline{p^*}) + \overline{q^*}^T g(\bar{x}) \geq -\overline{p^*}^T \bar{x}.$$

Summing these two relations one gets, taking also into account (2. 2),

$$0 \geq \overline{q^*}^T g(\bar{x}) = f^*(\overline{p^*}) + f(\bar{x}) + (\overline{q^*}^T g)^*_X(-\overline{p^*}) + \overline{q^*}^T g(\bar{x}) \geq 0,$$

where the first inequality is valid because $\bar{x}$ is feasible to (P) and $\overline{q^*} \in C^*$. It is clear that all the inequalities above must be fulfilled as equalities and this delivers immediately the optimality conditions $(i) - (iii)$.

(b) All the calculations presented above can be carried out in reverse order, so the assertion holds. $\square$

Remark 2.2 (cf. [9,15]) We want to mention that (b) applies without any convexity assumption as well as constraint qualification. So the sufficiency of the optimality conditions $(i) - (iii)$ is true in the most general case.

We give now a concrete problem that shows that a relaxation of (CQ) by considering in its formulation the whole set X instead of its relative interior does not guarantee strong duality.

Example 2.1 Let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by $f(x_1, x_2) = x_2$ and $g : X \to \mathbb{R}$, $g(x_1, x_2) = x_1$, where

$$X = \left\{ x = (x_1, x_2) \in \mathbb{R}^2 : 0 \leq x_1 \leq 2, \begin{array}{ll} 3 \leq x_2 \leq 4, & \text{if } x_1 = 0, \\ 1 < x_2 \leq 4, & \text{if } x_1 > 0 \end{array} \right\}.$$

It is obvious that f is a convex function and g is convex on X. Formulate the optimization problem

$$(P_e) \qquad \inf_{\substack{x \in X, \\ g(x)=0}} f(x).$$

This problem fits into our scheme for $C = \{0\}$. The constraint qualification (CQ) becomes in this case

$$\exists x' \in \text{ri}(X) \text{ such that } g(x') \in -C,$$

which means, by Lemma 2.1, $0 \in \text{ri}(g(X) + C)$, i.e. $0 \in \text{ri}([0, 2] + 0) = (0, 2)$, that is false. One may notice that also (CQ_R) fails in this case, as it coincides here with (CQ). On the other hand the condition $0 \in g(X) + \text{ri}(C)$ is fulfilled, being in this case $0 \in [0, 2]$, that is true.

As in [32], where this example has been borrowed from, the optimal objective value of (P_e) turns out to be $v(P_e) = 3$, while the one of its Lagrange dual problem is 1. Because of the convexity of the functions and sets involved the optimal objective value of the Lagrange dual coincides in this case (see the proof of Theorem 2.2) to the one of the Fenchel - Lagrange dual to (P_e).

Therefore we see that a relaxation of (CQ) by considering $x' \in X$ instead of $x' \in \text{ri}(X)$ does not guarantee strong duality.

We would also like to mention that Frenk and Kassay have shown in [36] that if there is an $y_0 \in \text{aff}(g(X))$ such that $g(X) \subseteq y_0 + \text{aff}(C)$ then $0 \in g(X) + \text{ri}(C)$ becomes equivalent to $0 \in \text{ri}(g(X) + C)$.

2.2.3 The ordinary convex programs as special case

The ordinary convex programs (cf. [72]) are among the problems to which the duality assertions formulated earlier are applicable. Consider such an ordinary convex program

$$(P_o) \qquad \inf_{\substack{x \in X, \\ g_i(x) \leq 0, i=1,\ldots,r, \\ g_j(x)=0, j=r+1,\ldots,m}} f(x),.$$

where $X \subseteq \mathbb{R}^n$ is a non - empty convex set, $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is a convex function with $\text{dom}(f) = X$, $g_i : X \to \mathbb{R}$, $i = 1, \ldots, r$, are functions convex on X and $g_j : X \to \mathbb{R}$, $j = r + 1, \ldots, m$, are the restrictions to X of some affine functions on $\mathbb{R}^n$. Denote

$g = (g_1, \ldots, g_m)^T$. This problem is a special case of (P) when we consider the cone $C = \mathbb{R}^r_+ \times \{0\}^{m-r}$. The Fenchel - Lagrange dual problem to (P_o) is

$$(D_o^{FL}) \qquad \sup_{\substack{q \in \mathbb{R}^r_+ \times \mathbb{R}^{m-r},\\ p \in \mathbb{R}^n}} \Big\{ - f^*(p) - \big(q^T g\big)^*_X(-p) \Big\}.$$

The constraint qualification that assures strong duality is in this case

$$(CQ_o) \qquad 0 \in \operatorname{ri}\big(g(X) + \mathbb{R}^r_+ \times \{0\}^{m-r}\big),$$

equivalent to $0 \in g(\operatorname{ri}(X)) + \operatorname{ri}(\mathbb{R}^r_+ \times \{0\}^{m-r})$, i.e.

$$(CQ_o) \qquad \exists x' \in \operatorname{ri}(X) : \begin{cases} g_i(x') < 0, & \text{if } i = 1, \ldots, r,\\ g_j(x') = 0, & \text{if } j = r+1, \ldots, m, \end{cases}$$

which is exactly the sufficient condition given in [72] to state strong duality between (P_o) and its Lagrange dual problem

$$(D_o^L) \qquad \sup_{q \in \mathbb{R}^r_+ \times \mathbb{R}^{m-r}} \inf_{x \in X} \Big[f(x) + q^T g(x) \Big].$$

As $\operatorname{ri}(\operatorname{dom}(f)) = \operatorname{ri}(X) \neq \emptyset$, we have that the value of the inner infimum in (D_o^L), as a convex optimization problem, is equal to the optimal value of its Fenchel dual, which leads us to the objective function in (D_o^{FL}). The strong duality statement concerning the problems (P_o) and (D_o^{FL}) follows.

Theorem 2.4 *(strong duality) Consider the constraint qualification (CQ_o) fulfilled. Then there is strong duality between the problem (P_o) and its dual (D_o^{FL}) and the latter has an optimal solution if $v(P_o) > -\infty$.*

Remark 2.3 Some authors take as ordinary convex program the following problem, where f, g and X are defined as before,

$$(P_o') \qquad \inf_{\substack{x \in X,\\ g(x) \leqq 0}} f(x).$$

For this problem the strong duality is attained provided the fulfillment of the constraint qualification (cf. [9, 32])

$$(CQ_o') \qquad \exists x' \in \operatorname{ri}(X) : \begin{cases} g_i(x') < 0, & \text{if } i = 1, \ldots, r,\\ g_j(x') \leq 0, & \text{if } j = r+1, \ldots, m. \end{cases}$$

A first look would make someone think that (P_o') is a special case of (P) by taking $C = \mathbb{R}^m_+$ and (CQ) requires in this case the existence of an $x' \in \operatorname{ri}(X)$ such that $g(x') \in -\operatorname{ri}\big(\mathbb{R}^m_+\big)$, i.e. for all $i = 1, \ldots, m$, $g_i(x') < 0$, condition that is more restrictive than (CQ_o'). Let us prove that there is another possible choice of the cone C such that (CQ_o') implies the fulfilment of (CQ), namely $0 \in g(\operatorname{ri}(X)) + \operatorname{ri}(C)$ for the primal problem rewritten as

$$(P_o') \qquad \inf_{\substack{x \in X,\\ g(x) \in -C}} f(x).$$

Consider (CQ_o') fulfilled and take the set

$$I = \Big\{ i \in \{r+1, \ldots, m\} : x \in X \text{ such that } g(x) \leqq 0 \Rightarrow g_i(x) = 0 \Big\}.$$

When $I = \emptyset$ then for each $i \in \{r+1, \ldots, m\}$ there is an $x^i \in X$ feasible to (P_o')

such that $g_i(x^i) < 0$. Take the cone $C = \mathbb{R}^m_+$. Introducing

$$x^0 = \sum_{i=r+1}^{m} \frac{1}{m-r+1} x^i + \frac{1}{m-r+1} x',$$

we show that it belongs to $\mathrm{ri}(X)$. First,

$$\sum_{i=r+1}^{m} \frac{1}{m-r+1} x^i = \frac{m-r}{m-r+1} \sum_{i=r+1}^{m} \frac{1}{m-r} x^i$$

and $\sum_{i=r+1}^{m} \big(1/(m-r)\big) x^i \in X$. Applying Theorem 6.1 in [72] it follows that $x^0 \in \mathrm{ri}(X)$. For any $j \in \{1, \ldots, m\}$ we have

$$g_j(x^0) \leq \sum_{i=r+1}^{m} \frac{1}{m-r+1} g_j(x^i) + \frac{1}{m-r+1} g_j(x') < 0.$$

Therefore there exists $x^0 \in \mathrm{ri}(X)$ such that $0 \in g(x^0) + \mathrm{ri}(C)$, which is the desired result.

When $I \neq \emptyset$, without loss of generality as we perform at most a reindexing of the functions g_j, $r+1 \leq j \leq m$, let $I = \{r+l, \ldots, m\}$, where l is a positive integer smaller than $m-r$. This means that for $j \in \{r+l, \ldots, m\}$ follows $g_j(x) = 0$ if $x \in X$ and $g(x) \leqq 0$. Then (P_0') is a special case of (P) for $C = \mathbb{R}^{r+l-1}_+ \times \{0\}^{m-r-l+1}$. For each $j \in \{r+1, \ldots, r+l-1\}$ there is an x^j feasible to (P_o') such that $g_j(x^j) < 0$. Taking

$$x^0 = \sum_{i=r+1}^{r+l-1} \frac{1}{l} x^i + \frac{1}{l} x',$$

we have as above that $x^0 \in \mathrm{ri}(X)$ and $g_j(x^0) < 0$ for any $j \in \{1, \ldots, r+l-1\}$ and $g_j(x^0) = 0$ for $j \in I$ (because of the affinity of the functions g_j, $r+1 \leq j \leq m$), which is exactly what (CQ) asserts.

Therefore there is always a choice of the cone C which guarantees that for the reformulated problem (CQ) stands.

2.3 Fenchel duality under weaker requirements

As we have seen, in order to prove strong duality for the Fenchel - Lagrange dual problem under more general conditions than known so far we have weakened the constraint qualification. Our research included also the classical Fenchel duality and we show in the following that it holds under more general conditions than given in the literature known to us. Unlike the previous section, here we do not weaken the constraint qualification but the convexity assumptions imposed on the functions involved. Thus we give some new results involving almost convex and nearly convex functions. They culminate with proving that the classical Fenchel duality theorem is valid when the functions involved are almost convex or nearly convex, too. Known being the applications of Fenchel duality in game theory (see, for instance, [7]) and the connections between this area and Lagrange duality given also for generalized convex functions (cf. [68, 69]), we give an application of our results in the latter field. By this we are trying to open the gate into the direction of Fenchel duality for games which can be written by using optimization problems involving generalized convex functions.

2.3.1 Motivation

Since ROCKAFELLAR's book [72], convex analysis started to spread into various directions, and Fenchel's duality theorem has been called in various contexts. Given initially for convex functions, it has been extended to some other classes of problems involving different types of functions as the need for such statements arose from both theoretical and practical needs. We mention here KANNIAPPAN who has given in [55] a Fenchel type duality theorem for non - convex and non - differentiable maximization problems, BEONI who extended Fenchel's statement to fractional programming in [5] and PENOT AND VOLLE who considered it for quasiconvex problems in [70].

As suggested by the latter paper, a direction to generalize the duality statements is to consider various generalizations of the convexity instead of convexity for the functions and sets involved. For instance FRENK AND KASSAY (cf. [36]) extended Lagrange duality for nearly convex functions and BOŢ, KASSAY AND WANKA gave strong duality for a primal optimization problem and its Fenchel - Lagrange dual when the functions involved were nearly convex in [17].

In the following we prove that some properties of the conjugate functions can be given also for almost convex and nearly convex functions. Since there are several notions of almost convexity in the literature, let us mention that we consider the one introduced by FRENK AND KASSAY in [36], while for nearly convex functions we use the definition usually encountered in the literature due to ALEMAN (cf. [1]). Our results regarding the conjugates of nearly convex and, respectively, almost convex functions bring towards the end of the chapter the proofs that the Fenchel duality theorem is valid when the functions involved are almost convex or nearly convex, too. Therefore we generalize this classical result in duality in a new direction.

2.3.2 Preliminaries on nearly and almost convex functions

We begin with some definitions and some useful results.

Definition 2.2 *A set $X \subseteq \mathbb{R}^n$ is called nearly convex if there is a constant $\alpha \in (0,1)$ such that for any x and y belonging to X one has $\alpha x + (1-\alpha)y \in X$.*

An example of a nearly convex set which is not convex is $\mathbb{Q}$. Important properties of the nearly convex sets follow.

Lemma 2.2 *(cf. [1]) For every nearly convex set $X \subseteq \mathbb{R}^n$ the following properties are valid*

- (i) $\mathrm{ri}(X)$ *is convex (may be empty),*
- (ii) $\mathrm{cl}(X)$ *is convex,*
- (iii) *for every $x \in \mathrm{cl}(X)$ and $y \in \mathrm{ri}(X)$ we have $tx + (1-t)y \in \mathrm{ri}(X)$ for each $0 \leq t < 1$.*

Definition 2.3 *(cf. [23, 36]) A function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is called*

- (i) *almost convex if $\bar{f}$ is convex and $\mathrm{ri}(\mathrm{epi}(\bar{f})) \subseteq \mathrm{epi}(f)$,*
- (ii) *nearly convex if $\mathrm{epi}(f)$ is nearly convex,*
- (iii) *closely convex if $\mathrm{epi}(\bar{f})$ is convex (i.e. $\bar{f}$ is convex).*

Connections between these kinds of functions arise from the following remarks, while to show that there are differences between them we give later an example.

Remark 2.4 Any almost convex function is also closely convex.

Remark 2.5 Any nearly convex function has a nearly convex effective domain. Moreover, as its epigraph is nearly convex, the function is also closely convex, according to Lemma 2.2(ii).

Although cited from the literature, the following auxiliary results are not so widely known, thus we have included them here.

Remark 2.6 Given any function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ we have $\mathrm{dom}(f) \subseteq \mathrm{dom}(\bar{f}) \subseteq \mathrm{cl}(\mathrm{dom}(f))$, which implies $\mathrm{cl}(\mathrm{dom}(f)) = \mathrm{cl}(\mathrm{dom}(\bar{f}))$.

Lemma 2.3 *(cf. [17, 36]) For a convex set $C \subseteq \mathbb{R}^n$ and any non - empty set $X \subseteq \mathbb{R}^n$ satisfying $X \subseteq C$ we have $\mathrm{ri}(C) \subseteq X$ if and only if $\mathrm{ri}(C) = \mathrm{ri}(X)$.*

Lemma 2.4 *(cf. [17]) Let $X \subseteq \mathbb{R}^n$ be a nearly convex set. Then $\mathrm{ri}(X) \neq \emptyset$ if and only if $\mathrm{ri}(\mathrm{cl}(X)) \subseteq X$.*

Lemma 2.5 *(cf. [17]) For a non - empty nearly convex set $X \subseteq \mathbb{R}^n$, $\mathrm{ri}(X) \neq \emptyset$ if and only if $\mathrm{ri}(X) = \mathrm{ri}(\mathrm{cl}(X))$.*

Using Remark 2.5 and Lemma 2.4 we deduce the following statement.

Proposition 2.1 *If $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is a nearly convex function satisfying $\mathrm{ri}(\mathrm{epi}(f)) \neq \emptyset$, then it is almost convex.*

Remark 2.7 Each convex function is nearly convex and almost convex.

The first observation is obvious, while the second can be easily proven. Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a convex function. If $f(x) = +\infty$ everywhere then $\mathrm{epi}(f) = \emptyset$, which is closed, so $\bar{f} = f$ and it follows f almost convex. Otherwise, $\mathrm{epi}(f)$ is non - empty and, being convex because of f's convexity, it has a non - empty relative interior (cf. Theorem 6.2 in [72]) so, by Proposition 2.1, it is almost convex.

2.3.3 Properties of the almost convex functions

Next we present some properties of the almost convex functions and some examples that underline the differences between this class of functions and the nearly convex functions.

Theorem 2.5 *(cf. [36]) Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ having non - empty domain. The function f is almost convex if and only if $\bar{f}$ is convex and $\bar{f}(x) = f(x)\ \forall x \in \mathrm{ri}(\mathrm{dom}(\bar{f}))$.*

Proof. "$\Rightarrow$" When f is almost convex, $\bar{f}$ is convex. As $\mathrm{dom}(f) \neq \emptyset$, we have $\mathrm{dom}(\bar{f}) \neq \emptyset$. It is known (cf. [72]) that

$$\mathrm{ri}(\mathrm{epi}(\bar{f})) = \left\{(x, r) : \bar{f}(x) < r, x \in \mathrm{ri}\left(\mathrm{dom}\left(\bar{f}\right)\right)\right\} \tag{2. 3}$$

so, as the definition of the almost convexity includes $\mathrm{ri}\left(\mathrm{epi}\left(\bar{f}\right)\right) \subseteq \mathrm{epi}(f)$, it follows that for any $x \in \mathrm{ri}\left(\mathrm{dom}\left(\bar{f}\right)\right)$ and $\varepsilon > 0$ one has $\left(x, \bar{f}(x) + \varepsilon\right) \in \mathrm{epi}(f)$. Thus $\bar{f}(x) \geq f(x)\ \forall x \in \mathrm{ri}(\mathrm{dom}(\bar{f}))$ and the definition of $\bar{f}$ yields the coincidence of f and $\bar{f}$ over $\mathrm{ri}\left(\mathrm{dom}\left(\bar{f}\right)\right)$.

"$\Leftarrow$" We have $\bar{f}$ convex and $\bar{f}(x) = f(x)\ \forall x \in \mathrm{ri}\big(\mathrm{dom}\,(\bar{f})\big)$. Thus $\mathrm{ri}\big(\mathrm{dom}\,(\bar{f})\big) \subseteq \mathrm{dom}(f)$. By Lemma 2.3 and Remark 2.5 one gets $\mathrm{ri}\big(\mathrm{dom}\,(\bar{f})\big) \subseteq \mathrm{dom}(f)$ if and only if $\mathrm{ri}\big(\mathrm{dom}\,(\bar{f})\big) = \mathrm{ri}(\mathrm{dom}(f))$, therefore this last equality holds. Using this and (2. 3) it follows $\mathrm{ri}\big(\mathrm{epi}\,(\bar{f})\big) = \big\{(x, r) : f(x) < r, x \in \mathrm{ri}(\mathrm{dom}(f))\big\}$, so $\mathrm{ri}\big(\mathrm{epi}\,(\bar{f})\big) \subseteq \mathrm{epi}(f)$. This and the hypothesis $\bar{f}$ convex yield that f is almost convex. $\square$

Remark 2.8 From the previous proof we obtain also that if f is almost convex and has a non - empty domain then $\mathrm{ri}(\mathrm{dom}(f)) = \mathrm{ri}(\mathrm{dom}(\bar{f})) \neq \emptyset$. We have also $\mathrm{ri}(\mathrm{epi}(\bar{f})) \subseteq \mathrm{epi}(f)$, from which, by the definition of $\bar{f}$, follows

$$\mathrm{ri}(\mathrm{cl}(\mathrm{epi}(f))) \subseteq \mathrm{epi}(f) \subseteq \mathrm{cl}(\mathrm{epi}(f)).$$

Applying Lemma 2.3 we get $\mathrm{ri}(\mathrm{epi}(f)) = \mathrm{ri}(\mathrm{cl}(\mathrm{epi}(f))) = \mathrm{ri}(\mathrm{epi}(\bar{f}))$.

In order to avoid confusions between the nearly convex functions and the almost convex functions we give below some examples that show that there is no inclusion between these two classes of functions. Their intersection is not empty, though, as Remark 2.7 states that the convex functions are concomitantly almost convex and nearly convex.

Example 2.2 (i) Let $f : \mathbb{R} \to \mathbb{R}$ be any discontinuous solution of Cauchy's functional equation $f(x + y) = f(x) + f(y)\ \forall x, y \in \mathbb{R}$. For each of these functions, whose existence is guaranteed in [42], one has

$$f\Big(\frac{x+y}{2}\Big) = \frac{f(x) + f(y)}{2}\ \forall x, y \in \mathbb{R},$$

i.e. these functions are nearly convex. None of these functions is convex because of the absence of continuity. We have that $\mathrm{dom}(f) = \mathbb{R} = \mathrm{ri}(\mathrm{dom}(f))$. Suppose f almost convex. Then Theorem 2.5 yields $\bar{f}$ convex and $f(x) = \bar{f}(x)\ \forall x \in \mathbb{R}$. Thus f is convex, but this is false. Therefore f is nearly convex, but not almost convex.

(ii) Consider the set $X = \big([0, 2] \times [0, 2]\big) \setminus \big(\{0\} \times (0, 1)\big)$ and let $g : \mathbb{R}^2 \to \overline{\mathbb{R}}$, $g = \delta_X$. We have $\mathrm{epi}(g) = X \times [0, +\infty)$, so $\mathrm{epi}(\bar{g}) = \mathrm{cl}(\mathrm{epi}(g)) = [0, 2] \times [0, 2] \times [0, +\infty)$. As this is a convex set, $\bar{g}$ is a convex function. We also have $\mathrm{ri}(\mathrm{epi}(\bar{g})) = (0, 2) \times (0, 2) \times (0, +\infty)$, which is clearly contained inside $\mathrm{epi}(g)$. Thus g is almost convex. On the other hand, $\mathrm{dom}(g) = X$ and X is not a nearly convex set, because for any $\alpha \in (0, 1)$ we have $\alpha(0, 1) + (1 - \alpha)(0, 0) = (0, \alpha) \notin X$. By Remark 2.5 it follows that the almost convex function g is not nearly convex.

Using Remark 2.7 and the facts above we see that there are almost convex and nearly functions which are not convex, i.e. both these classes are larger than the one of convex functions.

The following assertion states an interesting and important property of the almost convex functions that is not applicable in general for nearly convex functions.

Theorem 2.6 *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ and $g : \mathbb{R}^m \to \overline{\mathbb{R}}$ be proper almost convex functions. Then the function $F : \mathbb{R}^n \times \mathbb{R}^m \to \overline{\mathbb{R}}$ defined by $F(x, y) = f(x) + g(y)$ is almost convex, too.*

Proof. Consider the linear operator $L : (\mathbb{R}^n \times \mathbb{R}) \times (\mathbb{R}^m \times \mathbb{R}) \to \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}$ defined as $L(x, r, y, s) = (x, y, r + s)$. Let us show first that $L(\mathrm{epi}(f) \times \mathrm{epi}(g)) = \mathrm{epi}(F)$.

Taking the pairs $(x, r) \in \mathrm{epi}(f)$ and $(y, s) \in \mathrm{epi}(g)$ we have $f(x) \leq r$ and $g(y) \leq s$, so $F(x, y) = f(x) + g(y) \leq r + s$, i.e. $(x, y, r + s) \in \mathrm{epi}(F)$. Thus $L(\mathrm{epi}(f) \times \mathrm{epi}(g)) \subseteq \mathrm{epi}(F)$.

On the other hand, for $(x, y, t) \in \text{epi}(F)$ one has $F(x, y) = f(x) + g(y) \leq t$, so $f(x)$ and $g(y)$ are real numbers. It follows $(x, f(x), y, t - f(x)) \in \text{epi}(f) \times \text{epi}(g)$, i.e. $(x, y, t) \in L(\text{epi}(f) \times \text{epi}(g))$ meaning $\text{epi}(F) \subseteq L(\text{epi}(f) \times \text{epi}(g))$.

Therefore $L(\text{epi}(f) \times \text{epi}(g)) = \text{epi}(F)$. We prove that $\text{cl}(\text{epi}(F))$ is convex, which means $\bar{F}$ convex.

Let (x, y, r) and (u, v, s) in $\text{cl}(\text{epi}(F))$. There are two sequences, $(x_k, y_k, r_k)_{k\geq 1}$ and $(u_k, v_k, s_k)_{k\geq 1}$ in $\text{epi}(F)$, the first converging towards (x, y, r) and the second to (u, v, s). Then we have also the sequences of reals $(r_k^1)_{k\geq 1}$, $(r_k^2)_{k\geq 1}$, $(s_k^1)_{k\geq 1}$ and $(s_k^2)_{k\geq 1}$ fulfilling for each $k \geq 1$ the following $r_k^1 + r_k^2 = r_k$, $s_k^1 + s_k^2 = s_k$, $(x_k, r_k^1) \in \text{epi}(f)$, $(y_k, r_k^2) \in \text{epi}(g)$, $(u_k, s_k^1) \in \text{epi}(f)$ and $(v_k, s_k^2) \in \text{epi}(g)$. Let $\lambda \in [0, 1]$. We have, due to the convexity of the lower-semicontinuous hulls of f and g, $(\lambda x_k + (1-\lambda)u_k, \lambda r_k^1 + (1-\lambda)s_k^1) \in \text{cl}(\text{epi}(f)) = \text{epi}(\bar{f})$ and $(\lambda y_k + (1-\lambda)v_k, \lambda r_k^2 + (1-\lambda)s_k^2) \in \text{cl}(\text{epi}(g)) = \text{epi}(\bar{g})$. Further, $(\lambda x_k + (1-\lambda)u_k, \lambda y_k + (1-\lambda)v_k, \lambda r_k + (1-\lambda)s_k) \in L(\text{cl}(\text{epi}(f)) \times \text{cl}(\text{epi}(g))) = L(\text{cl}(\text{epi}(f) \times \text{epi}(g))) \subseteq \text{cl}(L(\text{epi}(f) \times \text{epi}(g)))$ for all $k \geq 1$. Letting k converge towards $+\infty$ we get $(\lambda x + (1-\lambda)u, \lambda y + (1-\lambda)v, \lambda r + (1-\lambda)s) \in \text{cl}(L(\text{epi}(f) \times \text{epi}(g))) = \text{cl}(\text{epi}(F))$. As this happens for any $\lambda \in [0, 1]$ it follows $\text{cl}(\text{epi}(F))$ convex, so $\text{epi}(\bar{F})$ is convex, i.e. $\bar{F}$ is a convex function.

Therefore, in order to obtain that F is almost convex we have to prove only that $\text{ri}(\text{cl}(\text{epi}(F))) \subseteq \text{epi}(F)$. Using some basic properties of the closures and relative interiors and also that f and g are almost convex we have $\text{ri}(\text{cl}(\text{epi}(f) \times \text{epi}(g))) = \text{ri}(\text{cl}(\text{epi}(f)) \times \text{cl}(\text{epi}(g))) = \text{ri}(\text{cl}(\text{epi}(f))) \times \text{ri}(\text{cl}(\text{epi}(g))) \subseteq \text{epi}(f) \times \text{epi}(g)$. Applying the linear operator L to both sides we get $L(\text{ri}(\text{cl}(\text{epi}(f) \times \text{epi}(g)))) \subseteq L(\text{epi}(f) \times \text{epi}(g)) = \text{epi}(F)$. One has $\text{cl}(\text{epi}(f) \times \text{epi}(g)) = \text{cl}(\text{epi}(f)) \times \text{cl}(\text{epi}(g)) = \text{epi}(\bar{f}) \times \text{epi}(\bar{g})$, which is a convex set, so also $L(\text{cl}(\text{epi}(f) \times \text{epi}(g)))$ is convex. As for any linear operator $A : \mathbb{R}^n \to \mathbb{R}^m$ and any convex set $X \subseteq \mathbb{R}^n$ one has $A(\text{ri}(X)) = \text{ri}(A(X))$ (see for instance Theorem 6.6 in [72]), it follows

$$\text{ri}(L(\text{cl}(\text{epi}(f) \times \text{epi}(g)))) = L(\text{ri}(\text{cl}(\text{epi}(f) \times \text{epi}(g)))) \subseteq \text{epi}(F). \tag{2. 4}$$

On the other hand,

$$\text{epi}(F) = L(\text{epi}(f) \times \text{epi}(g)) \subseteq L(\text{cl}(\text{epi}(f) \times \text{epi}(g))) \subseteq \text{cl}(L(\text{epi}(f) \times \text{epi}(g))),$$

so $\text{cl}(L(\text{epi}(f) \times \text{epi}(g))) = \text{cl}(L(\text{cl}(\text{epi}(f) \times \text{epi}(g))))$ and further

$$\text{ri}(\text{cl}(L(\text{epi}(f) \times \text{epi}(g)))) = \text{ri}(\text{cl}(L(\text{cl}(\text{epi}(f) \times \text{epi}(g))))).$$

As for any convex set $X \subseteq \mathbb{R}^n$ $\text{ri}(\text{cl}(X)) = \text{ri}(X)$ (see Theorem 6.3 in [72]), we have $\text{ri}(\text{cl}(L(\text{cl}(\text{epi}(f) \times \text{epi}(g))))) = \text{ri}(L(\text{cl}(\text{epi}\ (f) \times \text{epi}(g))))$, which implies $\text{ri}(\text{cl}(L(\text{epi}(f) \times \text{epi}(g)))) = \text{ri}(L(\text{cl}(\text{epi}(f) \times \text{epi}(g))))$. Using (2. 4) follows $\text{ri}(\text{epi}(\bar{F}))$ $= \text{ri}(\text{cl}(\text{epi}(F))) = \text{ri}(L(\text{cl}(\text{epi}(f) \times \text{epi}(g)))) \subseteq \text{epi}(F)$. Because $\bar{F}$ is a convex function it follows by definition that F is almost convex. □

Corollary 2.1 *Using the previous statement it can be shown that if $f_i : \mathbb{R}^{n_i} \to \overline{\mathbb{R}}$, $i = 1, \ldots, k$, are proper almost convex functions, then $F : R^{n_1} \times \ldots \times \mathbb{R}^{n_k} \to \overline{\mathbb{R}}$, $F(x^1, \ldots, x^k) = \sum_{i=1}^k f_i(x^i)$ is almost convex, too.*

Next we give an example that shows that the property just proven to hold for almost convex functions does not apply for nearly convex functions.

Example 2.3 Consider the sets

$$X_1 = \underset{n\geq 1}{\cup} \left\{\tfrac{k}{2^n} : 0 \leq k \leq 2^n\right\} \qquad \text{and} \qquad X_2 = \underset{n\geq 1}{\cup} \left\{\tfrac{k}{3^n} : 0 \leq k \leq 3^n\right\}.$$

They are both nearly convex, X_1 for $\alpha = 1/2$ and X_2 for $\alpha = 1/3$, for instance. It is easy to notice that δ_{X_1} and δ_{X_2} are nearly convex functions. Taking $F : \mathbb{R}^2 \to \overline{\mathbb{R}}$,

$F(x_1, x_2) = \delta_{X_1}(x_1) + \delta_{X_2}(x_2)$, we have $\mathrm{dom}(F) = X_1 \times X_2$, which is not nearly convex, thus F is not a nearly convex function. We have $(0,0) \in \mathrm{dom}(F)$ and assuming $\mathrm{dom}(F)$ nearly convex with the constant $\bar{\alpha} \in (0,1)$, for all $n \in \mathbb{N}$ and all k satisfying $0 \leq k \leq 2^n$ one gets $(\bar{\alpha}k/2^n, \bar{\alpha}k/3^n) \in \mathrm{dom}(F)$. This yields $\bar{\alpha} \in \mathbb{Q} \cap (0,1)$, so let $\bar{\alpha} = u/v$, with $u < v$ and $u, v \in \mathbb{N}$ having no common divisors. Because $(u/v)(k/2^n) \in X_1$ $\forall n \in \mathbb{N}$ and $\forall k$ such that $0 \leq k \leq 2^n$, it follows that there is some $m \in \mathbb{N}$ such that $m \geq 1$ and $v = 2^m$. Take $k = 1$ and $n = 1$. We also have $(u/2^m)(1/3) \in X_2$, i.e. $u/(3 \cdot 2^m) \in X_2$, which is false. Therefore F is not nearly convex.

2.3.4 Conjugacy and Fenchel duality for almost convex functions

Now we generalize some well - known results concerning the conjugates of convex functions. We prove that they keep their validity when the functions involved are taken almost convex, too. Moreover, these results are proven to stand also when the functions are nearly convex and their epigraphs have non - empty relative interiors.

First we deal with the conjugate of the precomposition with a linear operator (see, for instance, Theorem 16.3 in [72]).

Theorem 2.7 *Let $f : \mathbb{R}^m \to \overline{\mathbb{R}}$ be an almost convex function and $A : \mathbb{R}^n \to \mathbb{R}^m$ a linear operator such that there is some $x' \in \mathbb{R}^n$ satisfying $Ax' \in \mathrm{ri}(\mathrm{dom}(f))$. Then for any $p \in \mathbb{R}^m$ one has*

$$(f \circ A)^*(p) = \min\big\{f^*(q) : A^*q = p\big\}.$$

Proof. We prove first that $(f \circ A)^*(p) = (\bar{f} \circ A)^*(p)$ $\forall p \in \mathbb{R}^n$. By Remark 2.8 we get $Ax' \in \mathrm{ri}(\mathrm{dom}(\bar{f}))$. Assume first that $\bar{f}$ is not proper. Corollary 7.2.1 in [72] yields $\bar{f}(y) = -\infty$ $\forall y \in \mathrm{dom}(\bar{f})$. As $\mathrm{ri}(\mathrm{dom}(\bar{f})) = \mathrm{ri}(\mathrm{dom}(f))$ and $\bar{f}(y) = f(y)$ $\forall y \in \mathrm{ri}(\mathrm{dom}(\bar{f}))$, one has $\bar{f}(Ax') = f(Ax') = -\infty$. It follows easily $(\bar{f} \circ A)^*(p) = (f \circ A)^*(p) = +\infty$.

Now take $\bar{f}$ proper. By the way it is defined one has $(\bar{f} \circ A)(x) \leq (f \circ A)(x)$ $\forall x \in \mathbb{R}^n$ and, by simple calculations, one gets $(\bar{f} \circ A)^*(p) \geq (f \circ A)^*(p)$ for any $p \in \mathbb{R}^n$. Take some $p \in \mathbb{R}^n$ and denote $\beta = (\bar{f} \circ A)^*(p) \in (-\infty, +\infty]$. Assume $\beta \in \mathbb{R}$. We have $\beta = \sup_{x \in \mathbb{R}^n} \big\{p^T x - \bar{f} \circ A(x)\big\}$. Let $\varepsilon > 0$. Then there is an $\bar{x} \in \mathbb{R}^n$ such that $p^T \bar{x} - \bar{f} \circ A(\bar{x}) \geq \beta - \varepsilon$, so $A\bar{x} \in \mathrm{dom}(\bar{f})$. As $Ax' \in \mathrm{ri}(\mathrm{dom}(\bar{f}))$, we get, because of the linearity of A and of the convexity of $\mathrm{dom}(\bar{f})$, by Theorem 6.1 in [72] that for any $\lambda \in (0,1]$ it holds $A((1-\lambda)\bar{x} + \lambda x') = (1-\lambda)A\bar{x} + \lambda Ax' \in \mathrm{ri}(\mathrm{dom}(\bar{f}))$. Applying Theorem 2.5 and using the convexity of $\bar{f}$ we have

$$p^T((1-\lambda)\bar{x} + \lambda x') - f(A((1-\lambda)\bar{x} + \lambda x')) = p^T((1-\lambda)\bar{x} + \lambda x')$$

$$-\bar{f}(A((1-\lambda)\bar{x} + \lambda x')) \geq p^T((1-\lambda)\bar{x} + \lambda x') - (1-\lambda)\bar{f} \circ A(\bar{x})$$

$$-\lambda \bar{f} \circ A(x') = p^T \bar{x} - \bar{f} \circ A(\bar{x}) + \lambda\big[p^T(x' - \bar{x}) - (\bar{f} \circ A(x') - \bar{f} \circ A(\bar{x}))\big].$$

As Ax' and $A\bar{x}$ belong to the domain of the proper function $\bar{f}$, there is a $\bar{\lambda} \in (0,1]$ such that $\bar{\lambda}\big[p^T(x' - \bar{x}) - (\bar{f} \circ A(x') - \bar{f} \circ A(\bar{x}))\big] > -\varepsilon$.

The calculations above lead to

$$(f \circ A)^*(p) \geq p^T((1-\bar{\lambda})\bar{x} + \bar{\lambda}x') - (\bar{f} \circ A)((1-\bar{\lambda})\bar{x} + \bar{\lambda}x') \geq \beta - 2\varepsilon.$$

As ε is an arbitrarily chosen positive number, let it converge towards 0. We get $(f \circ A)^*(p) \geq \beta = (\bar{f} \circ A)^*(p)$. Because the opposite inequality is always true, we get $(f \circ A)^*(p) = (\bar{f} \circ A)^*(p)$.

Consider now the last possible situation, $\beta = +\infty$. Then for any $k \geq 1$ there is an $x_k \in \mathbb{R}^n$ such that $p^T x_k - \bar{f}(Ax_k) \geq k+1$. Thus $Ax_k \in \operatorname{dom}(\bar{f})$ and by Theorem 6.1 in [72] we have, for any $\lambda \in (0,1]$,

$$p^T((1-\lambda)x_k + \lambda x') - f \circ A((1-\lambda)x_k + \lambda x') = p^T((1-\lambda)x_k + \lambda x')$$
$$-\bar{f} \circ A((1-\lambda)x_k + \lambda x') \geq p^T((1-\lambda)x_k + \lambda x') - (1-\lambda)\bar{f} \circ A(x_k)$$
$$-\lambda \bar{f} \circ A(x') \geq p^T x_k - \bar{f} \circ A(x_k) + \lambda\big[p^T(x' - x_k) - (\bar{f} \circ A(x') - \bar{f} \circ A(x_k))\big].$$

Like before, there is some $\bar{\lambda} \in (0,1)$ such that

$$\bar{\lambda}\big[p^T(x' - x_k) - (\bar{f} \circ A(x') - \bar{f} \circ A(x_k))\big] \geq -1.$$

Denoting $z_k = (1-\bar{\lambda})x_k + \bar{\lambda}x'$ we have $z_k \in \mathbb{R}^n$ and $p^T z_k - f \circ A(z_k) \geq k+1-1 = k$. As $k \geq 1$ is arbitrarily chosen, one gets

$$(f \circ A)^*(p) = \sup_{x \in \mathbb{R}^n} \Big\{ p^T x - f \circ A(x) \Big\} = +\infty,$$

so $(f \circ A)^*(p) = +\infty = (\bar{f} \circ A)^*(p)$. Therefore, as $p \in \mathbb{R}^n$ has been arbitrary chosen, we get

$$(f \circ A)^*(p) = (\bar{f} \circ A)^*(p) \ \forall p \in \mathbb{R}^n. \tag{2. 5}$$

By Theorem 16.3 in [72] we have, as $\bar{f}$ is convex and $Ax' \in \operatorname{ri}(\operatorname{dom}(\bar{f})) = \operatorname{ri}(\operatorname{dom}(f))$,

$$(\bar{f} \circ A)^*(p) = \min\big\{(\bar{f})^*(q) : A^*q = p\big\},$$

with the minimum attained at some $\bar{q}$. But $f^* = (\bar{f})^*$ (cf. [72]), so the relation above gives

$$(\bar{f} \circ A)^*(p) = \min\big\{f^*(q) : A^*q = p\big\}.$$

Finally, by (2. 5), this turns into

$$(f \circ A)^*(p) = \min\big\{f^*(q) : A^*q = p\big\},$$

and the minimum is attained at $\bar{q}$. □

The following statement follows from Theorem 2.7 immediately by Proposition 2.1.

Corollary 2.2 *If $f : \mathbb{R}^m \to \overline{\mathbb{R}}$ is a nearly convex function satisfying $\operatorname{ri}(\operatorname{epi}(f)) \neq \emptyset$ and $A : \mathbb{R}^n \to \mathbb{R}^m$ is a linear operator such that there is some $x' \in \mathbb{R}^n$ fulfilling $Ax' \in \operatorname{ri}(\operatorname{dom}(f))$, then for any $p \in \mathbb{R}^m$ one has*

$$(f \circ A)^*(p) = \min\big\{f^*(q) : A^*q = p\big\}.$$

Now comes a statement concerning the conjugate of the sum of finitely many proper functions, which is actually the infimal convolution of their conjugates also when the functions are almost convex functions, provided that the relative interiors of their domains have a point in common.

Theorem 2.8 *(infimal convolution) Let $f_i : \mathbb{R}^n \to \overline{\mathbb{R}}$, $i = 1, \ldots, k$, be proper and almost convex functions whose domains satisfy $\cap_{i=1}^k \operatorname{ri}(\operatorname{dom}(f_i)) \neq \emptyset$. Then for any $p \in \mathbb{R}^n$ we have*

$$(f_1 + \ldots + f_k)^*(p) = \min\left\{ \sum_{i=1}^k f_i^*(p^i) : \sum_{i=1}^k p^i = p \right\}. \tag{2. 6}$$

Proof. Let $F : \mathbb{R}^n \times \ldots \times \mathbb{R}^n \to \overline{\mathbb{R}}$, $F(x^1, \ldots, x^k) = \sum_{i=1}^k f_i(x^i)$. By Corollary 2.1 we know that F is almost convex. We have $\operatorname{dom}(F) = \operatorname{dom}(f_1) \times \ldots \times \operatorname{dom}(f_k)$, so $\operatorname{ri}(\operatorname{dom}(F)) = \operatorname{ri}(\operatorname{dom}(f_1)) \times \ldots \times \operatorname{ri}(\operatorname{dom}(f_k))$. Consider also the linear operator $A : \mathbb{R}^n \to \underbrace{\mathbb{R}^n \times \ldots \times \mathbb{R}^n}_{k}$, $Ax = (\underbrace{x, \ldots, x}_{k})$. The existence of the element $x' \in \cap_{i=1}^k \operatorname{ri}(\operatorname{dom}(f_i))$ gives $(x', \ldots, x')^T \in \operatorname{ri}(\operatorname{dom}(F))$, so $Ax' \in \operatorname{ri}(\operatorname{dom}(F))$. By Theorem 2.7 we have for any $p \in \mathbb{R}^n$

$$(F \circ A)^*(p) = \min\{F^*(q) : A^*q = p\}, \tag{2. 7}$$

with the minimum attained at some $\bar{q} \in \mathbb{R}^n \times \ldots \mathbb{R}^n$. For the conjugates above we have for any $p \in \mathbb{R}^n$

$$(F \circ A)^*(p) = \sup_{x \in \mathbb{R}^n} \left\{ p^T x - \sum_{i=1}^k f_i(x) \right\} = \left(\sum_{i=1}^k f_i \right)^* (p)$$

and for every $q = (p^1, \ldots, p^k) \in \mathbb{R}^n \times \ldots \times \mathbb{R}^n$,

$$F^*(q) = \sup_{\substack{x^i \in \mathbb{R}^n, \\ i=1,\ldots,k}} \left\{ \sum_{i=1}^k (p^i)^T x^i - \sum_{i=1}^k f_i(x^i) \right\} = \sum_{i=1}^k f_i^*(p^i),$$

so, as $A^*q = \sum_{i=1}^k p^i$, (2. 7) delivers (2. 6). □

In [72] the formula (2. 6) is given assuming the functions f_i, $i = 1, \ldots, k$, proper and convex and the intersection of the relative interiors of their domains non - empty. We have proven above that it holds even under the much weaker than convexity assumption of almost convexity imposed on these functions, when the other two conditions, i.e. their properness and the non - emptiness of the intersection of the relative interiors of their domains, stand. As the following assertion states, the formula is valid under the assumption regarding the domains also when the functions are proper and nearly convex, provided that the relative interiors of their epigraphs are non - empty.

Corollary 2.3 *If $f_i : \mathbb{R}^n \to \overline{\mathbb{R}}$, $i = 1, \ldots, k$, are proper nearly convex functions whose epigraphs have non - empty relative interiors and with their domains satisfying $\cap_{i=1}^k \operatorname{ri}(\operatorname{dom}(f_i)) \neq \emptyset$, then for any $p \in \mathbb{R}^n$ one has*

$$(f_1 + \ldots + f_k)^*(p) = \min \left\{ \sum_{i=1}^k f_i^*(p_i) : \sum_{i=1}^k p_i = p \right\}.$$

Next we show that another important conjugacy formula remains true when imposing almost convexity (or near convexity) instead of convexity for the functions in discussion.

Theorem 2.9 *Given two proper almost convex functions $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ and $g : \mathbb{R}^m \to \overline{\mathbb{R}}$ and the linear operator $A : \mathbb{R}^n \to \mathbb{R}^m$ for which is guaranteed the existence of some $x' \in \operatorname{dom}(f)$ satisfying $Ax' \in \operatorname{ri}(\operatorname{dom}(g))$, one has for all $p \in \mathbb{R}^n$*

$$(f + g \circ A)^*(p) = \min \big\{ f^*(p - A^*q) + g^*(q) : q \in \mathbb{R}^m \big\}. \tag{2. 8}$$

Proof. Consider the linear operator $B : \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^m$ defined by $Bz = (z, Az)$ and the function $F : \mathbb{R}^n \times \mathbb{R}^m \to \overline{\mathbb{R}}$, $F(x, y) = f(x) + g(y)$. By Theorem 2.7 F is an almost convex function and we have $\operatorname{dom}(F) = \operatorname{dom}(f) \times \operatorname{dom}(g)$. From the hypothesis one gets

$$Bx' = (x', Ax') \in \operatorname{ri}(\operatorname{dom}(f)) \times \operatorname{ri}(\operatorname{dom}(g)) = \operatorname{ri}(\operatorname{dom}(f) \times \operatorname{dom}(g)) = \operatorname{ri}(\operatorname{dom}(F)),$$

thus $Bx' \in \text{ri}(\text{dom}(F))$. Theorem 2.7 is applicable, leading to

$$(F \circ B)^*(p) = \min\big\{F^*(q_1, q_2) : B^*(q_1, q_2) = p, (q_1, q_2) \in \mathbb{R}^n \times \mathbb{R}^m\big\}$$

where the minimum is attained for any $p \in \mathbb{R}^n$. Because

$$\begin{aligned}(F \circ B)^*(p) &= \sup_{x\in\mathbb{R}^n} \big\{p^T x - F(B(x))\big\} = \sup_{x\in\mathbb{R}^n} \big\{p^T x - F(x, Ax)\big\} \\ &= \sup_{x\in\mathbb{R}^n} \big\{p^T x - f(x) - g(Ax)\big\} = (f + g \circ A)^*(p)\ \forall p \in \mathbb{R}^n,\end{aligned}$$

$$F^*(q_1, q_2) = f^*(q_1) + g^*(q_2)\ \forall (q_1, q_2) \in \mathbb{R}^n \times \mathbb{R}^m$$

and

$$B^*(q_1, q_2) = q_1 + A^*q_2\ \forall (q_1, q_2) \in \mathbb{R}^n \times \mathbb{R}^m,$$

the relation above becomes

$$\begin{aligned}(f + g \circ A)^*(p) &= \inf\big\{f^*(q_1) + g^*(q_2) : q_1 + A^*q_2 = p\big\} \\ &= \min\big\{f^*(p - A^*q_2) + g^*(q_2) : q_2 \in \mathbb{R}^m\big\},\end{aligned}$$

where the minimum is attained for any $p \in \mathbb{R}^n$, i.e. (2. 8) stands. □

Corollary 2.4 *Let the proper nearly convex functions $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ and $g : \mathbb{R}^m \to \overline{\mathbb{R}}$ satisfying* $\text{ri}(\text{epi}(f)) \neq \emptyset$ *and* $\text{ri}(\text{epi}(g)) \neq \emptyset$ *and the linear operator $A : \mathbb{R}^n \to \mathbb{R}^m$ such that there is some $x' \in \text{dom}(f)$ fulfilling $Ax' \in \text{ri}(\text{dom}(g))$. Then* (2. 8) *holds for any $p \in \mathbb{R}^n$ and the minimum is always attained.*

After weakening the conditions under which some widely - used formulae concerning the conjugation of functions take place, we switch to duality where we found important results which hold even when replacing the convexity with almost convexity or near convexity.

The following duality theorem is an immediate consequence of Theorem 2.9 for $p = 0$ in formula (2. 8).

Theorem 2.10 *Given two proper almost convex functions $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ and $g : \mathbb{R}^m \to \overline{\mathbb{R}}$ and the linear operator $A : \mathbb{R}^n \to \mathbb{R}^m$ for which is guaranteed the existence of some $x' \in \text{dom}(f)$ satisfying $Ax' \in \text{ri}(\text{dom}(g))$, one has*

$$\inf_{x\in\mathbb{R}^n} \big[f(x) + g(Ax)\big] = -(f + g \circ A)^*(0) = \max_{q\in\mathbb{R}^m} \big\{ - f^*(A^*q) - g^*(-q)\big\}. \quad (2.\ 9)$$

Remark 2.9 This statement generalizes Corollary 31.2.1 in [72] as we take the functions f and g almost convex instead of convex and, moreover, we remove the lower semicontinuity assumption required for them in the mentioned book. It is easy to notice that when f and g are convex there is no need to consider them moreover lower semicontinuous in order to obtain the formula (2. 9). Let us remind that a proper convex lower semicontinuous function is called in [72] *closed.*

Remark 2.10 Theorem 2.10 states actually the strong duality between the primal problem

$$(P_A) \qquad \inf_{x\in\mathbb{R}^n} \big[f(x) + g(Ax)\big]$$

and its Fenchel dual

$$(D_A) \qquad \sup_{q\in\mathbb{R}^m} \big\{ - f^*(A^*q) - g^*(-q)\big\}.$$

Using Proposition 2.1 and Theorem 2.10 we rediscover the assertion in Theorem 4.1 in [14], which follows.

Corollary 2.5 *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ and $g : \mathbb{R}^m \to \overline{\mathbb{R}}$ two proper nearly convex functions whose epigraphs have non - empty relative interiors and consider the linear operator $A : \mathbb{R}^n \to \mathbb{R}^m$. If there is an $x' \in \mathrm{dom}(f)$ such that $Ax' \in \mathrm{ri}(\mathrm{dom}(g))$, then (2. 9) holds, i.e. $v(P_A) = v(D_A)$, and the dual problem (D_A) has a solution.*

In the end we give a generalization of the well - known Fenchel's duality theorem (Theorem 31.1 in [72]). It follows immediately from Theorem 2.10, when A is the identity mapping.

Theorem 2.11 *Let f and g be proper almost convex functions on $\mathbb{R}^n$ with values in $\overline{\mathbb{R}}$. If $\mathrm{ri}(\mathrm{dom}(f)) \cap \mathrm{ri}(\mathrm{dom}(g)) \neq \emptyset$, one has*

$$\inf_{x\in\mathbb{R}^n} \big[f(x) + g(x)\big] = \max_{q\in\mathbb{R}^n} \big\{ - f^*(q) - g^*(-q)\big\}.$$

When f and g are nearly convex functions we have, as in Theorem 3.1 in [14], the following statement.

Corollary 2.6 *Let f and g be proper nearly convex functions on $\mathbb{R}^n$ with values in $\overline{\mathbb{R}}$. If $\mathrm{ri}(\mathrm{epi}(f)) \neq \emptyset$, $\mathrm{ri}(\mathrm{epi}(g)) \neq \emptyset$ and $\mathrm{ri}(\mathrm{dom}(f)) \cap \mathrm{ri}(\mathrm{dom}(g)) \neq \emptyset$, one has*

$$\inf_{x\in\mathbb{R}^n} \big[f(x) + g(x)\big] = \max_{q\in\mathbb{R}^n} \big\{ - f^*(q) - g^*(-q)\big\}$$

Remark 2.11 The last two assertions give actually the strong duality between the primal problem

$$(P_F) \qquad \inf_{x\in\mathbb{R}^n} \big[f(x) + g(x)\big],$$

and its Fenchel dual

$$(D_F) \qquad \sup_{q\in\mathbb{R}^m} \big\{ - f^*(q) - g^*(-q)\big\}.$$

In both cases we have weakened the initial assumptions required in [72] to guarantee strong duality between (P_F) and (D_F) by asking the functions f and g to be almost convex, respectively nearly convex, instead of convex.

Remark 2.12 Let us notice that the relative interior of the epigraph of a proper nearly convex function f with $\mathrm{ri}(\mathrm{dom}(f)) \neq \emptyset$ may be empty. Take for instance the nearly convex function f in Example 2.2(i) whose effective domain is $\mathbb{R}$. If the relative interior of the epigraph were non - empty, by Proposition 2.1 would follow that f is almost convex, but this does not happen.

Remark 2.13 One may notice that the assumption of near convexity applied to f and g simultaneously does not require the same near convexity constant to be attached to both of these functions.

The following example contains a situation where the classical duality theorem due to Fenchel is not applicable, unlike one of our extensions to it.

Example 2.4 (cf. [14]) Consider the sets

$$\begin{aligned}\mathcal{F} &= \big\{(x_1,x_2)\in\mathbb{R}^2 : x_1>0, x_2>0\big\} \cup \big\{(x_1,0)\in\mathbb{R}^2 : x_1\in\mathbb{Q}, x_1\geq 0\big\}\\ &\cup \big\{(0,x_2)\in\mathbb{R}^2 : x_2\in\mathbb{Q}, x_2\geq 0\big\}\end{aligned}$$

and

$$\mathcal{G} = \big\{(x_1,x_2)\in\mathbb{R}^2 : x_1+x_2<3\big\} \cup \big\{(x_1,x_2)\in\mathbb{R}^2 : x_1,x_2\in\mathbb{Q}, x_1+x_2=3\big\}$$

and some real - valued convex functions defined on $\mathbb{R}^2$, f and g. Both $\mathcal{F}$ and $\mathcal{G}$ are nearly convex sets with $\alpha = 1/2$ playing the role of the constant required in the definition, but not convex. We are interested in treating by Fenchel duality the problem

$$(P_q) \qquad \inf_{x=(x_1,x_2)\in\mathcal{F}\cap\mathcal{G}} \big[f(x)+g(x)\big],$$

i.e. we would like to obtain the infimal objective value of (P_q) by using the conjugate functions of f and g. A Fenchel type dual problem may be attached to this problem, but the conditions under which the primal and the dual have equal optimal objective values are not known to us as we cannot apply Fenchel's duality theorem because $\mathcal{F}\cap\mathcal{G}$ is not convex. Let us define now the functions

$$\tilde{f}:\mathbb{R}^2\to\overline{\mathbb{R}},\ \tilde{f}(x)=\begin{cases} f(x), & \text{if } x\in\mathcal{F},\\ +\infty, & \text{otherwise,}\end{cases}$$

and

$$\tilde{g}:\mathbb{R}^2\to\overline{\mathbb{R}},\ \tilde{g}(x)=\begin{cases} g(x), & \text{if } x\in\mathcal{G},\\ +\infty, & \text{otherwise.}\end{cases}$$

The function $\tilde{f}$ and $\tilde{g}$ are clearly nearly convex (with the constant 1/2), but not convex since $\operatorname{dom}(\tilde{f})=\mathcal{F}$ is not convex. Therefore we are not yet in the situation to apply Fenchel's duality theorem for the problem

$$(P_q') \qquad \inf_{x=(x_1,x_2)\in\mathcal{F}\cap\mathcal{G}} \big[\tilde{f}(x)+\tilde{g}(x)\big],$$

which is actually equivalent to (P_q), but let us check whether the extension we have given in Corollary 2.6 is applicable. The condition concerning the non - emptiness of the intersection of the relative interiors of the domains of the functions involved is satisfied in this case since

$$\begin{aligned} \operatorname{ri}(\operatorname{dom}(\tilde{f}))\cap\operatorname{ri}(\operatorname{dom}(\tilde{g})) &= \operatorname{ri}(\mathcal{F})\cap\operatorname{ri}(\mathcal{G})\\ &= (0,+\infty)\times(0,+\infty)\cap\big\{(x,y)\in\mathbb{R}^2 : x+y<3\big\}, \end{aligned}$$

which is non - empty since the element $(1,1)$, for instance, is contained in both sets.

Regarding the relative interiors of the epigraphs of $\tilde{f}$ and $\tilde{g}$, it is not difficult to check that $\big((1,1), f(1,1)+1\big)\in\operatorname{int}(\operatorname{epi}(\tilde{f}))=\operatorname{ri}(\operatorname{epi}(\tilde{f}))$ and $\big((1,1), g(1,1)+1\big)\in\operatorname{int}(\operatorname{epi}(\tilde{g}))=\operatorname{ri}(\operatorname{epi}(\tilde{g}))$.

Therefore the conditions in the hypothesis of Corollary 2.6 are fulfilled for $\tilde{f}$ and $\tilde{g}$, respectively. So we can apply the statement and we get that

$$\begin{aligned} v(P_q) &= \inf_{x\in\mathcal{F}\cap\mathcal{G}}\big[f(x)+g(x)\big] = \inf_{x\in\mathbb{R}^2}\big[\tilde{f}(x)+\tilde{g}(x)\big]\\ &= \max_{u^*\in\mathbb{R}^2}\big\{-\tilde{f}^*(u^*)-\tilde{g}^*(-u^*)\big\} = \max_{u^*\in\mathbb{R}^n}\big\{-f^*_{\mathcal{F}}(u^*)-g^*_{\mathcal{G}}(-u^*)\big\}. \end{aligned}$$

As proven in Example 2.2 there are almost convex functions which are not convex, so our Theorems 2.7 - 2.11 really extend some results in [72]. An example given in [14] and cited above shows that also the Corollaries 2.2 - 2.6 generalize indeed the corresponding results from Rockafellar's book [72], as a nearly convex function whose epigraph has non - empty interior is not necessarily convex.

We finish this chapter with an example in game theory, where we found a small application of one of our results. Applications of Fenchel's duality theorem in game theory were already found, see for instance [7]. Knowing also the connections between Lagrange duality involving generalized convex functions and game theory (see, for instance, [68] or [69]), we give an application of Corollary 2.6 in this field opening the gate into the direction of Fenchel duality.

Example 2.5 (cf. [11]) Consider a two-person zero-sum game, where D and C are the sets of strategies for the players I and II, respectively, and $L : C \times D \to \mathbb{R}$ is the so - called payoff - function. By $\beta^L = \sup_{d\in D} \inf_{c\in C} L(c,\ d)$ and $\alpha^L = \inf_{c\in C} \sup_{d\in D} L(c, d)$ we denote the lower, respectively the upper values of the game (C, D, L). As the minmax inequality

$$\beta^L = \sup_{d\in D} \inf_{c\in C} L(c, d) \leq \inf_{c\in C} \sup_{d\in D} L(c, d) = \alpha^L, \tag{2. 10}$$

is always fulfilled, the challenge is to find weak conditions which guarantee equality in the relation above. Near convexity and its generalizations played an important role in it, as Paeck's paper [68] or his book [69] show. Having an optimization problem with geometrical and inequality constraints,

$$(P_b) \qquad \inf_{\substack{x\in X,\\ w(x)\leqq 0}} v(x),$$

where $X \subseteq \mathbb{R}^n$, $w : \mathbb{R}^n \to \mathbb{R}^k$, $v : \mathbb{R}^n \to \overline{\mathbb{R}}$, the Lagrangian attached to it, considered as a pay - off function of some two - person zero - sum game, is $L : X \times \mathbb{R}^k_+ \to \overline{\mathbb{R}}$, $L(x, \lambda) = v(x) + \lambda^T w(x)$. In the works cited above there are some results where sufficient conditions under which the strong duality occurs between (P_b) and its Lagrange dual

$$(D_b) \qquad \sup_{\lambda\geqq 0} \inf_{x\in X} L(x, \lambda) = \sup_{\lambda\geqq 0} \inf_{x\in X} [v(x) + \lambda^T w(x)],$$

which is nothing else than the equality in (2. 10). Let us mention that within these statements the usual convexity assumptions are replaced by near convexity.

Coming to the problem treated in Corollary 2.6, for f and g proper nearly convex functions,

$$(P_F) \qquad \inf_{x\in\mathbb{R}^n} \big[f(x) + g(x)\big],$$

one can define the Lagrangian attached to it by (cf. [31])

$$L : \mathbb{R}^n \times \mathbb{R}^n \to \overline{\mathbb{R}},\ L(x, u) = u^T x + g(x) - f^*(u).$$

As

$$\sup_{u\in\mathbb{R}^n} \inf_{x\in\mathbb{R}^n} L(x, u) = \sup_{u\in\mathbb{R}^n} \big\{ - g^*(-u) - f^*(u)\big\}$$

and

$$\inf_{x\in\mathbb{R}^n} \sup_{u\in\mathbb{R}^n} L(x, u) = \inf_{x\in\mathbb{R}^n} \big[g(x) + f^{**}(x)\big] \leq \inf_{x\in\mathbb{R}^n} [f(x) + g(x)],$$

and since under the hypotheses of Corollary 2.6 one has

$$\max_{u\in\mathbb{R}^n} \big\{ - g^*(-u) - f^*(u)\big\} = \inf_{x\in\mathbb{R}^n} [f(x) + g(x)],$$

we get, taking also into consideration (2. 10),

$$\max_{u\in\mathbb{R}^n} \inf_{x\in\mathbb{R}^n} L(x, u) = \inf_{x\in\mathbb{R}^n} \sup_{u\in\mathbb{R}^n} L(x, u).$$

The solution to the dual problem can be seen as an optimal strategy for the game having this Lagrangian as payoff - function.

Chapter 3

Fenchel - Lagrange duality versus geometric duality

Geometric programming, established during the late 1960's due to PETERSON together with DUFFIN AND ZENER has enjoyed an increasing popularity and usage that continues up to today, as papers based on it still get published (see, for instance, [79]). The most important works in geometric programming are the original book of DUFFIN, PETERSON AND ZENER [30] which deals with the so - called posynomial geometric programming and PETERSON's seminal article [71] on generalized geometric programming. Because the posynomial geometric programming is a special case of the latter and it can be applied only to some very special classes of problems and the objective function of the generalized geometric primal problem is very complicated, JEFFERSON AND SCOTT considered in [48] a simplified version of the generalized geometric programming. Then they and some other authors treated by means of geometric programming various optimization problems, see [48–52, 75–82]. An extension of the posynomial geometric programming is the so - called signomial programming, which is a special case of the generalized geometric programming, too, thus not so relevant from the theoretical point of view, but still used in various applications. Applications of geometric programming can be found in various fields, from the theoretical problems treated by JEFFERSON AND SCOTT in [48–52, 75–82] to the practical applications mentioned, for instance, in [30] or [34].

About Fenchel - Lagrange duality one could read for the first time in WANKA AND BOŢ's article [92], then its applications in multiobjective optimization were investigated by the same authors in works like [8, 18, 19, 89, 91, 93]. Despite being recently introduced, the areas of applicability of the Fenchel - Lagrange duality cover alongside the multiobjective optimization also Farkas type results, theorems of the alternative, set containment (cf. [22]), DC programming (cf. [21]), semidefinite programming (cf. [93]), convex composite programming (cf. [9]) or fractional programming (cf. [12, 91]). One of the most interesting features of the Fenchel - Lagrange duality is that it contains as a special case the classical geometric duality, improving its results. Thus all the problems treated by geometric programming can be dealt with via Fenchel - Lagrange duality, easier and with better results. Moreover when the generalized geometric programming problem is treated via perturbations, one obtains the dual problem easier and strong duality under more general conditions.

3.1 The geometric dual problem obtained via perturbations

Peterson's classical work [71] presents a complete duality treatment for geometric programs. These are convex optimization problems whose objective and constraint functions have some special forms. We show further that the duals he introduced there by using the geometric Lagrangian and geometric inequalities can be obtained also by using the perturbation approach already mentioned in the previous chapter. Moreover, the assertions concerning strong duality and optimality conditions are given here under weaker conditions than in the papers dealing with geometric duality.

3.1.1 Motivation

The generalized geometric programming due to Peterson deals with primal optimization problems having very complicated objective functions. Moreover, the functions and sets involved are asked to be convex and lower semicontinuous, respectively convex and closed. On the other hand the perturbation theory already presented on short in the previous chapter guarantees a dual problem and strong duality under some constraint qualification also when the functions and sets involved are only convex, not also lower semicontinuous, respectively closed. This brought the idea of trying to determine a dual problem to the initial geometric primal problem via perturbations (see [13]). As this dual turned out to be exactly the geometric dual introduced in [71] by some geometric inequalities and using the geometric Lagrangian, the next step was to compare the conditions that bring strong duality. Not surprisingly, the constraint qualification we obtained turned out to be slightly weaker than the one in [71]. Therefore we have proven that when treated via perturbations instead of using geometric inequalities, the primal geometric problem gets the same dual problem, subject to simpler calculations and, moreover, the strong duality arises under more general conditions.

3.1.2 The unconstrained case

First we treat the general unconstrained geometric programming problem. Let the proper function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, with $\mathrm{dom}(f) = X \subseteq \mathbb{R}^n$. There is given also a closed cone $N \subseteq \mathbb{R}^n$. The unconstrained geometric programming problem (here called primal problem) is

$$(A_{gu}) \qquad \inf_{x \in X \cap N} f(x).$$

Peterson attached in [71] the following dual to the problem (A_{gu}),

$$(B_{gu}) \qquad \inf_{y \in D \cap N^*} f^*(y),$$

where $D = \mathrm{dom}(f^*)$. Let us consider the perturbation function

$$\Phi : \mathbb{R}^n \times \mathbb{R}^n \to \overline{\mathbb{R}},\ \Phi(x,p) = \begin{cases} f(x+p), & \text{if } x \in N, x+p \in X, p \in \mathbb{R}^n, \\ +\infty, & \text{otherwise.} \end{cases}$$

As the perturbation function Φ fulfills (2. 1), according to the theory sketched in the previous section, the dual problem to (A_{gu}) is

$$(D_{gu}) \qquad \sup_{p^* \in \mathbb{R}^n} \{-\Phi^*(0,p^*)\}.$$

The conjugate function of the perturbation function is, at some $(x^*, p^*) \in \mathbb{R}^n \times \mathbb{R}^n$,

$$\begin{aligned}
\Phi^*(x^*, p^*) &= \sup_{\substack{x \in \mathbb{R}^n, \\ p \in \mathbb{R}^n}} \Big\{ {x^*}^T x + {p^*}^T p - \Phi(x, p) \Big\}, \\
&= \sup_{\substack{x \in N, \\ p \in \mathbb{R}^n, \\ x+p \in X}} \Big\{ {x^*}^T x + {p^*}^T p - f(x+p) \Big\}, \\
&= \sup_{\substack{x \in N, \\ t \in X}} \Big\{ {x^*}^T x + {p^*}^T (t - x) - f(t) \Big\}.
\end{aligned}$$

As we have to take $x^* = 0$ in order to calculate the dual problem, it follows, for $p^* \in \mathbb{R}^n$,

$$\Phi^*(0, p^*) = \sup_{\substack{x \in N, \\ t \in X}} \Big\{ {p^*}^T t - {p^*}^T x - f(t) \Big\} = \sup_{x \in N} -{p^*}^T x + \sup_{t \in X} \Big\{ {p^*}^T t - f(t) \Big\},$$

whence, as $\mathrm{dom}(f) = X$,

$$\Phi^*(0, p^*) = f^*(p^*) + \begin{cases} 0, & \text{if } p^* \in N^*, \\ +\infty, & \text{otherwise.} \end{cases}$$

Therefore the dual problem we obtain is

$$(D_{gu}) \qquad \sup_{p^* \in N^* \cap D} \{-f^*(p^*)\}.$$

which, transformed into a minimization problem turns out to be, when removing the leading minus, exactly Peterson's dual (B_{gu}) (see [71]). As mentioned before, the weak duality regarding the problems (A_{gu}) and (D_{gu}) always holds, while for the strong duality we have the following statement (see also Theorem 2.11).

Theorem 3.1 *(strong duality) If X is a convex set, f a function convex on X, N a closed convex cone and the condition $\mathrm{ri}(N) \cap \mathrm{ri}(X) \neq \emptyset$ is fulfilled, then the strong duality between (A_{gu}) and (D_{gu}) holds, i.e. (D_{gu}) has an optimal solution and the optimal objective values of the primal and dual problem coincide.*

Remark 3.1 In [71] the conditions regarding the strong duality are posed on the dual problem, in which case f has to be, moreover, lower semicontinuous and the dual problem's infimum must be finite.

Let us also present necessary and sufficient optimality conditions regarding the unconstrained geometric program.

Theorem 3.2 *(optimality conditions)*
(a) *Assume the hypotheses of Theorem 3.1 fulfilled and let $\bar{x}$ be an optimal solution to (A_{gu}). Then the strong duality between the primal problem and its dual holds and there exists an optimal solution $\overline{p^*}$ to (D_{gu}) satisfying the following optimality conditions*

(i) $f(\bar{x}) + f^*(\overline{p^*}) = \overline{p^*}^T \bar{x}$,

(ii) $\overline{p^*}^T \bar{x} = 0$.

(b) *Let $\bar{x}$ be a feasible solution to (A_{gu}) and $\overline{p^*}$ one to (D_{gu}) satisfying the optimality conditions (i) and (ii). Then $\bar{x}$ turns out to be an optimal solution to the primal problem, $\overline{p^*}$ one to the dual and the strong duality between the two problems holds.*

Proof. (a) From Theorem 3.1 we know that the strong duality holds and the dual problem has an optimal solution Let it be $\overline{p^*} \in N^* \cap D$. Therefore, it holds

$$f(\bar{x}) + f^*(\overline{p^*}) = 0.$$

From the Fenchel - Young inequality it is known that

$$f(\bar{x}) + f^*(\overline{p^*}) \geq \overline{p^*}^T \bar{x},$$

while

$$\overline{p^*}^T \bar{x} \geq 0$$

since $\overline{p^*} \in N^*$ and $\bar{x} \in N$. Hence it holds

$$f(\bar{x}) + f^*(\overline{p^*}) \geq \overline{p^*}^T \bar{x} \geq 0,$$

but, since we have equality between the first and the last member of the expression above, both these inequalities must be fulfilled as equalities. So the equality must hold in the previous two expressions, i.e. the optimality conditions are true.

(b) The optimality conditions imply

$$f(\bar{x}) + f^*(\overline{p^*}) = \overline{p^*}^T \bar{x} = 0.$$

So the assertion holds. □

3.1.3 The constrained case

Further the primal problem becomes more complicated, as some constraints appear and also the objective function is not so simple anymore. The following preliminaries are required.

Let there be the finite index sets I and J. For $t \in \{0\} \cup I \cup J$, the following functions are considered

$$g_t : X_t \to \mathbb{R},$$

with $\emptyset \neq X_t \subseteq \mathbb{R}^{n_t}$, as well as the independent vector variables $x^t \in \mathbb{R}^{n_t}$. There are also the sets

$$D_t = \Big\{ y^t \in \mathbb{R}^{n_t} : \sup_{x^t \in X_t} \Big\{ {y^t}^T x^t - g_t(x^t) \Big\} < +\infty \Big\},$$

which are the domains of the conjugates regarding the sets they are defined on, namely X_t, of the functions $g_t, t \in \{0\} \cup I \cup J$, respectively, and an independent vector variable $k = (k_1, \ldots, k_{|J|})^T$. With x^I one denotes the Cartesian product of the vector variables $x^i, i \in I$, while x^J denotes the same thing for $x^j, j \in J$. Hence, $x = (x^0, x^I, x^J)$ is an independent vector variable in $\mathbb{R}^n$, where $n = n_0 + \sum_{i \in I} n_i + \sum_{j \in J} n_j$. Finally, let there be a non - empty closed cone $N \subseteq \mathbb{R}^n$, the sets

$$X_j^+ = \Big\{ (x^j, k_j) \quad : \quad \text{either } k_j = 0 \text{ and } \sup_{d^j \in D_j} {d^j}^T x^j < \infty$$
$$\text{or } k_j > 0 \text{ and } x^j \in k_j X_j \Big\}, \quad j \in J,$$

and

$$X = \big\{ (x, k) : x^t \in X_t, t \in \{0\} \cup I, (x^j, k_j) \in X_j^+, j \in J \big\},$$

and, for $j \in J$, the functions $g_j : X_j^+ \to \mathbb{R}$,

$$g_j^+(x^j, k_j) \quad = \quad \begin{cases} \sup\limits_{d^j \in D_j} {d^j}^T x^j, & \text{if } k_j = 0 \text{ and } \sup\limits_{d^j \in D_j} {d^j}^T x^j < \infty, \\ k_j g_j\big(\frac{1}{k_j} x^j\big), & \text{if } k_j > 0 \text{ and } x^j \in k_j X_j, \end{cases}$$

which are the so - called *homogenous extensions* of the functions g_j, $j \in J$.

Peterson (cf. [71]) studied the minimization of the following objective function

$$f : X \to \mathbb{R}, \; f(x,k) = g_0(x^0) + \sum_{j \in J} g_j^+(x^j, k_j)$$

with the variables lying in the feasible set

$$S = \{(x,k) \in X : x \in N, \; g_i(x^i) \leq 0 \; \forall i \in I\}.$$

Thus, the problem he studied is

$$(A_{gc}) \qquad \inf_{(x,k) \in S} f(x,k),$$

further referred to as the *primal generalized geometric programming problem.*

To introduce a dual problem to it, one has to introduce the sets

$$D = \big\{(y^0, y^I, y^J, \lambda) : y^t \in D_t, t \in \{0\} \cup J, (y^i, \lambda_i) \in D_i^+, i \in I\big\},$$

and

$$\begin{aligned} D_i^+ \;\; &= \Big\{(y^i, \lambda_i) : \;\; \text{either } \lambda_i = 0 \text{ and } \sup_{c^i \in X_i} {y^i}^T c^i < \infty, \\ &\qquad\qquad\qquad\quad \text{or } \lambda_i > 0 \text{ and } y^i \in \lambda_i D_i \Big\}, \quad i \in I, \end{aligned}$$

and the family of functions $\big\{h_t : D_t \to \mathbb{R} : t \in \{0\} \cup I \cup J\big\}$, each of its members being the restriction to the domain of the conjugate regarding the set X_t of the function g_t, i.e. $h_t(y^t) = g_{X_t}^*(y^t)$ for all $y^t \in D_t$, $t \in \{0\} \cup I \cup J$, and for $i \in I$ we consider also the functions $h_i^+ : D_i^+ \to \mathbb{R}$,

$$h_i^+(y^i, \lambda_i) \;\; = \;\; \begin{cases} \sup\limits_{c^i \in X_i} {y^i}^T c^i, & \text{if } \lambda_i = 0 \text{ and } \sup\limits_{c^i \in X_i} {y^i}^T c^i < \infty, \\ \lambda_i h_i\big(\frac{1}{\lambda_i} y^i\big), & \text{if } \lambda_i > 0 \text{ and } y^i \in \lambda_i D_i. \end{cases}$$

In [71] there is introduced the following dual problem to (A_{gc}),

$$(B_{gc}) \qquad \inf_{(y,\lambda) \in T} h(y, \lambda),$$

with the feasible set

$$T = \{(y, \lambda) \in D : y \in N^*, h_j(y^j) \leq 0, j \in J\},$$

where the objective function is

$$h : D \to \mathbb{R}, \;\; h(y, \lambda) = h_0(y^0) + \sum_{i \in I} h_i^+(y^i, \lambda_i).$$

In the following part we demonstrate that this dual problem can be developed also by using the method based on perturbations already presented within this thesis. Like before, we introduce the following extension of the objective function

$$F : \mathbb{R}^n \times \mathbb{R}^{|J|} \to \overline{\mathbb{R}},$$

$$F(x,k) \;\; = \;\; \begin{cases} f(x,k), & \text{if } \; (x,k) \in X, x \in N, g_i(x^i) \leq 0, i \in I, \\ +\infty, & \text{otherwise.} \end{cases}$$

Thus we can write the problem (A_{gc}) equivalently as

$$(A'_{gc}) \qquad \inf_{(x,k)\in\mathbb{R}^n\times\mathbb{R}^{|J|}} F(x,k).$$

Let us introduce now the perturbation function associated to our problem,

$$\Phi : \mathbb{R}^n \times \mathbb{R}^{|J|} \times \mathbb{R}^n \times \mathbb{R}^{|I|} \to \overline{\mathbb{R}},$$

$$\Phi(x,k,p,v) = \begin{cases} f(x+p,k), & \text{if } x \in N, (x+p,k) \in X, \text{ and} \\ & \quad g_i(x^i+p^i) \leq v^i, i \in I, \\ +\infty, & \text{otherwise.} \end{cases}$$

It is obvious that $\Phi(x,k,0,0) = F(x,k) \ \ \forall (x,k) \in \mathbb{R}^n \times \mathbb{R}^{|J|}$, thus the dual problem to (A'_{gc}), so also to (A_{gc}), is

$$(D_{gc}) \qquad \sup_{\substack{p^*\in\mathbb{R}^n,\\ v^*\in\mathbb{R}^{|I|}}} \big\{ -\Phi^*(0,0,p^*,v^*)\big\},$$

where

$$\Phi^*(x^*,k^*,p^*,v^*) = \sup_{\substack{x,p\in\mathbb{R}^n,\\ k\in\mathbb{R}^{|J|},\\ v\in\mathbb{R}^{|I|}}} \Big\{ {x^*}^T x + {k^*}^T k + {p^*}^T p + {v^*}^T v - \Phi(x,k,p,v)\Big\}$$

$$= \sup_{\substack{x\in N, k\in\mathbb{R}^{|J|}_+,\\ p\in\mathbb{R}^n, v\in\mathbb{R}^{|I|},\\ g_i(x^i+p^i)\leq v^i, i\in I,\\ (x+p,k)\in X}} \Big\{ {x^*}^T x + {k^*}^T k + {p^*}^T p + {v^*}^T v - g_0(x^0+p^0) - \sum_{j\in J} g_j^+(x^j+p^j,k_j)\Big\},$$

with the dual variables $x^* = (x^{*0}, x^{*I}, x^{*J})$ and $p^* = (p^{*0}, p^{*I}, p^{*J})$. Introducing the new variables $z = x + p$ and $y = v - g_I(z^I)$, with $g_I(z^I) = (g_i(z^i))^T_{i\in I}$, there follows

$$\begin{aligned} \Phi^*(x^*,k^*,p^*,v^*) &= \sup_{\substack{x\in N, k\in\mathbb{R}^{|J|}_+,\\ (z,k)\in X, y\in\mathbb{R}^{|I|},\\ y^i\geq 0, i\in I}} \Big\{ {x^*}^T x + {k^*}^T k + {p^*}^T (z-x) \\ &+ {v^*}^T \big(y + g_I(z^I)\big) - g_0(x^0) - \sum_{j\in J} g_j^+(z^j,k_j)\Big\} \\ &= \sum_{i\in I} \sup_{y^i\geq 0} {v^{*i}}^T y^i + \sup_{z^0\in X_0} \Big\{ {p^{*0}}^T z^0 - g_0(z^0)\Big\} \\ &+ \sum_{i\in I} \sup_{z^i\in X_i} \Big\{ {p^{*i}}^T z^i + v^{*i} g_i(z^i)\Big\} + \sup_{x\in N} (x^* - p^*)^T x \\ &+ \sum_{j\in J} \sup_{(z^j,k_j)\in X_j^+} \Big\{ {p^{*j}}^T z^j + {k_j^*}^T k_j - g_j^+(z^j,k_j)\Big\}. \end{aligned}$$

In order to determine the dual problem (D_{gc}) according to the general theory we must take further $x^* = 0$ and $k^* = 0$. Also, we use the following results that arise from definitions or simple calculations. We have

$$\begin{aligned} \sup_{z^i\in X_i} \Big\{ {p^{*i}}^T z^i + v^{*i} g_i(z^i)\Big\} &= \begin{cases} \sup\limits_{z^i\in X_i} {p^{*i}}^T z^i, & \text{if } v^{*i} = 0, {p^{*i}}^T z^i < \infty, \\ -v^{*i} h_i\big(\frac{1}{-v^{*i}} p^{*i}\big), & \text{if } v^{*i} \neq 0, p^{*i} \in -v^{*i} D_i, \end{cases} \\ &= h_i^+(p^{*i}, -v^{*i}), i \in I, \end{aligned}$$

and

$$\sup_{z^0\in X_0}\left\{{p^{*0}}^T z^0 - g_0(z^0)\right\} = h_0(p^{*0}).$$

It is also clear that it holds

$$\sup_{y^i\geq 0} {v^{*i}}^T y^i = \begin{cases} 0, & \text{if } v^{*i}\leq 0, \\ +\infty, & \text{otherwise,}\end{cases}, i\in I,$$

while from the definition of the dual cone we have

$$\sup_{x\in N} -{p^*}^T x = \begin{cases} 0, & \text{if } p^*\in N^*, \\ +\infty, & \text{otherwise.}\end{cases}$$

Further we calculate the values of the terms summed after $j\in J$ in the last stage of the formula of $\Phi^*(0,0,p^*,v^*)$, splitting the calculations into two branches. When $k_j>0$ we have

$$\begin{aligned}\sup_{(z^j,k_j)\in X_j^+}\left\{{p^{*j}}^T z^j - g_j^+(z^j,k_j)\right\} &= \sup_{(z^j,k_j)\in X_j^+}\left\{{p^{*j}}^T z^j - k_j g_j\left(\frac{1}{k_j}z^j\right)\right\}, \\ &= \sup_{k_j>0} k_j h_j(p^{*j}) \\ &= \begin{cases} 0, & \text{if } h_j(p^{*j})\leq 0, p^{*j}\in D_j, \\ +\infty, & \text{otherwise.}\end{cases}\end{aligned}$$

When $h_j(p^{*j})\leq 0$ and $p^{*j}\in D_j$, the case $k_j=0$ leads to

$$\sup_{(z^j,k_j)\in X_j^+}\left\{{p^{*j}}^T z^j - g_j^+(z^j,k_j)\right\} = \sup_{(z^j,0)\in X_j^+}\left\{{p^{*j}}^T z^j - \sup_{d^j\in D_j} {d^j}^T z^j\right\} = 0,$$

so we can conclude that for every $j\in J$ it holds

$$\sup_{(z^j,k_j)\in X_j^+}\left\{{p^*}^T z - g_j^+(z^j,k_j)\right\} = \begin{cases} 0, & \text{if } h_j(p^{*j})\leq 0 \text{ and } p^{*j}\in D_j, \\ +\infty, & \text{otherwise.}\end{cases}$$

The dual problem can be simplified, denoting $\lambda = -v^*$, to

$$(D_{gc}) \qquad \sup_{\substack{p^{t*}\in D_t, t\in\{0\}\cup J, \\ (p^{*i},\lambda^i)\in D_i^+, i\in I, \\ h_j(p^{*j})\leq 0, j\in J, \\ p^*\in N^*}} \left\{ - h_0(p^{*0}) - \sum_{i\in I} h_i^+(p_i^*,\lambda_i)\right\},$$

which, transformed into a minimization problem turns, after removing the leading minus, into (using the notations in [71]),

$$(D'_{gc}) \qquad \inf_{(p^*,\lambda)\in T}\left\{h_0(p^{*0}) + \sum_{i\in I} h_i^+(p^{*i},\lambda_i)\right\},$$

i.e. exactly the dual introduced by Peterson, (B_{gc}).

Weak duality regarding the problems (A_{gc}) and (D_{gc}) always holds, while for strong duality we need to introduce some supplementary conditions.

First, let us consider that the sets X_t, $t\in\{0\}\cup I\cup J$ are convex. Each function g_t is taken convex on X_t, $t\in\{0\}\cup I\cup J$. The cone N needs to be closed and convex, too. We have to consider also that the sets X_j are closed and the functions g_j, $j\in J$, are lower semicontinuous. This last property, alongside the convexity, assures (cf. [72]) that, for each $j\in J$, the functions g_j and h_j are a pair of conjugate functions, each regarding the other's definition domain, convex on the sets they are defined on and lower semicontinuous. This fact allows us to characterize the sets X_j in the following way

$$X_j = \left\{ x^j \in \mathbb{R}^{n_j} : \sup_{d^j \in D_j} \left\{ {d^j}^T x^j - h_j(d^j) \right\} < +\infty \right\}, \quad j \in J.$$

Using this characterization, it follows that the set X_j^+ is convex and the function g_j^+ is convex on X_t^+, for all $j \in J$.

The constraint qualification we use in order to achieve strong duality for the mentioned pair of dual problems is

$$(CQ_{gc}) \qquad \exists (x', k') \in \mathrm{ri}(N) \times \mathrm{int}(\mathbb{R}_+^{|J|}) : \begin{cases} x'^0 \in \mathrm{ri}(X_0), \\ x'^i \in \mathrm{ri}(X_i), \\ g_i(x'^i) \leq 0, i \in L_{gc}, \\ g_i(x'^i) < 0, i \in I \backslash L_{gc}, \\ x'^j \in k'_j \,\mathrm{ri}(X_j), j \in J, \end{cases}$$

where $i \in L_{gc}$ if $i \in I$ and g_i is the restriction to X_i of an affine function. We are now ready to formulate the strong duality theorem, whose proof is similar to the one of Theorem 2.2.

Theorem 3.3 *(strong duality) If the conditions introduced above concerning the sets X_t, $t \in \{0\} \cup I \cup J$, the functions g_t, $t \in \{0\} \cup I \cup J$, and the cone N are fulfilled and the constraint qualification (CQ_{gc}) holds, then we have strong duality between (A_{gc}) and (D_{gc}).*

Remark 3.2 In [71] the constraint qualification regarding the strong duality is posed on the dual problem, while we choose to consider it on the primal problem. An advantage brought by our approach is that the sets X_t and the functions $g_t, t \in \{0\} \cup I$, do not have to be assumed moreover lower semicontinuous like in [71].

Remark 3.3 The closeness property of g_j and X_j, $j \in J$, is necessary in order to prove that g_j^+ and X_j^+ are convex, $\forall j \in J$, as Theorem 31.1 in [72] requires the existence of convexity for all the functions and sets involved in the primal problem.

From this strong duality statement we can conclude necessary and sufficient optimality conditions for the generalized geometric programming problem (A_{gc}).

Theorem 3.4 *(optimality conditions)*
(a) Assume the hypotheses of Theorem 3.3 fulfilled and let $(\bar{x}, \bar{k})$ be an optimal solution to (A_{gc}), where $\bar{x} = \left(\bar{x}^0, \bar{x}^I, \bar{x}^J\right)$. Then the strong duality between the primal problem and its dual holds and there exists an optimal solution $\left(\overline{p^*}, \bar{\lambda}\right)$ to (D_{gc}), with $\overline{p^*} = \left(\overline{p^*}^0, \overline{p^*}^I, \overline{p^*}^J\right)$, satisfying the following optimality conditions

(i) $g_0(\bar{x}^0) + h_0(\overline{p^*}^0) = {\overline{p^*}^0}^T \bar{x}^0,$

$$\text{(ii)} \begin{cases} g_j\left(\frac{1}{\bar{k}_j} \bar{x}^j\right) + h_j(\overline{p^*}^j) = \left(\overline{p^*}^j\right)^T \left(\frac{1}{\bar{k}_j} \bar{x}^j\right) \text{ and } h_j(\overline{p^*}^j) = 0, & \text{if } \bar{k}_j \neq 0, \\ \sup\limits_{d^j \in D_j} {d^j}^T \bar{x}^j = {\overline{p^*}^j}^T \bar{x}^j, & \text{if } \bar{k}_j = 0, \end{cases} \quad j \in J,$$

$$\text{(iii)} \begin{cases} g(\bar{x}^i) + h_i\left(\frac{1}{\bar{\lambda}_i} \overline{p^*}^i\right) = \left(\frac{1}{\bar{\lambda}_i} \overline{p^*}^i\right)^T \bar{x}^i \text{ and } g_i(\bar{x}^i) = 0, & \text{if } \bar{\lambda}_i \neq 0, \\ \sup\limits_{c^i \in X_i} {\overline{p^*}^i}^T c^i = {\overline{p^*}^i}^T \bar{x}^i, & \text{if } \bar{\lambda}_i = 0, \end{cases} \quad i \in I,$$

(iv) ${\overline{p^*}}^T \bar{x} = 0.$

(b) Let $(\bar{x}, \bar{k})$ be a feasible solution to (A_{gc}), with $\bar{x} = \left(\bar{x}^0, \bar{x}^I, \bar{x}^J\right)$ and $\left(\overline{p^*}, \bar{\lambda}\right)$ one to (D_{gc}), with $\overline{p^*} = \left(\overline{p^*}^0, \overline{p^*}^I, \overline{p^*}^J\right)$, satisfying the optimality conditions $(i) - (iv)$. Then $(\bar{x}, \bar{k})$ turns out to be an optimal solution to the primal problem, $\left(\overline{p^*}, \bar{\lambda}\right)$ one to the dual and the strong duality between the two problems holds.

Proof. (a) From Theorem 3.3 we know that the strong duality holds and the dual problem has an optimal solution. Let it be $(\overline{p^*}^0, \overline{p^*}^I, \overline{p^*}^J, \bar{\lambda})$ and we denote $\overline{p^*} = \left(\overline{p^*}^0, \overline{p^*}^I, \overline{p^*}^J\right)$.

Therefore, it holds

$$g_0(\bar{x}^0) + \sum_{j \in J} g_j^+(\bar{x}^j, \bar{k}_j) + h_0(\overline{p^*}^0) + \sum_{i \in I} h_i^+(\overline{p^*}^i, \bar{\lambda}_i) = 0,$$

rewritable as

$$g_0(\bar{x}^0) + h_0(\overline{p^*}^0) + \sum_{\substack{i \in I, \\ \bar{\lambda}_i \neq 0}} \bar{\lambda}_i h_i\left(\frac{1}{\bar{\lambda}_i}\overline{p^*}^i\right) + \sum_{\substack{i \in I, \\ \bar{\lambda}_i = 0}} \sup_{c^i \in X_i} \overline{p^*}^{i^T} c^i$$
$$+ \sum_{\substack{j \in J, \\ \bar{k}_j \neq 0}} \bar{k}_j g_j\left(\frac{1}{\bar{k}_j}\bar{x}^j\right) + \sum_{\substack{j \in J, \\ \bar{k}_j = 0}} \sup_{d^j \in D_j} d^{j^T} \bar{x}^j = 0.$$

Adding and subtracting some terms in the left - hand side and inverting the members of the equality, we obtain

$$\begin{aligned}
0 \;=\; & [g_0(\bar{x}^0) + h_0(\overline{p^*}^0) - \overline{p^*}^{0^T}\bar{x}^0] + \sum_{\substack{i \in I, \\ \bar{\lambda}_i \neq 0}} \left[\bar{\lambda}_i h_i\left(\frac{1}{\bar{\lambda}_i}\overline{p^*}^i\right) + \bar{\lambda}_i g_i(\bar{x}^i) - \overline{p^*}^{i^T}\bar{x}^i\right] \\
+ & \sum_{\substack{i \in I, \\ \bar{\lambda}_i = 0}} \left[\sup_{c^i \in X_i} \overline{p^*}^{i^T} c^i - \overline{p^*}^{i^T}\bar{x}^i\right] + \sum_{\substack{j \in J, \\ \bar{k}_j \neq 0}} \left[\bar{k}_j g_j\left(\frac{1}{\bar{k}_j}\bar{x}^j\right) + \bar{k}_j h_j(\overline{p^*}^j) - \overline{p^*}^{j^T}\bar{x}^j\right] \\
+ & \sum_{\substack{j \in J, \\ \bar{k}_j = 0}} \left[\sup_{d^j \in D_j} d^{j^T}\bar{x}^j - \overline{p^*}^{j^T}\bar{x}^j\right] \\
+ & (\overline{p^*}^0, \overline{p^*}^I, \overline{p^*}^J)^T(\bar{x}^0, \bar{x}^I, \bar{x}^J) - \sum_{\substack{i \in I, \\ \bar{\lambda}_i \neq 0}} \bar{\lambda}_i g_i(\bar{x}^i) - \sum_{\substack{j \in J, \\ \bar{k}_j \neq 0}} \bar{k}_j h_j(\overline{p^*}^j). \qquad (3.\ 1)
\end{aligned}$$

Let us prove now that all the terms summed in the right - hand side of (3. 1) are non - negative.

Applying the Fenchel - Young inequality, we get

$$g_0(\bar{x}^0) + h_0(\overline{p^*}^0) \geq \overline{p^*}^{0^T}\bar{x}^0,$$

$$\bar{k}_j g_j\left(\frac{1}{\bar{k}_j}\bar{x}^j\right) + \bar{k}_j h_j(\overline{p^*}^j) \geq \overline{p^*}^{j^T}\left(\bar{k}_j \frac{1}{\bar{k}_j}\bar{x}^j\right) = \overline{p^*}^{j^T}\bar{x}^j, j \in J : \bar{k}_j \neq 0,$$

and

$$\bar{\lambda}_i g_i(\bar{x}^i) + \bar{\lambda}_i h_i\left(\frac{1}{\bar{\lambda}_i}\overline{p^*}^i\right) \geq \bar{\lambda}_i \frac{1}{\bar{\lambda}_i}\overline{p^*}^{i^T}\bar{x}^i = \overline{p^*}^{i^T}\bar{x}^i, i \in I : \bar{\lambda}_i \neq 0.$$

On the other hand, it is obvious that

$$\sup_{d^j \in D_j} d^{j^T}\bar{x}^j \geq \overline{p^*}^{j^T}\bar{x}^j, j \in J : \bar{k}_j = 0,$$

and

$$\sup_{c^i \in X_i} \overline{p^*}^{iT} c^i \geq \overline{p^*}^{iT} \bar{x}^i, i \in I : \bar{\lambda}_i = 0.$$

Since $\overline{p^*} \in N^*$, it follows also that $\overline{p^*}^T \bar{x} \geq 0$. Moreover, from the feasibility conditions it follows that $g_i(\bar{x}^i) \leq 0, \bar{\lambda}_i \geq 0, i \in I$, so $-\sum_{\substack{i \in I, \\ \bar{\lambda}_i \neq 0}} \bar{\lambda}_i g_i(\bar{x}^i) \geq 0$. Also, $h_j(\overline{p^*}^j) \leq 0, \bar{k}_j \neq 0, j \in J$, implies $-\sum_{\substack{j \in J, \\ \bar{k}_j \neq 0}} \bar{k}_j h_j(\overline{p^*}^j) \geq 0$. Therefore it follows that in (3. 1) all the terms are greater than or equal to zero, while their sum is zero, so all of them must be equal to zero, i.e. the inequalities obtained above are fulfilled as equalities. So, (iv) is true and the other optimality conditions, $(i) - (iii)$, hold, too.

(b) The calculations above can be carried out in reverse order and the assertion arises easily. □

Remark 3.4 We mention that (b) applies without any convexity assumption as well as constraint qualification. So, the sufficiency of the optimality conditions $(i) - (iv)$ is true in the most general case.

3.2 Geometric programming duality as a special case of Fenchel - Lagrange duality

In the previous section we dealt with the generalized geometric duality. Because of its very intricate formulation of the objective function it has been and is still used in practice in a simpler form, namely by taking the index set J empty. This simplified geometric duality has been used by many authors to treat various convex optimization problems. Among the authors who dealt extensively in their works with this type of duality we mention Scott and Jefferson who co - wrote more than twenty papers where geometric duality is employed in different purposes. We cite here some of them, namely [48–52, 75–82] and in the next section we show how these problems can be easier treated via Fenchel - Lagrange duality.

3.2.1 Motivation

The problems treated via geometric programming in the literature are not always suitable for this. See for instance the papers we have already mentioned, [48–52, 75–82], where various optimization problems are trapped into the format of the primal geometric problem by very artificial reformulations. These reformulations bring additional variables to the primal problem and all the variables of the reformulated problem have to take values also inside some complicated cones, that need to be constructed, too. We prove that the geometric duality is a special case of the Fenchel - Lagrange duality, i.e. the geometric dual problem is the Fenchel - Lagrange dual of the geometric primal, obtained easier. Moreover, in the mentioned papers the functions involved are taken convex and lower semicontinuous, while the sets dealt with in the problems are considered non - empty, closed and convex for optimality purposes, and we show that strong duality and necessary and sufficient optimality conditions may be obtained under weaker assumptions, i.e. when the sets are taken only non - empty convex and the functions convex on the sets where they are defined on.

We took seven problems treated by Scott and Jefferson via geometric programming duality and we dealt with them via Fenchel - Lagrange duality. One may notice in each case that our results are obtained in a simpler manner than in the original papers, being moreover more general and complete than the ones due to the mentioned authors. Therefore it is not improper to claim that all the applications of geometric programming duality can be considered also as ones of the Fenchel - Lagrange duality, considerably enlarging its areas of applicability.

3.2.2 Fenchel - Lagrange duality for the geometric program

The primal geometric optimization problem is

$$(P_g) \qquad \inf_{\substack{x=(x^0,x^1,\dots,x^k)\in X_0\times X_1\times\ldots\times X_k,\\ g^i(x^i)\leq 0, i=1,\dots,k,\\ x\in N}} g^0(x^0),$$

where $X_i \subseteq \mathbb{R}^{l_i}$, $i=0,\dots,k$, $\sum_{i=0}^k l_i = n$, are convex sets, $g^i : X_i \to \mathbb{R}$, $i=0,\dots,k$, are functions convex on the sets they are defined on and $N \subseteq \mathbb{R}^n$ is a non - empty closed convex cone.

We consider first a special case of the primal problem (P) which is still more general than the geometric primal problem (P_g)

$$(P_N) \qquad \inf_{\substack{x\in X,\\ g(x)\leqq 0,\\ x\in N}} f(x),$$

where N is a non - empty closed convex cone in $\mathbb{R}^n$, X a non - empty convex subset of $\mathbb{R}^n$, $f:\mathbb{R}^n \to \overline{\mathbb{R}}$ is convex and $g: X\to\mathbb{R}^k$ with $g=(g_1,\dots,g_k)^T$ and each g_j is convex on X, $j=1,\dots,k$. Its Fenchel - Lagrange dual problem is

$$(D_N) \qquad \sup_{\substack{p\in\mathbb{R}^n,\\ q\in\mathbb{R}^k_+}} \big\{ -f^*(p) - (q^Tg)^*_{X\cap N}(-p)\big\}.$$

The constraint qualification that is sufficient for the existence of strong duality in this case is (cf. (CQ'_o)), with the notations introduced before,

$$(CQ_N) \quad \exists x' \in \mathrm{ri}(X\cap N)\cap \mathrm{ri}(\mathrm{dom}(f)) : \begin{cases} g_i(x')\leq 0, & \text{if } i\in L_N,\\ g_i(x')<0, & \text{if } i\in\{1,\dots,k\}\backslash L_N,\end{cases}$$

where L_N is defined analogously to L_{gc}, but for the problem in discussion here, (P_N). Because the presence of the cone N in the formula of the dual is not so desired, we need to find an alternative formulation for this dual problem, given in the following strong duality statement. In order to do this we also use a stronger constraint qualification, namely

$$(CQ'_N) \ \exists x' \in \mathrm{ri}(X)\cap\mathrm{ri}(N)\cap \mathrm{ri}(\mathrm{dom}(f)) : \begin{cases} g_i(x')\leq 0, & \text{if } i\in L_N,\\ g_i(x')<0, & \text{if } i\in\{1,\dots,k\}\backslash L_N,\end{cases}$$

whose fulfillment guarantees the satisfaction of (CQ_N), too.

Theorem 3.5 *(strong duality) When the constraint qualification (CQ'_N) is satisfied and $v(P_N)\in\mathbb{R}$, there is strong duality between the primal problem (P_N) and the equivalent formulation of its dual*

$$(D'_N) \qquad \sup_{\substack{p\in\mathbb{R}^n, q\in\mathbb{R}^k_+,\\ t\in N^*}} \Big\{ -f^*(p) + \inf_{x\in X}\Big[(p-t)^Tx + q^Tg(x)\Big]\Big\}.$$

Proof. (P_N) being a special case of the problem (P), like (D_N) of its dual (D) and because under (CQ'_N) one has $\mathrm{ri}(X)\cap\mathrm{ri}(N)=\mathrm{ri}(X\cap N)$, strong duality is valid for (P_N) and (D_N). Thus the problem (D_N) has the optimal solution $(\bar{p},\bar{q})$.

Let us rewrite the term containing N in the formulation of the dual in the following way

$$(\bar{q}^T g)^*_{X\cap N}(-\bar{p}) = -\inf_{x\in X\cap N}\Big[\bar{p}^T x + \bar{q}^T g(x)\Big] = -\inf_{x\in\mathbb{R}^n}\Big[\bar{p}^T x + \widetilde{\bar{q}^T g}(x) + \delta_N(x)\Big],$$

where the function $\widetilde{\bar{q}^T g}$ is defined like in the proof of Theorem 2.2.

By the definition of the conjugate function, the right - hand side of the relation above is equal to $\Big((\bar{p}^T\,\cdot\,) + \widetilde{\bar{q}^T g} + \delta_N\Big)^*(0)$, which, applying Theorem 20.1 in [72], can be written as

$$\Big((\bar{p}^T\,\cdot\,) + \widetilde{\bar{q}^T g} + \delta_N\Big)^*(0) = \min_{t\in\mathbb{R}^n}\Big[\Big((\bar{p}^T\,\cdot\,) + \widetilde{\bar{q}^T g}\Big)^*(t) + \delta_N^*(-t)\Big].$$

Since $\delta_N^*(-t) = 0$ if $t\in N^*$ and $\delta_N^*(-t) = +\infty$ otherwise, it follows, using moreover the definition of the conjugate function and taking into consideration the way the function $\widetilde{\bar{q}^T g}$ was given and that $v(P_N) = v(D_N)\in\mathbb{R}$,

$$\min_{t\in\mathbb{R}^n}\Big[\Big((\bar{p}^T\,\cdot\,) + \widetilde{\bar{q}^T g}\Big)^*(t) + \delta_N^*(-t)\Big] = \min_{t\in N^*}\sup_{x\in X}\Big\{t^T x - \bar{p}^T x - \bar{q}^T g(x)\Big\}.$$

The expression in the right - hand side can be rewritten as $-\max_{t\in N^*}\inf_{x\in X}\big[(\bar{p}-t)^T x + \bar{q}^T g(x)\big]$ and the calculations above lead to

$$(\bar{q}^T g)^*_{X\cap N}(-\bar{p}) = -\max_{t\in N^*}\inf_{x\in X}\Big[(\bar{p}-t)^T x + \bar{q}^T g(x)\Big] = -\inf_{x\in X}\Big[(\bar{p}-\bar{t})^T x + \bar{q}^T g(x)\Big],$$

where the maximum in the expression above is attained at $\bar{t}\in N^*$. It is clear that the dual problem (D_N) becomes (D'_N) and strong duality between (P_N) and (D'_N) is certain, i.e. $v(P_N) = v(D'_N)$ and (D'_N) has an optimal solution $(\bar{p},\bar{q},\bar{t})$. Let us mention that when (CQ'_N) fails it is possible to appear a gap between the optimal objective values of the problems (D_N) and (D'_N). □

Remark 3.5 One can obtain the dual problem (D'_N) also by perturbations, in a similar way we obtained (D) in the previous chapter by including the cone constraint in the constraint function (cf. [17, 89, 90]). We refer further to (D'_N) as the dual problem of (P_N). Moreover, it can be equivalently written as

$$(D''_N)\qquad \sup_{\substack{p\in\mathbb{R}^n, q\in\mathbb{R}^k_+,\\ t\in N^*}}\Big\{-f^*(p) - (q^T g)^*_X(t-p)\Big\},$$

but because of what will follow we prefer to have it formulated as in Theorem 3.5.

The necessary and sufficient optimality conditions are derived from the ones obtained in the general case, thus we skip the proof of the following statement.

Theorem 3.6 *(optimality conditions)*
(a) If the constraint qualification (CQ'_N) is fulfilled and the primal problem (P_N) has an optimal solution $\bar{x}$, then the dual problem (D'_N) has an optimal solution $(\bar{p},\bar{q},\bar{t})$ and the following optimality conditions are fulfilled

(i) $f(\bar{x}) + f^*(\bar{p}) = \bar{p}^T\bar{x}$,

(ii) $\inf_{x\in X}\Big[(\bar{p}-\bar{t})^T x + \bar{q}^T g(x)\Big] = \bar{p}^T\bar{x}$,

(iii) $\bar{q}^T g(\bar{x}) = 0$,

(iv) $\bar{t}^T\bar{x} = 0.$

(b) *If $\bar{x}$ is a feasible point to the primal problem (P_N) and $(\bar{p}, \bar{q}, \bar{t})$ is feasible to the dual problem (D'_N) fulfilling the optimality conditions $(i) - (iv)$, then there is strong duality between (P_N) and (D'_N) and the mentioned feasible points turn out to be optimal solutions.*

Now we show how these results can be applied in geometric programming. With a suitable choice of the functions and the sets involved in the problem (P_N) it becomes (P_g). The proper selection of the mentioned elements follows

$$\begin{cases} X = \mathbb{R}^{l_0} \times X_1 \times \ldots \times X_k, \\ f : \mathbb{R}^n \to \overline{\mathbb{R}}, \quad g_i : X \to \mathbb{R}, i = 1, \ldots, k, \\ f(x) = \begin{cases} g^0(x^0), & \text{if } x \in X_0 \times \mathbb{R}^{n-l_0}, \\ +\infty, & \text{otherwise}, \end{cases} \\ g_i(x) = g^i(x^i), i = 1, \ldots, k, x = (x^0, x^1, \ldots, x^k) \in X. \end{cases}$$

Now we can write the Fenchel - Lagrange dual problem to (P_g) (cf. (D'_N))

$$(D_g) \qquad \sup_{\substack{p \in \mathbb{R}^n, q \in \mathbb{R}^k_+, \\ t \in N^*}} \left\{ - f^*(p) + \inf_{x \in \mathbb{R}^{l_0} \times X_1 \times \ldots \times X_k} \left[(p-t)^T x + \sum_{i=1}^k q_i g^i(x^i) \right] \right\}.$$

The conjugate of f is

$$\begin{aligned} f^*(p) &= \sup_{x \in \mathbb{R}^n} \left\{ p^T x - f(x) \right\} \\ &= \sup_{x = (x^0, x^1, \ldots, x^k) \in X_0 \times \mathbb{R}^{l^1} \times \ldots \times \mathbb{R}^{l_k}} \left\{ \sum_{i=0}^k {p^i}^T x^i - g^0(x^0) \right\} \\ &= \begin{cases} g^{0*}_{X_0}(p^0), & \text{if } p^i = 0, i = 1, \ldots, k, \\ +\infty, & \text{otherwise}, \end{cases} \end{aligned}$$

if we consider $p = (p^0, p^1, \ldots, p^k) \in \mathbb{R}^{l_0} \times \mathbb{R}^{l_1} \times \ldots \times \mathbb{R}^{l_k}$. As the infimum that appears is separable into a sum of infima, the dual becomes

$$(D_g) \quad \sup_{\substack{p^0 \in \mathbb{R}^{l_0}, \\ q \in \mathbb{R}^k_+, \\ t \in N^*}} \left\{ - g^{0*}_{X_0}(p^0) + \inf_{x^0 \in \mathbb{R}^{l_0}} (p^0 - t^0)^T x^0 + \sum_{i=1}^k \inf_{x^i \in X_i} \left[- {t^i}^T x^i + q_i g^i(x^i) \right] \right\},$$

if we consider $t = (t^0, t^1, \ldots, t^k)^T \in \mathbb{R}^{l_0} \times \mathbb{R}^{l_1} \times \ldots \times \mathbb{R}^{l_k}$. As

$$\inf_{x^0 \in \mathbb{R}^{l_0}} (p^0 - t^0)^T x^0 = \begin{cases} 0, & \text{if } p^0 = t^0, \\ -\infty, & \text{otherwise}, \end{cases}$$

the dual problem to (P_g) turns into

$$(D_g) \qquad \sup_{\substack{q \in \mathbb{R}^k_+, \\ t \in N^*}} \left\{ - g^{0*}_{X_0}(t^0) - \sum_{i=1}^k \sup_{x^i \in X_i} \left[{t^i}^T x^i - q_i g^i(x^i) \right] \right\}.$$

This is exactly the geometric dual problem encountered in all the cited papers due to Scott and Jefferson, written without resorting to the homogenous extension of the conjugate functions that can replace the suprema in (D_g).

The constraint qualification sufficient to guarantee the validity of strong duality for this pair of problems, derived from (CQ'_N), is

$$(CQ_g)\ \ \exists x' = \left(x'^0, x'^1, \ldots, x'^k\right) \in \mathrm{ri}(N) : \begin{cases} g^i(x'^i) \leq 0, & \text{if } i \in L_g, \\ g^i(x'^i) < 0, & \text{if } i \in \{1, \ldots, k\} \backslash L_g, \\ x'^i \in \mathrm{ri}(X_i), & \text{for } i = 0, \ldots, k, \end{cases}$$

where L_g is defined analogously to L_{gc}. The strong duality statement concerning the primal geometric programming problem and its dual follows.

Theorem 3.7 *(strong duality) The validity of the constraint qualification (CQ_g) is sufficient to guarantee strong duality regarding (P_g) and (D_g) when $v(P_g)$ is finite.*

Remark 3.6 The cited papers of the mentioned authors do not assert trenchantly any strong duality statement, containing just the optimality conditions, while for the background of their achievement the reader is referred to [71]. There all the functions are taken moreover lower semicontinuous and the sets involved are postulated as being closed, alongside their convexity assumptions that proved to be sufficient in our proofs when the constraint qualification is fulfilled. Moreover, the possibility to impose a milder constraint qualification regarding the affine functions whose restrictions to the considered set are among the constraint functions is not taken into consideration at all.

The necessary and sufficient optimality conditions concerning (P_g) and (D_g) spring directly from the previous statement.

Theorem 3.8 *(optimality conditions)*
(a) *If the constraint qualification (CQ_g) is fulfilled and the primal problem (P_g) has an optimal solution $\bar{x} = \left(\bar{x}^0, \bar{x}^1, \ldots, \bar{x}^k\right)$, then the dual problem (D_g) has an optimal solution $(\bar{q}, \bar{t})$, with $\bar{q} = \left(\bar{q}_1, \ldots, \bar{q}_k\right)^T$ and $\bar{t} = \left(\bar{t}_0, \ldots, \bar{t}_k\right)^T$ and the following optimality conditions are fulfilled*

(i) $g^0(\bar{x}^0) + g_{X_0}^{0*}(\bar{t}^0) = \bar{t}^{0T}\bar{x}^0$,

(ii) $\left(\bar{q}_i g_i\right)_{X_i}^*(\bar{t}_i) = \bar{t}^{iT}\bar{x}^i$, $i = 1, \ldots, k$,

(iii) $\bar{q}_i g^i(\bar{x}^i) = 0$, $i = 0, \ldots, k$,

(iv) $\bar{t}^T\bar{x} = 0$.

(b) *If $\bar{x}$ is a feasible point to the primal problem (P_g) and $(\bar{q}, \bar{t})$ is feasible to the dual problem (D_g) fulfilling the optimality conditions $(i) - (iv)$, then there is strong duality between (P_g) and (D_g) and the mentioned feasible points turn out to be optimal solutions.*

Remark 3.7 The optimality conditions we derived are equivalent to the ones displayed by Scott and Jefferson in the cited papers.

Before going further, let us sum up the statements regarding (P_g). It is the primal geometric problem used by Scott and Jefferson in their cited papers and, more, it is a particular case of a special case of the initial primal problem (P) we considered. In all the invoked papers the mentioned authors present the geometric dual problem to the primal and give the necessary and sufficient optimality conditions that are true under assumptions of convexity and lower semicontinuity regarding the functions, respectively of non - emptiness, convexity and closedness concerning the sets involved there, together with a constraint qualification. We established the same dual problem to the primal exploiting the Fenchel - Lagrange duality we presented earlier. Strong duality and optimality conditions are revealed to stand in much weaker circumstances, i. e. the lower semicontinuity can be removed from the initial assumptions concerning the functions involved and sets they are defined on do not have to be taken closed. Moreover, the constraint qualifications can be generalized and weakened, respectively.

3.3 Comparisons on various applications given in the literature

Now let us review some of the problems treated during the last quarter of century by Scott and Jefferson, sometimes together with Jorjani or Wang, by means of geometric programming. All these problems were artificially trapped into the framework required by geometric programming by introducing new variables in order to separate implicit and explicit constraints and building some cone where the new vector-variable is forced to lie. Then their dual problems arose from the general theory developed by Peterson (see [71]) and the optimality conditions came out from the same place. We determine the Fenchel - Lagrange dual problem for each problem, then we specialize the adequate constraint qualification and state the strong duality assertion followed by the optimality conditions all without proofs as they are direct consequences of Theorems 3.7 and 3.8. One may notice that even if the functions are taken lower semicontinuous and the sets are considered closed in the original papers we removed these redundant properties, as we have proven that strong duality and optimality conditions stand even without their presence when the corresponding convexity assumptions are made and the sufficient constraint qualification is valid. We have chosen six problems that we have considered more interesting, but also the problems in [49–52, 75, 81] may benefit from the same treatment. We mention moreover that the last subsection is dedicated to the well - known posynomial geometric programming which is undertaken into our duality theory, too. Other papers of Scott and Jefferson treat some problems by means of posynomial geometric programming, so we might have included some of these problems here, too.

3.3.1 Minmax programs (cf. [78])

The first problem we deal with is the minmax program

$$(P_1) \qquad \inf_{\substack{x\in X,\\ b\leqq Ax,\\ g(x)\leqq 0}} \max_{i=1,\dots,I} f_i(x),$$

with the convex functions $f_i : \mathbb{R}^n \to \overline{\mathbb{R}}$, $\mathrm{dom}(f_i) = X$, $i = 1, \dots, I$, the non - empty convex set $X \subseteq \mathbb{R}^n$, the vector function $g = (g_1, \dots, g_J)^T : X \to \mathbb{R}^J$ where $g_j : X \to \mathbb{R}$ is convex on X for any $j \in \{1, \dots, J\}$, the matrix $A \in \mathbb{R}^{m\times n}$ and the vector $b \in \mathbb{R}^m$. In the original paper the functions $f_i, i = 1, \dots, I$, and g are taken also lower semicontinuous and the set X is required to be moreover closed, but strong duality is valid in more general circumstances, i.e. without these assumptions. To treat the problem (P_1) with the method presented in the second section, it is rewritten as

$$(P_1) \qquad \inf_{\substack{x\in X, s\in\mathbb{R},\\ b-Ax\leqq 0, g(x)\leqq 0,\\ f_i(x)-s\leq 0, i=1,\dots,I}} s.$$

The Fenchel - Lagrange dual problem to (P_1) is, considering the objective function $u : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}, u(x, s) = s$ and $q^f = (q^f_1, \dots, q^f_I)^T$

$$(D_1) \qquad \sup_{\substack{p^x\in\mathbb{R}^n, p^s\in\mathbb{R},\\ q^l\in\mathbb{R}^m_+, q^f\in\mathbb{R}^I_+, q^g\in\mathbb{R}^J_+}} \Bigg\{ -u^*(p^x, p^s) + \inf_{\substack{x\in X,\\ s\in\mathbb{R}}} \Big[{p^x}^T x + {p^s}^T s$$

$$+\sum_{i\in I} q_i^f (f_i(x)-s) + {q^g}^T g(x) + {q^l}^T (b-Ax)\Big]\Big\}.$$

Computing the conjugate of the objective function we get

$$u^*(p^x,p^s) = \sup_{\substack{x\in\mathbb{R}^n,\\ s\in\mathbb{R}}} \left\{(p^x,p^s)^T(x,s)-s\right\} = \begin{cases} 0, & \text{if } p^s=1, p^x=0,\\ +\infty, & \text{otherwise.}\end{cases}$$

Noticing that the infimum in (D_1) is separable into a sum of two infima, one concerning $s\in\mathbb{R}$, the other $x\in X$, the dual problem turns into

$$(D_1) \quad \sup_{\substack{q^l\in\mathbb{R}^m_+, q^f\in\mathbb{R}^I_+,\\ q^g\in\mathbb{R}^J_+}} \left\{ \inf_{x\in X}\left[\sum_{i\in I} q_i^f f_i(x) + {q^g}^T g(x) - {q^l}^T Ax\right] + \inf_{s\in\mathbb{R}}\left[s - s\sum_{i\in I} q_i^f\right] + {q^l}^T b\right\}.$$

The second infimum is equal to 0 when $\sum_{i\in I} q_i^f = 1$, otherwise having the value $-\infty$, while the first, transformed into a supremum, can be viewed as a conjugate function regarding to the set X. Applying Theorem 20.1 in [72] and denoting $q^g = (q_1^g,\dots,q_J^g)^T$, the dual problem becomes

$$(D_1) \quad \sup_{\substack{q^l\in\mathbb{R}^m_+, q^g\in\mathbb{R}^J_+,\\ q^f\in\mathbb{R}^I_+, \sum_{i=1}^I q_i^f=1,\\ \sum_{i=1}^I u_i + \sum_{j=1}^J v_j = A^T q^l}} \left\{ {q^l}^T b - \sum_{i=1}^I (q_i^f f_i)^*(u_i) - \sum_{j=1}^J (q_j^g g_j)^*_X(v_j)\right\}.$$

identical to the dual problem found in [78]. A sufficient circumstance to be able to formulate the strong duality assertion is the following constraint qualification, where the set L_1 is considered analogously to L_{gc} before,

$$(CQ_1) \quad \exists x' \in \mathrm{ri}(X): \begin{cases} b \leqq Ax', \\ g_j(x') \le 0, & \text{if } j\in L_1,\\ g_j(x') < 0, & \text{if } j\in\{1,\dots,J\}\backslash L_1.\end{cases}$$

Theorem 3.9 *(strong duality) If the constraint qualification (CQ_1) is satisfied, then the strong duality between (P_1) and (D_1) is assured.*

Since the optimality conditions are not delivered in [78], here they are, determined via our method.

Theorem 3.10 *(optimality conditions)*
(a) If the constraint qualification (CQ_1) is fulfilled and $\bar{x}$ is an optimal solution to (P_1), then strong duality between the problems (P_1) and (D_1) is attained and the dual problem has an optimal solution $(\bar{q}^l,\bar{q}^f,\bar{q}^g,\bar{u},\bar{v})$, where $\bar{u}=(\bar{u}_1,\dots,\bar{u}_I)^T$ and $\bar{v}=(\bar{v}_1,\dots,\bar{v}_J)^T$, satisfying the following optimality conditions

(i) $f_i(\bar{x}) - \max_{i=1,\dots,I} f_i(\bar{x}) = 0$ *if* $\bar{q}_i^f > 0$, $i=1,\dots,I$,

(ii) $\bar{q}^{lT}(b-A\bar{x}) = 0$,

(iii) $\bar{q}^{gT} g(\bar{x}) = 0$,

(iv) $\left(\bar{q}_i^f f_i\right)^*(\bar{u}_i) + \bar{q}_i^f f_i(\bar{x}) = \bar{u}_i^T\bar{x}, i=1,\dots,I$,

(v) $\left(\bar{q}_j^g g_j\right)^*_X(\bar{v}_j) + \bar{q}_j^g g_j(\bar{x}) = \bar{v}_j^T\bar{x}, j=1,\dots,J$.

(b) Having a feasible solution $\bar{x}$ to the primal problem and one $(\bar{q}^l, \bar{q}^f, \bar{q}^g, \bar{u}, \bar{v})$ to the dual satisfying the optimality conditions $(i)-(v)$, then the mentioned feasible solutions turn out to be optimal solutions to the corresponding problems and strong duality stands.

3.3.2 Entropy constrained linear programs (cf. [82])

A minute exposition of the way how the Fenchel - Lagrange duality is applicable to the problem treated in [82] is available in [13]. In the following we present the most important facts concerning this matter. The entropy inequality constrained optimization problem

$$(P_2) \qquad \inf_{\substack{b \leqq Ax, \\ -\sum\limits_{i=1}^{n} x_i \ln x_i \geq H, \\ \sum\limits_{i=1}^{n} x_i = 1, x \geqq 0}} c^T x,$$

where $x = (x_1, \dots, x_n)^T \in \mathbb{R}^n, c = (c_1, \dots, c_n)^T \in \mathbb{R}^n, A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$, prompts the following Fenchel - Lagrange dual problem

$$(D_2) \qquad \sup_{\substack{p \in \mathbb{R}^n, q^x \in \mathbb{R}, \\ q^l \in \mathbb{R}^m_+, q^H \in \mathbb{R}_+}} \Bigg\{ -(c^T \cdot)^*(p) + \inf_{x \geqq 0} \Bigg[p^T x + {q^l}^T (b - Ax) + q^H \left(H + \sum_{i=1}^{n} x_i \ln x_i \right) + q^x \left(\sum_{i=1}^{n} x_i - 1 \right) \Bigg] \Bigg\}.$$

It is known that $(c^T \cdot)^*(p) = 0$ if $p = c$, otherwise being equal to $+\infty$. In [13] we prove that in the constraints of the problem (D_2) one can consider $q^H > 0$ instead of $q^H \in \mathbb{R}_+$. Also, the infimum over $x \geqq 0$ is separable into a sum of infima concerning $x_i \geq 0$, $i = 1, \dots, n$. Denoting also by $a_{ji}, j = 1, \dots, m, i = 1, \dots, n$, the entries of the matrix A and $q^l = (q^l_1, \dots, q^l_m)^T$, the dual problem turns into

$$(D_2) \sup_{\substack{q^x \in \mathbb{R}, q^l \in \mathbb{R}^m_+, \\ q^H > 0}} \left\{ q^H H + {q^l}^T b - q^x + \sum_{i=1}^{n} \inf_{x_i \geq 0} \left[c_i x_i + q^H x_i \ln x_i + \left(q^x - \sum_{j=1}^{m} q^l_j a_{ji} \right) x_i \right] \right\}.$$

These infima can be easily computed (cf. [13]) and the dual becomes

$$(D_2) \qquad \sup_{\substack{q^x \in \mathbb{R}, q^l \in \mathbb{R}^m_+, \\ q^H > 0}} \left\{ q^H H + {q^l}^T b - q^x - q^H \sum_{i=1}^{n} e^{\left(\sum\limits_{j=1}^{m} q^l_j a_{ji} - c_i + q^x - q^H \right) / q^H} \right\}.$$

The supremum over $q^x \in \mathbb{R}$ is also computable using elementary knowledge regarding the extreme points of functions, so the dual problem turns into its final version

$$(D_2) \qquad \sup_{\substack{q^l \in \mathbb{R}^m_+, \\ q^H > 0}} \left\{ b^T q^l - q^H \ln \left(\sum_{i=1}^{n} e^{\left((A^T q^l - c)_i / q^H \right) + q^H H} \right) \right\},$$

almost identical to the dual problem found in [82]. The difference consists in that the interval variable q^H lies in is $\mathbb{R}_+ \backslash \{0\}$ instead of $\mathbb{R}_+$. By Lemma 2.2 in [13] we know that this does not affect the optimal objective value of the dual problem. We have denoted the i-th entry of the vector $A^T q^l - c$ by $(A^T q^l - c)_i$. With the help of the constraint qualification

$$(CQ_2) \qquad \exists x' \in \operatorname{int}(\mathbb{R}^n_+) : \begin{cases} H + \sum\limits_{i=1}^{n} x'_i \ln x'_i < 0, \\ b - Ax' \leqq 0, \\ \sum\limits_{i=1}^{n} x'_i = 1, \end{cases}$$

the strong duality affirmation is ready to be formulated, followed by the optimality conditions, equivalent to the ones in the original paper.

Theorem 3.11 *(strong duality) If the constraint qualification* (CQ_2) *is satisfied, then the strong duality between* (P_2) *and* (D_2) *is assured.*

Theorem 3.12 *(optimality conditions)*
(a) If the constraint qualification (CQ_2) *is fulfilled and* $\bar{x}$ *is an optimal solution to* (P_2)*, then strong duality between the problems* (P_2) *and* (D_2) *is attained and the dual problem has an optimal solution* $(\bar{q}^l, \bar{q}^H)$ *satisfying the following optimality conditions*

(i) $\bar{q}^{lT}(A\bar{x} - b) = 0,$

(ii) $\bar{q}^H \left(H + \sum\limits_{i=1}^{n} \bar{x}_i \ln \bar{x}_i \right) = 0,$

(iii) $\bar{q}^H \left(\sum\limits_{i=1}^{n} \bar{x}_i \ln \bar{x}_i + \ln \left(\sum\limits_{i=1}^{n} e^{(A^T \bar{q}^l - c)_i / \bar{q}^H} \right) \right) = \bar{x}^T (A^T \bar{q}^l - c).$

(b) Having a feasible solution $\bar{x}$ *to the primal problem and one to the dual* $(\bar{q}^l, \bar{q}^H)$ *satisfying the optimality conditions* $(i) - (iii)$*, then the mentioned feasible solutions turn out to be optimal solutions to the corresponding problems and strong duality stands.*

3.3.3 Facility location problem (cf. [80])

In [80] the authors calculate the geometric duals for some problems involving norms. We have chosen one of them to be presented here, namely

$$(P_3) \qquad \inf_{\substack{\|x - a_j\| \le d_j, \\ j = 1, \dots, m}} \left\{ \sum_{j=1}^{m} w_j \|x - a_j\| \right\},$$

where $a_j \in \mathbb{R}^n$, $w_j > 0$, $d_j > 0$, for $j = 1, \dots, m$. The raw version of its Fenchel - Lagrange dual problem is

$$(D_3) \quad \sup_{\substack{p \in \mathbb{R}^n, \\ q \in \mathbb{R}^m_+}} \left\{ - \left(\sum_{j=1}^{m} w_j \| \cdot - a_j \| \right)^* (p) + \inf_{x \in \mathbb{R}^n} \left[p^T x + \sum_{j=1}^{m} q_j \big(\|x - a_j\| - d_j \big) \right] \right\}.$$

By Theorem 20.1 in [72] it turns into

$$(D_3) \quad \sup_{\substack{p^j \in \mathbb{R}^n, \\ \sum_{j=1}^{m} p^j = p, \\ q \in \mathbb{R}^m_+}} \left\{ - \sum_{j=1}^{m} \Big(w_j \| \cdot - a_j \| \Big)^* (p^j) + \inf_{x \in \mathbb{R}^n} \left[p^T x + \sum_{j=1}^{m} q_j \big(\|x - a_j\| - d_j \big) \right] \right\}.$$

Knowing that

$$\big(w_j \| \cdot - a_j \| \big)^* (p^j) = \begin{cases} a_j^T p^j, & \text{if } \|p^j\| \le w_j, \\ +\infty, & \text{otherwise,} \end{cases} \quad j = 1, \dots, m,$$

and turning the infimum into supremum, we get, applying again Theorem 20.1 in [72], the following equivalent formulation of the dual problem, given also in [80],

$$(D_3) \qquad \sup_{\substack{p^j \in \mathbb{R}^n, r^j \in \mathbb{R}^n, \\ \|p^j\| \leq w_j, \|r^j\| \leq q_j, \\ j=1,\dots,m, \\ \sum\limits_{j=1}^m \left(p^j + r^j\right) = 0, q \in \mathbb{R}^m_+}} \left\{ -q^T d - \sum_{j=1}^m a_j^T p^j - \sum_{j=1}^m a_j^T r^j \right\},$$

rewritable as

$$(D_3) \qquad \sup_{\substack{p^j \in \mathbb{R}^n, r^j \in \mathbb{R}^n, \\ \|p^j\| \leq w_j, \|r^j\| \leq q_j, \\ j=1,\dots,m, \\ \sum\limits_{j=1}^m \left(p^j + r^j\right) = 0, q \in \mathbb{R}^m_+}} \left\{ -q^T d - \sum_{j=1}^m a_j^T (p^j + r^j) \right\}.$$

Of course, we have set here $d = (d_1, \dots, d_m)^T \in \mathbb{R}^m$ and $q = (q_1, \dots, q_m)^T \in \mathbb{R}^m$.

A sufficient background for the existence of strong duality is in this case

$$(CQ_3) \qquad \exists x' \in \mathbb{R}^n : \quad \|x' - a_j\| < d_j, j = 1, \dots, m.$$

Theorem 3.13 *(strong duality) If the constraint qualification (CQ_3) is satisfied, then the strong duality between (P_3) and (D_3) is assured.*

Although there is no mention of the optimality conditions in [80] for this pair of dual problems we have derived the following result.

Theorem 3.14 *(optimality conditions)*
(a) If the constraint qualification (CQ_3) is fulfilled and $\bar{x}$ is an optimal solution to (P_3), then strong duality between the problems (P_3) and (D_3) is achieved and the dual problem has an optimal solution $(\bar{p}^1, \dots, \bar{p}^m, \bar{r}^1, \dots, \bar{r}^m, \bar{q}_1, \dots, \bar{q}_m)$ satisfying the following optimality conditions

(i) $w_j \|\bar{x} - a_j\| = \bar{p}^{jT}(\bar{x} - a_j)$ *and* $\|\bar{p}^j\| = w_j$ *when* $\bar{x} \neq a_j$, $j = 1, \dots, m$,

(ii) $\bar{q}_j \|\bar{x} - a_j\| = \bar{r}^{jT}(\bar{x} - a_j)$, $\bar{q}_j \geq 0$, $j = 1, \dots, m$, *and if* $\bar{q}_j > 0$ *so is* $\|\bar{r}^j\| = \bar{q}_j$. *For* $\bar{q}_j = 0$ *there is also* $\bar{r}^j = 0$. *If in particular* $\bar{x} = a_j$ *for any* $j \in \{1, \dots, m\}$, *then* $\bar{q}_j = 0$ *and* $\bar{r}^j = 0$,

(iii) $\|\bar{x} - a_j\| = d_j$, *for* $j \in \{1, \dots, m\}$ *such that* $\bar{q}_j > 0$,

(iv) $\sum\limits_{j=1}^m \left(\bar{p}^j + \bar{r}^j\right) = 0.$

(b) Given $\bar{x}$ a feasible solution to (P_3) and $(\bar{p}^1, \dots, \bar{p}^m, \bar{r}^1, \dots, \bar{r}^m, \bar{q}_1, \dots, \bar{q}_m)$ feasible to (D_3) which satisfy the optimality conditions $(i)-(iv)$, then the mentioned feasible solutions turn out to be optimal solutions to the corresponding problems and strong duality holds.

3.3.4 Quadratic concave fractional programs (cf. [76])

Another problem artificially pressed into the selective framework of geometric programming by the mentioned authors is

$$(P_4) \qquad \inf_{Cx \leqq b} \frac{Q(x)}{f(x)},$$

where $Q(x) = (1/2)x^T Ax, x \in \mathbb{R}^n$, $A \in \mathbb{R}^{n\times n}$ is a symmetric positive definite matrix, $C \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$ and $f : \mathbb{R}^n \to \mathbb{R}$ a concave function having strictly positive values over the feasible set of the problem. Because no analytic representation of the conjugate of the objective function is available, the problem is rewritten as

$$(P_4) \qquad \inf_{\substack{sQ\left(\frac{1}{s}x\right)-f(x)\leq 0,\\ Cx\leqq b, s\in\mathbb{R}_+\setminus\{0\}}} s.$$

To compute the Fenchel - Lagrange dual problem to (P_4), we need first the conjugate of the objective function $u : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}, u(x, s) = s$. Using the results presented before for the same objective function, the dual problem becomes

$$(D_4) \qquad \sup_{\substack{q^x\in\mathbb{R}^m_+,\\ q^s\in\mathbb{R}_+}} \left\{ \inf_{\substack{x\in\mathbb{R}^n,\\ s>0}} \left[s + q^s\left(sQ\left(\frac{1}{s}x\right) - f(x)\right) + q^{xT}(Cx - b)\right]\right\}.$$

The infimum over (x, s), transformed into a supremum, can be viewed as a conjugate function that is determined after some standard calculations. The formula that results for the dual problem is identical to the geometric dual obtained by the cited authors,

$$(D_4) \qquad \sup_{\substack{q^x\in\mathbb{R}^m_+, q^s\in\mathbb{R}_+,\\ \frac{1}{2}u^TA^{-1}u\leq q^s,\\ u+v=-C^Tq^x}} \left\{ - b^T q^x - (-q^s f)^*(v)\right\},$$

moreover simplifiable even to

$$(D_4) \qquad \sup_{\substack{q^x\in\mathbb{R}^m_+, q^s\in\mathbb{R}_+,\\ \frac{1}{2}(-v-C^Tq^x)^TA^{-1}(-v-C^Tq^x)\leq q^s}} \left\{ - b^T q^x - (-q^s f)^*(v)\right\}.$$

Of course we have removed the assumption of lower semicontinuity that has been imposed on the function $-f$ before. The constraint qualification required in this case would be

$$(CQ_4) \qquad \exists(x', s') \in \mathbb{R}^n \times (0, +\infty) : \begin{cases} s'Q\left(\frac{1}{s'}x'\right) - f(x') < 0,\\ Cx' \leqq b.\end{cases}$$

It is not difficult to notice that if (P_4) has a feasible point x' then $f(x') > 0$. Taking any $s' > Q(x')/f(x')$, the pair (x', s') satisfies (CQ_4).

Theorem 3.15 *(strong duality) Provided that the primal problem has at least a feasible point, strong duality between problems (P_4) and (D_4) is assured.*

The optimality conditions, equivalent to the ones given in [76], are presented in the following statement.

Theorem 3.16 *(optimality conditions)*
(a) If the problem (P_4) has an optimal solution $\bar{x}$ then strong duality between the problems (P_4) and (D_4) is attained and the dual problem has an optimal solution $(\bar{v}, \bar{q}^x, \bar{q}^s)$ satisfying the following optimality conditions

(i) $(-\bar{q}^s f)^*(\bar{v}) - \bar{q}^s f(\bar{x}) = \bar{v}^T \bar{x}$,

(ii) $\frac{1}{2}(-\bar{v} - \bar{C}^T q^x)^T A^{-1}(-\bar{v} - \bar{C}^T q^x) + \frac{1}{2}\bar{x}^T A\bar{x} = (-\bar{v} - \bar{C}^T q^x)^T \bar{x}$,

(iii) $\bar{q}^{xT}(b - C\bar{x}) = 0,$

(iv) $\frac{1}{2}\big(- \bar{v} - \bar{C}^T q^x\big)^T A^{-1}\big(- \bar{v} - \bar{C}^T q^x\big) = \bar{q}^s.$

(b) *Having a feasible solution $\bar{x}$ to the primal problem and one to the dual $(\bar{v}, \bar{q}^x, \bar{q}^s)$ satisfying the optimality conditions $(i) - (iv)$ the mentioned feasible solutions turn out to be optimal solutions to the corresponding problems and strong duality stands.*

3.3.5 Sum of convex ratios (cf. [77])

An extension to vector optimization of the problem treated here can be found in [91]. Here we consider as primal problem

$$(P_5) \qquad \inf_{Cx \leqq b} \left[h(x) + \sum_{i=1}^{J} \frac{f_i^2(x)}{g_i(x)} \right].$$

where $f_i, h : \mathbb{R}^n \to \overline{\mathbb{R}}$ are proper convex functions, $g_i : \mathbb{R}^n \to \mathbb{R}$ concave, $f_i(x) \geq 0$, $g_i(x) > 0$, $i = 1, \ldots, J$, for all x feasible to (P_5), $b \in \mathbb{R}^m$, $C \in \mathbb{R}^{m \times n}$, that is equivalent to

$$(P_5) \qquad \inf_{\substack{f_i(x) \leq s_i, s_i \in \mathbb{R}_+, \\ g_i(x) \geq t_i, t_i \in \mathbb{R}_+ \setminus \{0\}, \\ i=1,\ldots,J, Cx \leqq b}} \left[h(x) + \sum_{i=1}^{J} \frac{s_i^2}{t_i} \right].$$

The Fenchel - Lagrange dual problem arises naturally from its basic formula, where we denote the objective function by $u(x, s, t)$, with the variables x, $s = (s_1, \ldots, s_J)^T$, $t = (t_1, \ldots, t_J)^T$ and also the functions $f = (f_1, \ldots, f_J)^T$ and $g = (g_1, \ldots, g_J)^T$,

$$(D_5) \qquad \sup_{\substack{p^x \in \mathbb{R}^n, p^s, p^t \in \mathbb{R}^J, \\ q^x \in \mathbb{R}^m_+, q^s, q^t \in \mathbb{R}^J_+}} \Bigg\{ - u^*(p^x, p^s, p^t) + \inf_{\substack{x \in \mathbb{R}^n, s \in \mathbb{R}^J_+, \\ t \in \operatorname{int}(\mathbb{R}^J_+)}} \Big[{p^x}^T x + {p^s}^T s + {p^t}^T t$$
$$+ {q^s}^T (f(x) - s) + {q^t}^T (t - g(x)) + {q^x}^T (Cx - b) \Big] \Bigg\}.$$

For the conjugate function one has (consult [91] for computational details), denoting $p^s = (p^s_1, \ldots, p^s_J)^T$ and $p^t = (p^t_1, \ldots, p^t_J)^T$,

$$u^*(p^x, p^s, p^t) = \begin{cases} h^*(p^x), & \text{if } (p^s_i)^2 + 4p^t_i \leq 0, i = 1, \ldots, J, \\ +\infty, & \text{otherwise,} \end{cases}$$

while the infimum over (x, s, t) is separable into a sum of three infima each of them concerning a variable. The dual problem becomes

$$(D_5) \qquad \sup_{\substack{p^x \in \mathbb{R}^n, p^s, p^t \in \mathbb{R}^J, \\ (p^s_i)^2 + 4p^t_i \leq 0, i=1,\ldots,J, \\ q^x \in \mathbb{R}^m_+, q^s, q^t \in \mathbb{R}^J_+}} \Bigg\{ - h^*(p^x) + \inf_{s \in \mathbb{R}^J_+} \Big[{p^s}^T s - {q^s}^T s \Big] - {q^x}^T b$$

$$+ \inf_{x \in \mathbb{R}^n} \Big[(p^x + C^T q^x)^T x + {q^s}^T f(x) - {q^t}^T g(x) \Big] + \inf_{t \in \mathbb{R}^J_+ \setminus \{0\}} \Big[{p^t}^T t + {q^t}^T t \Big] \Bigg\}.$$

The infimum regarding $s \in \mathbb{R}^J_+$ has a negative infinite value unless $p^s - q^s \geqq 0$, when it nullifies itself, while the one regarding $t \in \mathbb{R}^J_+ \setminus \{0\}$ is zero when $p^t + q^t \geqq 0$, otherwise being equal to $-\infty$. The infimum regarding $x \in \mathbb{R}^n$ can be turned into a supremum and computed as a conjugate of a sum of functions at $-(p^x + C^T q^x)$. Applying Theorem 20.1 in [72] to this conjugate, the dual develops denoting $q^s = (q_1^s, \ldots, q_J^s)^T$ and $q^t = (q_1^t, \ldots, q_J^t)^T$ into

$$(D_5) \qquad \sup_{\substack{q^x \in \mathbb{R}^m_+, q^s, q^t \in \mathbb{R}^J_+, \\ p^s, p^t \in \mathbb{R}^J, q^s \leqq p^s, -q^t \leqq p^t, \\ p_i^t \leq -\left(\frac{p_i^s}{2}\right)^2, i=1,\ldots,J, \\ a_i, d_i, p^x \in \mathbb{R}^n, i=1,\ldots,J, \\ \sum_{i=1}^J (a_i + d_i) = -p^x - C^T q^x}} \left\{ -h^*(p^x) - \sum_{i=1}^J (q_i^s f_i)^*(a_i) - \sum_{i=1}^J (-q_i^t g_i)^*(d_i) - q^{xT} b \right\},$$

that can be simplified, renouncing the variables p^s and p^t, to

$$(D_5) \qquad \sup_{\substack{q^x \in \mathbb{R}^m_+, q^s, q^t \in \mathbb{R}^J_+, \\ q_i^t \geq \left(\frac{q_i^s}{2}\right)^2, i=1,\ldots,J, \\ a_i, d_i, p^x \in \mathbb{R}^n, i=1,\ldots,J, \\ \sum_{i=1}^J (a_i + d_i) = -p^x - C^T q^x}} \left\{ -h^*(p^x) - \sum_{i=1}^J (q_i^s f_i)^*(a_i) - \sum_{i=1}^J (-q_i^t g_i)^*(d_i) - q^{xT} b \right\}.$$

Writing the homogenous extensions of the conjugate functions one gets the dual problem obtained in the original paper. Let us stress that we have ignored the hypotheses of lower semicontinuity associated to the functions f_i, $-g_i$, $i = 1, \ldots, J$, and h in [77], as strong duality can be proven as valid even in their absence.

Theorem 3.17 *(strong duality) Provided that the primal problem (P_5) has a finite optimal objective value, strong duality between problems (P_5) and (D_5) is assured.*

The optimality conditions we determined in this case are richer than the ones presented in [77].

Theorem 3.18 *(optimality conditions)*
(a) If the problem (P_5) has an optimal solution $\bar{x}$ where its objective function is finite, then strong duality between the problems (P_5) and (D_5) is attained and the dual problem has an optimal solution $(\bar{p}^x, \bar{q}^x, \bar{q}^s, \bar{q}^t, \bar{a}, \bar{d})$ with $\bar{a} = (\bar{a}_1, \ldots, \bar{a}_J)$ and $\bar{d} = (\bar{d}_1, \ldots, \bar{d}_J)$ satisfying the following optimality conditions

(i) $\left(\bar{q}_i^s f_i\right)^*(\bar{a}_i) + \bar{q}_i^s f_i(\bar{x}) = \bar{a}_i^T \bar{x}, i = 1, \ldots, J,$

(ii) $\left(-\bar{q}_i^t g_i\right)^*(\bar{d}_i) - \bar{q}_i^t g_i(\bar{x}) = \bar{d}_i^T \bar{x}, i = 1, \ldots, J,$

(iii) $h^*(\bar{p}^x) + h(\bar{x}) = \bar{p}^{xT} \bar{x},$

(iv) $\bar{q}_i^s = 2\frac{f_i(\bar{x})}{g_i(\bar{x})}, i = 1, \ldots, J,$

(v) $\bar{q}_i^t = \frac{f_i^2(\bar{x})}{g_i^2(\bar{x})}, i = 1, \ldots, J,$

(vi) $\bar{q}^{xT}(b - C^T \bar{x}) = 0.$

(b) Having a feasible solution $\bar{x}$ to the primal problem and one to the dual $(\bar{p}^x, \bar{q}^x, \bar{q}^s, \bar{q}^t, \bar{a}, \bar{d})$ satisfying the optimality conditions $(i) - (vi)$, the mentioned feasible solutions turn out to be optimal solutions to the corresponding problems and strong duality holds.

3.3.6 Quasiconcave multiplicative programs (cf. [79])

Despite its intricateness and limits, geometric programming duality seems to be yet very popular, as its direct applications still get published. One of the newest we found is on a class of quasiconcave multiplicative programs that originally look like

$$(P_6) \qquad \sup_{Ax \leqq b} \left\{ \prod_{i=1}^{k} [f_i(x)]^{a_i} \right\},$$

with $f_i : \mathbb{R}^n \to \mathbb{R}$ concave functions, positive over the feasible set of the problem, $a_i > 0$, $i = 1, \dots, k$, $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$. Denote moreover $f = (f_1, \dots, f_k)^T$. The problem is brought into another layout in order to be properly treated,

$$(\widetilde{P}_6) \qquad \inf_{\substack{f_i(x) \geq s_i, i=1,\dots,k, \\ Ax \leqq b, s \in \operatorname{int}(\mathbb{R}^k_+)}} \left\{ -\sum_{i=1}^{k} a_i \ln s_i \right\}.$$

Proposition 3.1 *We have* $\ln(v(P_6)) = -v(\widetilde{P}_6)$. *Moreover,* $\bar{x}$ *is an optimal solution to* (P_6) *if and only if* $(\bar{x}, f(\bar{x}))$ *is an optimal solution to* $(\widetilde{P}_6)$.

Denoting $s = (s_1, \dots, s^k)^T$ and $u : \mathbb{R}^n \times \mathbb{R}^k \to \overline{\mathbb{R}}$,

$$u(x, s) = \begin{cases} -\sum\limits_{i=1}^{k} a_i \ln s_i, & \text{if } (x, s) \in \mathbb{R}^n \times \operatorname{int}(\mathbb{R}^k_+), \\ +\infty, & \text{otherwise}, \end{cases}$$

the raw formula of the Fenchel - Lagrange dual to $(\widetilde{P}_6)$ is

$$(\widetilde{D}_6) \quad \sup_{\substack{p^x \in \mathbb{R}^n, p^s \in \mathbb{R}^k, \\ q^l \in \mathbb{R}^m_+, q^f \in \mathbb{R}^k_+}} \left\{ \inf_{\substack{x \in \mathbb{R}^n, \\ s \in \operatorname{int}(\mathbb{R}^k_+)}} \left[{p^x}^T x + {p^s}^T s + {q^l}^T (Ax - b) + {q^f}^T (s - f(x)) \right] - u^*(p^x, p^s) \right\}.$$

Regarding the conjugate of the objective function the following result is available for $p^s = (p^s_1, \dots, p^s_k)^T$

$$u^*(p^x, p^s) = \begin{cases} -\sum\limits_{i=1}^{k} a_i \Big(1 - \ln\Big(\frac{a_i}{-p^s_i}\Big)\Big), & \text{if } p^x = 0, p^s < 0, \\ +\infty, & \text{otherwise}. \end{cases}$$

The infimum in the dual problem can also be separated into a sum of two infima, one concerning $s \in \operatorname{int}(\mathbb{R}^k_+)$, the other $x \in \mathbb{R}^n$. Let us write again the dual using the last observations and denoting $-p^s_i$ by p^s_i, $i = 1, \dots, k$,

$$(\widetilde{D}_6) \qquad \sup_{\substack{p^s \in \operatorname{int}(\mathbb{R}^k_+), \\ q^l \in \mathbb{R}^m_+, q^f \in \mathbb{R}^k_+}} \left\{ \sum_{i=1}^{k} a_i \Big(1 - \ln\Big(\frac{a_i}{p^s_i}\Big)\Big) + \inf_{s \in \operatorname{int}(\mathbb{R}^k_+)} (q^f - p^s)^T s \right.$$
$$\left. + \inf_{x \in \mathbb{R}^n} \left[{q^l}^T (Ax) - {q^f}^T f(x) \right] - {q^l}^T b \right\}.$$

The infimum regarding s is equal to 0 when $q^f - p^s \geqq 0$, otherwise being $-\infty$, while the one over $x \in \mathbb{R}^n$ can be rewritten as a supremum and viewed as a conjugate of a sum of functions. The dual problem becomes

$$(\widetilde{D}_6) \qquad \sup_{\substack{p^s \in \operatorname{int}(\mathbb{R}^k_+), p^s \leqq q^f, \\ q^l \in \mathbb{R}^m_+, \sum\limits_{i=1}^k v_i = -A^T q^l}} \left\{ \sum_{i=1}^k \left(a_i - a_i \ln \left(\frac{a_i}{p_i^s} \right) \right) - \sum_{i=1}^k q_i^f (-f_i)^* \left(\frac{1}{q_i^f} v_i \right) - b^T q^l \right\},$$

where $q^f = \left(q_1^f, \ldots, q_k^f\right)^T$, and as the supremum regarding the variable p^s can be easily computed, being attained for $p^s = -q^f$, we get the following final version of the dual, equivalent to the one found in [79],

$$(\widetilde{D}_6) \qquad \sup_{\substack{q^l \in \mathbb{R}^m_+, q^f \in \operatorname{int}(\mathbb{R}^k_+), \\ \sum\limits_{i=1}^k v_i = -A^T q^l}} \left\{ \sum_{i=1}^k \left(a_i - a_i \ln \left(\frac{a_i}{q_i^f} \right) \right) - \sum_{i=1}^k q_i^f (-f_i)^* \left(\frac{1}{q_i^f} v_i \right) - b^T q^l \right\}.$$

For strong duality a constraint qualification would normally be required because within the constraints of $(\widetilde{P}_6)$ there are affine as well as non - affine functions. But when the feasible set of the problem (P_6) is non - empty there is some $x' \in \mathbb{R}^n$ such that $Ax' \leqq b$ and $f_i(x') > 0, i = 1, \ldots, k$. Consequently there is also an $s' > 0$ such that $f_i(x') > s' > 0$, $i = 1, \ldots, k$, too. So the constraint qualification that comes from the general case for $(\widetilde{P}_6)$ is automatically fulfilled provided that the feasible set of (P_6) is not empty. Without any additional assumption one may formulate the strong duality statement.

Theorem 3.19 *(strong duality) Provided that the primal problem (P_6) has a feasible point, there is strong duality between the problems $(\widetilde{P}_6)$ and $(\widetilde{D}_6)$.*

Corollary 3.1 *Provided that the primal problem (P_6) has a feasible point, the dual problem $(\widetilde{D}_6)$ has an optimal solution and one has $v(P_6) = e^{-v(\widetilde{D}_6)}$.*

No surprises appear when we derive the optimality conditions concerning the pair of dual problems in discussion. The proof takes also into consideration Proposition 3.1.

Theorem 3.20 *(optimality conditions)*
(a) If the problem (P_6) has an optimal solution $\bar{x}$, then strong duality between the problems $(\widetilde{P}_6)$ and $(\widetilde{D}_6)$ is attained and the dual problem has an optimal solution $(\bar{v}_1, \ldots, \bar{v}_k, \bar{q}^l, \bar{q}^f)$ satisfying the following optimality conditions

(i) $(-f_i)^* \left(\frac{1}{\bar{q}_i^f} \bar{v}_i \right) - f_i(\bar{x}) = \frac{1}{\bar{q}_i^f} \bar{v}_i^T \bar{x}$, $i = 1, \ldots, k$,

(ii) $(A^T \bar{x} - b)^T \bar{q}^l = 0$,

(iii) $\sum\limits_{i=1}^k \bar{v}_i = -\bar{A}^T q^l$,

(iv) $\ln(f_i(\bar{x})) + \frac{\bar{q}_i^f}{a_i} f_i(\bar{x}) = \ln \left(\frac{a_i}{\bar{q}_i^f} \right) - 1, i = 1, \ldots, k.$

(b) Having a feasible solution $\bar{x}$ to the problem (P_6) and one to the problem $(\widetilde{D}_6)$ $(\bar{v}_1, \ldots, \bar{v}_k, \bar{q}^l, \bar{q}^f)$ satisfying the optimality conditions $(i)-(iv)$, the mentioned feasible solutions turn out to be optimal solutions to the corresponding problems and strong duality holds between the problems $(\widetilde{P}_6)$ and $(\widetilde{D}_6)$.

3.3.7 Posynomial geometric programming (cf. [30])

We are going to prove now that also the posynomial geometric programming duality can be viewed as a special case of the Fenchel - Lagrange duality. As it has been already proven (cf. [48]) that the generalized geometric programming includes the posynomial instance as a special case, our result is not so surprising within the framework of this thesis. The primal-dual pair of posynomial geometric problems is composed by

$$(P_7) \qquad \inf_{\substack{t=(t_1,\ldots,t_m)^T\in \operatorname{int}(\mathbb{R}^m_+),\\ g_j(t)\leq 1, j=1,\ldots,s}} g_0(t),$$

where
$$g_k(t) = \sum_{i\in J[k]} c_i \prod_{j=1}^{m} (t_j)^{a_{ij}}, k = 0,\ldots,s,$$

$$a_{ij} \in \mathbb{R}, j = 1,\ldots,m, c_i > 0, i = 1,\ldots,n,$$

$$J[k] = \{m_k, m_k + 1,\ldots,n_k\}, k = 0,\ldots,s,$$

$$m_0 = 1, m_1 = n_0 + 1,\ldots,m_k = n_{k-1} + 1,\ldots,n_s = n,$$

and

$$(D_7) \qquad \sup_{\substack{\delta=(\delta_1,\ldots,\delta_n)^T\in\mathbb{R}^n_+,\\ \sum\limits_{i\in J[0]}\delta_i=1,\\ \sum\limits_{i=1}^{n}\delta_i a_{ij}=0,\\ j=1,\ldots,m}} \left[\prod_{i=1}^{n} \frac{c_i}{\delta_i}^{\delta_i}\right] \prod_{k=1}^{s} \lambda_k(\delta)^{\lambda_k(\delta)},$$

with
$$\lambda_k(\delta) = \sum_{i\in J[k]} \delta_i, k = 1,\ldots,s.$$

To the primal posynomial problem we attach the following problem (cf. [30,48])

$$(\widetilde{P}_7) \qquad \inf_{\substack{\ln\left(\sum\limits_{i\in J[k]} c_i e^{x_i}\right)\leq 0,\\ k=1,\ldots,s, x\in\mathcal{U}}} \left\{\ln\left(\sum_{i\in J[0]} c_i e^{x_i}\right)\right\},$$

where $\mathcal{U}$ denotes the linear subspace generated by the columns of the exponent matrix $(a_{ij})_{\substack{i=1,\ldots,n,\\ j=1,\ldots,m}}$. Let us name also $u(x)$ the primal objective function of the problem $(\widetilde{P}_7)$. These two problems are connected (cf. [48]) through the following result.

Proposition 3.2 *One has $v(\widetilde{P}_7) = \ln(v(P_7))$. Moreover, $\bar{t}$ is an optimal solution to (P_7) if and only if $\bar{x}$ is an optimal solution to $(\widetilde{P}_7)$, where $\bar{x} = (\bar{x}_1,\ldots,\bar{x}_n)^T$ and $\bar{x}_i = \sum_{j=1}^{m} a_{ij}\ln(\bar{t}_j)$ $\forall i = 1,\ldots,n$.*

We determine the Fenchel - Lagrange dual problem to $(\widetilde{P}_7)$ from the formula of (D'_N), with $\mathcal{U}^\perp$ indicating the orthogonal subspace of $\mathcal{U}$, for $q = (q_1,\ldots,q_s)^T$ and $t = (t_1,\ldots,t_n)^T$

$$(\widetilde{P}_7) \qquad \sup_{\substack{p\in\mathbb{R}^n, t\in\mathcal{U}^\perp,\\ q\in\mathbb{R}^s_+}} \left\{ - u^*(p) + \inf_{x\in\mathcal{U}} \left[(p-t)^T x + \sum_{k=1}^{s} q_k \ln\left(\sum_{i\in J[k]} c_i e^{x_i}\right)\right]\right\}.$$

For the conjugate of the objective function we have

$$u^*(p) = \begin{cases} \sum\limits_{i \in J[0]} p_i \ln\left(\frac{p_i}{c_i}\right), & \text{if } p_j = 0, j \in J[k], k = 1, \ldots, s, \\ & \sum\limits_{i \in J[0]} p_i = 1, p = (p_1, \ldots, p_n)^T \in \mathbb{R}^n_+, \\ +\infty, & \text{otherwise}, \end{cases}$$

and similar results can be derived if we write the infimum within the dual as a sum of suprema over $(x_i)_{i \in J[k]}$, $k = 1, \ldots, s$, just with the changed constraints $\sum_{i \in J[k]} t_i = q_k$. Also there follows $p_i = t_i, i \in J[0]$. As usual in entropy optimization we consider $0 \ln(0/c_i) = 0, c_i > 0, i = 1, \ldots, n$. After these, the dual problem becomes

$$(\widetilde{D}_7) \qquad \sup_{\substack{t \in \mathcal{U}^\perp, q \in \mathbb{R}^s_+, \\ t \geqq 0, \sum\limits_{i \in J[0]} t_i = 1, \\ \sum\limits_{i \in J[k]} t_i = q_k, k = 1, \ldots, s}} \left\{ \sum_{i=1}^n t_i \ln\left(\frac{c_i}{t_i}\right) + \sum_{k=1}^s q_k \ln q_k \right\}.$$

Finally, the condition that guarantees strong duality, derived from the constraint qualification (CQ), is actually the so - called *superconsistency* introduced in [30], i.e.

$$(CQ_7) \qquad \exists t' > 0 : g_k(t') < 1, k = 1, \ldots, s.$$

Theorem 3.21 *(strong duality) If the constraint qualification (CQ_7) is satisfied, then the strong duality between (P_7) and (D_7) is assured.*

Consequently we present also the optimality conditions concerning the pair of problems $(\widetilde{P}_7)$ and $(\widetilde{D}_7)$. The ones concerning the problems (P_7) and (D_7) can be derived from these and are available in the literature (see for instance [48]).

Theorem 3.22 *(optimality conditions)*
(a) If the constraint qualification (CQ_7) is fulfilled and $\bar{x}$ is an optimal solution to $(\widetilde{P}_7)$, then $v(\widetilde{P}_7) = v(\widetilde{D}_7)$ and $(\widetilde{D}_7)$ has an optimal solution $(\bar{t}, \bar{q})$ satisfying the following optimality conditions

(i) $\ln\left(\sum\limits_{i \in J[0]} c_i e^{\bar{x}_i}\right) + \sum\limits_{i \in J[0]} \bar{t}_i \ln\left(\frac{\bar{t}_i}{c_i}\right) = \left(\bar{t}^{J[0]}\right)^T \bar{x}^{J[0]}$,

(ii) $\bar{q}_k \ln\left(\sum\limits_{i \in J[k]} c_i e^{\bar{x}_i}\right) + \sum\limits_{i \in J[k]} \bar{t}_i \ln\left(\frac{\bar{t}_i}{c_i}\right) - \bar{q}_k \ln \bar{q}_k = \left(\bar{t}^{J[k]}\right)^T \bar{x}^{J[k]}$, $k = 1, \ldots, s$,

(iii) $\bar{x}^T \bar{t} = 0$,

(iv) $\sum\limits_{i \in J[k]} c_i e^{\bar{x}_i} = 1$ *when* $\bar{q}_k > 0$, $k = 1, \ldots, s$,

where $x^{J[k]} = (x_{m_k}, \ldots, x_{n_k})$ *and* $t^{J[k]} = (t_{m_k}, \ldots, t_{n_k})$, $k = 0, \ldots, s$.

(b) Having a feasible solution $\bar{x}$ to the primal problem and one to the dual $(\bar{t}, \bar{q})$ satisfying the optimality conditions $(i) - (iv)$, the mentioned feasible solutions turn out to be optimal solutions to the corresponding problems and strong duality holds.

Remark 3.8 This shows that the rich literature on posynomial programming can be easier treated from the point of view of the Fenchel - Lagrange duality, especially when the problems studied there are forced to agree with the strict framework of the posynomial programming.

Chapter 4

Extensions to different classes of optimization problems

Within the fourth chapter of this thesis we have included several interesting extensions and applications of the duality results presented earlier.

A first class where we apply the results given in the previous chapters consists of the so - called composed convex optimization problems. Their study is important for its theoretical aspect, but also for its practical applications. Let us mention only some papers dealing with optimization problems involving composed functions, namely [20, 27, 61, 62, 88]. Closely related to such optimization problems is the calculation of the conjugate of the composition of two functions. Papers like [27, 45, 60] and the famous books [44] are usually cited when this comes in discussion. As an application of the duality statements we provide for a composed convex optimization problem and its Fenchel - Lagrange dual we present a general duality framework for dealing with convex multiobjective programming problems. This uses the scalarization of the primal vector minimization problem with a strongly increasing function and we show that some other scalarization methods widely - used in the literature (cf. [8,18,19,35,43,47,54,58,63,65,73,74,84–87,89–91,93–95], among others) arise as special cases. The scalarization we present has already been used in the literature by Göpfert and Gerth (cf. [40]), Gerstewitz and Iwanow (cf. [38]), Jahn (cf. [46,47]) and Miglierina and Molho (cf. [64]), but even when a dual has been assigned to the initial vector problem, the approach came from Lagrange duality.

Entropy optimization is another field with many various applications in practice, from transportation and location problems to image reconstruction and speech recognition. We refer to the books [34] and [57] for overviews on entropy optimization and its applications in various fields.

We have considered a convex optimization problem having as objective function an entropy - like sum of functions. After constructing a dual to this problem, giving the strong duality and necessary and sufficient optimality conditions, we show that the most important and used three entropy measures, namely those due to Shannon, Kullback and Leibler and, respectively, Burg, are particular instances of our entropy - like objective function, for some careful choices of the functions involved. Therefore the problems usually treated in the literature via geometric programming (see [33, 34]) can be easier dealt with when considered as special cases of the optimization problem with entropy - like objective function.

4.1 Composed convex optimization problems

The objective functions of the optimization problems may have different formulations. Many convex optimization problems arising from various directions may be formulated as minimizations of some compositions of functions subject to some constraints. We cite here [20, 27, 61, 62, 88] as articles dealing with composed convex optimization problems. Duality assertions for this kind of problems may be delivered in different ways, one of the most common consisting in considering an equivalent problem to the primal one, whose dual is easier determinable. If the desired duality results are based on conjugate functions, sometimes even a more direct way is available by obtaining a dual problem based on the conjugate function of the composed objective function of the primal, which is writable, in some situations, by using only the conjugates of the functions involved and the dual variables. Depending on the framework, the formula of the conjugate of the composed functions is taken mainly from [27, 44, 45, 60].

Here we bring weaker conditions under which the known formula of the conjugate of a composed function holds when one works in $\mathbb{R}^n$. No lower semicontinuity or continuity concerning the functions to be composed is necessary, while the regularity condition saying that the image set of the postcomposed function should contain an interior point of the domain of the other function is weakened to a relation involving relative interiors which is actually implied by the first one. This important result is presented as an application of the strong duality for the unconstrained composed convex optimization problem, followed by the concrete case of calculating the conjugate of $1/F$, when F is a concave strictly positive function defined over the set of strictly positive reals.

4.1.1 Strong duality for the composed convex optimization problem

Let K and C be non - empty closed convex cones in $\mathbb{R}^k$ and $\mathbb{R}^m$, respectively, and X a non - empty convex subset of $\mathbb{R}^n$. On $\mathbb{R}^k$ and $\mathbb{R}^m$ we consider the partial orderings induced by the cones K and C, respectively. Take $f : \mathbb{R}^k \to \overline{\mathbb{R}}$ to be a proper K - increasing convex function, $F : X \to \mathbb{R}^k$ a function K - convex on X and $g : X \to \mathbb{R}^m$ a function C - convex on X. Moreover, we impose the feasibility condition $\mathcal{A} \cap F^{-1}(\mathrm{dom}(f)) \neq \emptyset$, where $\mathcal{A} = \{x \in X : g(x) \in -C\}$ and for any set $U \subseteq \mathbb{R}^k$, $F^{-1}(U) = \{x \in X : F(x) \in U\}$. The problem we consider within this section is

$$(P_c) \qquad \inf_{\substack{x \in X, \\ g(x) \in -C}} f(F(x)).$$

We could formulate its dual as a special case of (P) since $f \circ F$, completed with plus infinite values outside X is a convex function, but the existing formulae which allow to separate the conjugate of $f \circ F$ into a combination of the conjugates of f and F ask the functions to be moreover lower semicontinuous even in some particular cases (cf. [44, 45]). To avoid this too strong requirement we formulate the following problem equivalent to (P_c) in the sense that their optimal objective values coincide,

$$(P_c') \qquad \inf_{\substack{x \in X, y \in \mathbb{R}^k, \\ g(x) \in -C, \\ F(x) - y \in -K}} f(y).$$

Proposition 4.1 $v(P_c) = v(P_c')$.

Proof. Let x be feasible to (P_c). Take $y = F(x)$. Then $F(x) - y = 0 \in -K$ (remember that the convex cone K is non - empty and closed). Thus (x, y) is feasible to (P_c') and $f(F(x)) = f(y) \geq v(P_c')$. Since this is valid for any x feasible to (P_c) it is straightforward that $v(P_c) \geq v(P_c')$.

On the other hand, for (x, y) feasible to (P_c') we have $x \in X$ and $g(x) \in -C$, so x is feasible to (P_c). Since f is K - increasing we get $v(P_c) \leq f(F(x)) \leq f(y)$. Taking the infimum on the right - hand side after (x, y) feasible to (P_c') we get $v(P_c) \leq v(P_c')$. Therefore $v(P_c) = v(P_c')$. □

The problem (P_c') is a special case of the initial primal optimization problem

$$(P) \qquad \inf_{\substack{x\in X,\\ g(x)\in -C}} f(x),$$

with the variable $(x, y) \in X \times \mathbb{R}^k$, the objective function $A : \mathbb{R}^n \times \mathbb{R}^k \to \overline{\mathbb{R}}$ defined by $A(x, y) = f(y)$, the constraint function $B : X \times \mathbb{R}^k \to \mathbb{R}^m \times \mathbb{R}^k$ defined by $B(x, y) = (g(x), F(x) - y)$ and the cone $C \times K$, which is a non - empty closed convex cone in $\mathbb{R}^m \times \mathbb{R}^k$. We also use that $(C \times K)^* = C^* \times K^*$. The Fenchel - Lagrange dual problem to (P_c') is

$$(D_c') \qquad \sup_{\substack{\alpha\in C^*,\beta\in K^*,\\ (p,s)\in\mathbb{R}^n\times\mathbb{R}^k}} \Big\{ - A^*(p, s) - \big((\alpha, \beta)^T B\big)^*_{X\times\mathbb{R}^k}(-p, -s)\Big\}.$$

Let us determine the values of these conjugates. We have

$$A^*(p, s) = \sup_{\substack{x\in\mathbb{R}^n,\\ y\in\mathbb{R}^k}} \Big\{p^T x + s^T y - f(y)\Big\} = f^*(s) + \begin{cases} 0, & \text{if } p = 0,\\ +\infty, & \text{otherwise,}\end{cases}$$

and

$$\begin{aligned}\big((\alpha, \beta)^T B\big)^*_{X\times\mathbb{R}^k}(-p, -s) &= \sup_{\substack{x\in X,\\ y\in\mathbb{R}^k}} \Big\{ - p^T x - s^T y - \alpha^T g(x) - \beta^T (F(x) - y)\Big\}\\ &= \big(\alpha^T g + \beta^T F\big)^*_X(-p) + \begin{cases} 0, & \text{if } s = \beta,\\ +\infty, & \text{otherwise.}\end{cases}\end{aligned}$$

Pasting these formulae into the objective function of the dual problem we get

$$-A^*(p, s) - \big((\alpha, \beta)^T B\big)^*_{X\times\mathbb{R}^k}(-p, -s) = -f^*(\beta) - \big(\alpha^T g + \beta F\big)^*_X(0),$$

if $p = 0$ and $s = \beta$, while otherwise $-A^*(p, s) - \big((\alpha, \beta)^T B\big)^*_{X\times\mathbb{R}^k}(-p, -s) = +\infty$.

This makes the dual problem turn into

$$(D_c') \qquad \sup_{\substack{\alpha\in C^*,\\ \beta\in K^*}} \Big\{ - f^*(\beta) - \big(\alpha^T g + \beta^T F\big)^*_X(0)\Big\}.$$

When $\alpha \in C^*$ and $\beta \in K^*$, applying Theorem 16.4 in [72] we get

$$\big(\alpha^T g + \beta^T F\big)^*_X(0) = \inf_{u\in\mathbb{R}^n} \Big\{\big(\beta^T F\big)^*_X(u) + \big(\alpha^T g\big)^*_X(-u)\Big\}$$

and this leads to the following formulation of the dual problem

$$(D_c) \qquad \sup_{\substack{\alpha\in C^*,\beta\in K^*,\\ u\in\mathbb{R}^n}} \Big\{ - f^*(\beta) - \big(\beta^T F\big)^*_X(u) - \big(\alpha^T g\big)^*_X(-u)\Big\}.$$

Thanks to Proposition 4.1 we may call (D_c) a dual problem to (P_c), too.

The weak duality statement follows directly from the one given earlier in the general case.

Theorem 4.1 $v(D_c) \leq v(P_c)$.

Now let us write what becomes the constraint qualification (CQ) in this case. We have

$$(CQ_c) \qquad \exists (x', y') \in \mathrm{ri}(\mathrm{dom}(A)) \cap \mathrm{ri}(X \times \mathbb{R}^k) : \; B(x', y') \in \mathrm{ri}(C \times K),$$

equivalent to

$$(CQ_c) \qquad \exists x' \in \mathrm{ri}(X) : \begin{cases} g(x') \in -\,\mathrm{ri}(C), \\ F(x') \in \mathrm{ri}(\mathrm{dom}(f)) - \mathrm{ri}(K). \end{cases}$$

The strong duality statement follows accompanied by the necessary and sufficient optimality conditions.

Theorem 4.2 *(strong duality) Consider the constraint qualification (CQ_c) fulfilled. Then there is strong duality between the problem (P_c) and its dual (D_c), i.e. $v(P_c) = v(D_c)$ and the latter has an optimal solution if $v(P_c) > -\infty$.*

Proof. According to Theorem 2.2, the fulfilment of (CQ_c) is sufficient to guarantee that $v(P_c') = v(D_c)$ and, if $v(P_c') > -\infty$, the existence of an optimal solution to the dual problem. Applying now Proposition 4.1 it follows $v(P_c) = v(D_c)$ and (D_c) must have an optimal solution if $v(P_c) > -\infty$. $\square$

Theorem 4.3 *(optimality conditions)*
(a) If the constraint qualification (CQ_c) is fulfilled and the primal problem (P_c) has an optimal solution $\bar{x}$, then the dual problem has an optimal solution $(\bar{u}, \bar{\alpha}, \bar{\beta})$ and the following optimality conditions are satisfied

(i) $f^*(\bar{\beta}) + f(F(\bar{x})) = \bar{\beta}^T F(\bar{x})$,

(ii) $\left(\bar{\beta}^T F\right)_X^*(\bar{u}) + \bar{\beta}^T F(\bar{x}) = \bar{u}^T \bar{x}$,

(iii) $\left(\bar{\alpha}^T g\right)_X^*(-\bar{u}) + \bar{\alpha}^T g(\bar{x}) = -\bar{u}^T \bar{x}$,

(iv) $\bar{\alpha}^T g(\bar{x}) = 0$.

(b) If $\bar{x}$ is a feasible point to the primal problem (P_c) and $(\bar{u}, \bar{\alpha}, \bar{\beta})$ is feasible to the dual problem (D_c) fulfilling the optimality conditions $(i) - (iv)$, then there is strong duality between (P_c) and (D_c) and the mentioned feasible points turn out to be optimal solutions of the corresponding problems.

Proof. The previous theorem yields the existence of an optimal solution $(\bar{u}, \bar{\alpha}, \bar{\beta})$ to the dual problem. Strong duality is also attained, i.e.

$$f(F(\bar{x})) = -f^*(\bar{\beta}) - \left(\bar{\beta}^T F\right)_X^*(\bar{u}) - \left(\bar{\alpha}^T g\right)_X^*(-\bar{u}),$$

which is equivalent to

$$f(F(\bar{x})) + f^*(\bar{\beta}) + \left(\bar{\beta}^T F\right)_X^*(\bar{u}) + \left(\bar{\alpha}^T g\right)_X^*(-\bar{u}) = 0.$$

The Fenchel - Young inequality asserts for the functions involved in the latter equality

$$f(F(\bar{x})) + f^*(\bar{\beta}) \geq \bar{\beta}^T F(\bar{x}), \tag{4. 1}$$

$$\bar{\beta}^T F(\bar{x}) + \left(\bar{\beta}^T F\right)_X^*(\bar{u}) \geq \bar{u}^T \bar{x} \tag{4. 2}$$

and

$$\bar{\alpha}^T g(\bar{x}) + \left(\bar{\alpha}^T g\right)_X^*(-\bar{u}) \geq -\bar{u}^T \bar{x}. \tag{4. 3}$$

The last four relations lead to

$$0 \geq \bar{\beta}^T F(\bar{x}) + \bar{u}^T \bar{x} - \bar{\beta}^T F(\bar{x}) - \bar{u}^T \bar{x} - \bar{\alpha}^T g(\bar{x}) = -\bar{\alpha}^T g(\bar{x}) \geq 0,$$

as $\bar{\alpha} \in C^*$ and $g(\bar{x}) \in -C$. Therefore the inequalities above must be fulfilled as equalities. The last one implies the optimality condition (iv), while (i) arises from (4. 1), (ii) from (4. 2) and (iii) from (4. 3).

The reverse assertion follows immediately, even without the fulfilment of (CQ_c) and of any convexity assumption we made concerning the involved functions and sets. □

Remark 4.1 One can notice that the results in this section remain valid if the initial assumption on K to be a non - empty convex closed cone is relaxed by taking it only a convex cone in $\mathbb{R}^k$ that contains $0_{\mathbb{R}^k}$.

4.1.2 Conjugate of the precomposition with a K - increasing convex function

This subsection is dedicated to an interesting and important application of the duality assertions presented in the previous one. We calculate the conjugate function of a postcomposition of a function that is K - convex on X with a K - increasing convex function, for K a non - empty closed convex cone, and we obtain the same formula as in some other works dealing with the same subject, [27, 44, 45, 60]. But let us mention that we obtain this formula under weaker conditions than known so far or deducible from the ones given in more general contexts (for instance in the books [44], where the authors also work in finite-dimensional spaces, the two functions are required to be moreover lower semicontinuous).

We find it useful to give here first the duality assertions regarding the unconstrained problem having as objective function the postcomposition of a K - increasing convex function to a function which is K - convex on X, where K is a non - empty closed convex cone and X a non - empty convex subset of $\mathbb{R}^n$, problem treated within different conditions in [20]. We maintain the notations from the preceding subsection and the initial feasibility assumption becomes $F(X) \cap \operatorname{dom}(f) \neq \emptyset$. The primal unconstrained optimization problem is

$$(P_u) \qquad \inf_{x \in X} f(F(x)).$$

It may be obtained from (P_c) by taking g the zero function and $C = \{0\}^m$. So its Fenchel - Lagrange dual problem becomes

$$(D_u) \qquad \sup_{\substack{\alpha \in C^*, \beta \in K^*, \\ u \in \mathbb{R}^n}} \Big\{ - f^*(\beta) - \big(\beta^T F\big)^*_X(u) - \big(\alpha^T 0\big)^*_X(-u) \Big\},$$

equivalent, as $\alpha \in C^*$ is no longer necessary, to

$$(D_u) \qquad \sup_{\substack{\beta \in K^*, \\ u \in \mathbb{R}^n}} \Big\{ - f^*(\beta) - \big(\beta^T F\big)^*_X(u) - \sup_{x \in X} -u^T x \Big\}.$$

It is easy to see that (the supremum is attained for $u = 0$)

$$\sup_{u \in \mathbb{R}^n} \Big\{ - \big(\beta^T F\big)^*_X(u) - \sup_{x \in X} -u^T x \Big\} = -\big(\beta^T F\big)^*_X(0).$$

Thus the dual problem becomes

$$(D_u) \qquad \sup_{\beta \in K^*} \Big\{ - f^*(\beta) - \big(\beta^T F\big)^*_X(0)\Big\},$$

while the constraint qualification that is sufficient to guarantee strong duality between this dual and the primal problem (P_u) is

$$(CQ_u) \qquad \exists x' \in \mathrm{ri}(X): \ F(x') \in \mathrm{ri}(\mathrm{dom}(f)) - \mathrm{ri}(K).$$

The weak and strong duality statement follow, directly from the general case.

Theorem 4.4 $v(D_u) \leq v(P_u)$.

Theorem 4.5 *(strong duality) Consider the constraint qualification* (CQ_u) *fulfilled. Then there is strong duality between the problem* (P_u) *and its dual* (D_u) *and the latter has an optimal solution if* $v(P_u) > -\infty$.

We actually want to determine the formula of the conjugate function $(f \circ F)^*_X$ as a function of f^* and F^*_X. We have for some $p \in \mathbb{R}^n$

$$(f \circ F)^*_X(p) = \sup_{x \in X} \big\{p^T x - f(F(x))\big\} = - \inf_{x \in X} \big\{f(F(x)) - p^T x\big\}.$$

We are interested in writing the minimization problem above in the form of (P_u). Consider the functions

$$A : \mathbb{R}^k \times \mathbb{R}^n \to \overline{\mathbb{R}}, \ A(y, z) = f(y) - p^T z$$

and

$$B : X \to \mathbb{R}^k \times \mathbb{R}^n, \ B(x) = (F(x), x).$$

After a standard verification A turns out to be convex and $(K \times \{0\}^n)$ - increasing, while B is $(K \times \{0\}^n)$ - convex on X. It is not difficult to notice that

$$\inf_{x \in X} \big\{f(F(x)) - p^T x\big\} = \inf_{x \in X} A(B(x)).$$

According to Theorem 4.5, the values of these infima coincide with the optimal value of the Fenchel - Lagrange dual problem to the minimization problem in the right - hand side,

$$(P_a) \qquad \inf_{x \in X} A(B(x)),$$

when the constraint qualification (CQ_u) is fulfilled for B and the corresponding sets. Let us formulate its dual problem and the constraint qualification needed here. The first one arises from (D_u), being

$$(D_a) \qquad \sup_{\substack{\beta \in K^*,\\ \gamma \in \mathbb{R}^n}} \Big\{ - A^*(\beta, \gamma) - \big((\beta, \gamma)^T B\big)^*_X(0)\Big\},$$

while the constraint qualification is

$$(CQ_a) \qquad \exists x' \in \mathrm{ri}(X) : B(x') \in \mathrm{ri}(\mathrm{dom}(f) \times \mathbb{R}^n) - \mathrm{ri}(K \times \{0\}^n),$$

simplifiable to

$$(CQ_a) \qquad \exists x' \in \mathrm{ri}(X) : F(x') \in \mathrm{ri}(\mathrm{dom}(f)) - \mathrm{ri}(K),$$

or equivalently,

$$(CQ_a) \qquad 0 \in F(\mathrm{ri}(X)) - \mathrm{ri}(\mathrm{dom}(f)) + \mathrm{ri}(K).$$

By Lemma 2.1, (CQ_a) is actually

$$(CQ_a) \qquad \mathrm{ri}(F(X) + K) \cap \mathrm{ri}(\mathrm{dom}(f)) \neq \emptyset.$$

To determine a formulation of the dual problem that contains only the conjugates of f and F regarding X, we have to determine the following conjugate functions

$$\begin{aligned} A^*(\beta,\gamma) &= \sup_{\substack{y\in\mathbb{R}^k,\\ z\in\mathbb{R}^n}} \left\{\beta^T y + \gamma^T z - A(y,z)\right\} \\ &= \sup_{\substack{y\in\mathbb{R}^k,\\ z\in\mathbb{R}^n}} \left\{\beta^T y + \gamma^T z - f(y) + p^T z\right\} \\ &= \sup_{y\in\mathbb{R}^k} \left\{\beta^T y - f(y)\right\} + \sup_{z\in\mathbb{R}^n} \left\{\gamma^T z + p^T z\right\} \\ &= f^*(\beta) + \begin{cases} 0, & \text{if } \gamma = -p, \\ +\infty, & \text{otherwise,} \end{cases} \end{aligned}$$

and

$$\left((\beta,\gamma)^T B\right)^*_X(0) = \sup_{x\in X} \left\{0 - \beta^T F(x) - \gamma^T x\right\} = \left(\beta^T F\right)^*_X(-\gamma).$$

As the plus infinite value is not relevant for A^* in (D_a) which is a maximization problem where this function appears with a leading minus in front of, we take further $\gamma = -p$ and the dual problem becomes

$$(D_a) \qquad \sup_{\beta\in K^*} \left\{ - f^*(\beta) - \left(\beta^T F\right)^*_X(p)\right\}.$$

When the constraint qualification is satisfied, i.e. $\mathrm{ri}(F(X)+K)\cap\mathrm{ri}(\mathrm{dom}(f)) \neq \emptyset$, there is strong duality between (P_a) and (D_a), so we have

$$\begin{aligned} (f\circ F)^*_X(p) &= -\inf_{x\in X} \left[f(F(x)) - p^T x\right] = -\sup_{\beta\in K^*} \left\{ - f^*(\beta) - \left(\beta^T F\right)^*_X(p)\right\} \\ &= \inf_{\beta\in K^*} \left[f^*(\beta) + \left(\beta^T F\right)^*_X(p)\right] = \min_{\beta\in K^*} \left[f^*(\beta) + \left(\beta^T F\right)^*_X(p)\right]. \end{aligned}$$

Hence we have proven the following statement.

Theorem 4.6 *(the conjugate of the composition) Let K be a non - empty closed convex cone in $\mathbb{R}^k$ and X a non - empty convex subset of $\mathbb{R}^n$. Take $f : \mathbb{R}^k \to \overline{\mathbb{R}}$ to be a proper K - increasing convex function and $F : X \to \mathbb{R}^k$ a function K - convex on X such that $F(X) \cap \mathrm{dom}(f) \neq \emptyset$. Then the fulfillment of (CQ_a) yields*

$$(f\circ F)^*_X(p) = \min_{\beta\in K^*} \left[f^*(\beta) + \left(\beta^T F\right)^*_X(p)\right] \ \forall p \in \mathbb{R}^n. \tag{4. 4}$$

Unlike [44] no lower semicontinuity assumption regarding f or F is necessary for the validity of formula (4. 4). Let us prove now that the condition (CQ_a) is weaker than the one required in the work cited above, which is in our case

$$F(X) \cap \mathrm{int}(\mathrm{dom}(f)) \neq \emptyset. \tag{4. 5}$$

Assuming (4. 5) true let z' belong to the both sets involved there. It follows that $z' \in F(X)+K$ and $\mathrm{int}(\mathrm{dom}(f)) \neq \emptyset$, so $\mathrm{ri}(\mathrm{dom}(f)) = \mathrm{int}(\mathrm{dom}(f))$ which is an open set. On the other hand $F(X) + K$ is a convex set, so it has a non - empty relative interior. Take $z'' \in \mathrm{ri}(F(X) + K)$.

According to Theorem 6.1 in [72], for any $\lambda \in (0,1]$ one has $(1-\lambda)z' + \lambda z'' \in \mathrm{ri}\big(F(X)+K\big)$. As $z' \in \mathrm{int}(\mathrm{dom}(f))$ which is an open set, there is a $\bar{\lambda} \in (0,1]$ such that $\bar{z} = (1-\bar{\lambda})z' + \bar{\lambda}z'' \in \mathrm{int}(\mathrm{dom}(f)) = \mathrm{ri}(\mathrm{dom}(f))$. Therefore

$$\bar{z} \in \mathrm{ri}\big(F(X)+K\big) \cap \mathrm{ri}(\mathrm{dom}(f)),$$

i.e. (CQ_a) is fulfilled.

An example where our condition (CQ_a) is applicable, while (4. 5) fails follows.

Example 4.1 Take $k=2$, $X=\mathbb{R}$, $K=\{0\}\times\mathbb{R}_+$, $F:\mathbb{R}\to\mathbb{R}^2$, defined by $F(x)=(0,x)\ \forall x\in\mathbb{R}$ and $f:\mathbb{R}^2\to\overline{\mathbb{R}}$, given for any pair $(x,y)\in\mathbb{R}^2$ by

$$f(x,y) = \begin{cases} y, & \text{if } x=0, \\ +\infty, & \text{otherwise.} \end{cases}$$

It is easy to verify that F is K - convex, f is proper convex and K - increasing and $K^* = \mathbb{R}\times\mathbb{R}_+$. We also have $F(X)=\{0\}\times\mathbb{R}$, $\mathrm{dom}(f)=\{0\}\times\mathbb{R}$, $\mathrm{int}(\mathrm{dom}(f))=\emptyset$ and $\mathrm{ri}(\mathrm{dom}(f))=\{0\}\times\mathbb{R}$. The feasibility condition $F(X)\cap\mathrm{dom}(f)\neq\emptyset$ is satisfied, being in this case $\{0\}\times\mathbb{R}\neq\emptyset$. Let us notice also that since $X=\mathbb{R}$ the conjugates regarding X are actually the usual conjugate functions.

As $(f\circ F)(x) = f(0,x) = x\ \forall x\in\mathbb{R}$, it follows

$$(f\circ F)^*(p) = \begin{cases} 0, & \text{if } p=1, \\ +\infty, & \text{otherwise.} \end{cases}$$

We also have, for all $(a,b)\in\mathbb{R}\times\mathbb{R}_+$ and all $p\in\mathbb{R}$,

$$f^*(a,b) = \begin{cases} 0, & \text{if } b=1, \\ +\infty, & \text{otherwise,} \end{cases} \text{ and } \big((a,b)F\big)^*(p) = \begin{cases} 0, & \text{if } b=p, \\ +\infty, & \text{otherwise,} \end{cases}$$

which yields

$$\min_{(a,b)\in\mathbb{R}\times\mathbb{R}_+}\Big[f^*(a,b) + \big((a,b)F\big)^*(p)\Big] = \begin{cases} 0, & \text{if } p=1, \\ +\infty, & \text{otherwise.} \end{cases}$$

Therefore the formula (4. 4) is valid. Let us see what happens to (CQ_a) and (4. 5). Taking into consideration the things above, (CQ_a) means $\{0\}\times\mathbb{R}\neq\emptyset$, while (4. 5) is $\{0\}\times\mathbb{R}\cap\emptyset\neq\emptyset$. It is clear that the latter is false, while our new condition is satisfied. Therefore (CQ_a) is indeed weaker than (4. 5).

The formula of the conjugate of the postcomposition with an increasing convex function becomes for an appropriate choice of the functions and for $K=[0,+\infty)$ similar to the result given by Hiriart - Urruty and Lemaréchal in Theorem X.2.5.1 in [44]. As shown above, there is no need to impose lower semicontinuity on the functions involved and a so strong constraint qualification as there.

We conclude this section with a concrete problem where the results given in this section find a good application.

Example 4.2 (see also [45]) Let $F:X\to\mathbb{R}$ be a function concave on X with strictly positive values, where X is a non - empty convex subset of $\mathbb{R}^n$. We want to determine the value of the conjugate function of $1/F$ at some $p\in\mathbb{R}^n$. According to the preceding results, we write $\big(1/F\big)^*_X(p)$ as an unconstrained composed convex problem by taking $K=(-\infty,0]$, which is a non - empty closed convex cone and $f:\mathbb{R}\to\overline{\mathbb{R}}$ with $f(y)=1/y$ for $y\in(0,+\infty)$ and $+\infty$ otherwise. It is interesting to notice that the concave on X function F is actually K - convex on X for this K while f is K - increasing. Now let us see when the constrained qualification (CQ_a)

specialized for this problem is valid. It is in this case

$$(CQ_e) \qquad \operatorname{ri}\big(F(X)+(-\infty,0]\big)\cap(0,+\infty)\neq\emptyset,$$

which is nothing but

$$(CQ_e) \qquad \big(F(\operatorname{ri}(X))+(-\infty,0)\big)\cap(0,+\infty)\neq\emptyset,$$

which is always fulfilled since F has only strictly positive values.

So the formula (4. 4) obtained before can be applied without any additional assumption. We have

$$\left(\frac{1}{F}\right)^*_X(p)=\inf_{\beta\leq 0}\big[f^*(\beta)+(\beta F)^*_X(p)\big].$$

As f is known, we can calculate its conjugate function at some $\beta\leq 0$, which is actually

$$f^*(\beta)=\sup_{y>0}\left\{\beta y-\frac{1}{y}\right\}=\begin{cases}-2\sqrt{-\beta}, & \text{if } \beta<0,\\ 0, & \text{if } \beta=0.\end{cases}$$

Meanwhile, for $(\beta F)^*_X$ we have

$$(\beta F)^*_X(p)=\begin{cases}-\beta(-F)^*_X\left(\frac{1}{-\beta}p\right), & \text{if } \beta<0,\\ \delta^*_X(p), & \text{if } \beta=0,\end{cases}$$

We conclude after changing the sign of β that the formula of the conjugate of $1/F$ is

$$\left(\frac{1}{F}\right)^*_X(p)=\min\left\{\inf_{\beta>0}\left[\beta(-F)^*_X\left(\frac{1}{\beta}p\right)-2\sqrt{\beta}\right],\sigma_X(p)\right\}.$$

When the value of the conjugate is finite either it is equal to $\sigma_X(p)$ or there is a $\bar{\beta}>0$ for which the infimum in the right - hand side is attained. The value of the infimum gives in this latter case actually the formula of the conjugate.

4.1.3 Duality via a more general scalarization in multiobjective optimization

Multiobjective optimization is a modern and fruitful research field with many practical applications, concerning especially economy but also various algorithms, location and transports, even medicine. One of the methods one can use to deal with a vector - minimization problem is via duality, and this is realized mostly by attaching a scalarized problem to the initial one. Using the scalarized problem and its dual a multiobjective dual problem to the primal vector problem is constructed and the duality assertions follow. Different scalarization methods were proposed in the literature, using linear functions, norms or various other constructions. In the following we introduce a Fenchel - Lagrange duality for multiobjective programming problems in a more general framework, using the scalarization with a strongly K - increasing function. This type of scalarization has already been used in the literature by Gerstewitz and Iwanow (cf. [38]), Göpfert and Gerth (cf. [40]), Jahn (cf. [46,47]) and Miglierina and Molho (cf. [64]), among others, but without connecting it to conjugate duality. As many of the other scalarizations used in the literature use strongly increasing functions, too, they can be rediscovered as special cases in the framework we describe here. This happens for the linear scalarization, maximum scalarization, norm scalarization and other scalarizations involving special strongly K - increasing functions.

4.1.3.1 Duality for the multiobjective problem

Let the convex multiobjective optimization problem

$$(P_v) \qquad \operatorname*{v-min}_{\substack{x\in X,\\ g(x)\in -C}} F(x),$$

where $F = (F_1, \ldots, F_k)^T$, g, X, K and C are considered like in the beginning of the subsection. Like before, $\mathcal{A}$ denotes the feasible set of the problem (P_v). By a solution to (P_v) one can understand different notions, we rely in this part of the present thesis to the following ones.

Definition 4.1 *(cf. [38]) An element $\bar{x} \in \mathcal{A}$ is called efficient with respect to (P_v) if from $F(x) - F(\bar{x}) \in -K$ for $x \in \mathcal{A}$ follows $F(x) = F(\bar{x})$.*

Let us denote by $\mathcal{S}$ an arbitrary set of K - strongly increasing convex functions $s : \mathbb{R}^k \to \mathbb{R}$.

Definition 4.2 *(see also [38]) An element $\bar{x} \in \mathcal{A}$ is called properly efficient with respect to (P_v) when there is some $s \in \mathcal{S}$ fulfilling $s(F(\bar{x})) \leq s(F(x))$ $\forall x \in \mathcal{A}$.*

In order to deal with (P_v) via duality we introduce the following family of scalarized problems

$$(P_s) \qquad \inf_{x\in\mathcal{A}} s(F(x)),$$

where $s \in \mathcal{S}$. For any $s \in \mathcal{S}$, from the previous sections (see (D_c)) we know that the Fenchel-Lagrange dual problem to (P_s) is

$$(D_s) \qquad \sup_{\substack{\alpha\in C^*,\beta\in K^*,\\ u\in\mathbb{R}^n}} \Big\{ -s^*(\beta) - (\beta^T F)^*_X(u) - (\alpha^T g)^*_X(-u)\Big\}.$$

Using this, we introduce the following multiobjective dual problem to (P_v)

$$(D_v) \qquad \operatorname*{v-max}_{(z,s,\alpha,\beta,u)\in\mathcal{B}} z,$$

where

$$\begin{aligned}\mathcal{B} \;=\; \Big\{&(z, s, \alpha, \beta, u) \in \mathbb{R}^k \times \mathcal{S} \times C^* \times K^* \times \mathbb{R}^n :\\ &s(z) \leq -s^*(\beta) - \big(\beta^T F\big)^*_X(u) - \big(\alpha^T g\big)^*_X(-u)\Big\}.\end{aligned}$$

Theorem 4.7 *(weak duality) There is no $x \in \mathcal{A}$ and no $(z, s, \alpha, \beta, u) \in \mathcal{B}$ such that $F(x) - z \in -K$ and $F(x) \neq z$.*

Proof. Assume that there are some $x \in \mathcal{A}$ and $(z, s, \alpha, \beta, u) \in \mathcal{B}$ contradicting the assumption. As s is K - strongly increasing it follows

$$s(F(x)) < s(z).$$

On the other hand,

$$s(z) \leq -s^*(\beta) - (\beta^T F)^*_X(u) - (\alpha^T g)^*_X(-u),$$

so we get

$$s(F(x)) < -s^*(\beta) - (\beta^T F)^*_X(u) - (\alpha^T g)^*_X(-u).$$

This last relation contradicts the weak duality theorem for (P_s) and (D_s), therefore the supposition we made is false and weak duality holds. □

Definition 4.3 *An element* $(\bar{z}, \bar{s}, \bar{\alpha}, \bar{\beta}, \bar{u}) \in \mathcal{B}$ *is called efficient with respect to* (D_v) *if from* $\bar{z} - z \in -K$ *for* $(z, s, \alpha, \beta, u) \in \mathcal{B}$ *follows* $z = \bar{z}$.

The constraint qualification that guarantees strong duality between (P_v) and its dual (D_v) comes immediately from (CQ_c), being

$$(CQ_v) \qquad \exists x' \in \operatorname{ri}(X) : g(x') \in -\operatorname{ri}(C).$$

Theorem 4.8 *(strong duality) Assume* (CQ_v) *fulfilled and let* $\bar{x} \in \mathcal{A}$ *be a properly efficient solution to* (P_v). *Then the dual problem* (D_v) *has an efficient solution* $(\bar{z}, \bar{s}, \bar{\alpha}, \bar{\beta}, \bar{u})$ *such that* $F(\bar{x}) = \bar{z}$.

Proof. According to Definition 4.2 there is an $\bar{s} \in \mathcal{S}$ such that $\bar{s}(F(\bar{x})) \leq \bar{s}(F(x))$ $\forall x \in \mathcal{A}$. It is obvious that $\bar{x}$ is also an optimal solution to the scalarized problem $(P_{\bar{s}})$, therefore $v(P_{\bar{s}}) > -\infty$. As (CQ_v) is assumed valid there is strong duality between $(P_{\bar{s}})$ and $(D_{\bar{s}})$ because of Theorem 4.2 (notice that as $\bar{s}$ has only finite values (CQ_c) reduces to (CQ_v)). Therefore $(D_{\bar{s}})$ has an optimal solution, say $(\bar{\alpha}, \bar{\beta}, \bar{u}) \in C^* \times K^* \times \mathbb{R}^n$. We have

$$\bar{s}(F(\bar{x})) = -\bar{s}^*(\bar{\beta}) - (\bar{\beta}^T F)_X^*(\bar{u}) - \left(\bar{\alpha}^T g\right)_X^*(-\bar{u}).$$

Let also $\bar{z} = F(\bar{x}) \in \mathbb{R}^k$. It is obvious that $(\bar{z}, \bar{s}, \bar{\alpha}, \bar{\beta}, \bar{u}) \in \mathcal{B}$ and so we have found a feasible point to the dual problem. It remains to prove that $(\bar{z}, \bar{s}, \bar{\alpha}, \bar{\beta}, \bar{u})$ is efficient with respect to (D_v). Supposing that there is some $(z', s', \alpha', \beta', u') \in \mathcal{B}$ such that $\bar{z} - z' \in -K$ and $\bar{z} \neq z'$, it follows that $F(\bar{x}) - z' \in -K$ and $F(\bar{x}) \neq z'$, which contradicts Theorem 4.7. $\Box$

Theorem 4.9 *(optimality conditions)*
(a) *If the constraint qualification* (CQ_v) *is fulfilled and the primal problem* (P_v) *has a properly efficient solution* $\bar{x}$, *then the dual problem* (D_v) *has an efficient solution* $(\bar{z}, \bar{s}, \bar{\alpha}, \bar{\beta}, \bar{u})$ *and the following optimality conditions are satisfied*

(i) $F(\bar{x}) = \bar{z}$,

(ii) $s^*(\bar{\beta}) + s(F(\bar{x})) = \bar{\beta}^T F(\bar{x})$,

(iii) $\left(\bar{\beta}^T F\right)_X^*(\bar{u}) + \bar{\beta}^T F(\bar{x}) = \bar{u}^T \bar{x}$,

(iv) $\left(\bar{\alpha}^T g\right)_X^*(-\bar{u}) + \bar{\alpha}^T g(\bar{x}) = -\bar{u}^T \bar{x}$,

(v) $\bar{\alpha}^T g(\bar{x}) = 0$.

(b) *If* $\bar{x}$ *is a feasible point to the primal problem* (P_v) *and* $(\bar{z}, \bar{s}, \bar{\alpha}, \bar{\beta}, \bar{u})$ *is feasible to the dual problem* (D_v) *fulfilling the optimality conditions* $(i) - (v)$, *then* $\bar{x}$ *is a properly efficient solution to* (P_v) *and* $(\bar{z}, \bar{s}, \bar{\alpha}, \bar{\beta}, \bar{u})$ *is efficient to the dual problem* (D_v).

Remark 4.2 The optimality conditions regarding (P_v) and (D_v) follow immediately from the ones concerning the problems (P_s) and (D_s) and Theorem 4.8. The result in Theorem 4.9(b) is valid even without assuming (CQ_v) fulfilled.

Next we show how the duality statements given above can be applied for other scalarizations in the literature. We have chosen three of them to be included here, namely the linear scalarization, the maximum scalarization and the norm scalarization.

4.1.3.2 Special case 1: linear scalarization

The most usual scalarization in vector optimization is the one with strongly increasing linear functionals, called linear scalarization or weighted scalarization. From the large amount of papers dealing with this kind of scalarization we mention here the works of Boţ and Wanka [18, 19, 89–91, 93], as they worked with Fenchel - Lagrange duality, too.

Take $K = \mathbb{R}^k_+$. For some fixed $\lambda = (\lambda_1, \ldots, \lambda_K)^T \in \operatorname{int}(\mathbb{R}^k_+)$, the scalarized primal problem is

$$(P_\lambda) \qquad \inf_{x \in \mathcal{A}} \left[\sum_{j=1}^{k} \lambda_j F_j(x) \right].$$

The linear scalarization is a special case of the general framework we presented as the objective function in (P_λ) can be written as $s_\lambda(F(x))$, for $s_\lambda(y) = \lambda^T y$ and it is clear that s_λ is $\mathbb{R}^k_+$ - strongly increasing and convex for any $\lambda \in \operatorname{int}(\mathbb{R}^k_+)$. In this case let $\mathcal{S} = \mathcal{S}_l$, the latter being defined as follows

$$\mathcal{S}_l = \Big\{ s_\lambda : \mathbb{R}^k \to \mathbb{R} : \ s_\lambda(y) = \lambda^T y, \lambda \in \operatorname{int}(\mathbb{R}^k_+) \Big\}.$$

The following definition of the proper efficient elements to (P_v) is available in this case (cf. [18, 19, 89–91, 93], among others).

Definition 4.4 *An element $\bar{x} \in \mathcal{A}$ is called (l) properly efficient with respect to (P_v) when there is $\lambda \in \operatorname{int}(\mathbb{R}^k_+)$ fulfilling $\sum_{j=1}^k \lambda_j F_j(\bar{x}) \leq \sum_{j=1}^k \lambda_j F(x)$ $\forall x \in \mathcal{A}$.*

Let us write now the dual problem to (P_v) that arises by using the scalarization function $s \in \mathcal{S}_l$. One can easily notice that the dual variable $s_\lambda \in \mathcal{S}_l$ that fulfills $s_\lambda(y) = \lambda^T y$ $\forall y \in \mathbb{R}^k$, where $\lambda \in \operatorname{int}(\mathbb{R}^k_+)$, can be replaced by the variable $\lambda \in \operatorname{int}(\mathbb{R}^k_+)$. Moreover, as $s^*_\lambda(u^*) = 0$ if $u^* = \lambda$ and $s^*_\lambda(u^*) = +\infty$ otherwise, the variable $\beta \in K^*$ from (D_v) is no longer necessary since the inequality in the feasible set of the dual problem is not fulfilled unless $\beta = \lambda$. Therefore the dual problem obtained in this case to (P_v) is

$$(D_l) \qquad \operatorname*{v-max}_{(z,\lambda,\alpha,u) \in \mathcal{B}_l} z,$$

where

$$\begin{aligned} \mathcal{B}_l \ = \ & \bigg\{ (z, \lambda, \alpha, u) \in \mathbb{R}^k \times \operatorname{int}(\mathbb{R}^k_+) \times C^* \times \mathbb{R}^n : z = (z_1, \ldots, z_k)^T, \\ & \lambda = (\lambda_1, \ldots, \lambda_k)^T, \ \sum_{j=1}^k \lambda_j z_j \leq -\sum_{j=1}^k (\lambda^T F)^*_X(u) - (\alpha^T g)^*_X(-u) \bigg\}. \end{aligned}$$

By Theorem 16.4 in [72] we have for λ and u taken like in $\mathcal{B}_l$

$$(\lambda^T F)^*_X(u) = \min \left\{ (\lambda_j F_j)^*_X(p_j) : \sum_{j=1}^k p_j = u \right\}$$

and, as $\lambda_j > 0$, $j = 1, \ldots, k$, this becomes

$$(\lambda^T F)^*_X(u) = \min \left\{ \lambda_j {F_j}^*_X \Big(\frac{1}{\lambda_j} p_j \Big) : \sum_{j=1}^k p_j = u \right\}.$$

Denoting $y_j = (1/\lambda_j)p_j$ for $j = 1, \dots, k$, and $y = (y_1, \dots, y_k)$, the latter dual problem turns into

$$(D_l') \qquad \underset{(z,\lambda,\alpha,y)\in\mathcal{B}_l'}{\text{v-max}} z,$$

with

$$\mathcal{B}_l' = \bigg\{(z,\lambda,\alpha,y) \in \mathbb{R}^k \times \text{int}(\mathbb{R}^k_+) \times C^* \times (\mathbb{R}^n \times \dots \mathbb{R}^n) : y = (y_1, \dots, y_k), z = (z_1, \dots, z_k)^T, \sum_{j=1}^k \lambda_j z_j \leq - \sum_{j=1}^k \lambda_j (F_j)^*_X(y_j) - (\alpha^T g)^*_X\bigg(- \sum_{j=1}^k \lambda_j y_j\bigg)\bigg\},$$

which is exactly the dual problem obtained by Boţ and Wanka in [18] and [19].

Theorem 4.10 *(weak duality) There is no $x \in \mathcal{A}$ and no $(z, \lambda, \alpha, y) \in \mathcal{B}_l'$ such that $F(x) \leqq z$ and $F(x) \neq z$.*

Theorem 4.11 *(strong duality) Assume (CQ_v) fulfilled and let $\bar{x} \in \mathcal{A}$ be an (l) properly efficient solution to (P_v). Then the dual problem (D_l') has an efficient solution $(\bar{z}, \bar{\lambda}, \bar{\alpha}, \bar{y})$ such that $F(\bar{x}) = \bar{z}$.*

4.1.3.3 Special case 2: maximum scalarization

Another scalarization met especially in the applications of vector optimization is the so - called Tchebyshev scalarization or maximum scalarization, where the scalarized problem's objective function consists in the maximal entry of the vector - function at each point. Among the papers dealing with this kind of scalarization we cite here Mbunga's [63], mentioning also [35]. The weighted Tchebyshev scalarization in [47, 65, 87] is closely related to the scalarization we treat here, but as the calculations work like in the case we treat we present the simpler situation. Take $K = \text{int}(\mathbb{R}^k_+) \cup \{0\}$. In this case the scalarized primal problem is

$$(P_{\max}) \qquad \inf_{x\in\mathcal{A}} \max_{j=1,\dots,k} F_j(x).$$

The maximum scalarization is a special case of the general framework we presented as the objective function in $(P_{\max})$ is K - strongly increasing and convex (see also Remark 4.1). The set $\mathcal{S}$ is in this case

$$\mathcal{S}_m = \Big\{s : \mathbb{R}^k \to \mathbb{R},\ s(x) = \max_{j=1}^k x_j,\ x = (x_1, \dots, x_k)^T \in \mathbb{R}^k\Big\}.$$

The following definition of the proper efficient elements to (P_v) is available in this case.

Definition 4.5 *An element $\bar{x} \in \mathcal{A}$ is called (m) properly efficient with respect to (P_v) when $\max_{j=1}^k F_j(\bar{x}) \leq \max_{j=1}^k F_j(x)\ \forall x \in \mathcal{A}$.*

Let us write now the dual problem to (P_v) when the scalarization function $s \in \mathcal{S}_m$. First, the variable $s \in \mathcal{S}_m$ disappears, as $s = \max_{j=1}^k$. We also have $K^* = \mathbb{R}^k_+$ and $s^*(y^*) = 0$ if $y^* = (y_1^*, \dots, y_k^*)^T \in \mathbb{R}^k_+$ and $\sum_{j=1}^k y_j^* = 1$ and $s^*(y^*) = +\infty$ otherwise. Therefore the dual problem obtained in this case to (P_v) is

$$(D_m) \qquad \underset{(z,\alpha,\beta,u)\in\mathcal{B}_m}{\text{v-max}} z,$$

where

$$\mathcal{B}_m = \Big\{(z,\alpha,\beta,u) \in \mathbb{R}^k \times C^* \times \mathbb{R}^k_+ \times \mathbb{R}^n : z = (z_1,\dots,z_k)^T, \beta = (\beta_1,\dots,\beta_k)^T \\ \sum_{j=1}^k \beta_j = 1, \max_{j=1}^{k}\{z_j\} \leq -(\beta^T F)^*_X(u) - (\alpha^T g)^*_X(-u)\Big\}.$$

Theorem 4.12 *(weak duality) There is no $x \in \mathcal{A}$ and no $(z,\alpha,\beta,u) \in \mathcal{B}_m$ such that $z - F(x) \in \operatorname{int}(\mathbb{R}^k_+)$.*

Theorem 4.13 *(strong duality) Assume (CQ_v) fulfilled and let $\bar{x} \in \mathcal{A}$ be an (m) properly efficient solution to (P_v). Then the dual problem (D_m) has an efficient solution $(\bar{z},\bar{\alpha},\bar{\beta},\bar{u})$ such that $F(\bar{x}) = \bar{z}$.*

4.1.3.4 Special case 3: norm scalarization

Other scalarizations used in the literature deal with monotonically increasing norms and scalarization functions generated by norms. Works due to Kaliszewski (cf. [54]), Khánh (cf. [58]), Rubinov and Gasimov (cf. [73]), Schandl, Klamroth and Wiecek (cf. [74]), Tammer (cf. [84]), Tammer and Göpfert (cf. [85]) and others contain multiobjective optimization problems scalarized by techniques involving norms. The scalarization functions in these papers are strongly increasing on different cones and one can apply the results we gave for the general scalarized problem (P_s) in a similar way to the scalarized problem we treat here. In the following we attach to (P_v) a scalarized problem obtained with the scalarization function used by Tammer and Winkler in [86] and by Winkler in [96] (see also Weidner's papers [94,95]). In order to proceed we need to introduce some special classes of norms, about which more is available in [74] and some references therein. Take the cone K such that $K = \operatorname{int}(K) \cup \{0\}$.

Definition 4.6 *A subset $A \subseteq \mathbb{R}^k$ is called polyhedral if it can be expressed as the intersection of a finite collection of closed half-spaces.*

Definition 4.7 *A norm $\gamma : \mathbb{R}^k \to \mathbb{R}$ is called block norm if its unit ball B_γ is polyhedral.*

Definition 4.8 *A norm $\gamma : \mathbb{R}^k \to \mathbb{R}$ is called absolute if for any $\bar{y} \in \mathbb{R}^k$ one has $\gamma(y) = \gamma(\bar{y})$ for all $y \in \big\{z = (z_1,\dots,z_k)^T \in \mathbb{R}^k : |z_j| = |\bar{y}_j| \ \forall j = 1,\dots,k\big\}$.*

Definition 4.9 *A block norm $\gamma : \mathbb{R}^k \to \mathbb{R}$ is called oblique if it is absolute and satisfies, for all $y \in \mathbb{R}^k_+ \cap \operatorname{bd}(B_\gamma)$,*

$$\big(y - \mathbb{R}^k_+\big) \cap \mathbb{R}^k_+ \cap \operatorname{bd}(B_\gamma) = \{y\}.$$

According to [74] and [86] (see Definition 4.7), for an absolute norm γ on $\mathbb{R}^k$ there are some $w \in \mathbb{N}$, $a_i \in \mathbb{R}^k$ and $\eta_i \in \mathbb{R}$, $i = 1,\dots,w$, such that the unit ball generated by γ is

$$B_\gamma = \Big\{y \in \mathbb{R}^k : a_i^T y \leq \eta_i, i = 1,\dots,w\Big\}.$$

We need also the following sets

$$I_\gamma = \Big\{i \in \{1,\dots,w\} : \Big\{y \in \mathbb{R}^k : a_i^T y = \eta_i\Big\} \cap B_\gamma \cap \operatorname{int}(\mathbb{R}^k_+) \neq \emptyset\Big\}$$

and

$$E_\gamma = \Big\{y \in \mathbb{R}^k : a_i^T y \leq \eta_i \ \forall i \in I_\gamma\Big\}.$$

Theorem 4.14 *(cf. [86]) The function $\zeta_{\gamma,l,v} : \mathbb{R}^k \to \mathbb{R}$, defined by*

$$\zeta_{\gamma,l,v}(y) = \inf\big\{\tau \in \mathbb{R} : y \in \tau l + E_\gamma + v\big\},$$

where γ is an absolute norm on $\mathbb{R}^k$, $l \in \operatorname{int}(\mathbb{R}^k_+)$ and $v \in \mathbb{R}^k$, is convex and K - strongly increasing when $\operatorname{bd}(E_\gamma) - (K \backslash \{0\}) \in \operatorname{int}(E_\gamma)$.

Corollary 4.1 *(cf. [86]) When γ is an absolute block norm and $K = \operatorname{int}(\mathbb{R}^k_+) \cup \{0\}$ then $\zeta_{\gamma,l,v}$ is K - strongly increasing for any $l \in \operatorname{int}(\mathbb{R}^k_+)$ and $v \in \mathbb{R}^k$.*

Corollary 4.2 *(cf. [86]) When γ is an oblique norm and $K = \mathbb{R}^k_+$ then $\zeta_{\gamma,l,v}$ is K - strongly increasing for any $l \in \operatorname{int}(\mathbb{R}^k_+)$ and $v \in \mathbb{R}^k$.*

Denote by $\mathcal{O}$ the set of the absolute norms $\gamma : \mathbb{R}^k \to \mathbb{R}$ and consider the following set

$$\mathcal{S}_n = \Big\{\gamma \in \mathcal{O} : \operatorname{bd}(E_\gamma) - \operatorname{int}(K) \in \operatorname{int}(E_\gamma)\Big\} \times \operatorname{int}(\mathbb{R}^k_+) \times \mathbb{R}^k.$$

The family of scalarized problems attached to (P_v) in this case is

$$(P_{\gamma,l,v}) \qquad \inf_{x \in \mathcal{A}} \zeta_{\gamma,l,v}(F(x)),$$

where $(\gamma, l, v) \in \mathcal{S}_n$. According to the definitions above and Theorem 4.14 this fits into our framework, too, by taking $\mathcal{S} = \mathcal{S}_n$. The definition of the proper efficient elements comes as follows.

Definition 4.10 *An element $\bar{x} \in \mathcal{A}$ is called (n) properly efficient with respect to (P_v) when there is an absolute norm γ, some $l \in \operatorname{int}(\mathbb{R}^k_+)$ and $v \in \mathbb{R}^k$ such that $\zeta_{\gamma,l,v}(F(\bar{x})) \le \zeta_{\gamma,l,v}(F(x))\ \forall x \in \mathcal{A}$.*

To obtain the dual problem to (P_v) that arises by using the scalarization just presented, let us calculate the conjugate of the scalarization functions $\zeta_{\gamma,l,v}$, for some fixed $(\gamma, l, v) \in \mathcal{S}_n$. We have

$$\begin{aligned}
\zeta^*_{\gamma,l,v}(y^*) &= \sup_{y \in \mathbb{R}^k} \Big\{ {y^*}^T y - \inf\big[\tau \in \mathbb{R} : y \in \tau l + E_\gamma + v\big]\Big\} \\
&= \sup_{y \in \mathbb{R}^k} \Big\{ {y^*}^T y + \sup\big\{ -\tau \in \mathbb{R} : y \in \tau l + E_\gamma + v\big\}\Big\}.
\end{aligned}$$

Denoting $w = y - \tau l - v$, one gets

$$\begin{aligned}
\zeta^*_{\gamma,l,v}(y^*) &= \sup_{\tau \in \mathbb{R}} \bigg\{ -\tau + \sup_{w \in E_\gamma} \Big\{ {y^*}^T (w + \tau l + v)\Big\}\bigg\} \\
&= \sup_{\tau \in \mathbb{R}} \bigg\{ -\tau + \tau {y^*}^T l + \sup_{w \in E_\gamma} {y^*}^T w \bigg\} + {y^*}^T v \\
&= \sup_{\tau \in \mathbb{R}} \Big\{ \tau\Big({y^*}^T l - 1\Big)\Big\} + \sigma_{E_\gamma}(y^*) + {y^*}^T v \\
&= \begin{cases} \sigma_{E_\gamma}(y^*) + {y^*}^T v, & \text{if } {y^*}^T l = 1, \\ +\infty, & \text{otherwise.} \end{cases}
\end{aligned}$$

The dual problem to (P_v) obtained in this case is

$$(D_n) \qquad \underset{(z,\gamma,l,v,\alpha,\beta,u) \in \mathcal{B}_n}{\text{v-max}} z,$$

where

$$\begin{aligned}
\mathcal{B}_n = \Big\{ &(z, \gamma, l, v, \alpha, \beta, u) \in \mathbb{R}^k \times \mathcal{S}_n \times C^* \times K^* \times \mathbb{R}^n : \beta^T l = 1, \\
&\zeta_{\gamma,l,v}(z) \le -\sigma_{E_\gamma}(\beta) + \beta^T v - (\beta^T F)^*_X(u) - (\alpha^T g)^*_X(-u)\Big\}.
\end{aligned}$$

Theorem 4.15 *(weak duality) There is no $x \in \mathcal{A}$ and no $(z, \gamma, l, v, \alpha, \beta, u) \in \mathcal{B}_n$ such that $z - F(x) \in \operatorname{int}(K)$.*

Theorem 4.16 *(strong duality) Assume (CQ_v) fulfilled and let $\bar{x} \in \mathcal{A}$ be an (n) properly efficient solution to (P_v). Then the dual problem (D_n) has an efficient solution $(\bar{z}, \overline{\gamma}, \bar{l}, \bar{v}, \bar{\alpha}, \bar{\beta}, \bar{u})$ such that $F(\bar{x}) = \bar{z}$.*

4.2 Problems with convex entropy - like objective functions

Entropy optimization is a modern and fruitful research area for scientists having various backgrounds: mathematicians, physicists, engineers, even chemists or linguists. For comprehensive studies on its history and applications the reader is referred to the quite recent books [34, 57]. We mention just that the most important entropy measures are due to Shannon (cf. [83]), Kullback and Leibler (cf. [59]) and, respectively, Burg (cf. [24]).

Many papers, including two co - written by the author of the present thesis together with Boţ and Wanka (see [13, 16]), and books among which we have just mentioned two (see [34, 57]) deal with entropy optimization, especially with its multitude of applications in various fields such as transport and location problems, pattern and image recognition, text classification, image reconstruction, etc.

The problem we consider here cannot be classified as a pure entropy optimization problem. It is actually a generalization of the usual entropy optimization problems and we argue this statement by the special cases we will present in Subsection 4.2.2. These special cases cover a multitude of problems in entropy optimization, thus our results provide a good framework to deal with many entropy optimization problems. To remain connected to the previous results in this thesis let us mention that many entropy optimization problems were treated so far via posynomial geometric programming, as the objective function of the logarithmed dual problem in posynomial programming $(\widetilde{D}_7)$ is actually an entropy type function. We treat these problems via conjugate duality, as they occur as special cases of the general entropy - like problem we deal with.

4.2.1 Duality for problems with entropy - like objective functions

Consider the non - empty convex set $X \subseteq \mathbb{R}^n$, the affine functions $f_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, k$, the concave functions $g_i : \mathbb{R}^n \to \overline{\mathbb{R}}$, $i = 1, \ldots, k$, and the functions $h_j : X \to \mathbb{R}$, $j = 1, \ldots, m$, which are convex on X. Assume that for $i = 1, \ldots, k$, $f_i(x) \geq 0$ and $0 < g_i(x) \neq +\infty$ when $x \in X$ such that $h(x) \leqq 0$, where $h = (h_1, \ldots, h_m)^T$. Denote further $f = (f_1, \ldots, f_k)^T$ and $g = (g_1, \ldots, g_k)^T$.

The convex optimization problem we consider throughout this section is

$$(P_f) \qquad \inf_{\substack{x \in X,\\ h(x) \leqq 0}} \left[\sum_{i=1}^{k} f_i(x) \ln \left(\frac{f_i(x)}{g_i(x)} \right) \right].$$

As usual in entropy optimization we use further the convention $0 \ln 0 = 0$.

The objective function of problem (P_f) is a Kullback - Leibler type sum, but instead of probabilities we have as terms functions. To the best of our knowledge this kind of objective function has not been considered yet in the literature. There are some papers dealing with problems having as objective function expressions like $\int f(t) \ln(f(t)/g(t)) dt$, such as [4], but the results described there do not interfere with ours. Of course the functions involved in the objective function of the problem

(P_f) may take some particular shapes and (P_f) turns into an entropy optimization problem.

Using a special construction, similar to the one used by Wanka and Boţ in [91], we obtain another problem that is equivalent to (P_f), whose dual problem (D_f) is easier to determine. Let us introduce, for $i = 1, \dots, k$, the functions $\Phi_i : \mathbb{R}^2 \to \overline{\mathbb{R}}$,

$$\Phi_i(s_i, t_i) = \begin{cases} s_i \ln\left(\frac{s_i}{t_i}\right), & \text{if } s_i \geq 0 \text{ and } t_i > 0, \\ +\infty, & \text{otherwise}, \end{cases}$$

where $s = (s_1, \dots, s_k)^T$, $t = (t_1, \dots, t_k)^T$ and the set

$$\mathcal{E} = \left\{(x, s, t) \in X \times \mathbb{R}^k_+ \times \operatorname{int}(\mathbb{R}^k_+) : h(x) \leqq 0, f(x) = s, t \leqq g(x)\right\}.$$

Now we consider a new optimization problem

$$(P_\Phi) \qquad \inf_{(x,s,t)\in\mathcal{E}} \left[\sum_{i=1}^{k} \Phi_i(s_i, t_i)\right].$$

For each $i = 1, \dots, k$, the function Φ_i is convex, being the extension with positive infinite values to the whole space of a convex function (see [57]). The convexity of the set $\mathcal{E}$ follows from its definition.

Remark 4.3 The observations above assure the convexity of the problem (P_Φ).

Even if the problems (P_f) and (P_Φ) seem related, an accurate connection between their optimal objective values is required. The following assertion states it.

Proposition 4.2 *The problems (P_f) and (P_Φ) are equivalent in the sense that $v(P_f) = v(P_\Phi)$.*

Proof. Let us take first an element $x \in X$ such that $h(x) \leqq 0$. It is obvious that $(x, f(x), g(x)) \in \mathcal{E}$. Further,

$$\sum_{i=1}^{k} f_i(x) \ln\left(\frac{f_i(x)}{g_i(x)}\right) = \sum_{i=1}^{k} \Phi_i(f_i(x), g_i(x)) \geq v(P_\Phi).$$

As x is chosen arbitrarily in order to fulfill the constraints of the problem (P_f) we can conclude for the moment that $v(P_f) \geq v(P_\Phi)$.

Conversely, take a triplet $(x, s, t) \in \mathcal{E}$. This means that we have for each $i = 1, \dots, k$, $f_i(x) = s_i$ and $g_i(x) \geq t_i$. Further we have for all $i = 1, \dots, k$, $1/(g_i(x)) \leq 1/t_i$, followed by $(f_i(x))/(g_i(x)) \leq s_i/t_i$. Consequently, because ln is a monotonically increasing function, it holds $\ln((f_i(x))/(g_i(x))) \leq \ln(s_i/t_i)$, $i = 1, \dots, k$. Multiplying the terms in both sides by the corresponding $f_i(x) = s_i$, $i = 1, \dots, k$, and assembling the resulting relations it follows

$$\sum_{i=1}^{k} \Phi_i(s_i, t_i) \geq \sum_{i=1}^{k} f_i(x) \ln\left(\frac{f_i(x)}{g_i(x)}\right) \geq v(P_f).$$

As the element (x, s, t) has been taken arbitrarily in $\mathcal{E}$, it yields $v(P_\Phi) \geq v(P_f)$. Therefore, $v(P_f) = v(P_\Phi)$. □

Further we determine the Lagrange dual problem to (P_Φ), that is also a dual to the problem (P_f). It has the following raw formulation, where q^f, q^g and q^h are the Lagrange multipliers,

$$(D_\Phi)\ \sup_{\substack{q^f\in\mathbb{R}^k,\\ q^g\in\mathbb{R}^k_+,\\ q^h\in\mathbb{R}^m_+}} \inf_{\substack{x\in X,\\ s\in\mathbb{R}^k_+,\\ t\in\operatorname{int}(\mathbb{R}^k_+)}} \left[\sum_{i=1}^k s_i\ln\left(\frac{s_i}{t_i}\right)+(q^h)^T h(x)+(q^f)^T(f(x)-s)+(q^g)^T(t-g(x))\right].$$

Taking a closer look to the infimum that appears above one may notice that it is separable into a sum of infima in the following way

$$\begin{aligned}
&\inf_{\substack{x\in X,\\ s\in\mathbb{R}^k_+,\\ t\in\operatorname{int}(\mathbb{R}^k_+)}} \left[\sum_{i=1}^k s_i\ln\left(\frac{s_i}{t_i}\right)+(q^h)^T h(x)+(q^f)^T(f(x)-s)+(q^g)^T(t-g(x))\right]\\
=\ &\inf_{x\in X}\left[(q^h)^T h(x)+(q^f)^T f(x)-(q^g)^T g(x)\right]\\
+\ &\sum_{i=1}^k \inf_{\substack{s_i\geq 0,\\ t_i>0}}\left[s_i\ln\left(\frac{s_i}{t_i}\right)-q_i^f s_i+q_i^g t_i\right],
\end{aligned}$$

where $q^g=\left(q_1^g,\dots,q_k^g\right)^T$ and $q^f=\left(q_1^f,\dots,q_k^f\right)^T$. We can calculate the infima regarding $s_i\geq 0$ and $t_i>0$ for all $i=1,\dots,k$,

$$\inf_{\substack{s_i\geq 0,\\ t_i>0}}\left[s_i\ln\left(\frac{s_i}{t_i}\right)-q_i^f s_i+q_i^g t_i\right]=\inf_{s_i\geq 0}\left[s_i\ln s_i-q_i^f s_i+\inf_{t_i>0}\left[q_i^g t_i-s_i\ln t_i\right]\right].$$

In order to resolve the inner infimum, consider the function $\varphi:(0,+\infty)\to\mathbb{R}$, $\varphi(t)=\alpha t-\beta\ln t$, where $\alpha>0$ and $\beta\geq 0$. Its minimum is attained at $t=\beta/\alpha>0$, being $\varphi(\beta/\alpha)=\beta-\beta\ln\beta+\beta\ln\alpha$. Applying this result to the infima concerning t_i in the expressions above for $i=1,\dots,k$, there follows

$$\inf_{t_i>0}\left[q_i^g t_i-s_i\ln t_i\right]=\begin{cases} s_i-s_i\ln s_i+s_i\ln q_i^g, & \text{if } q_i^g>0,\\ 0, & \text{if } q_i^g=0 \text{ and } s_i=0,\\ -\infty, & \text{if } q_i^g=0 \text{ and } s_i>0.\end{cases}$$

Further we have to calculate for each $i=1,\dots,k$ the infimum above with respect to $s_i\geq 0$ after replacing the infimum concerning t_i with its value. In case $q_i^g>0$ we have

$$\begin{aligned}
\inf_{s_i\geq 0}\left[s_i\ln s_i-q_i^f s_i+s_i-s_i\ln s_i+s_i\ln q_i^g\right] &= \inf_{s_i\geq 0}\left[s_i(1-q_i^f+\ln q_i^g)\right]\\
&= \begin{cases} 0, & \text{if } 1-q_i^f+\ln q_i^g\geq 0,\\ -\infty, & \text{otherwise.}\end{cases}
\end{aligned}$$

When $q_i^g=0$ the infimum concerning s_i is equal to $-\infty$.

One may conclude for each $i=1,\dots,k$, the following

$$\inf_{\substack{s_i\geq 0,\\ t_i>0}}\left[s_i\ln\left(\frac{s_i}{t_i}\right)-q_i^f s_i+q_i^g t_i\right]=\begin{cases} 0, & \text{if } 1-q_i^f+\ln q_i^g\geq 0 \text{ and } q_i^g>0,\\ -\infty, & \text{otherwise.}\end{cases} \tag{4. 6}$$

The negative infinite values are not relevant to the dual problem we work on since after determining the inner infima one has to calculate the supremum of the obtained values, so we must consider further the cases where the infima with respect to $s_i\geq 0$ and $t_i>0$ are 0, i.e. the following constraints have to be fulfilled $1-q_i^f+\ln q_i^g\geq 0$ and $q_i^g>0$, $i=1,\dots,k$. The former additional constraints are equivalent to $q_i^g\geq e^{q_i^f-1}$ $\forall i=1,\dots,k$. Let us write now the final form of the dual problem to (P_Φ), after noticing that as $e^{q_i^f-1}>0$ the constraints $q^g\in\mathbb{R}^k_+$ and

$q_i^g > 0$, $i = 1, \dots, k$, become redundant and may be ignored,

$$(D_f) \qquad \sup_{\substack{q^f \in \mathbb{R}^k, q^h \in \mathbb{R}^m_+, \\ q_i^g \geq e^{q_i^f - 1}, i=1,\dots,k}} \inf_{x \in X} \Big[(q^h)^T h(x) + (q^f)^T f(x) - (q^g)^T g(x) \Big].$$

Although (D_f) has been obtained via Lagrange duality from (P_Φ) we refer to it further as the dual problem to (P_f) since (P_f) and (P_Φ) are equivalent. We could go even further, to the Fenchel - Lagrange dual problem to (P_f), but, as the Lagrange dual is suitable for our purposes we will use it for the moment. Then, in the special cases, the conjugates of the functions involved will appear.

Next we present the duality assertions regarding (P_f) and (D_f). Weak duality always holds, but we cannot say the same about strong duality. Alongside the initial convexity assumptions for X and h_i, $i = 1, \dots, m$, the concavity of g_i, $i = 1, \dots, k$, and the affinity of the functions f_i, $i = 1, \dots, k$, an additional constraint qualification is sufficient in order to achieve strong duality. The one we use here comes from (CQ_o), being

$$(CQ_f) \qquad \exists x' \in \mathrm{ri}(X) : \begin{cases} f(x') > 0, \\ h_j(x') \leq 0, & \text{if } j \in L_f, \\ h_j(x') < 0, & \text{if } j \in \{1, \dots, m\} \backslash L_f, \end{cases}$$

where we have divided the set $\{1, \dots, m\}$ into two disjunctive sets as before, L_f containing the indices of the functions h_j that are restrictions to X of affine functions, $j \in \{1, \dots, m\}$.

The strong duality statement arises naturally.

Theorem 4.17 *(strong duality) If the constraint qualification (CQ_f) is fulfilled then there is strong duality between problems (P_f) and (D_f), i.e. (D_f) has an optimal solution and $v(P_f) = v(P_\Phi) = v(D_f)$.*

Proof. Since $\Phi_i(s_i, t_i) \geq 0$ $\forall (x, s, t) \in \mathcal{E}$ $\forall i \in \{1, \dots, k\}$ (the most important properties of the Kullback - Leibler entropy measure are presented and proved in [57]) it follows

$$v(P_\Phi) \geq 0.$$

The constraint qualification (CQ_f) being fulfilled, there is a triplet $(x', s', t') \in \mathrm{ri}(X) \times \mathrm{int}(R^k_+) \times \mathrm{int}(R^k_+)$ such that

$$\begin{cases} h_j(x') \leq 0, & \text{if } j \in L_f, \\ h_j(x') < 0, & \text{if } j \in \{1, \dots, m\} \backslash L_f, \\ f(x') = s', \\ t_i' < g_i(x'), & \text{for } i = 1, \dots, k. \end{cases}$$

For instance take $s' = f(x')$ and $t' = (1/2)g(x')$.

The results above allow us to apply Theorem 5.7 in [32], so strong duality between (P_Φ) and (D_f) is certain, i.e. (D_f) has an optimal solution and $v(P_\Phi) = v(D_f)$. Proposition 4.2 yields $v(P_f) = v(D_f)$. □

Another step forward is to present some necessary and sufficient optimality conditions regarding the pair of dual problems we treat.

Theorem 4.18 *(optimality conditions)*
(a) Let the constraint qualification (CQ_f) be fulfilled and assume that the primal problem (P_f) has an optimal solution $\bar{x}$. Then the dual problem (D_f) has an optimal solution, too, let it be $(\bar{q}^f, \bar{q}^g, \bar{q}^h)$, and the following optimality conditions are true,

(i) $f_i(\bar{x}) \ln\left(\frac{f_i(\bar{x})}{g_i(\bar{x})}\right) = \bar{q}_i^f f_i(\bar{x}) - \bar{q}_i^g g_i(\bar{x}),\ i = 1, \dots, k,$

(ii) $\inf\limits_{x \in X} \left[(\bar{q}^h)^T h(x) + (\bar{q}^f)^T f(x) - (\bar{q}^g)^T g(x)\right] = (\bar{q}^f)^T f(\bar{x}) - (\bar{q}^g)^T g(\bar{x}),$

(iii) $\bar{q}_j^h h_j(\bar{x}) = 0,\ j = 1, \dots, m.$

(b) If $\bar{x}$ is a feasible point to (P_f) and $(\bar{q}^f, \bar{q}^g, \bar{q}^h)$ is feasible to (D_f) fulfilling the optimality conditions $(i) - (iii)$, then there is strong duality between (P_f) and (D_f). Moreover, $\bar{x}$ is an optimal solution to the primal problem and $(\bar{q}^f, \bar{q}^g, \bar{q}^h)$ an optimal solution to the dual.

Proof. (a) Under weaker assumptions than here Theorem 4.17 yields strong duality between (P_f) and (D_f). Therefore the existence of an optimal solution $(\bar{q}^f, \bar{q}^g, \bar{q}^h)$ to the dual problem is guaranteed. Moreover, $v(P_f) = v(D_f)$ and because (P_f) has an optimal solution its optimal objective value is attained at $\bar{x}$ and we have

$$\sum_{i=1}^k f_i(\bar{x}) \ln\left(\frac{f_i(\bar{x})}{g_i(\bar{x})}\right) = \inf_{x \in X} \left[(\bar{q}^h)^T h(x) + (\bar{q}^f)^T f(x) - (\bar{q}^g)^T g(x)\right]. \tag{4. 7}$$

Earlier we have proven the validity of (4. 6). Using it we can determine the conjugate function of Φ_i, $i = 1, \dots, k$, at $\left(\bar{q}_i^f, -\bar{q}_i^g\right)$ as follows

$$\begin{aligned}
\Phi_i^*(\bar{q}_i^f, -\bar{q}_i^g) &= \sup_{(s_i, t_i) \in \mathbb{R}^2} \left\{(\bar{q}_i^f, -\bar{q}_i^g)^T (s_i, t_i) - \Phi_i(s_i, t_i)\right\} \\
&= \sum_{i=1}^k \sup_{\substack{s_i \geq 0, \\ t_i > 0}} \left\{\bar{q}_i^f s_i - \bar{q}_i^g t_i - s_i \ln\left(\frac{s_i}{t_i}\right)\right\} \\
&= -\sum_{i=1}^k \inf_{\substack{s_i \geq 0, \\ t_i > 0}} \left[s_i \ln\left(\frac{s_i}{t_i}\right) - \bar{q}_i^f s_i + \bar{q}_i^g t_i\right] \\
&= \begin{cases} 0, & \text{if } 1 - \bar{q}_i^f + \ln \bar{q}_i^g \geq 0 \text{ and } \bar{q}_i^g > 0, \\ +\infty, & \text{otherwise.} \end{cases}
\end{aligned}$$

As $\bar{q}^f$ and $\bar{q}^g$ are feasible to (D_F) we have $\Phi_i^*\left(\bar{q}_i^f, -\bar{q}_i^g\right) = 0\ \forall i = 1, \dots, k$. Let us apply the Fenchel - Young inequality for $\Phi_i(f_i(\bar{x}), g_i(\bar{x}))$ and $\Phi_i^*\left(\bar{q}_i^f, -\bar{q}_i^g\right)$, when $i = 1, \dots, k$. We have

$$\Phi_i(f_i(\bar{x}), g_i(\bar{x})) + \Phi_i^*\left(\bar{q}_i^f, -\bar{q}_i^g\right) \geq \bar{q}_i^f f_i(\bar{x}) - \bar{q}_i^g g_i(\bar{x}),\ i = 1, \dots, k.$$

Summing these relations up one gets

$$\sum_{i=1}^k f_i(\bar{x}) \ln\left(\frac{f_i(\bar{x})}{g_i(\bar{x})}\right) \geq (\bar{q}^f)^T f(\bar{x}) - (\bar{q}^g)^T g(\bar{x}). \tag{4. 8}$$

On the other hand it is obvious that

$$\inf_{x \in X} \left[(\bar{q}^h)^T h(x) + (\bar{q}^f)^T f(x) - (\bar{q}^g)^T g(x)\right] \leq (\bar{q}^h)^T h(\bar{x}) + (\bar{q}^f)^T f(\bar{x}) - (\bar{q}^g)^T g(\bar{x}). \tag{4. 9}$$

Relations (4. 7) - (4. 9) yield

$$\begin{aligned}
0 &= \sum_{i=1}^k f_i(\bar{x}) \ln\left(\frac{f_i(\bar{x})}{g_i(\bar{x})}\right) - \inf_{x \in X} \left[(\bar{q}^h)^T h(x) + (\bar{q}^f)^T f(x) - (\bar{q}^g)^T g(x)\right] \\
&\geq (\bar{q}^f)^T f(\bar{x}) - (\bar{q}^g)^T g(\bar{x}) - \left[(\bar{q}^h)^T h(\bar{x}) + (\bar{q}^f)^T f(\bar{x}) - (\bar{q}^g)^T g(\bar{x})\right] \\
&= -(\bar{q}^h)^T h(\bar{x}) \geq 0.
\end{aligned}$$

The last inequality holds due to the fact that $\bar{x}$ is feasible to (P_f) and $\bar{q}^h$ to (D_f). Thus, of course all of these inequalities must be fulfilled as equalities. Therefore we immediately have (iii) and

$$\sum_{i=1}^{k}\left[\quad f_i(\bar{x})\ln\left(\frac{f_i(\bar{x})}{g_i(\bar{x})}\right) - \left((\bar{q}_i^f)^T f_i(\bar{x}) - (\bar{q}_i^g)^T g_i(\bar{x})\right)\right] + \left[(\bar{q}^h)^T h(\bar{x}) + (\bar{q}^f)^T f(\bar{x}) -(\bar{q}^g)^T g(\bar{x})\right] - \inf_{x\in X}\left[(\bar{q}^h)^T h(x) + (\bar{q}^f)^T f(x) - (\bar{q}^g)^T g(x)\right] = 0.$$

This yields the fulfillment of the above Fenchel - Young inequality for Φ_i and Φ_i^* as equality when $i = 1, \ldots, k$, that is nothing but (i). With (iii) then also (ii) is clear.

(b) The conclusion arises obviously following the proof above backwards. □

Remark 4.4 It is a natural question wonder what happens when the functions f_i, $i = 1, \ldots, k$, are taken not affine, but convex? In this situation the method we used to derive a dual problem to (P_f) would have been utilizable only if the additional constraints $f_i(x) \geq g_i(x)$ $\forall x \in X$ such that $h(x) \geqq 0$, $i = 1, \ldots, k$, were posed. We considered also treating the so - modified problem, but applications to it appear too seldom. For instance the three special cases we treated could not be trapped into such a form without particularizing them more.

4.2.2 Classical entropy optimization problems as particular cases

This subsection is dedicated to some interesting special cases of the problem (P_f). The first of them is the convex - constrained minimum cross - entropy problem, then follows a norm - constrained maximum entropy problem and as a third special case we present a so - called linearly constrained Burg entropy optimization problem.

For a definite choice of the functions f, g and h and taking $X = \mathbb{R}_+^n$ we obtain the entropy optimization problem with a Kullback - Leibler measure (cf. [59]) as objective function and convex constraint functions treated in [34]. From (D_f) we derive a dual problem to this particular one that turns out to be exactly the dual problem obtained via geometric programming in the original paper. When the convex constraint functions h_j, $j = 1, \ldots, m$, have some more particular properties, i.e. they are linear or quadratic, the dual problems turn into some more specific formulae. As a second special case we took a problem treated by Noll in [67]. After a suitable choice of particular formulae for f, g, h and X we obtain the maximum entropy optimization problem the author used in the applications described in [67], whose objective function is the Shannon entropy (cf. [83]) of a probability - like vector. The dual problem obtained using Lagrange duality there arises also when we derive a dual to this problem using (D_f). A third special case considered here is when the mentioned functions are chosen such that the objective function becomes the so - called Burg entropy (cf. [24]) minimization problem with linear constraints in [26]. The fact that the dual problems we obtain in the first two special cases are actually the ones determined in the original papers shows that the problem we treated is a generalization of the classical entropy optimization problems. For all the special cases we present the strong duality assertion and necessary and sufficient optimality conditions, derived from the general case.

4.2.2.1 The Kullback-Leibler entropy as objective function

The book [34] is a must for anyone interested in entropy optimization. Among many other interesting statements and applications, the authors consider the cross

- entropy minimization problem with convex constraint functions

$$(P_{KL}) \qquad \inf_{\substack{x\in\mathbb{R}^n_+,\sum\limits_{i=1}^n x_i=1,\\ l_j(A_jx)+b_j^Tx+c_j\le 0,\\ j=1,\dots,r}} \left[\sum_{i=1}^n x_i \ln\left(\frac{x_i}{q_i}\right)\right],$$

where A_j are $k_j \times n$ matrices with full row-rank, $b_j \in \mathbb{R}^n$, $j = 1,\dots,r$, $c = (c_1,\dots,c_r)^T \in \mathbb{R}^r$, $l_j : \mathbb{R}^{k_j} \to \mathbb{R}$, $j = 1,\dots,r$, are convex functions and there is also the probability distribution $q = (q_1,\dots,q_n)^T \in \operatorname{int}(\mathbb{R}^n_+)$, with $\sum_{i=1}^n q_i = 1$. We omit the additional assumptions of differentiability and co - finiteness for the functions l_j, $j = 1,\dots,r$, from the mentioned paper.

After dealing with the problem (P_{KL}) we particularize its constraints like in [34], first to become linear, then to obtain a quadratically - constrained cross - entropy optimization problem.

The problem (P_{KL}) is a special case of our problem (P_f) when the elements involved are taken as follows

$$\begin{cases} X = \mathbb{R}^n_+, k = n, m = r+2,\\ f_i(x) = x_i \ \forall x = (x_1,\dots,x_n)^T \in \mathbb{R}^n, i = 1,\dots,n,\\ g_i(x) = q_i \ \forall x = (x_1,\dots,x_n)^T \in \mathbb{R}^n, i = 1,\dots,n,\\ h_j(x) = l_j(A_jx) + b_j^Tx + c_j \ \forall x \in \mathbb{R}^n_+, j = 1,\dots,m-2,\\ h_{m-1}(x) = \sum\limits_{i=1}^n x_i - 1 \ \forall x = (x_1,\dots,x_n)^T \in \mathbb{R}^n_+,\\ h_m(x) = 1 - \sum\limits_{i=1}^n x_i \ \forall x = (x_1,\dots,x_n)^T \in \mathbb{R}^n_+. \end{cases}$$

We want to determine the dual problem to (P_{KL}) which is to be obtained from (D_f) by replacing the terms involved with the above-mentioned expressions. For $x \in \mathbb{R}^n$, respectively $x = (x_1,\dots,x_n)^T \in \mathbb{R}^n_+$ for h, one has

$$\begin{aligned} &(q^f)^T f(x) = (q^f)^T x, \quad (q^g)^T g(x) = (q^g)^T q,\\ &(\tilde{q}^h)^T h(x) = \sum_{j=1}^r q_j^h\big(l_j(A_jx) + b_j^Tx + c_j\big) + q_{m-1}^h\left(\sum_{i=1}^n x_i - 1\right) + q_m^h\left(1 - \sum_{i=1}^n x_i\right), \end{aligned}$$

where $\tilde{q}^h = \big(q_1^h,\dots,q_r^h,q_{m-1}^h,q_m^h\big)^T$. Denoting $w = q_{m-1}^h - q_m^h$ and $q^h = \big(q_1^h,\dots, q_k^h\big)$, the dual problem to (P_{KL}) is

$$\begin{aligned}(D_{KL}) \qquad \sup_{\substack{q_j^h\ge 0, j=1,\dots,r,\\ w\in\mathbb{R}, q_i^f\in\mathbb{R},\\ q_i^g\ge e^{q_i^f-1}, i=1,\dots,n}} \inf_{x\in\mathbb{R}^n_+} \Bigg[&(q^f)^Tx - (q^g)^Tq + \sum_{j=1}^r q_j^h l_j(A_jx)\\ &+\left(\sum_{j=1}^r q_j^h b_j\right)^T x + (q^h)^Tc + w\left(\sum_{i=1}^n x_i - 1\right)\Bigg].\end{aligned}$$

We can rearrange the terms and the dual problem becomes

$$\begin{aligned}(D_{KL}) \qquad \sup_{\substack{q_j^h\ge 0, j=1,\dots,r,\\ w\in\mathbb{R}, q_i^f\in\mathbb{R},\\ q_i^g\ge e^{q_i^f-1}, i=1,\dots,n}} \Bigg\{ &\inf_{x\in\mathbb{R}^n_+}\left[\sum_{i=1}^n x_i\left(q_i^f + \left(\sum_{j=1}^r q_j^h b_j\right)_i + w\right) + \sum_{j=1}^r q_j^h l_j(A_jx)\right]\\ &+(q^h)^Tc - w - (q^g)^Tq\Bigg\},\end{aligned}$$

where $\left(\sum_{j=1}^r q_j^h b_j\right)_i$, $i = 1, \ldots, n$, is the i-th entry of the vector $\sum_{j=1}^r q_j^h b_j$.

For $q^f \in \mathbb{R}^n$, $q^h \in \mathbb{R}^r_+$ and $w \in \mathbb{R}$ fixed, let us calculate the infimum over $x \in \mathbb{R}^n_+$ in the dual problem above. In order to do this we introduce the linear operators $\tilde{A}_j : \mathbb{R}^n \to \mathbb{R}^{k_j}$ defined by $\tilde{A}_j(x) = A_j x$, $j = 1, \ldots, m$. These infima become

$$\inf_{x \in \mathbb{R}^n_+} \left[\sum_{i=1}^n x_i \left(q_i^f + \left(\sum_{j=1}^r q_j^h b_j \right)_i + w \right) + \sum_{j=1}^r \left((q_j^h l_j) \circ \tilde{A}_j \right)(x) \right]. \tag{4. 10}$$

By Proposition 5.7 in [32] the expression in (4. 10) is equal to

$$\sup_{\gamma \in \mathbb{R}^n_+} \left\{ \inf_{x \in \mathbb{R}^n} \left[\sum_{i=1}^n x_i \left(q_i^f + \left(\sum_{j=1}^r q_j^h b_j \right)_i + w - \gamma_i \right) + \sum_{j=1}^r \left((q_j^h l_j) \circ \tilde{A}_j \right)(x) \right] \right\}, \tag{4. 11}$$

further equivalent to

$$\sup_{\gamma \in \mathbb{R}^n_+} \left\{ - \sup_{x \in \mathbb{R}^n} \left\{ \sum_{i=1}^n x_i \left(\gamma_i - q_i^f - \left(\sum_{j=1}^r q_j^h b_j \right)_i - w \right) - \sum_{j=1}^r \left((q_j^h l_j) \circ \tilde{A}_j \right)(x) \right\} \right\}.$$

The inner supremum may be written as a conjugate function, so the term above becomes

$$\sup_{\gamma \in \mathbb{R}^n_+} \left\{ - \left(\sum_{j=1}^r \left((q_j^h l_j) \circ \tilde{A}_j \right) \right)^* (\gamma - u) \right\},$$

where we have denoted by $u = (u_1, \ldots, u_n)^T$ the vector with the entries $u_i = q_i^f + \left(\sum_{j=1}^r q_j^h b_j \right)_i + w$, $i = 1, \ldots, n$. Theorem 16.4 in [72] yields

$$\left(\sum_{j=1}^r \left((q_j^h l_j) \circ \tilde{A}_j \right) \right)^* (\gamma - u) = \inf_{\substack{a_j \in \mathbb{R}^n, j=1,\ldots,r, \\ \sum_{j=1}^r a_j = \gamma - u}} \left[\sum_{j=1}^r \left((q_j^h l_j) \circ \tilde{A}_j \right)^* (a_j) \right]. \tag{4. 12}$$

The relation (4. 10) is now equivalent to

$$\sup_{\gamma \in \mathbb{R}^n_+} \left\{ - \inf_{\substack{a_j \in \mathbb{R}^n, j=1,\ldots,r, \\ \sum_{j=1}^r a_j = \gamma - u}} \left[\sum_{j=1}^r \left((q_j^h l_j) \circ \tilde{A}_j \right)^* (a_j) \right] \right\}$$

and furthermore to

$$\sup_{\substack{a_j \in \mathbb{R}^n, j=1,\ldots,r, \\ \gamma \in \mathbb{R}^n_+, \sum_{j=1}^r a_j = \gamma - u}} \left\{ - \sum_{j=1}^r \left((q_j^h l_j) \circ \tilde{A}_j \right)^* (a_j) \right\}.$$

As for any $j = 1, \ldots, r$, the image set of the operator $\tilde{A}_j$ is included into $\mathbb{R}^{k_j}$ that is the domain of the function $q_j^h l_j$ defined by $\left(q_j^h l_j\right)(x) = q_j^h l_j(x)$, we may apply Theorem 16.3 in [72] and the last expression becomes equivalent to

$$\sup_{\substack{a_j \in \mathbb{R}^n, j=1,\ldots,r, \\ \gamma \in \mathbb{R}^n_+, \sum_{j=1}^r a_j = \gamma - u}} \left\{ - \sum_{j=1}^r \inf_{\substack{\lambda_j \in \mathbb{R}^{k_j}, \\ \tilde{A}_j^* \lambda_j = a_j}} \left[\left(q_j^h l_j \right)^* (\lambda_j) \right] \right\}. \tag{4. 13}$$

Turning the inner infima into suprema and drawing all the variables under the leading supremum (4. 13) is equivalent, after applying the definition of the adjoint of a linear operator, to

$$\sup_{\substack{\gamma\in\mathbb{R}^n_+, a_j\in\mathbb{R}^n, \lambda_j\in\mathbb{R}^{k_j}, j=1,\dots,r,\\ \sum\limits_{j=1}^r a_j=\gamma-u, A_j^T\lambda_j=a_j}} \left\{ -\sum_{j=1}^r \left(q_j^h l_j\right)^*(\lambda_j)\right\}.$$

One may remark that the variables γ and a_j, $j=1,\dots,r$, are no longer necessary, so the expression is further simplifiable to

$$\sup_{\substack{\lambda_j\in\mathbb{R}^{k_j}, j=1,\dots,r,\\ \sum\limits_{j=1}^r A_j^T\lambda_j+u\in\mathbb{R}^n_+}} \left\{ -\sum_{j=1}^r \left(q_j^h l_j\right)^*(\lambda_j)\right\}.$$

Let us resume the calculations concerning the dual problem using the partial results obtained above. The dual problem to (P_{KL}) becomes

$$(D_{KL}) \qquad \sup_{\substack{q_i^f\in\mathbb{R}, q_i^g\geq e^{q_i^f-1}, i=1,\dots,n,\\ w\in\mathbb{R}, q_j^h\geq 0, \lambda_j\in\mathbb{R}^{k_j}, j=1,\dots,r,\\ q_i^f+\left(\sum\limits_{j=1}^r q_j^h b_j\right)_i+w+\left(\sum\limits_{j=1}^r A_j^T\lambda_j\right)_i\geq 0,\\ i=1,\dots,n}} \left\{(q^h)^T c - w - (q^g)^T q - \sum_{j=1}^r \left(q_j^h l_j\right)^*(\lambda_j)\right\},$$

rewritable as

$$(D_{KL}) \qquad \sup_{\substack{q^f\in\mathbb{R}^n, w\in\mathbb{R}, q^h\in\mathbb{R}^r_+, \lambda_j\in\mathbb{R}^{k_j}, j=1,\dots,r,\\ q_i^f+\left(\sum\limits_{j=1}^r q_j^h b_j\right)_i+w+\left(\sum\limits_{j=1}^r A_j^T\lambda_j\right)_i\geq 0,\\ i=1,\dots,n}} \left\{(q^h)^T c - w - \sum_{j=1}^r \left(q_j^h l_j\right)^*(\lambda_j) + \sum_{i=1}^n \sup_{q_i^g\geq e^{q_i^f-1}} -q_i^g q_i\right\}.$$

It is obvious that $\sup\left\{-q_i^g q_i : q_i^g \geq e^{q_i^f-1}\right\} = -q_i e^{q_i^f-1}$, $i=1,\dots,n$, so the dual problem turns into

$$(D_{KL}) \qquad \sup_{\substack{w\in\mathbb{R}, q_j^h\geq 0, \lambda_j\in\mathbb{R}^{k_j}, j=1,\dots,r, q_i^f\in\mathbb{R},\\ q_i^f+\left(\sum\limits_{j=1}^r q_j^h b_j\right)_i+w+\left(\sum\limits_{j=1}^r A_j^T\lambda_j\right)_i\geq 0,\\ i=1,\dots,n}} \left\{(q^h)^T c - w - \sum_{j=1}^r \left(q_j^h l_j\right)^*(\lambda_j) - \sum_{i=1}^n q_i e^{q_i^f-1}\right\}$$

The suprema after q_i^f, $i=1,\dots,n$, are easily computable since the constraints are linear inequalities and the objective functions are monotonically decreasing, i.e.

$$\sup\left\{-e^{q_i^f-1} : q_i^f + \left(\sum_{j=1}^r \left(q_j^h b_j + A_j^T\lambda_j\right)\right)_i + w \geq 0\right\} = -e^{-w-\left(\sum\limits_{j=1}^r \left(q_j^h b_j+A_j^T\lambda_j\right)\right)_i-1}.$$

Back to the dual problem, it becomes

$$(D_{KL}) \sup_{\substack{w\in\mathbb{R}, q_j^h\geq 0,\\ \lambda_j\in\mathbb{R}^{k_j},\\ j=1,\dots,r}} \left\{ (q^h)^T c - \sum_{j=1}^r \left(q_j^h l_j\right)^*(\lambda_j) - w - \sum_{i=1}^n q_i e^{-w-\left(\sum\limits_{j=1}^r \left(q_j^h b_j + A_j^T\lambda_j\right)\right)_i -1} \right\},$$

which is already a Fenchel - Lagrange type dual problem.

The next variable we want to renounce is w. In order to do this let us consider the function $\eta : \mathbb{R} \to \mathbb{R}$, $\eta(w) = -w - Be^{-w-1}$, $B > 0$. Its derivative is $\eta'(w) = Be^{-w-1} - 1$, $w \in \mathbb{R}$, a monotonically decreasing function that takes the value zero at $w = \ln B - 1$. So η attains its maximal value at $w = \ln B - 1$, that is $\eta(\ln B - 1) = -\ln B$. Applying these considerations to our dual problem for $B = \sum_{i=1}^n q_i e^{-\left(\sum_{j=1}^r \left(q_j^h b_j + A_j^T\lambda_j\right)\right)_i}$ we get rid of the variable $w \in \mathbb{R}$ and the simplified version of the dual problem is

$$(D_{KL}) \sup_{\substack{q_j^h\geq 0, \lambda_j\in\mathbb{R}^{k_j},\\ j=1,\dots,r}} \left\{ (q^h)^T c - \sum_{j=1}^r \left(q_j^h l_j\right)^*(\lambda_j) - \ln\left(\sum_{i=1}^n q_i e^{-\left(\sum\limits_{j=1}^r \left(q_j^h b_j + A_j^T\lambda_j\right)\right)_i} \right) \right\},$$

that turns out, after redenoting the variables, to be the dual problem obtained in [34] via geometric duality. This is not surprising taking into consideration the results given in the previous chapter.

As weak duality between (P_{KL}) and (D_{KL}) is certain, we focus on the strong duality. In order to achieve it we particularize the constraint qualification (CQ_f) as follows

$$(CQ_{KL}) \qquad \exists x' \in \mathbb{R}^n_+ : \begin{cases} x' = (x'_1, \dots, x'_n), & \\ \sum\limits_{i=1}^n x'_i = 1, & \\ x'_i > 0, & i = 1, \dots, n, \\ l_j(A_j x') + b_j^T x' + c_j \leq 0, & \text{if } j \in L_{KL}, \\ l_j(A_j x') + b_j^T x' + c_j < 0, & \text{if } j \in \{1, \dots, r\} \backslash L_{KL}, \end{cases}$$

where the set L_{KL} is defined analogously to L_{gc}, i.e.

$$L_{KL} = \big\{ j \in \{1, \dots, r\} : l_j \text{ is an affine function} \big\}.$$

We are ready now to enunciate the strong duality assertion.

Theorem 4.19 *(strong duality) If the constraint qualification (CQ_{KL}) is fulfilled, then there is strong duality between problems (P_{KL}) and (D_{KL}), i.e. (D_{KL}) has an optimal solution and $v(P_{KL}) = v(D_{KL})$.*

Proof. It is known (see [34]) that the objective function of (P_{KL}) takes only non - negative values, thus $v(P_{KL})$ is finite. From the general case we have strong duality between (P_{KL}) and the first formulation of the dual problem in this section. The equality $v(P_{KL}) = v(D_{KL})$ has been preserved after all the steps we performed in order to simplify the formulation of the dual, but there could be a problem regarding the existence of the solution to the dual problem. Fortunately, the results applied to obtain (4. 11) - (4. 13) mention also the existence of a solution to the resulting problems, respectively, so this property is preserved up to the final formulation of the dual problem. □

Furthermore we give also some necessary and sufficient optimality conditions in the following statement. They were obtained in the same way as in Theorem 4.18, so we have decided to omit the proof.

Theorem 4.20 *(optimality conditions)*
(a) Let the constraint qualification (CQ_{KL}) be fulfilled and assume that the primal problem (P_{KL}) has an optimal solution $\bar{x}$. Then the dual problem (D_{KL}) has an optimal solution $\left(\bar{q}^h, \bar{\lambda}_1, \dots, \bar{\lambda}_r\right)$ and the following optimality conditions hold

(i) $\sum\limits_{i=1}^{n} \bar{x}_i \ln\left(\frac{\bar{x}_i}{q_i}\right) + \ln\left(\sum\limits_{i=1}^{n} q_i e^{-\left(\sum\limits_{j=1}^{r}\left(\bar{q}_j^h b_j + A_j^T \bar{\lambda}_j\right)\right)_i}\right) = -\left(\sum\limits_{j=1}^{r}\left(\bar{q}_j^h b_j + A_j^T \bar{\lambda}_j\right)\right)^T \bar{x}$,

(ii) $\left(\bar{q}_j^h l_j\right)^*(\bar{\lambda}_j) + \left(\bar{q}_j^h l_j\right)\left(A_j \bar{x}\right) = \bar{\lambda}_j^T A_j \bar{x}$, $j = 1, \dots, r$,

(iii) $\bar{q}_j^h \Big(l_j(A_j \bar{x}) + b_j^T \bar{x} + c_j\Big) = 0$, $j = 1, \dots, r$.

(b) If $\bar{x}$ is a feasible point to (P_{KL}) and $\left(\bar{q}^h, \bar{\lambda}_1, \dots, \bar{\lambda}_r\right)$ is feasible to (D_{KL}) fulfilling the optimality conditions $(i) - (iii)$, then there is strong duality between (P_{KL}) and (D_{KL}). Moreover, $\bar{x}$ is an optimal solution to the primal problem and $\left(\bar{q}^h, \bar{\lambda}_1, \dots, \bar{\lambda}_r\right)$ an optimal solution to the dual.

The problem (P_{KL}) may be particularized even more, in order to fit a wide range of applications. We present further two special cases obtained from (P_{KL}) by assigning some particular values to the constraint functions, as indicated also in [34].

Special case 1: Kullback-Leibler entropy objective function and linear constraints

Taking $l_j(y_j) = 0$, $y_j \in \mathbb{R}^{k_j}$, $j = 1, \dots, r$, we have for the conjugates involved in the dual problem

$$\left(q_j^h l_j\right)^*(\lambda_j) = \sup_{y_j \in \mathbb{R}^{k_j}} \left\{\lambda_j^T y_j - 0\right\} = \begin{cases} 0, & \text{if } \lambda_j = 0, \\ +\infty, & \text{otherwise,} \end{cases} \quad j = 1, \dots, r.$$

Performing the necessary substitutions, we get the following pair of dual problems

$$(P_L) \qquad \inf_{\substack{x \in \mathbb{R}^n_+, \sum\limits_{i=1}^{n} x_i = 1, \\ b_j^T x + c_j \leq 0, j=1,\dots,r}} \left[\sum_{i=1}^{n} x_i \ln\left(\frac{x_i}{q_i}\right)\right],$$

and

$$(D_L) \qquad \sup_{\substack{q_j^h \geq 0, \\ j=1,\dots,r}} \left\{(q^h)^T c - \ln\left(\sum_{i=1}^{n} q_i e^{-\left(\sum\limits_{j=1}^{r} q_j^h b_j\right)_i}\right)\right\}.$$

In [34] there is treated a similar problem to (P_L), but instead of inequality constraints Fang, Rajasekera and Tsao use equality constraints. The dual problem they obtain is also similar to (D_L), the only difference consisting of the feasible set, $\mathbb{R}^r_+$ to (D_L), respectively $\mathbb{R}^r$ in [34]. Let us mention further that an interesting application of the optimization problem with Kullback - Leibler entropy objective function and linear constraints can be found in [39]. In order to achieve strong duality the sufficient constraint qualification is

$$(CQ_L) \qquad \exists x' = (x'_1, \dots, x'_n) \in \mathbb{R}^n_+ : \begin{cases} \sum\limits_{i=1}^{n} x'_i = 1, & \\ x'_i > 0, & i = 1, \dots, n, \\ b_j^T x' + c_j \leq 0, & \text{for } j = 1, \dots, r. \end{cases}$$

Theorem 4.21 *(strong duality) If the constraint qualification* (CQ_L) *is valid, then there is strong duality between problems* (P_L) *and* (D_L)*, i.e.* (D_L) *has an optimal solution and* $v(P_L) = v(D_L)$*.*

As this assertion is a special case of Theorem 4.19 we omit its proof. The optimality conditions arise also easily from Theorem 4.20.

Theorem 4.22 *(optimality conditions)*
(a) *Assume that the primal problem* (P_L) *has an optimal solution* $\bar{x}$ *and that the constraint qualification* (CQ_L) *is fulfilled. Then the dual problem* (D_L) *has an optimal solution* $\bar{q}^h = \left(q_1^h, \ldots, q_r^h\right)^T$ *and the following optimality conditions hold*

(i) $\sum\limits_{i=1}^{n} \bar{x}_i \ln\left(\frac{\bar{x}_i}{q_i}\right) + \ln\left(\sum\limits_{i=1}^{n} q_i e^{-\left(\sum\limits_{j=1}^{r}\left(\bar{q}_j^h b_j\right)\right)_i}\right) = -\left(\sum\limits_{j=1}^{r} \bar{q}_j^h b_j\right)^T \bar{x},$

(ii) $\bar{q}_j^h\left(b_j^T \bar{x} + c_j\right) = 0,\ j = 1, \ldots, r.$

(b) *If* $\bar{x}$ *is a feasible point to* (P_L) *and* $\bar{q}^h$ *a feasible point to* (D_L) *fulfilling the optimality conditions* (i) *and* (ii)*, then there is strong duality between* (P_L) *and* (D_L)*. Moreover,* $\bar{x}$ *is an optimal solution to the primal problem and* $\bar{q}^h$ *one to the dual.*

Special case 2: Kullback - Leibler entropy objective function and quadratic constraints

Take now $l_j(y_j) = \frac{1}{2} y_j^T y_j$, $y_j \in \mathbb{R}^{k_j}$, $j = 1, \ldots, r$. We have (cf. [72])

$$\left(q_j^h l_j\right)^*(\lambda_j) = \begin{cases} \frac{\|\lambda_j\|^2}{2q_j^h}, & \text{if } q_j^h \neq 0, \\ 0, & \text{otherwise.} \end{cases}$$

The pair of dual problems is in this case consists of (cf. [12])

$$(P_Q) \qquad \inf_{\substack{x \in \mathbb{R}^n_+,\ \sum\limits_{i=1}^{n} x_i = 1, \\ \frac{1}{2} x^T A_j^T A_j x + b_j^T x + c_j \leq 0, \\ j=1,\ldots,r}} \left[\sum_{i=1}^{n} x_i \ln\left(\frac{x_i}{q_i}\right)\right]$$

and

$$(D_Q) \qquad \sup_{\substack{q_j^h > 0, \lambda_j \in \mathbb{R}^{k_j}, \\ j=1,\ldots,r}} \left\{ (q^h)^T c - \ln\left(\sum_{i=1}^{n} q_i e^{-\left(\sum\limits_{j=1}^{r}\left(q_j^h b_j + A_j^T \lambda_j\right)\right)_i}\right) - \frac{1}{2}\sum_{j=1}^{r} \frac{\|\lambda_j\|^2}{q_j^h} \right\},$$

like in [34]. The following constraint qualification is sufficient in order to assure strong duality

$$(CQ_Q) \qquad \exists x' \in \mathbb{R}^n_+ : \begin{cases} x' = (x'_1, \ldots, x'_n), & \\ \sum\limits_{i=1}^{n} x'_i = 1, & \\ x'_i > 0, & i = 1, \ldots, n, \\ \frac{1}{2} x'^T A_j^T A_j x' + b_j^T x' + c_j < 0, & \text{for } j = 1, \ldots, r. \end{cases}$$

Theorem 4.23 *(strong duality) If the constraint qualification* (CQ_Q) *is fulfilled, then there is strong duality between problems* (P_Q) *and* (D_Q)*, i.e.* (D_Q) *has an optimal solution and* $v(P_Q) = v(D_Q)$*.*

Furthermore, we give without proof also some necessary and sufficient optimality conditions in the following statement.

Theorem 4.24 *(optimality conditions)*
(a) *Let the constraint qualification (CQ_Q) be fulfilled and assume that the primal problem (P_Q) has an optimal solution $\bar{x}$. Then the dual problem (D_Q) has an optimal solution $\left(\bar{q}^h, \bar{\lambda}_1, \ldots, \bar{\lambda}_r\right)$ and the following optimality conditions hold*

$$\text{(i)}\ \sum_{i=1}^{n} \bar{x}_i \ln\left(\frac{\bar{x}_i}{q_i}\right) + \ln\left(\sum_{i=1}^{n} q_i e^{-\left(\sum\limits_{j=1}^{r}\left(\bar{q}_j^h b_j + A_j^T \bar{\lambda}_j\right)\right)_i}\right) = -\left(\sum_{j=1}^{r}\left(\bar{q}_j^h b_j + A_j^T \bar{\lambda}_j\right)\right)^T \bar{x},$$

$$\text{(ii)}\ \frac{1}{2}\bar{q}_j^h \bar{x}^T A_j^T A_j \bar{x} + \frac{\|\bar{\lambda}_j\|^2}{2\bar{q}_j^h} = \bar{\lambda}_j^T A_j \bar{x},\ j = 1, \ldots, r,$$

$$\text{(iii)}\ \bar{q}_j^h \left(\bar{x}^T A_j^T A_j \bar{x} + b_j^T \bar{x} + c_j\right) = 0,\ j = 1, \ldots, r.$$

(b) *If $\bar{x}$ is a feasible point to (P_Q) and $\left(\bar{q}^h, \bar{\lambda}_1, \ldots, \bar{\lambda}_r\right)$ is feasible to (D_Q) fulfilling the optimality conditions $(i) - (iii)$, then there is strong duality between (P_Q) and (D_Q). Moreover, $\bar{x}$ is an optimal solution to the primal problem, while $\left(\bar{q}^h, \bar{\lambda}_1, \ldots, \bar{\lambda}_r\right)$ turns out to be an optimal solution to the dual.*

4.2.2.2 The Shannon entropy as objective function

Noll presents in [67] an interesting application of the maximum entropy optimization in image reconstruction considering the following problem

$$(P_S) \qquad \inf_{\substack{x_{ij} \geq 0, i=1,\ldots,n, j=1,\ldots,m,\\ \sum\limits_{i=1}^{n}\sum\limits_{j=1}^{m} x_{ij} = T,\ \|Ax-y\| \leq \varepsilon}} \left[\sum_{i=1}^{n}\sum_{j=1}^{m} x_{ij} \ln x_{ij}\right],$$

where $x \in \mathbb{R}^{n\times m}$ with the entries x_{ij}, $i = 1, \ldots, n$, $j = 1, \ldots, m$, $A \in \mathbb{R}^{n\times n}$, $y \in \mathbb{R}^{n\times m}$, with the entries y_{ij}, $i = 1, \ldots, n$, $j = 1, \ldots, m$, $\varepsilon > 0$ and $T = \sum_{i=1}^{n}\sum_{j=1}^{m} y_{ij} > 0$. The norm is the Euclidean one. It is easy to notice that the objective function in this problem is the well - known Shannon entropy measure with variables x_{ij}, $i = 1, \ldots, n$, $j = 1, \ldots, m$, so (P_S) is actually equivalent to the following maximum entropy optimization problem

$$(P_S') \qquad - \sup_{\substack{\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{m} x_{ij} = T,\ \|Ax-y\| \leq \varepsilon,\\ x_{ij} \geq 0, i=1,\ldots,n, j=1,\ldots,m}} \left\{-\sum_{i=1}^{n}\sum_{j=1}^{m} x_{ij} \ln x_{ij}\right\}.$$

However, (P_S) is viewable as a special case of problem (P_f) by assigning to the sets and functions involved there the following values

$$\begin{cases} X = \mathbb{R}_+^{n\times m}, x = (x_{ij})_{\substack{i=1,\ldots,n,\\ j=1,\ldots,m}} \in \mathbb{R}_+^{n\times m}, \\ f_{ij}(x) = x_{ij}\ \forall x \in \mathbb{R}^{n\times m}, i = 1, \ldots, n, j = 1, \ldots, m, \\ g_{ij}(x) = 1\ \forall x \in \mathbb{R}^{n\times m}, i = 1, \ldots, n, j = 1, \ldots, m, \\ h_1(x) = \sum\limits_{i=1}^{n}\sum\limits_{j=1}^{m} x_{ij} - T\ \forall x \in \mathbb{R}_+^{n\times m}, \\ h_2(x) = T - \sum\limits_{i=1}^{n}\sum\limits_{j=1}^{m} x_{ij}\ \forall x \in \mathbb{R}_+^{n\times m}, \\ h_3(x) = \|Ax - y\| - \varepsilon\ \forall x \in \mathbb{R}_+^{n\times m}. \end{cases}$$

Remark 4.5 Some may object that (P_S) is not a pure special case of (P_f) because the variable x is not an n - dimensional vector as in (P_f), but a $n \times m$ matrix. As matrices can be viewed also as vectors, in this case the variable becomes an $n \times m$ - dimensional vector, so we may apply the results obtained for (P_f) also to (P_S).

To obtain the dual problem to (P_S) from (D_f) we calculate the following expressions, where the Lagrange multipliers are now $q^f = (q^f_{ij})_{\substack{i=1,\dots,n,\\ j=1,\dots,m}} \in \mathbb{R}^{n\times m}$, $q^g = (q^g_{ij})_{\substack{i=1,\dots,n,\\ j=1,\dots,m}} \in \mathbb{R}^{n\times m}_+$ and $q^h = \left(q^h_1, q^h_2, q^h_3\right) \in \mathbb{R}^3_+$,

$$\sum_{i=1}^n\sum_{j=1}^m q^f_{ij} f_{ij}(x) = \sum_{i=1}^n\sum_{j=1}^m q^f_{ij} x_{ij}, \quad \sum_{i=1}^n\sum_{j=1}^m q^g_{ij} g_{ij}(x) = \sum_{i=1}^n\sum_{j=1}^m q^g_{ij},$$
$$\sum_{j=1}^3 q^h_j h_j(x) = \left(q^h_1 - q^h_2\right)\left(\sum_{i=1}^n\sum_{j=1}^m x_{ij} - T\right) + q^h_3\left(\|Ax-y\| - \varepsilon\right).$$

The multipliers q^h_1 and q^h_2 appear only together, so we may replace both of them, i.e. their difference, with a new variable $w = q^h_1 - q^h_2 \in \mathbb{R}$. The dual problem to (P_S) becomes

$$(D_S) \qquad \sup_{\substack{q^f\in\mathbb{R}^{n\times m},\\ q^h_3\geq 0, w\in\mathbb{R},\\ q^g_{ij}\geq e^{q^f_{ij}-1},\\ i=1,\dots,n, j=1,\dots,m}} \inf_{x=(x_{ij})_{ij}\in\mathbb{R}^{n\times m}_+} \left[\sum_{i=1}^n\sum_{j=1}^m q^f_{ij}x_{ij} - \sum_{i=1}^n\sum_{j=1}^m q^g_{ij} + w\left(\sum_{i=1}^n\sum_{j=1}^m x_{ij} - T\right) + q^h_3\left(\|Ax-y\|-\varepsilon\right)\right],$$

rewritable as

$$(D_S) \qquad \sup_{\substack{q^f\in\mathbb{R}^{n\times m}, q^h_3\geq 0,\\ w\in\mathbb{R}, q^g_{ij}\geq e^{q^f_{ij}-1},\\ i=1,\dots,n, j=1,\dots,m}} \left\{ -wT - \sum_{i=1}^n\sum_{j=1}^m q^g_{ij} - q^h_3\varepsilon + \inf_{x\in\mathbb{R}^{n\times m}_+}\left[\sum_{i=1}^n\sum_{j=1}^m x_{ij}\left(q^f_{ij}+w\right) + q^h_3\|Ax-y\|\right]\right\}.$$

We transform now the infimum concerning $x \in \mathbb{R}^{n\times m}_+$ as in the previous special case into a conjugate function, turning the dual into a Fenchel - Lagrange dual problem. By Theorem 16.3 in [72] it turns out to be equal to

$$-\sup_{\substack{q^f_{ij}+w+\left(A^TB\right)_{ij}\geq 0,\\ i=1,\dots,n, j=1,\dots,m, B\in\mathbb{R}^{n\times m}}} \left\{\left(q^h_3\|\cdot - y\|\right)^*(B)\right\}.$$

For the conjugate of the norm function we have (cf. [80])

$$\left(q^h_3\|\cdot - y\|\right)^*(B) = \begin{cases} -y^TB, & \text{if } \|B\|\leq q^h_3,\\ -\infty, & \text{otherwise.}\end{cases}$$

As negative infinite values are not relevant to our problem since there is a leading supremum to be determined, the dual problem becomes

$$(D_S) \qquad \sup_{\substack{q^f\in\mathbb{R}^{n\times m}, q^h_3\geq 0, w\in\mathbb{R}, B\in\mathbb{R}^{n\times m},\\ \|B\|\leq q^h_3, q^f_{ij}+w+\left(A^TB\right)_{ij}\geq 0,\\ q^g_{ij}\geq e^{q^f_{ij}-1}, i=1,\dots,n, j=1,\dots,m}} \left\{ -wT - \sum_{i=1}^n\sum_{j=1}^m q^g_{ij} - q^h_3\varepsilon - y^TB\right\},$$

equivalent to

$$(D_S) \qquad \sup_{\substack{q^f\in\mathbb{R}^{n\times m},\\ w\in\mathbb{R}, B\in\mathbb{R}^{n\times m},\\ q^f_{ij}+w+\left(A^TB\right)_{ij}\geq 0,\\ i=1,\dots,n, j=1,\dots,m}} \left\{ -wT - y^TB + \sum_{i=1}^n\sum_{j=1}^m \sup_{q^g_{ij}\geq e^{q^f_{ij}-1}} -q^g_{ij} + \varepsilon \sup_{q^h_3\geq \|B\|} -q^h_3 \right\}.$$

The suprema from inside are trivially determinable, so we obtain for the dual problem the following expression

$$(D_S) \qquad \sup_{\substack{q^f\in\mathbb{R}^{n\times m}, w\in\mathbb{R}, B\in\mathbb{R}^{n\times m},\\ q^f_{ij}+w+\left(A^TB\right)_{ij}\geq 0, i=1,\dots,n, j=1,\dots,m}} \left\{ -wT - y^TB - \sum_{i=1}^n\sum_{j=1}^m e^{q^f_{ij}-1} - \varepsilon\|B\| \right\},$$

further equivalent to

$$(D_S) \qquad \sup_{\substack{w\in\mathbb{R},\\ B\in\mathbb{R}^{n\times m}}} \left\{ -wT - y^TB - \varepsilon\|B\| + \sum_{i=1}^n\sum_{j=1}^m \sup_{q^f_{ij}+w+\left(A^TB\right)_{ij}\geq 0} -e^{q^f_{ij}-1} \right\}.$$

Regarding the inner suprema we have for all $i=1,\dots,n$, and $j=1,\dots,m$,

$$\sup\left\{ -e^{q^f_{ij}-1} : q^f_{ij} + w + \left(A^TB\right)_{ij} \geq 0 \right\} = -e^{-w-(A^TB)_{ij}-1},$$

so the dual problem is simplifiable even to

$$(D_S) \qquad \sup_{\substack{w\in\mathbb{R},\\ B\in\mathbb{R}^{n\times m}}} \left\{ -wT - y^TB - \varepsilon\|B\| - e^{-w-1}\sum_{i=1}^n\sum_{j=1}^m e^{-(A^TB)_{ij}} \right\},$$

that is exactly the dual problem obtained via Lagrange duality in [67].

Moreover, one may notice that also the variable $w\in\mathbb{R}$ could be eradicated. Using the results regarding the maximal value of the function η introduced before, we have

$$\sup_{w\in\mathbb{R}} \left\{ -wT - e^{-w-1}\sum_{i=1}^n\sum_{j=1}^m e^{-(A^TB)_{ij}} \right\} = T\left(\ln T - \ln\left(\sum_{i=1}^n\sum_{j=1}^m e^{-(A^TB)_{ij}} \right)\right).$$

The last version of the dual problem we reach is

$$(D_S) \qquad \sup_{B\in\mathbb{R}^{n\times m}} \left\{ T\left(\ln T - \ln\left(\sum_{i=1}^n\sum_{j=1}^m e^{-(A^TB)_{ij}} \right)\right) - y^TB - \varepsilon\|B\| \right\}.$$

As weak duality is certain, we skip stating it explicitly and focus on the strong duality. In order to achieve it the following constraint qualification is sufficient

$$(CQ_S) \quad \exists x' = (x'_{ij})_{\substack{i=1,\dots,n,\\ j=1,\dots,m}} \in \mathbb{R}^{n\times m} : \begin{cases} x'_{ij} > 0 \ \forall i=1,\dots,n \ \forall j=1,\dots,m,\\ \sum_{i=1}^n\sum_{j=1}^m x'_{ij} = T,\\ \|Ax'-y\| < \varepsilon. \end{cases}$$

The strong duality assertion comes immediately and the necessary and sufficient optimality conditions follow thereafter. Even if the original paper does not contain such statements, we omit the both proofs because they arise simply from the former proofs in the present thesis.

Theorem 4.25 *(strong duality) Assume the constraint qualification (CQ_S) fulfilled. Then strong duality between (P_S) and (D_S) is valid, i.e. (D_S) has an optimal solution and $v(P_S) = v(D_S)$.*

Theorem 4.26 *(optimality conditions)*
(a) Assume the constraint qualification (CQ_S) fulfilled and let $\bar{x}$ be an optimal solution to (P_S). Then the dual problem (D_S) has an optimal solution $\bar{B}$ and the following optimality conditions hold

$$(i)\ \sum_{i=1}^{n}\sum_{j=1}^{m}\bar{x}_{ij}\ln\bar{x}_{ij}+T\left(\ln\left(\sum_{i=1}^{n}\sum_{j=1}^{m}e^{-(A^T\bar{B})_{ij}}\right)-\ln T\right)=(A^T\bar{B})^T\bar{x},$$

$$(ii)\ \|A\bar{x}-y\|=\varepsilon,$$

$$(iii)\ \bar{B}^T(A\bar{x}-y)=\|\bar{B}\|\|A\bar{x}-y\|.$$

(b) If $\bar{x}$ is a feasible point to (P_S) and $\bar{B}$ one to (D_S) satisfying the optimality conditions $(i)-(iii)$, then they are actually optimal solutions to the corresponding problems that enjoy moreover strong duality.

4.2.2.3 The Burg entropy as objective function

A third widely - used entropy measure is the one introduced by Burg in [24]. Although there are some others in the literature, we confine ourselves to the most used three, as they have proven to be the most important from the viewpoint of applications. The Burg entropy problem we have chosen as the third application comes from Censor and Lent's paper [26] having Burg entropy as objective function and linear equality constraints,

$$(P_B)\qquad \sup_{\substack{x=(x_1,\dots,x_n)^T,\\ x_i>0,i=1,\dots,n,\\ Ax=b}}\left\{\sum_{i=1}^{n}\ln x_i\right\},$$

where $A\in\mathbb{R}^{m\times n}$ and $b\in\mathbb{R}^m$.

Some other problems with Burg entropy objective function and linear constraints that slightly differ from the one we treat are available, for example, in [25] and [56]. None of these papers contain explicitly a dual to the Burg entropy problem they consider.

The problem (P_B) may be equivalently rewritten as a minimization problem as follows

$$(P_B')\qquad -\inf_{\substack{x=(x_1,\dots,x_n)^T,\\ x_i>0,i=1,\dots,n,\\ Ax=b}}\left[-\sum_{i=1}^{n}\ln x_i\right].$$

Denoting (P_B'') the problem (P_B') after eluding the leading minus, it may be trapped as a special case of (P_f) by taking

$$\begin{cases} X=\operatorname{int}(\mathbb{R}^n_+), k=n,\\ f_i(x)=1\ \forall x\in\mathbb{R}^n, i=1,\dots,n,\\ g_i(x)=x_i\ \forall x\in\mathbb{R}^n, i=1,\dots,n,\\ h_1(x)=Ax-b\ \forall x\in\operatorname{int}(\mathbb{R}^n_+),\\ h_2(x)=b-Ax\ \forall x\in\operatorname{int}(\mathbb{R}^n_+).\end{cases}$$

To calculate the dual problem to (P_B'') let us replace the values above in (D_f). We get

$$(D_B)\qquad \sup_{\substack{q_1^h,q_2^h\in\mathbb{R}^m_+,q_i^f\in\mathbb{R},\\ q_i^g\geq e^{q_i^f-1},i=1,\dots,n}}\ \inf_{x>0}\left[\left(q_1^h-q_2^h\right)^T\left(Ax-b\right)+\sum_{i=1}^{n}\left(q_i^f-q_i^g x_i\right)\right].$$

Again, we introduce a new variable $w = q_1^h - q_2^h \in \mathbb{R}^m$ to replace the difference of the two non - negative ones that appear only together. After rearranging the terms the dual becomes

$$(D_B) \qquad \sup_{\substack{w\in\mathbb{R}^m, q_i^f\in\mathbb{R},\\ q_i^g\geq e^{q_i^f-1}, i=1,\dots,n}} \left\{ \sum_{i=1}^n q_i^f - w^T b + \sum_{i=1}^n \inf_{x_i>0} \left[\left(\left(w^T A\right)_i - q_i^g \right) x_i \right] \right\}.$$

For the infima inside we have for $i = 1, \dots, n$,

$$\inf_{x_i>0} \left[\left(\left(w^T A\right)_i - q_i^g \right) x_i \right] = \begin{cases} 0, & \text{if } \left(w^T A\right)_i - q_i^g \geq 0, \\ -\infty, & \text{otherwise.} \end{cases}$$

Let us drag these results along the dual problem, that is now

$$(D_B) \qquad \sup_{\substack{w\in\mathbb{R}^m, q_i^f\in\mathbb{R}, q_i^g\geq e^{q_i^f-1},\\ \left(w^T A\right)_i - q_i^g\geq 0, i=1,\dots,n}} \left\{ \sum_{i=1}^n q_i^f - w^T b \right\},$$

rewritable as

$$(D_B) \qquad \sup_{\substack{w\in\mathbb{R}^m, q_i^g>0,\\ \left(w^T A\right)_i - q_i^g\geq 0, i=1,\dots,n}} \left\{ - w^T b + \sum_{i=1}^n \sup_{q_i^f\leq 1+\ln q_i^g} q_i^f \right\}.$$

As the suprema after q_i^f, $i = 1, \dots, n$, are trivially computable we get for the dual problem the following continuation

$$(D_B) \qquad \sup_{\substack{w\in\mathbb{R}^m, q_i^g>0,\\ \left(w^T A\right)_i - q_i^g\geq 0, i=1,\dots,n}} \left\{ - w^T b + \sum_{i=1}^n \left(1 + \ln q_i^g\right) \right\}.$$

The variable q^g may also be retired, but in this case another constraint appears, namely $w^T A > 0$. For the sake of simplicity let us perform this step, too. The following problem is the ultimate dual problem to (P_B'')

$$(D_B) \qquad \sup_{\substack{w\in\mathbb{R}^m,\\ w^T A>0}} \left\{ n - w^T b + \sum_{i=1}^n \ln \left(w^T A\right)_i \right\}.$$

Since the constraints of the primal problem (P_B) are linear and all the feasible points x are in $\text{int}(\mathbb{R}_+^n) = \text{ri}(\mathbb{R}_+^n)$, no constraint qualification is required in this case. We can formulate the strong duality and optimality conditions statements right away. These assertions do not appear in the cited article, but we give them without proofs since these are similar to the ones already presented in the paper. There is a difference between the strong duality notion used here and the previous ones because normally we would present strong duality between (P_B'') and (D_B). But since the starting problem is (P_B) we modify a bit the statements using the obvious result $v(P_B) = -v(P_B'')$.

Theorem 4.27 *(strong duality) Provided that the primal problem (P_B) has a feasible point, the dual problem (D_B) has an optimal solution where it attains its maximal value and the sum of the optimal objective values of the two problems is zero, i.e. $v(P_B) + v(D_B) = 0$.*

Theorem 4.28 *(optimality conditions)*
(a) *If the primal problem* (P_B) *has an optimal solution* $\bar{x}$*, then the dual problem* (D_B) *has also an optimal solution* $\bar{w}$ *and the following optimality conditions hold*

(i) $\ln \bar{x}_i + \ln \left(\bar{w}^T A\right)_i + n = \bar{w}^T A\bar{x}$,

(ii) $\bar{w}^T(A\bar{x} - b) = 0$.

(b) *If* $\bar{x}$ *is a feasible point to* (P_B) *and* $\bar{w}$ *is feasible to* (D_B) *such that the optimality conditions* (i) *and* (ii) *are true, then* $v(P_B)+v(D_B) = 0$, $\bar{x}$ *is an optimal solution to* (P_B) *and* $\bar{w}$ *an optimal solution to* (D_B).

4.2.3 An application in text classification

After presenting all these theoretical results, let us give a concrete application of the entropy optimization in text classification.

Here we rigorously apply the maximum entropy optimization to a text classification problem, correcting the errors encountered in some other papers regarding the subject, whose authors make some compromises in order to obtain some "good-looking" results (see [6, 66]). We have a set of documents which must be classified into some given classes. A small amount of them have been a priori labelled by an expert and we have also some real-valued functions linking all the documents and the classes, called features functions. Our goal is to obtain a distribution of probabilities assigning to each document the chances to belong to the given classes.

Therefore, we impose the condition that the expected value of each features function over the whole set of documents shall be equal to its expected value over the training sample. Using this information as constraints, we formulate the so - called maximum entropy optimization problem. Its solutions are consistent with all the constraints, but otherwise are as uniform as possible (cf. [34, 41, 57]).

To our maximum entropy optimization problem we attach the Lagrange dual. As a consequence of the optimality conditions, we write the solutions of the primal problem as functions having as variables the solutions of the dual problem. The last ones are determined using the so - called iterative scaling algorithm developed from the one introduced by Darroch and Ratcliff (cf. [28, 66]).

Finally, by the use of the solutions of the dual, we find the desired distribution of probabilities.

4.2.3.1 The formulation of the problem

Let us consider a set of documents $\mathcal{D}$ and the set of classes $\mathcal{C}$ where they must be classified into. There is also a given subset of $\mathcal{D}$, denoted $\mathcal{D}'$, whose elements have been labelled by an expert as to belong to a certain class from $\mathcal{C}$. To have information about all the classes, we need to consider that each class contains at least an element from $\mathcal{D}'$. One may notice that between the sets $\mathcal{C}$ and $\mathcal{D}'$, it must hold $|\mathcal{C}| \leq |\mathcal{D}'|$, where $|\mathcal{C}|$ is the cardinality of the set $\mathcal{C}$ and $|\mathcal{D}'|$ is the cardinality of the set $\mathcal{D}'$. The set of pairs $\left\{(d', c(d')), d' \in \mathcal{D}'\right\}$, obtained above, is called the *training data* and $c(d') \in \mathcal{C}$ denotes the class which is assigned to d' by the expert.

The labelled training data set is used to derive a set of constraints for the model that characterize the class-specific expectations for the distribution. The constraints are represented as expected values of so - called features functions, which may be any positive real-valued functions defined over $\mathcal{D} \times \mathcal{C}$. Let us denote by $f_i, i \in I$, the features functions for the problem of text classification treated here.

As an example, we will present the set of features functions considered in [66] for the same problem of text classification. Denoting by $\mathcal{W}$ the set of the words which appear in the whole family of documents $\mathcal{D}$, the set I is defined by

$$I = \mathcal{W} \times \mathcal{C}.$$

For each word - class combination $(w, c') \in \mathcal{W} \times \mathcal{C}$, one can consider the features function $f_{w,c'} : \mathcal{W} \times \mathcal{C} \to \mathbb{R}$,

$$f_{w,c'}(d, c) = \begin{cases} 0, & \text{if } c \neq c', \\ \frac{N(d,w)}{N(d)}, & \text{otherwise,} \end{cases}$$

where $N(d, w)$ is the number of times word w occurs in the document d, $N(d)$ is the number of words in d, and c, c' are classes in $\mathcal{C}$.

Other ways to consider features functions can be found for instance in [2] and [53].

In order to build a mathematical model of the problem we need the expected values of the features functions. For $i \in I$, the *expected value* of the features function f_i *regarding the whole set* $\mathcal{D} \times \mathcal{C}$ is

$$E(f_i) = \sum_{d \in \mathcal{D}} \sum_{c \in \mathcal{C}} p(d, c) f_i(d, c),$$

where $p(d, c)$ denotes the joint probability of c and d, $c \in \mathcal{C}$, $d \in \mathcal{D}$, while its *expected value regarding the training sample* comes from the following formula

$$\tilde{E}(f_i) = \sum_{d' \in \mathcal{D}'} \sum_{c \in \mathcal{C}} p(d', c) f_i(d', c). \tag{4. 14}$$

The joint probability can be decomposed as

$$p(d, c) = p(d) p(c|d),$$

with $p(d)$ being the probability of the document d to be chosen from the set $\mathcal{D}$ and $p(c|d)$ the conditional probability of the class $c \in \mathcal{C}$ with respect to the document $d \in \mathcal{D}$. In (4. 14) instead of $\mathcal{D}$ we work with the training sample $\mathcal{D}'$.

Using the information given by this training data and the features functions, we want to obtain the distribution of probabilities of each document $d \in \mathcal{D}$ among the given classes. The way we do this is quite heuristical (cf. [2, 53]), consisting in generalizing some facts that hold for the training sample to the whole set of documents. The expected value of each features function over all the documents and classes will be forced to coincide to its expected value over the training sample

$$\tilde{E}(f_i) = E(f_i) \ \forall i \in I. \tag{4. 15}$$

The probability of the document d' to be chosen from the training data is

$$p(d') = \frac{1}{|\mathcal{D}'|}, \text{ for } d' \in \mathcal{D}'.$$

On the other hand, as we know that each document from the training data has been a priori labelled, it is clear that

$$p(c|d') = \begin{cases} 1, & \text{if} \quad c = c(d'), \\ 0, & \text{if} \quad c \neq c(d'), \end{cases}$$

for every $c \in \mathcal{C}$ and $d' \in \mathcal{D}'$.

By (4. 14), we have then

$$\tilde{E}(f_i) = \frac{1}{|\mathcal{D}'|} \sum_{d' \in \mathcal{D}'} f_i(d', c(d')), i \in I. \tag{4. 16}$$

It is clear that the probability to choose the document d from $\mathcal{D}$ is $p(d) = \frac{1}{|\mathcal{D}|}$. Then one has

$$E(f_i) = \frac{1}{|\mathcal{D}|} \sum_{d \in \mathcal{D}} \sum_{c \in \mathcal{C}} p(c|d) f_i(d,c), i \in I. \tag{4. 17}$$

For each features function $f_i, i \in I$, we will constrain now the model to have the same expected value for it over the whole set of documents as the one obtained from the training set. From (4. 15), (4. 16) and (4. 17), we obtain

$$\frac{1}{|\mathcal{D}'|} \sum_{d' \in \mathcal{D}'} f_i(d', c(d')) = \frac{1}{|\mathcal{D}|} \sum_{d \in \mathcal{D}} \sum_{c \in \mathcal{C}} p(c|d) f_i(d,c), i \in I. \tag{4. 18}$$

Moreover, from the basic properties of the probability distributions, it holds

$$p(c|d) \geq 0 \ \forall c \in \mathcal{C} \ \forall d \in \mathcal{D}, \tag{4. 19}$$

and

$$\sum_{c \in \mathcal{C}} p(c|d) = 1 \ \forall d \in \mathcal{D}. \tag{4. 20}$$

The problem that we have to solve now is to find a probability distribution which fulfills the constraints (4. 18), (4. 19) and (4. 20). Therefore, we will use a technique which is based on theory of maximum entropy (cf. [34, 41, 57]). The over - riding principle in maximum entropy is that when nothing else is known, the distribution of probabilities should be as uniform as possible.

This is exactly what results by solving the following so - called maximum entropy optimization problem

$$(P_t) \qquad \sup H(p),$$

subject to

$$\frac{|\mathcal{D}|}{|\mathcal{D}'|} \sum_{d' \in \mathcal{D}'} f_i(d', c(d')) = \sum_{d \in \mathcal{D}} \sum_{c \in \mathcal{C}} p(c|d) f_i(d,c) \ \forall i \in I,$$

$$\sum_{c \in \mathcal{C}} p(c|d) = 1 \ \forall d \in \mathcal{D},$$

and

$$p(c|d) \geq 0, \forall c \in \mathcal{C} \ \forall d \in \mathcal{D}.$$

Here, $H : \mathbb{R}^{|C| \cdot |D|} \to \overline{\mathbb{R}}$ is the entropy function and it is defined, for $p = (p(c|d))_{c \in \mathcal{C}, d \in \mathcal{D}}$, by

$$H(p) = \begin{cases} -\sum\limits_{d \in \mathcal{D}} \sum\limits_{c \in \mathcal{C}} p(c|d) \ln p(c|d), & \text{if } p(c|d) \geq 0 \ \forall c \in \mathcal{C} \ \forall d \in \mathcal{D}, \\ -\infty, & \text{otherwise.} \end{cases}$$

It is obvious that H is a concave function.

4.2.3.2 Duality for the maximum entropy optimization problem

The goal of this chapter is to formulate a dual problem to the maximum entropy optimization problem

$$(P_t) \qquad \sup \left\{ -\sum_{d \in \mathcal{D}} \sum_{c \in \mathcal{C}} p(c|d) \ln p(c|d) \right\},$$

subject to

$$\frac{|\mathcal{D}|}{|\mathcal{D}'|} \sum_{d' \in \mathcal{D}'} f_i(d', c(d')) = \sum_{d \in \mathcal{D}} \sum_{c \in \mathcal{C}} p(c|d) f_i(d, c) \ \forall i \in I,$$

$$\sum_{c \in \mathcal{C}} p(c|d) = 1 \ \forall d \in \mathcal{D},$$

$$p(c|d) \geq 0 \ \forall c \in \mathcal{C} \ \forall d \in \mathcal{D},$$

and to derive, by means of strong duality, the optimality conditions for (P_t) and its dual. As this is a maximization problem and our approach works for minimization problems, we need to consider another optimization problem

$$(P_t') \qquad \inf \sum_{d \in \mathcal{D}} \sum_{c \in \mathcal{C}} p(c|d) \ln p(c|d),$$

subject to

$$\frac{|\mathcal{D}|}{|\mathcal{D}'|} \sum_{d' \in \mathcal{D}'} f_i(d', c(d')) = \sum_{d \in \mathcal{D}} \sum_{c \in \mathcal{C}} p(c|d) f_i(d, c) \ \forall i \in I,$$

$$\sum_{c \in \mathcal{C}} p(c|d) = 1 \ \forall d \in \mathcal{D},$$

$$p \in X,$$

with $X = \{p = (p(c|d))_{\substack{c \in \mathcal{C}, \\ d \in \mathcal{D}}} : p(c|d) \geq 0 \ \forall c \in \mathcal{C} \ \forall d \in \mathcal{D}\}$. The problem (P_t') fits in the scheme already presented and has the same optimal solutions as (P_t) so that it holds $v(P_t) = -v(P_t')$. According to the general case, its Lagrange dual problem is

$$(D_t') \qquad \sup_{\substack{\lambda_i \in \mathbb{R}, i \in I, \\ \lambda_d \in \mathbb{R}, d \in \mathcal{D}}} \inf_{\substack{p(c|d) \geq 0, \\ (d,c) \in \mathcal{D} \times \mathcal{C}}} \Bigg[\sum_{d \in \mathcal{D}} \sum_{c \in \mathcal{C}} p(c|d) \ln p(c|d) + \sum_{d \in \mathcal{D}} \lambda_d \left(\sum_{c \in \mathcal{C}} p(c|d) - 1 \right)$$
$$+ \sum_{i \in I} \lambda_i \left(\frac{|\mathcal{D}|}{|\mathcal{D}'|} \sum_{d' \in \mathcal{D}'} f_i(d', c(d')) - \sum_{d \in \mathcal{D}} \sum_{c \in \mathcal{C}} p(c|d) f_i(d, c) \right) \Bigg].$$

Like above, we can find another problem, (D_t''), which has the same solutions as (D_t') so that $v(D_t') = -v(D_t'')$,

$$(D_t'') \qquad \inf_{\substack{\lambda_i \in \mathbb{R}, i \in I, \\ \lambda_d \in \mathbb{R}, d \in \mathcal{D}}} \sup_{\substack{p(c|d) \geq 0, \\ (d,c) \in \mathcal{D} \times \mathcal{C}}} \Bigg\{ - \sum_{d \in \mathcal{D}} \sum_{c \in \mathcal{C}} p(c|d) \ln p(c|d) - \sum_{d \in \mathcal{D}} \lambda_d \left(\sum_{c \in \mathcal{C}} p(c|d) - 1 \right)$$
$$- \sum_{i \in I} \lambda_i \left(\frac{|\mathcal{D}|}{|\mathcal{D}'|} \sum_{d' \in \mathcal{D}'} f_i(d', c(d')) - \sum_{d \in \mathcal{D}} \sum_{c \in \mathcal{C}} p(c|d) f_i(d, c) \right) \Bigg\}.$$

The latter can be rewritten as

$$(D_t'') \qquad \inf_{\substack{\lambda_i \in \mathbb{R}, i \in I, \\ \lambda_d \in \mathbb{R}, d \in \mathcal{D}}} \Bigg[\sum_{d \in \mathcal{D}} \lambda_d - \frac{|\mathcal{D}|}{|\mathcal{D}'|} \sum_{i \in I} \lambda_i \sum_{d' \in \mathcal{D}'} f_i(d', c(d')) +$$
$$\sum_{d \in \mathcal{D}} \sum_{c \in \mathcal{C}} \sup_{p(c|d) \geq 0} \Bigg\{ - p(c|d) \ln p(c|d) - \lambda_d p(c|d) + \sum_{i \in I} \lambda_i p(c|d) f_i(d, c) \Bigg\} \Bigg].$$

To calculate the suprema which appear in the above formula, we consider the function

$$u : \mathbb{R}_+ \to \mathbb{R}, u(x) = -x \ln x - \lambda_d x + x \sum_{i \in I} \lambda_i f_i(d, c), x \in \mathbb{R}_+,$$

for some fixed $(d,c) \in \mathcal{D} \times \mathcal{C}$.

Its derivative is

$$u'(x) = -\ln x - 1 - \lambda_d + \sum_{i \in I} \lambda_i f_i(d,c),$$

and it holds

$$u'(x) = 0 \Leftrightarrow \ln x = -\lambda_d - 1 + \sum_{i \in I} \lambda_i f_i(d,c) \Leftrightarrow x = e^{-\lambda_d - 1 + \sum\limits_{i \in I} \lambda_i f_i(d,c)} > 0.$$

The function u being concave, it follows that at $x = e^{-\lambda_d - 1 + \sum\limits_{i \in I} \lambda_i f_i(d,c)}$ it attains its maximal value. So

$$\begin{aligned}
\max_{x \geq 0} u(x) &= u\left(e^{-\lambda_d - 1 + \sum\limits_{i \in I} \lambda_i f_i(d,c)}\right) \\
&= -e^{-\lambda_d - 1 + \sum\limits_{i \in I} \lambda_i f_i(d,c)} \left(\ln e^{-\lambda_d - 1 + \sum\limits_{i \in I} \lambda_i f_i(d,c)} + \lambda_d - \sum_{i \in I} \lambda_i f_i(d,c)\right) \\
&= -e^{-\lambda_d - 1 + \sum\limits_{i \in I} \lambda_i f_i(d,c)} \left(\sum_{i \in I} \lambda_i f_i(d,c) - \lambda_d - 1 + \lambda_d - \sum_{i \in I} \lambda_i f_i(d,c)\right) \\
&= e^{-\lambda_d - 1 + \sum\limits_{i \in I} \lambda_i f_i(d,c)}.
\end{aligned}$$

The dual problem becomes then

$$(D_t'') \quad \inf_{\substack{\lambda_i \in \mathbb{R}, i \in I, \\ \lambda_d \in \mathbb{R}, d \in \mathcal{D}}} \left[\sum_{d \in \mathcal{D}} \lambda_d - \frac{|\mathcal{D}|}{|\mathcal{D}'|} \sum_{i \in I} \lambda_i \sum_{d' \in \mathcal{D}'} f_i(d', c(d')) + \sum_{d \in \mathcal{D}} \sum_{c \in \mathcal{C}} e^{\sum\limits_{i \in I} \lambda_i f_i(d,c) - \lambda_d - 1} \right].$$

In the next part of the section we will make some assertions concerning the duality between (P_t) and (D_t''). For this, we will apply the results formulated in the general case. We need to introduce a constraint qualification

$$(CQ_t) \exists p' = (p'(c|d))_{\substack{c \in \mathcal{C}, \\ d \in \mathcal{D}}} : \begin{cases} p'(c|d) > 0 \ \forall c \in \mathcal{C} \ \forall d \in \mathcal{D}, \\ \frac{|\mathcal{D}|}{|\mathcal{D}'|} \sum\limits_{d' \in \mathcal{D}'} f_i(d', c(d')) = \sum\limits_{d \in \mathcal{D}} \sum\limits_{c \in \mathcal{C}} p'(c|d) f_i(d,c) \ \forall i \in I, \\ \sum\limits_{c \in \mathcal{C}} p'(c|d) = 1 \ \forall d \in \mathcal{D}. \end{cases}$$

By this, we can state now the desired strong duality theorem and the optimality conditions for (P_t) and (D_t'').

Theorem 4.29 *(strong duality) Let (CQ_t) be fulfilled. Then there is strong duality between (P_t) and (D_t''), i.e. $v(P_t) = v(D_t'')$ and (D_t'') has an optimal solution.*

Theorem 4.30 *(optimality conditions) Let us assume that the constraint qualification (CQ_t) is fulfilled. Then $\bar{p} = ((\bar{p}(c|d))_{\substack{c \in \mathcal{C}, \\ d \in \mathcal{D}}}$ is a solution to (P_t) if and only if $\bar{p}$ is feasible to (P_t) and there exist $\bar{\lambda}_i \in \mathbb{R}, i \in I$, and $\bar{\lambda}_d \in \mathbb{R}, d \in \mathcal{D}$, such that the following optimality conditions are satisfied*

$$(i) \quad \inf_{\substack{p(c|d)\geq 0,\\ d\in\mathcal{D},c\in\mathcal{C}}} \Bigg[\sum_{d\in\mathcal{D}}\sum_{c\in\mathcal{C}} p(c|d)\ln p(c|d) + \sum_{d\in\mathcal{D}}\bar{\lambda}_d\left(\sum_{c\in\mathcal{C}} p(c|d) - 1\right)$$

$$+ \sum_{i\in I}\bar{\lambda}_i\left(\frac{|\mathcal{D}|}{|\mathcal{D}'|}\sum_{d'\in\mathcal{D}'} f_i(d',c(d')) - \sum_{d\in\mathcal{D}}\sum_{c\in\mathcal{C}} p(c|d)f_i(d,c)\right)\Bigg]$$

$$= \sum_{d\in\mathcal{D}}\sum_{c\in\mathcal{C}}\bar{p}(c|d)\ln\bar{p}(c|d),$$

$$(ii) \quad 0 = \sum_{i\in I}\bar{\lambda}_i\left(\frac{|\mathcal{D}|}{|\mathcal{D}'|}\sum_{d'\in\mathcal{D}'} f_i(d',c(d')) - \sum_{d\in\mathcal{D}}\sum_{c\in\mathcal{C}}\bar{p}(c|d)f_i(d,c)\right)$$

$$+ \sum_{d\in\mathcal{D}}\bar{\lambda}_d\left(\sum_{c\in\mathcal{C}}\bar{p}(c|d) - 1\right).$$

Remark 4.6 Let us point out that all the functions involved in the formulation of the primal problem are differentiable. This implies that the equality (i) in Theorem 4.30 can be, equivalently, written as

$$\ln p(c|d) + 1 + \bar{\lambda}_d - \sum_{i\in I}\bar{\lambda}_i f_i(d,c) = 0 \ \forall d\in\mathcal{D}\ \forall c\in\mathcal{C},$$

or

$$p(c|d) = \frac{e^{\sum\limits_{i\in I}\bar{\lambda}_i f_i(d,c)}}{e^{\bar{\lambda}_d+1}} \ \forall d\in\mathcal{D}\ \forall c\in\mathcal{C}. \tag{4. 21}$$

Getting now back to the problem (D_t''), one may observe that it can be decomposed into

$$(D_t'') \inf_{\substack{\lambda_i\in\mathbb{R},\\ i\in I}}\left[\sum_{d\in\mathcal{D}}\inf_{\lambda_d\in\mathbb{R}}\left[\sum_{c\in\mathcal{C}} e^{\sum\limits_{i\in I}\lambda_i f_i(d,c)-\lambda_d-1} + \lambda_d\right] - \frac{|\mathcal{D}|}{|\mathcal{D}'|}\sum_{i\in I}\lambda_i\sum_{d'\in\mathcal{D}'} f_i(d',c(d'))\right].$$

We can calculate the infima inside the parentheses, using another auxiliary function, namely

$$v:\mathbb{R}\to\mathbb{R}, v(x) = e^{-x-1}a + x, x\in\mathbb{R}, a>0.$$

It is convex and derivable, its derivative

$$v'(x) = 1 - ae^{-x-1}$$

fulfilling

$$v'(x) = 0 \Leftrightarrow e^{-x-1} = \frac{1}{a} \Leftrightarrow x = \ln a - 1.$$

So, v's minimum is attained at $\ln a - 1$, being

$$v(\ln a - 1) = \ln a.$$

Taking $a = \sum_{c\in\mathcal{C}} e^{\sum_{i\in I}\lambda_i f_i(d,c)} > 0$, the dual problem turns into

$$(D_t) \quad \inf_{\substack{\lambda_i\in\mathbb{R},\\ i\in I}}\left[\sum_{d\in\mathcal{D}}\ln\sum_{c\in\mathcal{C}} e^{\sum\limits_{i\in I}\lambda_i f_i(d,c)} - \frac{|\mathcal{D}|}{|\mathcal{D}'|}\sum_{i\in I}\lambda_i\sum_{d'\in\mathcal{D}'} f_i(d',c(d'))\right],$$

and, obviously, we have

$$v(D_t'') = v(D_t).$$

In fact, we have proven the following assertion concerning the solutions of the problems (D_t) and (D_t'').

Theorem 4.31 *The following equivalence holds*

$$((\bar{\lambda}_d)_{d\in\mathcal{D}}, (\bar{\lambda}_i)_{i\in I}) \text{ is a solution to } (D_t'') \Leftrightarrow \begin{cases} \bar{\lambda}_d = \ln \sum\limits_{c\in\mathcal{C}} e^{\sum\limits_{i\in I} \bar{\lambda}_i f_i(d,c)} - 1 \forall d \in \mathcal{D} \\ \text{and } ((\bar{\lambda}_i)_{i\in I}) \text{ is a solution to } (D_t). \end{cases}$$

Remark 4.7 By Remark 4.6 and Theorem 4.31 it follows that, in order to find a solution of the problem (P_t), it is enough to solve the dual problem (D_t). Getting $(\bar{\lambda}_i)_{i\in I}$, solution to (D_t), we obtain, for each $d \in \mathcal{D}$,

$$\bar{\lambda}_d = \ln \sum_{c\in\mathcal{C}} e^{\sum\limits_{i\in I} \bar{\lambda}_i f_i(d,c)} - 1 \tag{4. 22}$$

and, by (4. 21),

$$p(c|d) = \frac{e^{\sum\limits_{i\in I} \bar{\lambda}_i f_i(d,c)}}{e^{\bar{\lambda}_d + 1}} = \frac{e^{\sum\limits_{i\in I} \bar{\lambda}_i f_i(d,c)}}{\sum\limits_{c\in\mathcal{C}} e^{\sum\limits_{i\in I} \bar{\lambda}_i f_i(d,c)}} \ \forall (d,c) \in \mathcal{D} \times \mathcal{C}. \tag{4. 23}$$

4.2.3.3 Solving the dual problem

In the following we outline the derivation of an algorithm for finding a solution of the dual problem (D_t). The algorithm is called *improved iterative scaling* and some other versions of it have been described by different authors in connection with maximum entropy optimization problems (cf. [6] and [66]).

First, let us introduce the function $l : \mathbb{R}^{|I|} \to \mathbb{R}$, defined by

$$l(\lambda) = \sum_{i\in I} \lambda_i \left(\frac{|\mathcal{D}|}{|\mathcal{D}'|} \sum_{d'\in\mathcal{D}'} f_i(d', c(d')) \right) - \sum_{d\in\mathcal{D}} \ln \sum_{c\in\mathcal{C}} e^{\sum\limits_{i\in I} \lambda_i f_i(d,c)},$$

for $\lambda = (\lambda_i)_{i\in I}$.

Considering the optimization problem

$$(P_i) \qquad \max_{\lambda\in\mathbb{R}^{|I|}} l(\lambda),$$

it is obvious that $v(D_t) = -v(P_i)$ and the sets of the solutions of the two problems are nonempty and coincide. So, in order to obtain the desired results, it is enough to solve (P_i).

Let us calculate now, for $\lambda = (\lambda_i)_{i\in I}, \delta = (\delta_i)_{i\in I} \in \mathbb{R}^{|I|}$, the expression $\Delta l = l(\lambda + \delta) - l(\lambda)$. It holds

$$\begin{aligned} \Delta l &= \frac{|\mathcal{D}|}{|\mathcal{D}'|} \sum_{i\in I} \sum_{d'\in\mathcal{D}'} (\lambda_i + \delta_i) f_i(d', c(d')) - \sum_{d\in\mathcal{D}} \ln \sum_{c\in\mathcal{C}} e^{\sum\limits_{i\in I} (\lambda_i+\delta_i) f_i(d,c)} \\ &- \frac{|\mathcal{D}|}{|\mathcal{D}'|} \sum_{i\in I} \sum_{d'\in\mathcal{D}'} \lambda_i f_i(d', c(d')) + \sum_{d\in\mathcal{D}} \ln \sum_{c\in\mathcal{C}} e^{\sum\limits_{i\in I} \lambda_i f_i(d,c)} \\ &= \frac{|\mathcal{D}|}{|\mathcal{D}'|} \sum_{i\in I} \sum_{d'\in\mathcal{D}'} \delta_i f_i(d', c(d')) - \sum_{d\in\mathcal{D}} \ln \frac{\sum\limits_{c\in\mathcal{C}} e^{\sum\limits_{i\in I} (\lambda_i+\delta_i) f_i(d,c)}}{\sum\limits_{c\in\mathcal{C}} e^{\sum\limits_{i\in I} \lambda_i f_i(d,c)}}. \end{aligned}$$

As it is known that

$$-\ln(x) \geq 1 - x \ \forall x \in \mathbb{R}_+,$$

we have

$$\begin{aligned}\Delta l &\geq \frac{|\mathcal{D}|}{|\mathcal{D}'|}\sum_{i\in I}\sum_{d'\in\mathcal{D}'}\delta_i f_i(d',c(d')) + \sum_{d\in\mathcal{D}}\left(1-\frac{\sum\limits_{c\in\mathcal{C}} e^{\sum\limits_{i\in I}(\lambda_i+\delta_i)f_i(d,c)}}{\sum\limits_{c\in\mathcal{C}} e^{\sum\limits_{i\in I}\lambda_i f_i(d,c)}}\right)\\ &= \frac{|\mathcal{D}|}{|\mathcal{D}'|}\sum_{i\in I}\sum_{d'\in\mathcal{D}'}\delta_i f_i(d',c(d')) + \sum_{d\in\mathcal{D}}\left(1-\sum_{c\in\mathcal{C}}p(c|d)e^{\sum\limits_{i\in I}\delta_i f_i(d,c)}\right).\end{aligned}$$

Denoting

$$f^{\#}(d,c)=\sum_{i\in I} f_i(d,c),$$

we get

$$\Delta l \geq \frac{|\mathcal{D}|}{|\mathcal{D}'|}\sum_{i\in I}\sum_{d'\in\mathcal{D}'}\delta_i f_i(d',c(d')) + \sum_{d\in\mathcal{D}}\left(1-\sum_{c\in\mathcal{C}}p(c|d)e^{f^{\#}(d,c)\sum\limits_{i\in I}\delta_i\frac{f_i(d,c)}{f^{\#}(d,c)}}\right).$$

As the exponential function is convex, applying Jensen's inequality

$$e^{f^{\#}(d,c)\sum\limits_{i\in I}\delta_i\frac{f_i(d,c)}{f^{\#}(d,c)}} \leq \sum_{i\in I}\frac{f_i(d,c)}{f^{\#}(d,c)}e^{f^{\#}(d,c)\delta_i},$$

there follows

$$\Delta l \geq \frac{|\mathcal{D}|}{|\mathcal{D}'|}\sum_{i\in I}\sum_{d'\in\mathcal{D}'}\delta_i f_i(d',c(d')) - \sum_{d\in\mathcal{D}}\sum_{c\in\mathcal{C}}p(c|d)\sum_{i\in I}\frac{f_i(d,c)}{f^{\#}(d,c)}e^{f^{\#}(d,c)\delta_i}+|\mathcal{D}|.$$

Let $\mathcal{W}:\mathbb{R}^{|I|}\to\mathbb{R}$, be the following function

$$\mathcal{W}(\delta)=\frac{|\mathcal{D}|}{|\mathcal{D}'|}\sum_{i\in I}\sum_{d'\in\mathcal{D}'}\delta_i f_i(d',c(d')) - \sum_{d\in\mathcal{D}}\sum_{c\in\mathcal{C}}p(c|d)\sum_{i\in I}\frac{f_i(d,c)}{f^{\#}(d,c)}e^{f^{\#}(d,c)\delta_i}+|\mathcal{D}|,$$

for $\delta=(\delta_i)_{i\in I}$.

We can guarantee an increase of the value of the function l if we can find a δ such that $\mathcal{W}(\delta)$ is positive. $\mathcal{W}$ is a concave function since its first term is a linear function, the second contains a sum of concave functions and the third is a constant. Moreover, $\mathcal{W}$ is a differentiable function. So, to find the best δ, we need to differentiate $\mathcal{W}(\delta)$ with respect to the change in each parameter δ_i, $i\in I$, and to set

$$\frac{\partial\mathcal{W}}{\partial\delta_i}=0\ \forall i\in I.$$

We get

$$\frac{|\mathcal{D}|}{|\mathcal{D}'|}\sum_{d'\in\mathcal{D}'}f_i(d',c(d'))=\sum_{c\in\mathcal{C}}\sum_{d\in\mathcal{D}}p(c|d)f_i(d,c)e^{f^{\#}(d,c)\delta_i}\ \forall i\in I.$$

Solving these equations we obtain the values of $\delta_i, i\in I$. In the next section we present an algorithm which helps to determine the maximum of the function l.

Remark 4.8 We have to mention here that in [6] and [66] the function l has been identified with the so - called *maximum likelihood* function, whose formula is considered

$$L(\lambda)=\ln\prod_{d'\in\mathcal{D}'}p(c(d')|d').$$

This is possible only if one considers the sets $\mathcal{D}$ and $\mathcal{D}'$ identical. In this case, we have

$$\begin{aligned}
l(\lambda) &= \sum_{i\in I}\lambda_i\left(\frac{|\mathcal{D}|}{|\mathcal{D}'|}\sum_{d'\in\mathcal{D}'}f_i(d',c(d'))\right)-\sum_{d\in\mathcal{D}}\ln\sum_{c\in\mathcal{C}}e^{\sum\limits_{i\in I}\lambda_i f_i(d,c)}\\
&= \sum_{d'\in\mathcal{D}'}\left[\sum_{i\in I}\lambda_i f_i(d',c(d'))-\ln\sum_{c\in\mathcal{C}}e^{\sum\limits_{i\in I}\lambda_i f_i(d',c(d'))}\right]\\
&= \sum_{d'\in\mathcal{D}'}\left[\ln e^{\sum\limits_{i\in I}\lambda_i f_i(d',c(d'))}-\ln\sum_{c\in\mathcal{C}}e^{\sum\limits_{i\in I}\lambda_i f_i(d',c(d'))}\right]\\
&= \sum_{d'\in\mathcal{D}'}\left[\ln\frac{e^{\sum\limits_{i\in I}\lambda_i f_i(d',c(d'))}}{\sum\limits_{c\in\mathcal{C}}e^{\sum\limits_{i\in I}\lambda_i f_i(d',c(d'))}}\right].
\end{aligned}$$

Finally, using the relations given in (4. 23), the function l turns out to be in this case identical to the maximum likelihood function

$$l(\lambda)=\sum_{d'\in\mathcal{D}'}\left[\ln\frac{e^{\sum\limits_{i\in I}\lambda_i f_i(d',c(d'))}}{\sum\limits_{c\in\mathcal{C}}e^{\sum\limits_{i\in I}\lambda_i f_i(d',c(d'))}}\right]=\sum_{d'\in\mathcal{D}'}\ln p(c(d')|d')=\ln\prod_{d'\in\mathcal{D}'}p(c(d')|d').$$

We can conclude that the results obtained in [6] and [66] do not refer to the unclassified documents using the information given by that expert regarding the training sample, being just distributions of the same a priori labelled documents among all the classes. We consider that this compromise is not useful in our problem, as we have proven before that the algorithm works also without it.

4.2.3.4 An algorithm for solving the maximum entropy optimization problem

Making use of the results obtained in the previous sections we present now an algorithm for solving the dual of the maximum entropy optimization problem. Assuming that the constraint qualification (CQ_t) is fulfilled the solutions of the primal problem arise by calling (4. 23). This is a generalization of the algorithm introduced by Darroch and Ratcliff in [28].

Inputs: A collection $\mathcal{D}$ of documents, a subset of it $\mathcal{D}'$ of labelled documents, a set of classes $\mathcal{C}$ and a set of features functions $f_i, i\in I$, connecting the documents and the classes. Let $\varepsilon>0$ be the admitted error of the iterative process.

Step 1: Set the constraints. For every features function $f_i, i\in I$, estimate its expected value over the set of the documents and the set of classes.

Step 2: Set the initial values $\lambda_i=0, i\in I$.

Step 3:

- Using the equalities in (4. 23), calculate with the current parameters $(\lambda_i)_{i\in I}$ the values for $p(c|d), (d,c)\in\mathcal{D}\times\mathcal{C}$.
- For each $i\in I$:
  - find δ_i, a solution of the equation

$$\frac{|\mathcal{D}|}{|\mathcal{D}'|}\sum_{d'\in\mathcal{D}'}f_i(d',c(d'))=\sum_{c\in\mathcal{C}}\sum_{d\in\mathcal{D}}p(c|d)f_i(d,c)e^{f^{\#}(d,c)\delta_i},$$

· set $\lambda_i = \lambda_i + \delta_i$.

Step 4: If there exists an $i \in I$, such that $|\delta_i| > \varepsilon$, then go to **Step 3**.

Outputs: The approximate solutions to the dual problem λ_i, $i \in I$.

Remark 4.9

(a) By setting $\lambda_i = 0\ \forall i \in I$, the initial values for the probability distributions are

$$p(c|d) = \frac{1}{|\mathcal{C}|}, c \in \mathcal{C}, d \in \mathcal{D}.$$

(b) In the original algorithm Darroch and Ratcliff assumed in [28] that $f^{\#}(d,c)$ is constant. Denoting its value by M, one gets then

$$\delta_i = \frac{1}{M} \ln \left(\frac{|\mathcal{D}|}{|\mathcal{D}'|} \frac{\sum\limits_{d' \in \mathcal{D}'} f_i(d', c(d'))}{\sum\limits_{c \in \mathcal{C}} \sum\limits_{d \in \mathcal{D}} p(c|d) f_i(d,c)} \right) \quad \forall i \in I.$$

(c) A more detailed discussion regarding the iterative scaling algorithm, including a proof of its convergence, can be found in [3, 28, 29, 66].

Having obtained $\lambda_i, i \in I$, returned by the algorithm, we can determine by (4. 22) and (4. 23) the solutions of the primal problem, i.e. the probability distributions of each document among the given classes.

To assign each document with a certain class, one can consider more criteria, such as to choose the class whose probability is the highest, or to establish a minimal value of probability and to label the documents as belonging to all the classes that fit, and if neither does, to create an additional class for this document.

Theses

1. The general optimization problem

$$(P) \qquad \inf_{\substack{x \in X, \\ g(x) \in -C}} f(x),$$

is considered, where X is a non - empty convex subset of $\mathbb{R}^n$, C is a non - empty closed convex cone in $\mathbb{R}^m$, $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is a proper convex function and $g : X \to \mathbb{R}^m$ is a C - convex function on X. The Lagrange, Fenchel and Fenchel - Lagrange dual problems are attached to (P) via perturbations. The latter, extensively used later within this work, is

$$(D^{FL}) \qquad \sup_{\substack{p^* \in \mathbb{R}^n, \\ q^* \in C^*}} \Big\{ -f^*(p^*) - ({q^*}^T g)^*_X(-p^*) \Big\},$$

where ${q^*}^T g : X \to \mathbb{R}$ is defined as ${q^*}^T g(x) = \sum_{j=1}^m q_j^* g_j(x)$, with $q^* = \big(q_1^*, \ldots, q_m^*\big) \in C^*$. Weak and strong duality and necessary and sufficient optimality conditions are formulated for the pair of problems $(P) - (D^{FL})$. These results naturally generalize the ones known so far for the case $C = \mathbb{R}^m_+$ (see also [9]).

2. Consider the primal optimization problem

$$(P_F) \qquad \inf_{x \in \mathbb{R}^n} \big[f(x) + g(x) \big],$$

where $f, g : \mathbb{R}^n \to \overline{\mathbb{R}}$ and its Fenchel dual

$$(D_F) \qquad \sup_{q^* \in \mathbb{R}^m} \big\{ -f^*(q^*) - g^*(-q^*) \big\}.$$

It is proven that strong duality between these problems occurs, provided that $\mathrm{ri}(\mathrm{dom}(f)) \cap \mathrm{ri}(\mathrm{dom}(g)) \neq \emptyset$, also when f and g are almost convex functions, respectively they are nearly convex functions with the relative interiors of the epigraphs non - empty. These statements generalize the classical Fenchel duality theorem (see also [11] and [14]). Other results concerning conjugacy for almost convex and nearly convex functions are delivered, as well as an application in game theory.

3. The dual to the generalized primal geometric programming problem is proven to be obtainable also via perturbations. This approach provides the strong duality for this pair of problems under more general conditions than considered so far in the literature. When to the primal geometric programming problem

$$(P_g) \qquad \inf_{\substack{x=(x^0,x^1,\ldots,x^k)\in X_0\times X_1\times\ldots\times X_k,\\ g^i(x^i)\leq 0, i=1,\ldots,k,\ x\in N}} g^0(x^0),$$

where $X_i \subseteq \mathbb{R}^{l_i}, i = 0,\ldots,k$, $\sum_{i=0}^k l_i = n$, are convex sets, $g^i : X_i \to \mathbb{R}$, $i = 0,\ldots,k$, are functions convex on the sets they are defined on and $N \subseteq \mathbb{R}^n$ is a non - empty closed convex cone, one calculates the Fenchel - Lagrange dual problem, it turns out to coincide with the geometric dual problem

$$(D_g) \qquad \sup_{\substack{q_i^*\geq 0, i=1,\ldots,k,\\ t=(t^0,\ldots,t^k)\in N^*}} \left\{ -g^{0*}_{X_0}(t^0) - \sum_{i=1}^k \sup_{x^i\in X_i} \left[{t^i}^T x^i - q_i^* g^i(x^i) \right] \right\}.$$

The strong duality statement offered by the Fenchel - Lagrange duality is more general than the one stated in geometric programming duality, as the functions and sets involved are not required to be moreover lower semicontinuous, respectively closed, alongside the convexity assumptions imposed on them, and the constraint qualification treats more flexible the constraint functions that are restrictions to some sets of affine functions.

For seven problems treated in the literature via geometric programming duality, including the posynomial geometric problem - the starting point of geometric programming, the Fenchel - Lagrange dual problems are determined and strong duality and optimality conditions are delivered, underlining the advantages of this approach (see also [15]).

4. Consider the primal composite programming problem

$$(P_c) \qquad \inf_{\substack{x\in X,\\ g(x)\in -C}} f(F(x)),$$

where K and C are non - empty closed convex cones in $\mathbb{R}^k$ and $\mathbb{R}^m$, respectively, X is a non - empty convex subset of $\mathbb{R}^n$, $f : \mathbb{R}^k \to \overline{\mathbb{R}}$ is a K - increasing convex function, $F : X \to \mathbb{R}^k$ a function K - convex on X and $g : X \to \mathbb{R}^m$ a function C - convex on X. Moreover, it is imposed the feasibility condition $\mathcal{A} \cap F^{-1}(\mathrm{dom}(f)) \neq \emptyset$, where $\mathcal{A} = \{x \in X : g(x) \in -C\}$ is the feasible set of (P_c) (and also of (P)) and for any set $U \subseteq \mathbb{R}^k$, $F^{-1}(U) = \{x \in X : F(x) \in U\}$.

The Fenchel - Lagrange dual problem to (P_c) is

$$(D_c) \qquad \sup_{\substack{\alpha\in C^*,\beta\in K^*,\\ u\in\mathbb{R}^n}} \left\{ -f^*(\beta) - \left(\beta^T F\right)^*_X(u) - \left(\alpha^T g\right)^*_X(-u) \right\}.$$

Weak and strong duality and optimality conditions are delivered here, too. Using the unconstrained version of (P_c) the formula of the conjugate of $f \circ F$ regarding the set X is proven. The constraint qualification is proven to be weaker than what has been considered so far in the literature and the functions do not need to be taken moreover lower semicontinuous like there. As a special case the conjugate of $1/F$ regarding X is calculated, when $F : X \to \mathbb{R}$ is a strictly positive concave function on X, where X is a non - empty convex subset of $\mathbb{R}^n$ (see also [9]).

5. Let the convex multiobjective optimization problem

$$(P_v) \qquad \text{v-}\min_{\substack{x\in X,\\ g(x)\in -C}} F(x),$$

where $F = (F_1, \ldots, F_k)^T$, g, X, K and C are considered like above. Denoting by $\mathcal{S}$ some set containing K - strongly increasing convex functions $s : \mathbb{R}^k \to \mathbb{R}$, the following family of scalarized problems

$$(P_s) \qquad \inf_{\substack{x \in X, \\ g(x) \in -C}} s(F(x)),$$

where $s \in \mathcal{S}$, is attached to (P_v). Using the duality statements for (P_s), i.e. for (P_c), respectively, the following multiobjective dual problem is assigned to (P_v)

$$(D_v) \qquad \underset{(z,s,\alpha,\beta,u) \in \mathcal{B}}{\text{v-max}} z,$$

where

$$\begin{aligned} \mathcal{B} = \Big\{ & (z, s, \alpha, \beta, u) \in \mathbb{R}^k \times \mathcal{S} \times C^* \times K^* \times \mathbb{R}^n : \\ & s(z) \leq -s^*(\beta) - \left(\beta^T F\right)^*_X (u) - \left(\alpha^T g\right)^*_X (-u) \Big\}. \end{aligned}$$

Weak and strong duality are delivered for the pair of multiobjective dual problems. When the scalarization function $s \in \mathcal{S}$ is required to fulfill additional conditions other scalarizations widely - used in the literature occur as special cases. Here there are considered linear scalarization, maximum scalarization and norm scalarization (see also [10]).

6. Consider the non - empty convex set $X \subseteq \mathbb{R}^n$, the affine functions $f_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, k$, the concave functions $g_i : \mathbb{R}^n \to \overline{\mathbb{R}}$, $i = 1, \ldots, k$, and the functions $h_j : X \to \mathbb{R}$, $j = 1, \ldots, m$, convex on X. Assume that for $i = 1, \ldots, k$, $f_i(x) \geq 0$ and $0 < g_i(x) \neq +\infty$ when $x \in X$ such that $h(x) \leqq 0$, where $h = (h_1, \ldots, h_m)^T$. Denote further $f = (f_1, \ldots, f_k)^T$ and $g = (g_1, \ldots, g_k)^T$. As usual in entropy optimization the convention $0 \ln 0 = 0$ is used. Consider the convex optimization problem

$$(P_f) \qquad \inf_{\substack{x \in X, \\ h(x) \leqq 0}} \left[\sum_{i=1}^{k} f_i(x) \ln \left(\frac{f_i(x)}{g_i(x)} \right) \right].$$

This problem has as objective function an entropy - like sum of functions. The following dual problem is obtained for it

$$(D_f) \qquad \sup_{\substack{q^f \in \mathbb{R}^k, q^h \in \mathbb{R}^m_+, \\ q^g_i \geq e^{q^f_i - 1}, i=1,\ldots,k}} \inf_{x \in X} \left[\left(q^h\right)^T h(x) + \left(q^f\right)^T f(x) - \left(q^g\right)^T g(x) \right].$$

Weak and strong duality and optimality conditions are delivered here, too. When the objective function is specialized in order to become one of the most used entropy measures, namely the ones due to Shannon, Kullback and Leibler and, respectively, Burg, different results in entropy optimization are rediscovered as particular cases (see also [12]).

7. An application of maximum entropy optimization in text classification is presented, too, accompanied by an algorithm (see also [16]).

Index of notation

$\mathbb{N}$	the set of natural numbers
$\mathbb{Q}$	the set of rational numbers
$\mathbb{R}$	the set of real numbers
$\overline{\mathbb{R}}$	the extended set of real numbers
$\mathbb{R}^{m\times n}$	the set of $m \times n$ matrices with real entries
$\mathbb{R}^m_+$	the non - negative orthant of $\mathbb{R}^m$
$\leqq$	the partial ordering introduced on $\mathbb{R}^m$ by $\mathbb{R}^m_+$
K^*	the dual cone of the cone K
$\mathrm{int}(X)$	the interior of the set X
$\mathrm{ri}(X)$	the relative interior of the set X
$\mathrm{cl}(X)$	the closure of the set X
$\mathrm{bd}(X)$	the border of the set X
$\mathrm{aff}(X)$	the affine hull of the set X
$\lvert X\rvert$	is the cardinality of the set X
$x^T y$	the inner product of the vectors x and y
$\mathrm{dom}(f)$	the domain of the function f
$\mathrm{epi}(f)$	the epigraph of the function f
$\bar{f}$	the lower semicontinuous envelope of the function f
f^*	the conjugate of the function f
f^*_X	the conjugate of the function f regarding the set X
∂f	the subdifferential of the function f
A^*	the adjoint of the linear transformation A
δ_X	the indicator function of the set X
σ_X	the support function of the set X
$v(P)$	the optimal objective value of the optimization problem (P)
$v - min$	the notation for a multiobjective optimization problem in the sense of minimum
$v - max$	the notation for a multiobjective optimization problem in the sense of maximum

$\mathcal{S}$	some set of K - strongly increasing convex functions $s : \mathbb{R}^k \to \mathbb{R}$
$\mathcal{O}$	the set of the absolute norms $\gamma : \mathbb{R}^k \to \mathbb{R}$
B_γ	the unit ball corresponding to the norm γ
$\mathcal{U}^\perp$	the orthogonal subspace to the linear subspace U
$E(f)$	the expected value of the features function f
$p(d)$	the probability of the document d to be chosen from the set $\mathcal{D}$
$p(c\|d)$	the conditional probability of the class c with respect to the document d
$p(d,c)$	the joint probability of the document d and of the class c

Bibliography

[1] A. Aleman, *On some generalizations of convex sets and convex functions*, Mathematica - Revue d'Analyse Numérique et de la Théorie de l'Approximation **14 (1)**, 1–6, 1985.

[2] R. Baeza-Yates, B. Ribeiro-Neto, *Modern information retrieval*, Addison-Wesley Verlag, 1999.

[3] R. B. Bapat, T. E. S. Raghavan, *Nonnegative matrices and applications*, Cambridge University Press, Cambridge, 1997.

[4] A. Ben-Tal, M. Teboulle, A. Charnes, *The role of duality in optimization problems involving entropy functionals with applications to information theory*, Journal of Optimization Theory and Applications **58 (2)**, 209–223, 1988.

[5] C. Beoni, *A generalization of Fenchel duality theory*, Journal of Optimization Theory and Applications **49 (3)**, 375–387, 1986.

[6] A. L. Berger, S. A. Della Pietra, V. J. Della Pietra, *A maximum entropy approach to natural language processing*, Computational Linguistics **22 (1)**, 1–36, 1996.

[7] J. M. Bilbao, J. E. Martínez - Legaz, *Some applications of convex analysis to cooperative game theory*, Journal of Statistics & Management Systems **5 (1-3)**, 39–61, 2002.

[8] R. I. Boţ, *Duality and optimality in multiobjective optimization*, PhD Thesis, Technische Universität Chemnitz, 2003.

[9] R. I. Boţ, S.-M. Grad, G. Wanka, *A new constraint qualification and conjugate duality for composed convex optimization problems*, Preprint **2004-15**, Fakultät für Mathematik, Technische Universität Chemnitz, 2004.

[10] R. I. Boţ, S.-M. Grad, G. Wanka, *A new framework for studying duality in multiobjective optimization*, submitted to Mathematical Methods of Operations Research, 2006.

[11] R. I. Boţ, S.-M. Grad, G. Wanka, *Almost convex functions: conjugacy and duality*, Proceedings of the 8th International Symposium on Generalized Convexity and Generalized Monotonicity - Varese, Italy, 2005, to appear.

[12] R. I. Boţ, S.-M. Grad, G. Wanka, *Duality for optimization problems with entropy - like objective functions*, Journal of Information & Optimization Sciences **26 (2)**, 415–441, 2005.

[13] R. I. Boţ, S.-M. Grad, G. Wanka, *Entropy constrained programs and geometric duality obtained via Fenchel - Lagrange duality approach*, Nonlinear Analysis Forum **9 (1)**, 65–85, 2004.

[14] R. I. Boţ, S.-M. Grad, G. Wanka, *Fenchel's duality theorem for nearly convex functions*, Preprint **2005-13**, Fakultät für Mathematik, Technische Universität Chemnitz, 2005.

[15] R. I. Boţ, S.-M. Grad, G. Wanka, *Fenchel - Lagrange versus geometric programming in convex optimization*, Journal of Optimization Theory and Applications **129**, 2006.

[16] R. I. Boţ, S.-M. Grad, G. Wanka, *Maximum entropy optimization for text classification problems*, in: W. Habenicht, B. Scheubrein and R. Scheubein (eds.), "Multi-Criteria- und Fuzzy-Systeme in Theorie und Praxis", Deutscher Universitäts Verlag, Wiesbaden, 247–260, 2003.

[17] R. I. Boţ, G. Kassay, G. Wanka, *Strong duality for generalized convex optimization problems*, Journal of Optimization Theory and Applications **127 (1)**, 45–70, 2005.

[18] R. I. Boţ, G. Wanka, *An analysis of some dual problems in multiobjective optimization I*, Optimization **53 (3)**, 281–300, 2004.

[19] R. I. Boţ, G. Wanka, *An analysis of some dual problems in multiobjective optimization II*, Optimization **53 (3)**, 301–324, 2004.

[20] R. I. Boţ, G. Wanka, *Duality for composed convex functions with applications in location theory*, in: W. Habenicht, B. Scheubrein and R. Scheubein (eds.), "Multi-Criteria- und Fuzzy-Systeme in Theorie und Praxis", Deutscher Universitäts Verlag, Wiesbaden, 1–18, 2003.

[21] R. I. Boţ, G. Wanka, *Duality for multiobjective optimization problems with convex objective functions and D.C. constraints*, Journal of Mathematical Analysis and Applications **315 (2)**, 526–543, 2006.

[22] R. I. Boţ, G. Wanka, *Farkas - type results with conjugate functions*, SIAM Journal on Optimization **15 (2)**, 540–554, 2005.

[23] W. W. Breckner, G. Kassay, *A systematization of convexity concepts for sets and functions*, Journal of Convex Analysis **4 (1)**, 109–127, 1997.

[24] J. P. Burg, *Maximum entropy spectral analysis*, Proceedings of the 37-th Annual Meeting of the Society of Exploration Geophysicists, Oklahoma City, 1967.

[25] Y. Censor, A. R. De Pierro, A. N. Iusem, *Optimization of Burg's entropy over linear constraints*, Applied Numerical Mathematics **7 (2)**, 151–165, 1991.

[26] Y. Censor, A. Lent, *Optimization of "log x" entropy over linear equality constraints*, SIAM Journal on Control and Optimization **25 (4)**, 921–933, 1987.

[27] C. Combari, M. Laghdir, L. Thibault, *Sous-différentiels de fonctions convexes composées*, Annales des Sciences Mathématiques du Québec **18 (2)**, 119–148, 1994.

[28] J. N. Darroch, D. Ratcliff, *Generalized iterative scaling for log - linear models*, Annals of Mathematical Statistics **43**, 1470–1480, 1972.

[29] S. A. Della Pietra, V. J. Della Pietra, J. Lafferty, *Inducing features of random fields*, IEEE Transactions on Pattern Analysis and Machine Intelligence **19 (4)**, 1–13, 1997.

[30] R. J. Duffin, E. L. Peterson, C. Zener, *Geometric programming - theory and applications*, John Wiley and Sons, New York, 1967.

[31] I. Ekeland, R. Temam, *Convex analysis and variational problems*, North-Holland Publishing Company, Amsterdam, 1976.

[32] K. H. Elster, R. Reinhardt, M. Schäuble, G. Donath, *Einführung in die Nichtlineare Optimierung*, B. G. Teubner Verlag, Leipzig, 1977.

[33] S. Erlander, *Entropy in linear programs*, Mathematical Programming **21 (2)**, 137–151, 1981.

[34] S. C. Fang, J. R. Rajasekera, H.-S. J. Tsao, *Entropy optimization and mathematical programming*, Kluwer Academic Publishers, Boston, 1997.

[35] J. Fliege, A. Heseler, *Constructing approximations to the efficient set of convex quadratic multiobjective problems*, Ergebnisberichte Angewandte Mathematik **211**, Fachbereich Mathematik, Universität Dortmund, 2002.

[36] J. B. G. Frenk, G. Kassay, *Lagrangian duality and cone convexlike functions*, Journal of Optimization Theory and Applications, to appear.

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

[38] C. Gerstewitz, E. Iwanow, *Dualität für nichtkonvexe Vektoroptimierungsprobleme*, Workshop on Vector Optimization (Plauen, 1984), Wissenschaftliche Zeitschrift der Technischen Hochschule Ilmenau **31 (2)**, 61–81, 1985.

[39] A. Golan, V. Dose, *Tomographic reconstruction from noisy data*, in: "Bayesian inference and maximum entropy methods in science and engineering", AIP Conference Proceedings **617**, 248–258, 2002.

[40] A. Göpfert, C. Gerth, *Über die Skalarisierung und Dualisierung von Vektoroptimierungsproblemen*, Zeitschrift für Analysis und ihre Anwendungen **5 (4)**, 377–384, 1986.

[41] S. Guiaşu, A. Shenitzer, *The principle of maximum entropy*, The Mathematical Intelligencer **7 (1)**, 42–48, 1985.

[42] G. Hamel, *Eine Basis aller Zahlen und die unstetigen Lösungen der Funktionalgleichung: $f(x+y) = f(x) + f(y)$*, Mathematische Annalen **60 (3)**, 459–462, 1905.

[43] S. Helbig, *A scalarization for vector optimization problems in locally convex spaces*, Proceedings of the Annual Scientific Meeting of the GAMM (Vienna, 1988), Zeitschrift für Angewandte Mathematik und Mechanik **69 (4)**, T89–T91, 1989.

[44] J.-B. Hiriart-Urruty, C. Lemaréchal, *Convex analysis and minimization algorithms, I and II*, Springer Verlag, Berlin, 1993.

[45] J.-B. Hiriart-Urruty, J.-E. Martínez-Legaz, *New formulas for the Legendre-Fenchel transform*, Journal of Mathematical Analysis and Applications **288 (2)**, 544–555, 2003.

[46] J. Jahn, *Scalarization in vector optimization*, Mathematical Programming **29 (2)**, 203–218, 1984.

[47] J. Jahn, *Vector optimization - theory, applications, and extensions*, Springer Verlag, Berlin, 2004.

[48] T. R. Jefferson, C. H. Scott, *Avenues of geometric programming I. Theory*, New Zealand Operational Research **6 (2)**, 109–136, 1978.

[49] T. R. Jefferson, C. H. Scott, *Avenues of geometric programming II. Applications*, New Zealand Operational Research **7 (1)**, 51–68, 1978.

[50] T. R. Jefferson, C. H. Scott, *Quadratic geometric programming with application to machining economics*, Mathematical Programming **31 (2)**, 137–152, 1985.

[51] T. R. Jefferson, C. H. Scott, *Quality tolerancing and conjugate duality*, Annals of Operations Research **105**, 185–200, 2001.

[52] T. R. Jefferson, Y. P. Wang, C. H. Scott, *Composite geometric programming*, Journal of Optimization Theory and Applications, **64 (1)**, 101–118, 1990.

[53] D. Jurafsky, J. H. Martin, *Speech and language processing*, Prentice-Hall Inc., 2000.

[54] I. Kaliszewski, *Norm scalarization and proper efficiency in vector optimization*, Foundations of Control Engineering **11 (3)**, 117–131, 1986.

[55] P. Kanniappan, *Fenchel-Rockafellar type duality for a nonconvex nondifferential optimization problem*, Journal of Mathematical Analysis and Applications **97 (1)**, 266–276, 1983.

[56] J. N. Kapur, *Burg and Shannon's measures of entropy as limiting cases of two families of measures of entropy*, Proceedings of the National Academy of Sciences India **61A (3)**, 375–387, 1991.

[57] J. N. Kapur, H. K. Kesavan, *Entropy optimization principles with applications*, Academic Press Inc., San Diego, 1992.

[58] P. Q. Khánh, *Optimality conditions via norm scalarization in vector optimization*, SIAM Journal on Control and Optimization **31 (3)**, 646–658, 1993.

[59] S. Kullback, R. A. Leibler, *On information and sufficiency*, Annals of Mathematical Statistics **22**, 79–86, 1951.

[60] S. S. Kutateladže, *Changes of variables in the Young transformation*, Soviet Mathematics Doklady **18 (2)**, 1039–1041, 1977.

[61] B. Lemaire, *Application of a subdifferential of a convex composite functional to optimal control in variational inequalities*, in: Lecture Notes in Economics and Mathematical Systems **255**, Springer Verlag, Berlin, 103–117, 1985.

[62] V. L. Levin, *Sur le sous - différentiel de fonctions composées*, Doklady Akademia Nauk **194**, 28–29, 1970.

[63] P. Mbunga, *Structural stability of vector optimization problems*, Optimization and optimal control (Ulaanbaatar, 2002), Series on Computers and Operations Research **1**, World Scientific Publishing, River Edge, 175–183, 2003.

[64] E. Miglierina, E. Molho, *Scalarization and stability in vector optimization*, Journal of Optimization Theory and Applications **114 (3)**, 657–670, 2002.

[65] K. Mitani, H. Nakayama, *A multiobjective diet planning support system using the satisficing trade - off method*, Journal of Multi - Criteria Decision Analysis **6 (3)**, 131–139, 1997.

[66] K. Nigam, J. Lafferty, A. McCallum, *Using maximum entropy for text classification*, Proceedings of IJCAI-99 Workshop on Machine Learning for Information Filtering, 61–67, 1999.

[67] D. Noll, *Restoration of degraded images with maximum entropy*, Journal of Global Optimization **10 (1)**, 91–103, 1997.

[68] S. Paeck, *Convexlike and concavelike conditions in alternative, minimax, and minimization theorems*, Journal of Optimization Theory and Applications **74 (2)**, 317–332, 1992.

[69] S. Paeck, *Konvex- und konkav-ähnliche Bedingungen in der Spiel- und Optimierungstheorie*, Wissenschaft und Technik Verlag, Berlin, 1996.

[70] J. P. Penot, M. Volle, *On quasi-convex duality*, Mathematics of Operations Research **15 (4)**, 597–625, 1987.

[71] E.L. Peterson, *Geometric programming*, SIAM Review **18 (1)**, 1–51, 1976.

[72] R.T. Rockafellar, *Convex analysis*, Princeton University Press, Princeton, 1970.

[73] A. M. Rubinov, R. N. Gasimov, *Scalarization and nonlinear scalar duality for vector optimization with preferences that are not necessarily a pre-order relation*, Journal of Global Optimization **29 (4)**, 455–477, 2004.

[74] B. Schandl, K. Klamroth, M. M. Wiecek, *Norm - based approximation in multicriteria programming*, Global optimization, control, and games IV, Computers & Mathematics with Applications **44 (7)**, 925–942, 2002.

[75] C. H. Scott, T. R. Jefferson, *Conjugate duality in generalized fractional programming*, Journal of Optimization Theory and Applications **60 (3)**, 475–483, 1989.

[76] C. H. Scott, T. R. Jefferson, *Convex dual for quadratic concave fractional programs*, Journal of Optimization Theory and Applications **91 (1)**, 115–122, 1996.

[77] C. H. Scott, T. R. Jefferson, *Duality for a sum of convex ratios*, Optimization **40 (4)**, 303–312, 1997.

[78] C. H. Scott, T. R. Jefferson, *Duality for minmax programs*, Journal of Mathematical Analysis and Applications **100 (2)**, 385–393, 1984.

[79] C. H. Scott, T. R. Jefferson, *On duality for a class of quasiconcave multiplicative programs*, Journal of Optimization Theory and Applications **117 (3)**, 575–583, 2003.

[80] C. H. Scott, T. R. Jefferson, S. Jorjani, *Conjugate duality in facility location*, in: Z. Drezner (ed.), "Facility Location: A Survey of Applications and Methods", Springer Verlag, Berlin, 89–101, 1995.

[81] C. H. Scott, T. R. Jefferson, S. Jorjani, *Duality for programs involving non - separable objective functions*, International Journal of Systems Science, **19 (9)**, 1911–1918, 1988.

[82] C.H. Scott, T.R. Jefferson, S. Jorjani, *On duality for entropy constrained programs*, Information Sciences **35 (3)**, 211–221, 1985.

[83] C. E. Shannon, *A mathematical theory of communication*, Bell System Technical Journal **27**, 379–423 and 623–656, 1948.

[84] C. Tammer, *A variational principle and applications for vectorial control approximation problems*, Preprint **96-09**, Reports on Optimization and Stochastics, Martin-Luther-Universität Halle-Wittenberg, 1996.

[85] C. Tammer, A. Göpfert, *Theory of vector optimization*, in: "Multiple criteria optimization: state of the art annotated bibliographic surveys", International Series in Operations Research & Management Science **52**, Kluwer Academic Publishers, Boston, 1–70, 2002.

[86] C. Tammer, K. Winkler, *A new scalarization approach and applications in multicriteria d.c. optimization*, Journal of Nonlinear and Convex Analysis **4 (3)**, 365–380, 2003.

[87] T. Tanino, H. Kuk, *Nonlinear multiobjective programming*, in: "Multiple criteria optimization: state of the art annotated bibliographic surveys", International Series in Operations Research & Management Science **52**, Kluwer Academic Publishers, Boston, 71–128, 2002.

[88] M. Volle, *Duality principles for optimization problems dealing with the difference of vector-valued convex mappings*, Journal of Optimization Theory and Application **114 (1)**, 223–241, 2002.

[89] G. Wanka, R. I. Boţ, *A new duality approach for multiobjective convex optimization problems*, Journal of Nonlinear and Convex Analysis **3 (1)**, 41–57, 2001.

[90] G. Wanka, R. I. Boţ, *Multiobjective duality for convex - linear problems II*, Mathematical Methods of Operations Research **53 (3)**, 419–433, 2000.

[91] G. Wanka, R. I. Boţ, *Multiobjective duality for convex ratios*, Journal of Mathematical Analysis and Applications **275 (1)**, 354–368, 2002.

[92] G. Wanka, R. I. Boţ, *On the relations between different dual problems in convex mathematical programming*, in: P. Chamoni, R. Leisten, A. Martin, J. Minnemann and H. Stadtler (eds.), "Operations Research Proceedings 2001", Springer Verlag, Berlin, 255–262, 2002.

[93] G. Wanka, R. I. Boţ, S.-M. Grad, *Multiobjective duality for convex semidefinite programming problems*, Zeitschrift für Analysis und ihre Anwendungen (Journal for Analysis and its Applications) **22 (3)**, 711–728, 2003.

[94] P. Weidner, *An approach to different scalarizations in vector optimization*, Wissenschaftliche Zeitschrift der Technischen Hochschule Ilmenau **36 (3)**, 103–110, 1990.

[95] P. Weidner, *The influence of proper efficiency on optimal solutions of scalarizing problems in multicriteria optimization*, OR Spektrum **16 (4)**, 255–260, 1994.

[96] K. Winkler, *Skalarisierung mehrkriterieller Optimierungsprobleme mittels schiefer Normen*, in: W. Habenicht, B. Scheubrein and R. Scheubein (eds.), "Multi-Criteria- und Fuzzy-Systeme in Theorie und Praxis", Deutscher Universitäts Verlag, Wiesbaden, 173–190, 2003.

Lebenslauf

Persönliche Angaben

Name:	Sorin - Mihai Grad
Adresse:	Vettersstrasse 64/527 09126 Chemnitz
Geburtsdatum:	02.03.1979
Geburtsort:	Satu Mare, Rumänien

Schulausbildung

09/1985 - 06/1993	"Mircea Eliade" Schule in Satu Mare, Rumänien
09/1993 - 06/1997	"Mihai Eminescu" Kollegium in Satu Mare, Rumänien Abschluß: Abitur

Studium

10/1997 - 06/2001	"Babeş-Bolyai" - Universität Cluj-Napoca, Rumänien Fakultät für Mathematik und Informatik Fachbereich: Mathematik Abschluß: Diplom (Gesamtmittelnote: 9,33) Titel der Diplomarbeit: "Multiobjective duality for convex semidefinite programming problems"
10/2000 - 03/2001	Technische Universität Chemnitz: Teilzeitstudent Fakultät für Mathematik
10/2001 - 06/2002	"Babeş-Bolyai" - Universität Cluj-Napoca, Rumänien Fakultät für Mathematik und Informatik Masterstudium: Reelle und Komplexe Analysis Abschluß: Masterdiplom (Gesamtmittelnote: 9,36) Titel der Masterarbeit: "Duality for entropy programming problems and applications in text classification"
10/2001 - 09/2002	Technische Universität Chemnitz: Masterstudent Fakultät für Mathematik Internationaler Master- und Promotionsstudiengang
seit 10/2002	Technische Universität Chemnitz: Promotionsstudent Fakultät für Mathematik Internationaler Master- und Promotionsstudiengang

Erklärung gemäß §6 der Promotionsordnung

Hiermit erkläre ich an Eides Statt, dass ich die von mir eingereichte Arbeit "New insights into conjugate duality" selbstständig und nur unter Benutzung der in der Arbeit angegebenen Quellen und Hilfsmittel angefertigt habe.

Chemnitz, den 25.01.2006 Sorin - Mihai Grad