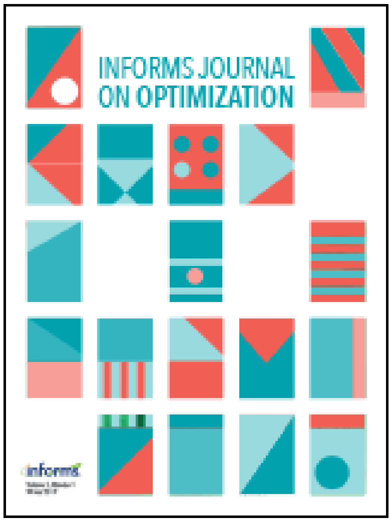


INFORMS Journal on Optimization

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

A Universal and Structured Way to Derive Dual Optimization Problem Formulations

Kees Roos, Marleen Balvert, Bram L. Gorissen, Dick den Hertog

informs
http://pubsonline.informs.org/journal/ijoo

INFORMS JOURNAL ON OPTIMIZATION
Vol. 2, No. 4, Fall 2020, pp. 229–255
ISSN 2575-1484 (print), ISSN 2575-1492 (online)

A Universal and Structured Way to Derive Dual Optimization Problem Formulations

Kees Roos,[a] Marleen Balvert,[b] Bram L. Gorissen,[c] Dick den Hertog[b]

[a] Department of Electrical Engineering, Mathematics and Computer Science, Technical University Delft, 2628 CD Delft, Netherlands; [b] Department of Econometrics and Operations Research, Tilburg University, 5000 LE Tilburg, Netherlands; [c] Department of Mathematics, Faculty of Exact Sciences, Free University, 1081 HV Amsterdam, Netherlands

Contact: c.roos@tudelft.nl, https://orcid.org/0000-0002-7963-9127 (KR); m.balvert@tilburguniversity.edu, https://orcid.org/0000-0002-2376-9301 (MB); bgorissen@broadinstitute.org, https://orcid.org/0000-0001-5992-0432 (BLG); d.denhertog@tilburguniversity.edu, https://orcid.org/0000-0002-1829-855X (DdenH)

Received: November 30, 2018
Revised: January 6, 2020
Accepted: January 10, 2020
Published Online in Articles in Advance: October 7, 2020

https://doi.org/10.1287/ijoo.2019.0034

Abstract. The dual problem of a convex optimization problem can be obtained in a relatively simple and structural way by using a well-known result in convex analysis, namely Fenchel's duality theorem. This alternative way of forming a strong dual problem is the subject of this paper. We recall some standard results from convex analysis and then discuss how the dual problem can be written in terms of the conjugates of the objective function and the constraint functions. This is a didactically valuable method to explicitly write the dual problem. We demonstrate the method by deriving dual problems for several classical problems and also for a practical model for radiotherapy treatment planning, for which deriving the dual problem using other methods is a more tedious task. Additional material is presented in the appendices, including useful tables for finding conjugate functions of many functions.

Keywords: convex optimization • duality theory • conjugate functions • support functions • Fenchel duality

1. Introduction

Nowadays, optimization is one of the most widely taught and applied subjects of mathematics. A mathematical optimization problem consists of an objective function that has to be minimized (or maximized) subject to some constraints. Every maximization problem can be written as a minimization problem by replacing the objective function by its opposite. In this paper, we always assume that the given problem is a minimization problem. To every such problem, one can assign another problem, called the *dual problem* of the original problem; the original problem is called the *primal problem*. In many cases, a dual problem exists that has the same optimal value as the primal problem. If this happens, then we say that we have *strong duality*, and otherwise, *weak duality*. It is well known that if the primal problem is convex, then there exists a strong dual problem under some mild conditions. In this paper, we consider only convex problems with a strong dual. Strong duality and obtaining explicit formulations of the dual problem are of great importance for at least four reasons:

1. Sometimes the dual problem is easier to solve than the primal problem.
2. If we have feasible solutions of the primal problem and the dual problem, respectively, and their objective values are equal, then we may conclude that both solutions are optimal. In that case, the primal solution is a certificate for the optimality of the dual solution, and vice versa.
3. If the two objective values are not equal, then the dual objective value is a lower bound for the optimal value of the primal problem, and the primal objective value is an upper bound for the optimal value of the dual problem. This provides information on the "quality" of the two solutions, that is, on how much their values may differ from the optimal value.
4. Some optimization problems benefit from being solved with the dual decomposition method. This method requires repeated evaluation of the dual objective value of two (smaller) problems for fixed dual variables. Our explicit formulation of the dual can potentially speed up the evaluation of the dual objective function.

In textbooks, different ways for deriving dual optimization problems are discussed. One option is to write the optimization problem as a conic optimization problem and then state the dual optimization problem in terms of the dual cone. However, rewriting general convex constraints as conic constraints may be awkward or even not possible. Another option is to derive the dual via Lagrange duality, which often leads to an unnecessarily long and tedious derivation process. We propose a universal and more structured way to derive dual problems, using basic results from convex analysis. The duals derived in this way are ordinary Lagrange duals. The value of this paper is therefore not to present new duality theory but rather to provide a unifying, pedagogically relevant framework for deriving the dual.

Let the primal optimization problem be given. Then, by using Fenchel's duality theorem, one can derive a dual problem in terms of so-called perspectives of the conjugates of the objective function and constraint functions. The main task is therefore to derive expressions for the conjugates of each of these functions separately (taking the perspective functions is easy). At first, this may seem to be a difficult task—and this may be the reason that in textbooks this is not further elaborated on. However, the key issue underlying this paper is that it is not necessary to derive "explicit" expressions for the conjugates of these functions: if one can derive an expression for the conjugate as an "inf over some variables," then this is sufficient to state the dual problem. And, indeed, by using standard composition rules for conjugates—well-known results in convex analysis—we arrive at such expressions with an "inf."

Therefore, the aim of this paper is to show that the dual problem can be derived by using the composition rules for conjugates and by using the well-known results for the conjugates of basic univariate and multivariate functions. Thus, the proposed method is similar in nature to the way Fourier and Laplace transforms are taught, where a table with known transforms of functions is combined with a table with identities for transformations of functions. We consider this the most structured way to teach deriving dual optimization problems. Using this structured approach, one might even computerize the derivation of the dual optimization problem. We must emphasize that the ingredients in this approach are not new. Approximately the same dual is presented in Boyd and Vandenberghe (2004, p. 256), where it is obtained via Lagrange duality, and in Hiriart-Urruty (2006, p. 45). We also mention Tomioka (2010), where three strategies for deriving a dual problem are discussed, and Fenchel's duality theorem is used to construct the dual problem of a specific second-order cone problem. As far as we know, no comprehensive survey as presented in this paper exists.

We should also mention that in special cases where one would like to derive a partial Lagrange dual problem (e.g., for relaxation purposes), Fenchel's duality formulation cannot be used. By contrast, the use of Fenchel's dual has found application in recent developments in the field of robust optimization (see, e.g., Ben-Tal et al. 2015, Gorissen and den Hertog 2015).

The outline of the paper is as follows. In Section 2, we recall some fundamental notions from convex analysis, as defined in Rockafellar (1970). In Section 3, we present the general form of the primal problem that we want to solve and Fenchel's dual problem. We also discuss under which condition strong duality is guaranteed. To make the paper self-supporting, we present a proof of this strong duality in Appendices C and D. In Section 4, we present a scheme that describes in four steps how to get Fenchel's dual problem. In this scheme, tables of conjugate functions and support functions in Appendix E are of crucial importance. In Appendix A, we demonstrate how conjugates for some functions are computed.

Use of the scheme in Section 4 is illustrated in Section 6, where we show how dual problems for several classical convex optimization problems can be obtained. In these classical problems, the functions that appear in the primal problem all have a similar structure. However, problems with a variety of constraint functions can also be handled systematically. In Section 6.7, we demonstrate this by applying our approach to a model for radiotherapy treatment.

2. Preliminaries

We recall in this section some notation and terminology that are common in the literature. The *effective domain* of a convex function $f : \mathbf{R}^n \to [-\infty, +\infty]$ is defined by

$$\operatorname{dom} f = \{x \mid f(x) < \infty\}.$$

A convex function f is *proper* if $f(x) > -\infty$ for all $x \in \mathbf{R}^n$ and $f(x) < \infty$ for at least one $x \in \mathbf{R}^n$. The function f is *closed* if the set $\{x \mid f(x) \le \alpha\}$ is closed for every $\alpha \in \mathbf{R}$.

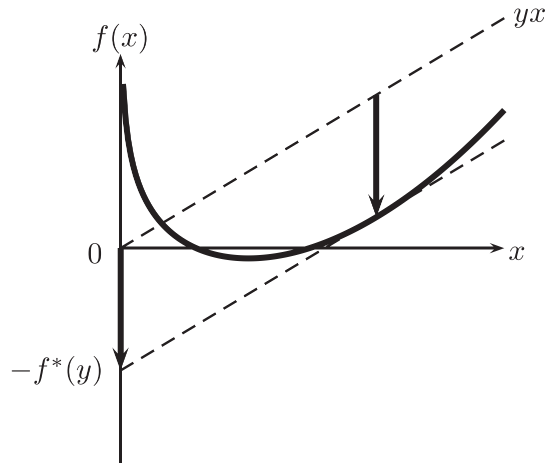
For any function f, its convex conjugate f^* is defined by

$$f^*(y) := \sup_{x \in \operatorname{dom} f} \{y^T x - f(x)\}. \tag{1}$$

Figure 1 illustrates this notion graphically for the case where $n = 1$.

Given y, the maximal value of $yx - f(x)$ occurs where the derivative of $f(x)$ equals y, and the tangent of the graph of f where this happens intersects the y-axis at the point $(0, -f^*(y))$.

Figure 1. Value of Convex Conjugate at y



Because $f^*(y)$ is the pointwise maximum of linear functions (linear in y), f^* is convex. Moreover, if f is convex, then f^* is closed, $(\mathrm{cl} f)^* = f^*$, and $f^{**} = \mathrm{cl} f$, which implies that $f^{**} = f$ if f is closed convex (see Rockafellar 2010, theorem 12.2).

The *indicator function* of a set $S \subseteq \mathbf{R}^n$ is denoted as $\delta_S(x)$ and defined by

$$\delta_S(x) = \begin{cases} 0, & x \in S, \\ \infty, & x \notin S. \end{cases}$$

Note that $\delta_S(x)$ is convex if and only if S is convex and proper if S is nonempty. Obviously, we have

$$x \in S \quad \Leftrightarrow \quad \delta_S(x) \leq 0.$$

This implies that any convex *set constraint* can be replaced with a convex *functional constraint*, and vice versa.

The conjugate function of $\delta_S(x)$ is called the *support function* of S and is given by

$$\delta_S^*(y) = \sup_{x \in S} \left\{ y^T x \right\}. \tag{2}$$

Our approach highly depends on a simple relation between the conjugate f^* of a convex function f and the support function of the set $S := \{x \mid f(x) \leq 0\}$. Provided that S is nonempty, and $\mathrm{ri}\, S$ is nonempty if f is nonlinear, we have (see Lemma D.1)

$$\delta_S^*(y) = \min_{\lambda \geq 0} \{ (\lambda f)^*(y) \}. \tag{3}$$

Here the function λf is defined in the natural way: $(\lambda f)(x) = \lambda f(x)$. If $\lambda > 0$, then one has

$$(\lambda f)^*(y) = \sup_x \{ y^T x - \lambda f(x) \} = \sup_x \left\{ \lambda \left(x^T \frac{y}{\lambda} - f(x) \right) \right\} = \lambda f^* \left(\frac{y}{\lambda} \right), \tag{4}$$

and for $\lambda = 0$, we get

$$(0f)^*(y) = \sup_x \; \{ y^T x \} = \begin{cases} 0, & \text{if } y = 0, \\ \infty, & \text{otherwise} \end{cases} = \delta_{\{0\}}(y). \tag{5}$$

The function $(\lambda f)^*$ is called the *perspective function* of f^* (with λ as the new variable). Because of the first equality in (4), the perspective function of f^* is convex (more precisely, jointly convex in y and λ). Although $(\lambda f)^*$ is closed for each fixed value of λ, it is not necessarily closed for λ. It may therefore occur that $(\lambda f)^*(y)$ is well defined even if λ approaches zero, whereas y stays away from zero. Closing the perspective function for λ

can therefore result in the wrong dual. An example of this phenomenon can be found in Section B.9 of Appendix B.

The concave conjugate g_* of a function g is defined by

$$g_*(y) := \inf_{x \in \operatorname{dom} g} \left\{ y^T x - g(x) \right\}.$$

Because $g_*(y)$ is the pointwise minimum of linear functions, g_* is concave. Putting $f = -g$, one easily verifies the following relation (see Rockafellar 1970, p. 308):

$$g_*(y) = -f^*(-y) = -(-g)^*(-y). \tag{6}$$

Hence, all properties of convex conjugates lead to similar properties of concave conjugates.

3. Fenchel's Dual Problem of a Convex Optimization Problem

In this section, we derive Fenchel's dual problem for a convex optimization problem by using conjugate functions. We will use some properties of conjugate functions without proving them at this stage; this will be the subject in the following sections and in the appendices. So, for the moment, we take these properties for granted.

Let us first consider the case of *unconstrained* minimization. Here we have a convex function f_0 and want to know its minimal value z^*. Because of definition (1) of f_0^*, we may write

$$z^* = \inf_{x \in \mathbf{R}} \{f_0(x)\} = -\sup_x \{-f_0(x)\} = -f_0^*(0). \tag{7}$$

This already reveals the relevance of the notion of a conjugate function for this relatively simple optimization problem: if we know the conjugate function $f_0^*(y)$ of f_0, then $-f_0^*(0)$ gives the minimal value. As we will see, this simple property is the keystone of Fenchel's duality theory.

Next, we consider the case where S is a convex subset of $\mathbf{R}$, and we want to find the minimal value of f_0 on the set S. So the problem we want to solve is

$$\inf_x \{f_0(x) \mid x \in S\}.$$

In this case, we consider the function $g(x)$ defined by

$$g(x) := f_0(x) + \delta_S(x).$$

Obviously, the effective domain of $g(x)$ is the set S, and on this set, g coincides with f_0. So we have transformed our *constrained* problem to an unconstrained problem, and therefore the optimal value will be $-g^*(0)$. The question arises as to how to find the conjugate function of g. It turns out that it can be derived easily from the conjugate functions of f_0 and δ_S. Using the formula in line 6 of Table E.2, we obtain

$$g^*(y) = \min_{y^0, y^1} \left\{ f_0^*(y^0) + \delta_S^*(y^1) \mid y^0 + y^1 = y \right\}.$$

We need to compute $g^*(y)$ at $y = 0$. Substitution of $y = 0$ in the preceding expression yields $y^0 + y^1 = 0$; that is, $y^1 = -y^0$. So we can eliminate y^1. Thus, we get

$$g^*(0) = \min_{y^0} \{ f_0^*(y^0) + \delta_S^*(-y^0) \}.$$

In general, the problem that we want to solve is the following (primal) convex optimization problem:

$$(P) \qquad \inf_x \{ f_0(x) \mid f_i(x) \le 0,\ i = 1, \ldots, m \},$$

where the functions $f_i : \mathbf{R}^n \to [-\infty, \infty]$ are proper convex functions, for $i = 0, \ldots, m$. In this case, the set S is the intersection of the sets

$$S_i = \{x \mid f_i(x) \le 0\}, \quad 1 \le i \le m.$$

Obviously, the indicator function of S satisfies

$$\delta_S(x) = \delta_{S_1}(x) + \ldots + \delta_{S_m}(x),$$

and hence, by line 6 in Table E.3,

$$\delta_S^*(-y^0) = \min_{y^1,\ldots,y^m}\left\{\delta_{S_1}^*(y^1) + \ldots + \delta_{S_m}^*(y^m) \mid y^1 + \ldots + y^m = -y^0\right\}.$$

By (3), we have

$$\delta_{S_i}^*(y^i) = \min_{u_i\geq 0}\{(u_i f_i)^*(y^i)\}, \quad 1 \leq i \leq m.$$

Substitution gives

$$\begin{aligned}\delta_S^*(-y^0) &= \min_{\{y^i\}_{i=1}^m}\left\{\sum_{i=1}^m \min_{u_i\geq 0}\{(u_i f_i)^*(y^i)\} \,\middle|\, \sum_{i=0}^m y^i = 0\right\}\\ &= \min_{\{y^i\}_{i=1}^m,\, u\geq 0}\left\{\sum_{i=1}^m (u_i f_i)^*(y^i) \,\middle|\, \sum_{i=0}^m y^i = 0\right\}.\end{aligned}$$

Substitution into the expression for $g^*(0)$ gives

$$g^*(0) = \min_{(y^i)_{i=0}^m,\, u\geq 0}\left\{f^*(y^0) + \sum_{i=1}^m (u_i f_i)^*(y^i) \,\middle|\, \sum_{i=0}^m y^i = 0\right\}.$$

Changing the sign on both sides implies that

$$-g^*(0) = \max_{(y^i)_{i=0}^m,\, u\geq 0}\left\{-f^*(y^0) - \sum_{i=1}^m (u_i f_i)^*(y^i) \,\middle|\, \sum_{i=0}^m y^i = 0\right\}.$$

Thus, we may conclude that the optimal value of (P) is the same as that of the problem

$$(D) \qquad \sup_{\{y^i\}_{i=0}^m,\, u}\left\{-f_0^*(y^0) - \sum_{i=1}^m (u_i f_i)^*(y^i) \,\middle|\, \sum_{i=0}^m y^i = 0,\ u \geq 0\right\}.$$

This is Fenchel's dual problem of (P).[2]

Note that the objective function in (D) is concave (because the perspective functions under the sum are convex), and the constraints are linear (and so convex as well). Hence, when writing (D) as a minimization problem, it becomes a convex problem.

It is clear that (D) is a practical formulation of the dual if the convex conjugates f_i^* can be easily computed, which turns out to be true in many cases. This important fact is the main motivation for this paper.

In what follows, we make the following assumption to satisfy the condition in line 6 of Table E.2.

Assumption 3.1. *Assume that* $\mathrm{ri.dom} f_0 \cap S \neq \emptyset$ *or* $\cap_{i=0}^m \mathrm{ri.dom} f_i \neq \emptyset$.

This is the same as requiring that (P) is *Slater regular* (see Slater 2014). In this case, the optimal values of (P) and (D) are equal; for a proof, see Appendix D. We should add that if f_0 is linear, then $x \in \mathit{ri.dom} f_0$ can be replaced by $x \in \mathrm{dom} f_0$ in the Slater condition.

In what follows, especially when the objective function has the same form as the constraint functions, we extend the vector u with the entry $u_0 = 1$. Then (D) can be written as

$$(D') \qquad \sup_{\{y^i\}_{i=0}^m,\, u}\left\{-\sum_{i=0}^m (u_i f_i)^*(y^i) \mid \sum_{i=0}^m y^i = 0,\ u \geq 0,\ u_0 = 1\right\}.$$

In the preceding expressions for (D) and (D'), we used the "sup" operator to indicate that we are dealing with a maximization problem. In what follows, we use the "max" operator only if the maximum value is attained. In this paper, it is not our goal to establish when this happens; our main focus is to find a dual problem yielding strong duality. Therefore, if we use the "sup" operator in a maximization problem later, it is not excluded that the maximal value is attained, and similarly for the "inf" operator.

4. Recipe for Setting Up Fenchel's Dual Problem

As became clear in the preceding section, Fenchel's dual gives us a straightforward method for finding the dual problem of (P). The method can be split into four steps, as discussed next. We will provide some demonstrations of the resulting recipe in Section 6.

Step 1: Formulate the Primal Problem. The primal problem should be a convex optimization problem. Moreover, all set constraints should be replaced by their indicator functions. Strong duality, that is, equality of the optimal values of (P) and (D), is guaranteed only if (P) is Slater regular.

Step 2: Derive Tractable Expressions for the Conjugates. In our approach, we need the conjugates of the functions that occur in the problem. For this purpose, we can use the three tables in Appendix E. Table E.1 contains expressions for the conjugates of some basic functions and the domains of these functions, Table E.2 provides some formulas for conjugates of function transformations, and Table E.3 shows some support functions of common sets. An overview of commonly used cones and their dual cones is given in Table E.4. Usually a conjugate function (or its perspective function) occurring in (D) or (D') is written as $\inf_y\{h(y,z)\}$, where $h(y,z)$ is jointly convex in y and z. A frequently used tool in our approach is the obvious equality

$$\sup_z\left\{-\inf_y h(y,z)\right\} = \sup_{y,z}\{-h(y,z)\}.$$

In other words, when writing down (D) or (D'), we can get rid of "inf" operators by moving the concerning variables under the "sup" operator. This helps to simplify the formulation of the dual problem. Deriving tractable expressions for the conjugates is further elaborated on and illustrated in Appendix A. The domain of a conjugate function is also important. For each i, the vector y^i belongs to the domain of $(u_i f_i)^*$ in (D) or (D'). This not only leads to convex constraints in Fenchel's dual problem, but in most cases it also enables the elimination of the variables y^i. In some cases, we get a conjugate function that consists of "branches"; that is, one obtains different formulas for $f^*(y)$ on several parts of the domain of f^*. This phenomenon is demonstrated in Section B.7 of Appendix B, where it is also shown how this can be prevented.

Step 3: Derive Tractable Convex Expressions for Perspectives of Conjugates. The conjugate functions found in Step 2 are substituted into the term $(uf)^*$ in (D) or (D'). For $u > 0$, this term is

$$uf^*\left(\frac{y}{u}\right), \text{ with } \frac{y}{u} \in \operatorname{dom} f^*.$$

One may be troubled by the fact that the quotient y/u is not jointly convex in y and u. Indeed, sometimes the resulting formulation of the dual problem may not be convex, even though the set $\{(y,u) : u > 0, y/u \in \operatorname{dom} f^*\}$ is convex. A simple substitution yields a convex formulation. For example, when using line 8b in Table E.2, one has

$$f^*(y) = \inf_z\left\{h^*(z) - b^T z \mid A^T z = y\right\}.$$

This gives rise to the term $(uf)^*(y)$ in the dual problem; that is,

$$\begin{aligned}(uf)^*(y) &= \inf_z \left\{uh^*(z) - ub^T z \mid A^T z = \frac{y}{u}\right\}, && u > 0,\\ &= \inf_z \left\{uh^*(z) - b^T(uz) \mid A^T(uz) = y\right\}, && u > 0.\end{aligned}$$

As a result of the introduction of the variable u, we are left with an expression that is not jointly convex in y and u because of the occurrence of the products $uh^*(z)$ and uz. However, by the substitution $\tilde{z} = uz$, we get

$$\begin{aligned}(uf)^*(y) &= \inf_{\tilde{z}} \left\{uh^*\left(\frac{\tilde{z}}{u}\right) - b^T\tilde{z} \mid A^T\tilde{z} = y\right\}\\ &= \inf_{\tilde{z}} \left\{(uh)^*(\tilde{z}) - b^T\tilde{z} \mid A^T\tilde{z} = y\right\}.\end{aligned}$$

Obviously, both the objective function and the constraint are now convex. Section 6 shows many substitutions of this type, which are necessary to obtain a convex (and hence computationally tractable) formulation of the dual problem.

Step 4: Formulate the Dual Problem. The results of Steps 1–3 are used in the dual formulation (D) or (D'). Often one can simplify the resulting formulation by eliminating variables.

5. Sensitivity Analysis

The existing theory for sensitivity analysis is based on Lagrange multipliers (see, e.g., Boyd and Vandenberghe 2004, section 5.6.3). We therefore show that the variables u in (D) are precisely the Lagrange multipliers.

The classical approach to sensitivity theory investigates how the objective value of the dual problem depends on the right-hand sides in the primal problem. We first recall the classical approach to sensitivity theory. The Lagrange dual of (P) is given by

$$(LD) \qquad \sup_{\lambda \geq 0} \inf_x \left\{ f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) \right\}.$$

Let $(x^\diamond, \lambda^\diamond)$ be an optimal solution. If the ith constraint $f_i(x) \leq 0$ is replaced by $f_i(x) + \varepsilon_i \leq 0$, with $\varepsilon_i \in \mathbf{R}$, then the optimal value $z^\diamond$ becomes

$$z^\diamond = f_0(x^\diamond) + \sum_{i=1}^m \lambda_i^\diamond \big(f_i(x^\diamond) + \varepsilon_i\big) = f_0(x^\diamond) + \sum_{i=1}^m \lambda_i^\diamond f_i(x^\diamond) + \sum_{i=1}^m \lambda_i^\diamond \varepsilon_i.$$

Because of this expression, the Lagrange multiplier λ_i is called the *shadow price* for the ith constraint. This is because one usually concludes that adding a small number ε_i to $f_i(x)$ will change the objective value with the amount $\varepsilon_i \lambda_i$. Though this might be true in many cases, it does not hold in general, especially not if either the primal or the dual problem is degenerate. For examples, we refer to Jansen et al. (1997).

We now show that the variables u_i in Fenchel's dual problem are also shadow prices in the preceding sense. Recall that Fenchel's dual of (P) is given by

$$(D) \qquad \sup_{y^i, u \geq 0} \left\{ -f_0^*(y^0) - \sum_{i=1}^m (u_i f_i)^*(y^i) \mid \sum_{i=0}^m y^i = 0 \right\}.$$

Now $(u_i(f_i + \varepsilon_i))^*(y^i) = (u_i f_i)^*(y^i) - u_i \varepsilon_i$, by line 1 in Table E.2. Hence, at optimality, the part of the objective value in the preceding problem that is due to the perturbation in the primal problem is just $\sum_{i=0}^m u_i^\diamond \varepsilon_i$, where $u^\diamond$ is optimal for (D). This proves that in sensitivity analysis, u takes over the role of the vector λ of Lagrange multipliers.

6. Some Applications

In this section, we demonstrate how Fenchel's dual can be obtained for some well-known convex optimization problems and a mathematical model for radiotherapy treatment planning.

6.1. Linear Optimization

We consider the standard linear optimization problem

$$\inf_x \{ c^T x \mid Ax = b,\ x \in \mathbf{R}^n_+ \},$$

where A is an $m \times n$ matrix, and $b \in \mathbf{R}^m$, $c \in \mathbf{R}^n$. For the conjugate f^* of the objective function $f(x) = c^T x$, we have (see Table E.1, line 1)

$$f^*(y) = 0, \quad \operatorname{dom} f^* = \{c\}.$$

Denoting the support function of the set $\{x | Ax = b\}$ as $g_1^*(y)$, we have (see Table E.3, line 1)

$$g_1^*(y) = \min_z \{ b^T z \mid A^T z = y \}.$$

Denoting the support function of the set constraint $x \in \mathbf{R}^n_+$ as $g_2^*(y)$, we have (see Table E.3, line 2)

$$g_2^*(y) = 0, \quad \operatorname{dom} g_2^* = \{ y \mid y \leq 0 \} = -\mathbf{R}^n_+.$$

It follows that Fenchel's dual problem is given by

$$\sup\left\{-f^*(y^0) - \sum_{i=1}^{2} g_i^*(y^i) \,\middle|\, \sum_{i=0}^{2} y^i = 0\right\}.$$

Substitution of the expressions for f^* and g_i^* $(i = 1, 2)$ yields the following formulation of the dual problem:

$$\sup_{z, y^i}\left\{-b^T z \mid y^0 + y^1 + y^2 = 0,\ y^0 = c,\ A^T z = y^1,\ y^2 \le 0\right\}.$$

We can eliminate the vectors y^i, which gives

$$\sup_z\{-b^T z \mid c + A^T z \ge 0\}.$$

Changing the sign of z leads to the well-known duality theorem for linear optimization, namely

$$\inf_x\{c^T x \mid Ax = b,\ x \ge 0\} = \sup_z\{b^T z \mid A^T z \le c\}.$$

6.2. Conic Optimization

Let $\mathcal{K}$ be a convex cone in $\mathbf{R}^n$. We consider the standard linear optimization problem over the cone $\mathcal{K}$:

$$\inf_x\{c^T x \mid Ax = b,\ x \in \mathcal{K}\},$$

with A, b, and c as in Section 6.1. This problem differs from the linear optimization problem in Section 6.1 only in the set constraint $x \in \mathcal{K}$, so it suffices to find the support function of this constraint. Denoting this function as g_2^*, we have (see Table E.3, line 3)

$$g_2^*(y) = 0, \quad \operatorname{dom} g_2^* = -\mathcal{K}_*.$$

Just as in the preceding section, we obtain that Fenchel's dual problem is given by

$$\sup_z\{-b^T z \mid A^T z + c \in \mathcal{K}_*\}.$$

Replacing z by $-z$, we get the usual form of the duality theorem for conic optimization, namely

$$\inf_x\{c^T x \mid Ax = b,\ x \in \mathcal{K}\} = \sup_z\{b^T z \mid c - A^T z \in \mathcal{K}_*\}.$$

This formulation is valid for any cone, including the positive semidefinite cone and the copositive cone. Assumption 3.1 simplifies to the typical Slater condition: there should be an x in the interior of $\mathcal{K}$ that satisfies $Ax = b$.

6.3. Quadratically Constrained Quadratic Optimization

Consider the following quadratically constrained quadratic optimization (QCQO) problem:

$$\inf_x\left\{\frac{1}{2}x^T P_0 x + q_0^T x + r_0 \mid \frac{1}{2}x^T P_i x + q_i^T x + r_i \le 0, \quad i = 1, \ldots, m\right\},$$

where P_i is positive semidefinite for $i = 0, \ldots, m$. Defining $f_i(x) := g_i(x) + h_i(x)$, where $g_i(x) = \frac{1}{2}x^T P_i x$ and $h_i(x) = q_i^T x + r_i$, we have (see Table E.1, lines 1 and 13)

$$\begin{aligned} g_i^*(y) &= \frac{1}{2}y^T P_i^{\dagger} y, \quad \operatorname{dom} g_i^* = \{y \mid y = P_i z,\ z \in \mathbf{R}^n\}, \\ h_i^*(x) &= -r_i,\ \operatorname{dom} h_i^* = \{q_i\}. \end{aligned}$$

Hence, by using the sum rule in Table E.2 (line 6), we get

$$\begin{aligned} f_i^*(y) &= \min_{y^1,y^2}\{g_i^*(y^1) + h_i^*(y^2) \mid y^1 + y^2 = y\} \\ &= \min_{y^1,y^2}\left\{\left\{\frac{1}{2}{y^1}^T P_i^\dagger y^1 \mid y^1 = P_i z\right\} + \{-r_i \mid y^2 = q_i\} \mid y^1 + y^2 = y\right\} \\ &= \min_z\left\{\frac{1}{2}z^T P_i P_i^\dagger P_i z - r_i \mid P_i z + q_i = y\right\} \\ &= \min_z\left\{\frac{1}{2}z^T P_i z - r_i \mid P_i z + q_i = y\right\}, \end{aligned}$$

where we used that the generalized inverse $P_i^\dagger$ of P_i satisfies $P_i P_i^\dagger P_i = P_i$ (see, e.g., Ben-Israel and Greville 2003). The dual problem (D') thus becomes

$$\begin{aligned} &\sup_{u\ge 0, y^i}\left\{-\sum_{i=0}^m (u_i f_i)^*(y^i) \,\middle|\, \sum_{i=0}^m y^i = 0,\ u_0 = 1\right\} \\ &= \sup_{u\ge 0, y^i}\left\{-\sum_{i=0}^m u_i \min_{z^i}\left\{\frac{1}{2}(z^i)^T P_i z^i - r_i \,\middle|\, P_i z^i + q_i = \frac{y^i}{u_i}\right\} \,\middle|\, \sum_{i=0}^m y^i = 0,\ u_0 = 1\right\} \\ &= \sup_{u, z^i}\left\{\sum_{i=0}^m u_i\left(r_i - \frac{1}{2}(z^i)^T P_i z^i\right) \,\middle|\, \sum_{i=0}^m u_i(P_i z^i + q_i) = 0,\ u \ge 0,\ u_0 = 1\right\}. \end{aligned}$$

Defining $\tilde{z}^i = u_i z^i$, this becomes

$$\sup_{u,\tilde{z}^i}\left\{u^T r - \frac{1}{2}\sum_{i=0}^m \frac{(\tilde{z}^i)^T P_i \tilde{z}^i}{u_i} \,\middle|\, \sum_{i=0}^m u_i q_i + \sum_{i=0}^m P_i \tilde{z}^i = 0,\ u \ge 0,\ u_0 = 1\right\}.$$

It may be noted that by using standard tricks, the dual problem can be written as a conic quadratic problem.

6.4. Second-Order Cone Optimization

Consider the following primal second-order cone optimization (SOCO) problem:

$$\min\{\|A_0 x - b_0\|_2 - p_0^T x + q_0 \mid \|A_i x - b_i\|_2 - p_i^T x + q_i \le 0,\ i = 1, \ldots, m\}.$$

The objective function and the constraint functions are written as $f_i(x) = g_i(x) + h_i(x)$, for $i = 0, \ldots, m$, where $g_i(x) = \|A_i x - b_i\|_2$ and $h_i(x) = -p_i^T x + q_i$. This enables us to find the conjugate of $f_i(x)$ by using the sum rule. The conjugate of $h_i(x)$ is given by

$$h_i^*(y) = -q_i, \quad \operatorname{dom} h_i^* = \{-p_i\}.$$

In order to derive the conjugate of $g_i(x)$, first note that this is a function of a linear transformation of x; that is, $g_i(x) = \sigma_i(A_i x - b_i)$, where $\sigma_i(x) = \|x\|_2$. Because of line 6 in Table E.1, we have $\sigma_i^*(y) = 0$, with $\operatorname{dom}\sigma^* = \{y \mid \|y\|_2 \le 1\}$. Hence, by the linear substitution rule in Table E.2 (line 8b), the conjugate of $g_i(x)$ is given by

$$g_i^*(y) = \inf_{z^i}\{\sigma_i^*(z^i) + b_i^T z^i \mid A_i^T z^i = y\} = \inf_{z^i}\left\{b_i^T z^i \mid \|z^i\|_2 \le 1,\ A_i^T z^i = y\right\}.$$

Putting the preceding results together, we obtain

$$\begin{aligned} f_i^*(y) &= \inf_{y^1,y^2}\{g_i^*(y^1) + h_i^*(y^2) \mid y^1 + y^2 = y\} \\ &= \inf_{y^1,y^2}\left\{\inf_{z^i}\{b_i^T z^i \mid \|z^i\|_2 \le 1,\ A_i^T z^i = y^1\} + \{-q_i \mid y^2 = -p_i\} \mid y^1 + y^2 = y\right\} \\ &= \inf_{z^i}\left\{b_i^T z^i - q_i \mid \|z^i\|_2 \le 1,\ A_i^T z^i - p_i = y\right\}. \end{aligned}$$

When plugging this expression for $f_i^*(y)$ into (D'), we obtain the following dual for the SOCO problem:

$$\sup_{y^i,\, u\geq 0}\left\{-\sum_{i=0}^m u_i \inf_{z^i}\left\{b_i^T z^i - q_i \mid \|z^i\|_2 \leq 1,\, A_i^T z^i - p_i = \frac{y^i}{u_i}\right\} \;\middle|\; \sum_{i=0}^m y^i = 0,\, u_0 = 1\right\}.$$

We eliminate the variables y^i and omit the "inf" operator, which gives

$$\sup_{u,\, z^i}\left\{\sum_{i=0}^m u_i\big(q_i - b_i^T z^i\big) \;\middle|\; \|z^i\|_2 \leq 1,\, \forall i,\, \sum_{i=0}^m u_i\big(A_i^T z^i - p_i\big) = 0,\, u \geq 0,\, u_0 = 1\right\}.$$

Introducing $\tilde{z}^i = u_i z^i$, we get

$$\sup_{u,\, \tilde{z}^i}\left\{u^T q - \sum_{i=0}^m b_i^T \tilde{z}^i \;\middle|\; \|\tilde{z}^i\|_2 \leq u_i,\, \forall i,\, -\sum_{i=0}^m u_i p_i + \sum_{i=0}^m A_i^T \tilde{z}^i = 0,\, u \geq 0,\, u_0 = 1\right\}.$$

Finally, defining matrices A and P according to

$$A^T = \left[A_0^T \;\ldots\; A_m^T\right], \quad P = [p_0 \ldots p_m],$$

and b and $\tilde{z}$ as the vectors that arise by concatenating the vectors b_i and $\tilde{z}^i$, respectively, the dual problem gets the form

$$\sup_{u,\, \tilde{z}^i}\{u^T q - b^T \tilde{z} \mid \|\tilde{z}^i\|_2 \leq u_i,\, \forall i,\, A^T \tilde{z} = Pu,\, u \geq 0,\, u_0 = 1\}.$$

This is the known dual for the SOCO problem, as written in, for example, Ben-Tal and Nemirovski (2001).

6.5. Geometric Optimization

Consider the geometric optimization problem (see Boyd and Vandenberghe 2004, p. 254)[3]

$$\min_x\left\{\log\sum_{k=1}^{K_0} e^{a_{0k}^T x + b_{0k}} \;\middle|\; \log\sum_{k=1}^{K_i} e^{a_{ik}^T x + b_{ik}} \leq 0, \quad i = 1, \ldots, m\right\}, \tag{8}$$

where $K_i \geq 1$, $a_{ik} \in \mathbf{R}^n$, $b_{ik} \in \mathbf{R}$, for $i = 0, \ldots, m$. We define A_i as the $n \times K_i$ matrix with columns $a_{i1}, \ldots, a_{iK_i}$ and b_i as the vector with entries $b_{i1}, \ldots, b_{iK_i}$, for $i = 0, \ldots, m$. In order to compute the conjugate functions of the objective function and the constraint functions we define, for $i = 0, \ldots, m$,

$$h_i(x) := \log\sum_{k=1}^{K_i} e^{x_k},$$

$$f_i(x) := h_i\big(A_i^T x + b_i\big) = \log\sum_{k=1}^{K_i} e^{a_{ik}^T x + b_{ik}}.$$

Using the linear substitution rule (Table E.2, line 8b) and the formula for the conjugate of $h_i(x)$ (Table E.1, line 5), the conjugate of $f_i(x)$ can be written as

$$\begin{aligned} f_i^*(y) &= \inf_z\{h_i^*(z) - b_i^T z \mid A_i^T z = y\} \\ &= \inf_z\left\{\sum_{k=1}^{K_i} z_k \log z_k - b_i^T z \;\middle|\; z \geq 0,\, \mathbf{1}^T z = 1,\, A_i^T z = y\right\}. \end{aligned}$$

To simplify notation, let p^* be the optimal value of (8). Using the dual problem as given by (D'), we obtain

$$\begin{aligned} p^* &= \sup_{y^i,\, u\geq 0}\left\{-\sum_{i=0}^m (u_i f_i)^*(y^i) \;\middle|\; \sum_{i=0}^m y^i = 0,\, u_0 = 1\right\} \\ &= \sup_{y^i,\, u\geq 0}\left\{-\sum_{i=0}^m u_i \inf_{z^i\geq 0}\{\sum_{k=1}^{K_i} z_k^i \log z_k^i - b_i^T z^i\right\}. \\ &\qquad \text{subject to (s.t.)} \sum_{i=0}^m y^i = 0,\, \mathbf{1}^T z^i = 1,\, A_i^T z^i = \frac{y^i}{u_i},\, \forall i,\, u_0 = 1\bigg\}. \end{aligned}$$

This is equivalent to

$$p^* = \sup_{y^i,\, u\geq 0,\, z^i\geq 0} \left\{ \sum_{i=0}^{m} u_i \left[b_i^T z^i - \sum_{k=1}^{K_i} z_k^i \log z_k^i \right] \right.$$
$$\left. \text{s.t.} \sum_{i=0}^{m} y^i = 0,\, \mathbf{1}^T z^i = 1,\, A_i^T z^i = \frac{y^i}{u_i},\, \forall i,\, u_0 = 1 \right\},$$
$$= \sup_{u\geq 0,\, z^i\geq 0} \left\{ \sum_{i=0}^{m} u_i \left[b_i^T z^i - \sum_{k=1}^{K_i} z_k^i \log z_k^i \right] \middle| \sum_{i=0}^{m} u_i A_i^T z^i = 0,\, \mathbf{1}^T z^i = 1,\, \forall i,\, u_0 = 1 \right\}.$$

Defining $\tilde{z}^i = u_i z^i$, $i = 0, \ldots, m$, we obtain a convex formulation of Fenchel's dual problem:

$$p^* = \sup_{u\geq 0,\, \tilde{z}^i\geq 0} \left\{ \sum_{i=0}^{m} \left[b_i^T \tilde{z}^i - \sum_{k=1}^{K_i} \tilde{z}_k^i \log \frac{\tilde{z}_k^i}{u_i} \right] \middle| \sum_{i=0}^{m} A_i^T \tilde{z}^i = 0,\, \mathbf{1}^T \tilde{z}^i = u_i,\, \forall i,\, u_0 = 1 \right\}$$
$$= \sup_{\tilde{z}^i\geq 0} \left\{ \sum_{i=0}^{m} \left[b_i^T \tilde{z}^i - \sum_{k=1}^{K_i} \tilde{z}_k^i \log \frac{\tilde{z}_k^i}{\mathbf{1}^T \tilde{z}^i} \right] \middle| \sum_{i=0}^{m} A_i^T \tilde{z}^i = 0,\, \mathbf{1}^T \tilde{z}^0 = 1 \right\}.$$

For $m = 0$, this is the same dual problem as the one given in Boyd and Vandenberghe (2004, p. 254). For earlier versions of this dual problem, we refer to de Hartog (1994), Duffin et al. (1967), and Fiacco and McCormick (1990).

6.6. ℓ_p-Norm Optimization

We consider the following ℓ_p-norm optimization problem (see den Hertog 1994, Terlaky 1985):

$$\max_x \left\{ \eta^T x \middle| \sum_{i\in I_k} \frac{1}{p_i} |a_i^T x - c_i|^{p_i} + b_k^T x - d_k \leq 0,\, k = 1, \ldots, m \right\}, \tag{9}$$

where $I_1, \ldots, I_m$ denotes a partition of $\{1, \ldots, r\}$, with $1 \leq m \leq r$. Moreover, $b_k \in \mathbf{R}^n$ and $d_k \in \mathbf{R}$ for each k, and $a_i \in \mathbf{R}^n$, $c_i \in \mathbf{R}$, and $p_i \geq 1$ for $1 \leq i \leq r$. In order to obtain Fenchel's dual problem, we consider the minimization problem

$$\min_x \left\{ -\eta^T x \middle| \sum_{i\in I_k} \frac{1}{p_i} |a_i^T x - c_i|^{p_i} + b_k^T x - d_k \leq 0,\, k = 1, \ldots, m \right\}, \tag{10}$$

whose optimal value is opposite to the optimal value of (9). We define

$$f_0(x) := -\eta^T x,$$
$$f_k(x) := g_k(x) + h_k(x), \quad k = 1, \ldots, m,$$

where

$$g_k(x) := \sum_{i\in I_k} \frac{1}{p_i} |a_i^T x - c_i|^{p_i}, \quad h_k(x) := b_k^T x - d_k.$$

To derive the dual problem of (10), we need the conjugates of $f_0, \ldots, f_m$. The functions f_0 and h_k are linear, so one has (Table E.1, line 1)

$$f_0^*(y) = 0, \quad \operatorname{dom} f^* = \{-\eta\},$$
$$h_k^*(y) = d_k, \quad \operatorname{dom} h_k^* = \{b_k\}.$$

The function $g_k(x)$ can be written as $g_k(x) = \sum_{i\in I_k} \sigma_i(a_i^T x - c_i)$, where $\sigma_i(x) = |x|^{p_i}/p_i$. Using the sum rule (Table E.2, line 6), the linear substitution rule (Table E.2, line 8b), and the convex conjugate of $\sigma_i(x)$ (Table E.1, line 10), we obtain the following conjugate function of $g_k(x)$:

$$g_k^*(y) = \inf_{y^i}\left\{\sum_{i\in I_k} \inf_{z_i\in\mathbf{R}}\left\{\frac{1}{q_i}|z_i|^{q_i} + c_i z_i \mid a_i z_i = y^i\right\} \;\middle|\; \sum_{i\in I_k} y^i = y\right\}$$
$$= \inf_{z_i}\left\{\sum_{i\in I_k}\left(\frac{1}{q_i}|z_i|^{q_i} + c_i z_i\right) \;\middle|\; \sum_{i\in I_k} a_i z_i = y\right\},$$

with q_i such that $1/p_i + 1/q_i = 1$. Now the convex conjugate of $f_k(x)$ can be written as

$$f_k^*(y) = \inf_{y^1, y^2}\left\{g_k^*(y^1) + h_k^*(y^2) \mid y^1 + y^2 = y\right\}$$
$$= \inf_{z_i}\left\{\left(\sum_{i\in I_k}\frac{1}{q_i}|z_i|^{q_i} + c_i z_i\right) + d_k \;\middle|\; \sum_{i\in I_k} a_i z_i + b_k = y\right\}.$$

Fenchel's dual of (10) is now given by

$$\max_{y^k,\, u\geq 0}\left\{-f_0^*(y^0) - \sum_{k=1}^m (u_k f_k)^*(y^k) \;\middle|\; \sum_{k=0}^m y^k = 0\right\}.$$

Using the formulas that we derived earlier, we get the problem

$$\max_{y^k,\, u\geq 0}\left\{-\sum_{k=1}^m u_k \inf_{z_i}\left\{\sum_{i\in I_k}\left(\frac{1}{q_i}|z_i|^{q_i} + c_i z_i\right) + d_k \;\middle|\; y^0 = -\eta,\ \sum_{i\in I_k} a_i z_i + b_k = \frac{y^k}{u_k}\right\}\right\},$$

subject to the condition $\sum_{k=0}^m y^k = 0$. This can be simplified to

$$\max_{u\geq 0,\, z_i}\left\{-\sum_{k=1}^m u_k\left[\sum_{i\in I_k}\left(\frac{1}{q_i}|z_i|^{q_i} + c_i z_i\right) + d_k\right] \;\middle|\; \sum_{k=1}^m u_k\left(\sum_{i\in I_k} a_i z_i + b_k\right) = \eta\right\}.$$

To further simplify this formulation, we define $\tilde{z}_i = u_k z_i$ for each i. Then we get

$$\max_{u\geq 0,\, \tilde{z}_i}\left\{-\sum_{k=1}^m u_k\left[\sum_{i\in I_k}\left(\frac{1}{q_i}\left|\frac{\tilde{z}_i}{u_k}\right|^{q_i} + c_i\frac{\tilde{z}_i}{u_k}\right) + d_k\right] \;\middle|\; \sum_{k=1}^m u_k\left(\sum_{i\in I_k} a_i\frac{\tilde{z}_i}{u_k} + b_k\right) = \eta\right\}.$$

By changing the sign of the optimal value, we arrive at

$$\min_{u\geq 0,\, \tilde{z}_i}\left\{c^T\tilde{z} + u^T d + \sum_{k=1}^m u_k \sum_{i\in I_k}\frac{1}{q_i}\left|\frac{\tilde{z}_i}{u_k}\right|^{q_i} \;\middle|\; \sum_{k=1}^m\sum_{i\in I_k} a_i\tilde{z}_i + \sum_{k=1}^m u_k b_k = \eta\right\},$$

which is the dual problem of (9) that we are looking for. Finally, defining matrices A and B according to

$$A = [a_1 \ldots a_n], \quad B = [b_1 \ldots b_r],$$

the last problem gets the form

$$\min_{u\geq 0,\, \tilde{z}_i}\left\{u^T d + c^T\tilde{z} + \sum_{k=1}^m u_k\sum_{i\in I_k}\frac{1}{q_i}\left|\frac{\tilde{z}_i}{u_k}\right|^{q_i} \;\middle|\; A\tilde{z} + Bu = \eta\right\},$$

which is the same dual as was obtained in Terlaky (1985).

6.7. Radiotherapy Treatment Planning

Optimization plays an important role in radiotherapy treatment planning. The resulting optimization problems contain many different nonlinear objective and constraint functions. To the best of our knowledge, the corresponding dual problems have not been derived. These dual problems might be useful for reasons already stated in the Introduction. In this section, we show that with the framework proposed in this paper, the dual problem can be derived in a structured and straightforward way.

Treatment planning for external beam radiotherapy aims at finding beam intensities such that the tumor is controlled while limiting the dose to the surrounding organs at risk (OARs). For planning purposes, the structures of interest (e.g., tumor, lung, heart) are virtually divided into a large number of small cubic volumes, so-called voxels. An important input parameter to all treatment planning models is the matrix D, whose (i, j)-element is the dose per unit intensity (dose rate) from *beamlet* j to *voxel* i (each beam is subdivided into beamlets). The dose from beamlet j to voxel i is equal to $D_{ij}x_j$, with x_j the intensity of beamlet j. The total dose to voxel i can be written as $d_i = D_i x$, where D_i is the ith row of D, and x is the vector of beamlet intensities. Treatment plans are evaluated based on the dose to each of the voxels. Treatment planning models employ constraints and objectives such as the minimum, mean, or maximum dose over all voxels within a structure; dose-volume metrics, such as the minimum dose to the hottest p% of the voxels; and biological metrics, such as the tumor control probability (TCP). Treatment planning models thus may contain different types of constraints and do usually not fit a generic optimization format (such as, e.g., linear or second-order cone optimization). A readily given dual problem is thus generally not available. Here we give an example of such a treatment planning model and derive its dual using the method proposed in this paper.

The objective of our model is to maximize the TCP (see Stavreva et al. 2003):

$$\mathrm{TCP}(x) = \prod_{i \in \mathcal{T}} \exp\left(-N_0 v_i e^{-\alpha D_i x}\right),$$

where $\mathcal{T}$ is the set of voxels in the target volume, N_0 is the number of clonogenic cells, v_i is the relative volume of voxel i, and α describes the radio resistance of cells in the target volume. This function is not concave. Taking the log of TCP gives the concave function LTCP:

$$\mathrm{LTCP}(x) = -\sum_{i \in \mathcal{T}} N_0 v_i e^{-\alpha D_i x},$$

which yields the same optimal solution for x as $\log(x)$ is increasing (see Romeijn et al. 2004). Therefore, LTCP is optimized instead.

Our planning model contains a constraint on the dose to the OARs. For simplicity, we assume that there is only one OAR, denoted as S. First, we consider the generalized equivalent uniform dose (gEUD), which is the generalized mean of the dose to a structure S (see Niemierko 1999):

$$\mathrm{gEUD}(x) = \left(\sum_{i \in \mathcal{S}} v_i (D_i x)^p\right)^{\frac{1}{p}},$$

where $\mathcal{S}$ denotes the set of voxels in structure S, and $p \geq 1$ is a structure-dependent parameter for OARs that indicates whether it is a serial or a parallel organ. We limit the normal tissue complication probability (NTCP) to this OAR (see Alber and Nüsslin 2001):

$$\mathrm{NTCP}(x) = 1 - \exp\left(-\left(\frac{\mathrm{gEUD}(x)}{\Delta}\right)^p\right),$$

where Δ is a structure-dependent parameter that indicates the uniform dose level for which there is a 63% probability on complications. Furthermore, $p > 1$, because we restrict ourselves to a nonparallel organ. We would like to constrain NTCP from above; however, NTCP is not a convex function. Therefore, following Romeijn et al. (2004), we use instead

$$-\ln(1 - \mathrm{NTCP}(x)) = \frac{1}{\Delta^p} \sum_{i \in \mathcal{S}} v_i (D_i x)^p,$$

which is convex. As was the case with the transformation from TCP to LTCP, this does not change the optimal solution. Note that we need to apply the same transformation to the upper bound on NTCP.

We can now formulate our treatment plan optimization model:

$$\min_x \left\{ N_0 \sum_{i \in \mathcal{T}} v_i e^{-\alpha D_i x} \;\middle|\; \left\{ \begin{array}{r} \frac{1}{\Delta^p} \sum_{i \in \mathcal{S}} v_i (D_i x)^p - \gamma \leq 0 \\ x \geq 0 \end{array} \right. \right\},$$

where γ is a predetermined upper bound on $-\ln(1-\text{NTCP})$. The nonnegativity constraint is included because a negative beam intensity is physically impossible.

Before giving the dual of this problem, we derive the conjugates of the objective function and the constraint function.

6.7.1. Objective Function. The objective function $f_0(x)$ is a constant multiplied by the sum of $v_i h_i(x)$, where $h_i(x) = g(-\alpha D_i x)$, $g(u) = e^u$. Using Table E.1 (line 2) and the linear substitution rule (Table E.2, line 8b), we obtain the convex conjugate of $h_i(x)$:

$$h_i^*(y) = \inf_{z_i}\{z_i \log z_i - z_i \mid -\alpha z_i D_i^T = y, z_i \geq 0\}.$$

Application of the sum rule for conjugates (Table E.2, line 6) now enables us to write the conjugate of $f_0(x)$ as follows:

$$\begin{aligned}
f_0^*(y) &= N_0\left(\sum_{i\in\mathcal{T}} v_i h_i\right)^*\left(\frac{y}{N_0}\right) \text{ (Product with a scalar)}\\
&= N_0 \inf_{\{y^i\}_{i\in\mathcal{T}}}\left\{\sum_{i\in\mathcal{T}}(v_i h_i)^*(y^i)\,\middle|\, \sum_{i\in\mathcal{T}} y^i = \frac{y}{N_0}\right\} \text{ (Sum rule)}\\
&= N_0 \inf_{\{y^i\}_{i\in\mathcal{T}}}\left\{\sum_{i\in\mathcal{T}} v_i h_i^*\left(\frac{y^i}{v_i}\right)\,\middle|\, \sum_{i\in\mathcal{T}} y^i = \frac{y}{N_0}\right\} \text{ (Product with a scalar)}\\
&= N_0 \inf_{\{y^i\}_{i\in\mathcal{T}}}\left\{\sum_{i\in\mathcal{T}} v_i \inf_{z_i}\left\{z_i\log z_i - z_i \mid -\alpha z_i D_i^T = \frac{y^i}{v_i}, z_i \geq 0\right\}\,\middle|\, \sum_{i\in\mathcal{T}} y^i = \frac{y}{N_0}\right\}\\
&= N_0 \inf_w\left\{\sum_{i\in\mathcal{T}} w_i \log\frac{w_i}{v_i} - w_i\,\middle|\, \sum_{i\in\mathcal{T}} -\alpha w_i D_i^T = \frac{y}{N_0}, w \geq 0\right\} \text{ } (w_i := v_i z_i)\\
&= N_0 \inf_w\left\{\sum_{i\in\mathcal{T}} w_i \log\frac{w_i}{v_i} - w_i\,\middle|\, -\alpha N_0 D_{\mathcal{T}}^T w = y,\ w \geq 0\right\},
\end{aligned}$$

where $D_{\mathcal{T}}$ consists of the rows in D related to the voxels in $\mathcal{T}$.

6.7.2. NTCP Constraint. The NTCP constraint function $f_1(x)$ is given by

$$f_1(x) = \frac{1}{\Delta^p}\sum_{i\in\mathcal{S}} v_i(D_i x)^p - \gamma = \frac{1}{\Delta^p}\sum_{i\in\mathcal{S}} v_i g_i(x) - \gamma,$$

where $g_i(x) = h_i(D_i x)$, $h_i(u) = u^p$ $(u \geq 0)$, with $p > 1$. The convex conjugates of $h_i(u)$ and $g_i(x)$ are (using Table E.1, line 8, and Table E.2, lines 4 and 8b)

$$\begin{aligned}
h_i^*(y) &= \inf_{t_i}\left\{\frac{p}{q}t_i^q \mid t_i \geq 0,\ t_i \geq \frac{y}{p}\right\},\\
g_i^*(y) &= \inf_{r_i}\left\{\inf_{t_i}\left\{\frac{p}{q}t_i^q \mid t_i \geq 0,\ t_i \geq \frac{r_i}{p}\right\} \mid D_i^T r_i = y\right\}\\
&= \inf_{r_i,t_i}\left\{\frac{p}{q}t_i^q \mid D_i^T r_i = y,\ t_i \geq 0,\ t_i \geq \frac{r_i}{p}\right\}.
\end{aligned}$$

By using the sum rule for conjugates, the multiply with a constant rule, and the add a constant rule (Table E.2, lines 6, 4, and 1, respectively) we can now compute $f_2^*(y)$:

$$
\begin{aligned}
f_2^*(y) &= \frac{1}{\Delta^p}\inf_{y^i}\left\{\sum_{i\in\mathcal{S}} v_i g_i^*\left(\frac{y^i}{v_i}\right) \,\middle|\, \sum_{i\in\mathcal{S}} y^i = \Delta^p y\right\} + \gamma \\
&= \frac{1}{\Delta^p}\inf_{y^i}\left\{\sum_{i\in\mathcal{S}} v_i \inf_{r_i,t_i}\left\{\frac{p}{q}t_i^q \mid D_i^T r_i = \frac{y^i}{v_i},\ t_i \ge 0,\ t_i \ge \frac{r_i}{p}\right\} \,\middle|\, \sum_{i\in\mathcal{S}} y^i = \Delta^p y\right\} + \gamma \\
&= \frac{p}{q\Delta^p}\inf_{r_i,t_i}\left\{\sum_{i\in\mathcal{S}} v_i t_i^q \,\middle|\, \sum_{i\in\mathcal{S}} v_i D_i^T r_i = \Delta^p y,\ t_i \ge 0,\ p\,t_i \ge r_i\right\} + \gamma \\
&= \inf_{r,t}\left\{\gamma + \frac{p}{q\Delta^p} v_{\mathcal{S}}^T t^q \mid D_{\mathcal{S}}^T V_{\mathcal{S}} r = \Delta^p y,\ t \ge 0,\ p\,t \ge r\right\},
\end{aligned}
$$

where t^q is the vector with entries t_i^q, $i \in \mathcal{S}$.

6.7.3. Nonnegativity Constraint. We finally have to deal with the nonnegativity constraint $x \ge 0$. Denoting the indicator function of $\mathbf{R}_+^n$ as $f_2(x)$, the corresponding support function is (see Table E.3, line 2)

$$f_1^*(y) = 0, \quad \operatorname{dom} f_1^* = \{y \mid y \le 0\}.$$

As we know, Fenchel's dual problem is given by

$$\sup_{\{y^i\}_{i=0}^2, u}\left\{-f_0^*(y^0) - u f_1^*\left(\frac{y^1}{u}\right) - f_2^*(y^2) \,\middle|\, \sum_{k=0}^2 y^k = 0,\ u \ge 0\right\}.$$

We can already eliminate y^2, yielding the equivalent problem

$$\sup_{y^0,y^1,u}\left\{-f_0^*(y^0) - u f_1^*\left(\frac{y^1}{u}\right) \,\middle|\, y^0 + y^1 \ge 0,\ u \ge 0\right\}.$$

Thus, we obtain the following dual for the treatment planning optimization problem:

$$
\begin{aligned}
\sup_{y^0,y^1,u}\Bigg\{ & -N_0\{\inf_w \sum_{i\in\mathcal{T}}\left(w_i \log\frac{w_i}{v_i} - w_i\right) \,\Big|\, -\alpha N_0 D_{\mathcal{T}}^T w = y^0,\ w \ge 0\Big\}. \\
& - u\inf_{r,t}\left\{\gamma + \frac{p}{q\Delta^p} v_{\mathcal{S}}^T t^q \,\middle|\, D_{\mathcal{S}}^T V_{\mathcal{S}} r = \Delta^p \frac{y^1}{u},\ t \ge 0,\ p\,t \ge r\right\} \\
& \Big|\ y^0 + y^1 \ge 0,\ u \ge 0\Bigg\}.
\end{aligned}
$$

By omitting the "inf" operators, we get

$$
\begin{aligned}
\sup_{y^0,y^1,w,z,r,t,s,u}\Bigg\{ & -N_0\{\sum_{i\in\mathcal{T}}\left(w_i \log\frac{w_i}{v_i} - w_i\right) \,\Big|\, -\alpha N_0 D_{\mathcal{T}}^T w = y^0, w \ge 0\Big\}. \\
& - u\left\{\gamma + \frac{p}{q\Delta^p} v_{\mathcal{S}}^T t^q \,\middle|\, D_{\mathcal{S}}^T V_{\mathcal{S}} r = \Delta^p \frac{y^1}{u},\ t \ge 0,\ pt \ge r\right\} \\
& \Big|\ y^0 + y^1 \ge 0,\ u \ge 0\Bigg\}.
\end{aligned}
$$

This is equivalent to

$$
\begin{aligned}
\sup_{w,r,t,u}\Bigg\{ & -N_0 \sum_{i\in\mathcal{T}}\left(w_i \log\frac{w_i}{v_i} - w_i\right) - u\gamma - \frac{upv_{\mathcal{S}}^T t^q}{q\Delta^p} \,\Big| \\
& w \ge 0,\ t \ge 0,\ p\,t \ge r,\ u \ge 0, \\
& -\alpha N_0 D_T^T w + \frac{1}{\Delta^p} D_{\mathcal{S}}^T V_{\mathcal{S}}(ur) \ge 0\Bigg\}.
\end{aligned}
$$

By redefining s according to $r := ur$, $t := ut$, this becomes

$$\sup_{w,r,t,u}\Bigg\{-N_0\sum_{i\in\mathcal{T}}\left(w_i\log\frac{w_i}{v_i}-w_i\right)-u\gamma-\frac{up}{q\Delta^p}v_{\mathcal{G}}^T\left(\frac{t}{u}\right)^q \Bigg| \\ w\geq 0,\ t\geq 0,\ p\,t\geq r,\ u\geq 0, \\ -\alpha N_0 D_T^T w+\frac{1}{\Delta^p}D_{\mathcal{G}}^T V_{\mathcal{G}} r+s\geq 0\Bigg\}.$$

One easily verifies that this dual problem of the treatment plan optimization problem is indeed a convex optimization problem.

7. Concluding Remarks

Mathematical education usually includes the use of tables (e.g., tables of derivatives), primitive functions, Laplace transforms, Fourier transforms, and so on. It may have become clear from this paper that the availability of tables as presented in Appendix E simplifies the process of dualizing an optimization problem. Combined with the use of Fenchel's duality theorem, it yields a universal and structured way for deriving dual problems for a wide spectrum of convex optimization problems.

Acknowledgment

The authors thank Lieven Vandenberghe (UCLA) for useful suggestions.

Appendix A. Some Examples of Conjugate Functions

By way of example, we demonstrate how some of the results in Table E.1 are obtained; this table also gives appropriate references to the existing literature whenever available. The examples in this section are used in Section 6, where we compute the dual problems of some more or less standard optimization problems.

A.1. Linear Function

We start by proving the formula in line 1 of Table E.1. With $f(x) = a^T x + b$, $a \in \mathbf{R}^n$ and $b \in \mathbf{R}$, we have

$$f^*(y) = \sup_x\{y^T x - (a^T x + b)\} = \sup_x\Big\{(y-a)^T x - b\Big\}.$$

Obviously, the "sup" equals $-b$ if $y = a$ and infinity otherwise (take $x = \lambda(y-a)$, $\lambda > 0$). Hence, we obtain

$$f^*(y) = -b, \quad \operatorname{dom} f^* = \{a\}.$$

A.2. Log-Sum-Exp Function

We next deal with line 5 in Table E.1. Let $f(x) = \log\left(\sum_{i=1}^n e^{x_i}\right)$, $x \in \mathbf{R}^n$. We then have

$$f^*(y) = \sup_x\left\{y^T x - \log\left(\sum_{i=1}^n e^{x_i}\right)\right\}.$$

Define $g(x) = y^T x - f(x)$. If $y_i < 0$ for some i, then we take $x = -\lambda e_i$, where $\lambda > 0$ and e_i is the ith unit vector. Then $g(x) = -\lambda y_i - \log(n - 1 + e^{-\lambda})$, which goes to infinity if λ goes to infinity. If $y \geq 0$ but $\mathbf{1}^T y \neq 1$, take $x = \lambda\mathbf{1}$. Then $g(x) = \lambda\mathbf{1}^T y - \log(ne^{\lambda}) = \lambda(\mathbf{1}^T y - 1) - \log n$, which goes to infinity both if $\mathbf{1}^T y > 1$ (when λ goes to infinity) and if $\mathbf{1}^T y < 1$ (when λ goes to minus infinity). Hence, $y \in \operatorname{dom} f^*$ holds only if $\mathbf{1}^T y = 1$ and $y \geq 0$. We proceed by computing the maximal value of $g(x)$ under these conditions. We have

$$\frac{\partial g(x)}{\partial x_i} = y_i - \frac{e^{x_i}}{\sum_{i=1}^n e^{x_i}}.$$

By setting these partial derivatives equal to zero, we obtain the condition

$$y_i = \frac{e^{x_i}}{\sum_{i=1}^n e^{x_i}}, \quad i = 1, \ldots, n.$$

Substitution of these values of y into $g(x)$ yields

$$f^*(y) = \sum_{i=1}^{n} y_i \log y_i, \quad y > 0, \mathbf{1}^T y = 1.$$

By interpreting $0 \log 0 = 0$, this expression remains valid if some entries of y vanish. Thus, we conclude that

$$f^*(y) = \sum_{i=1}^{n} y_i \log y_i, \quad \operatorname{dom} f^* = \{y \mid \mathbf{1}^T y = 1, \, y \geq 0\}.$$

A.3. Arbitrary Norm

In this section, we deal with line 6 in Table E.1. Let $\|x\|$ be an arbitrary norm on $\mathbf{R}^n$ and $f(x) := \|x\|$. We then have

$$f^*(y) = \sup_x \{y^T x - \|x\|\}.$$

Recall that the dual norm is defined by

$$\|y\|_* = \sup_x \{y^T x \mid \|x\| \leq 1\}.$$

Because $\|\frac{x}{\|x\|}\| = 1$ if $x \neq 0$, we have, for any nonzero $x \in \mathbf{R}^n$,

$$\|y\|_* \geq y^T \frac{x}{\|x\|} = \frac{y^T x}{\|x\|},$$

which gives

$$y^T x \leq \|y\|_* \|x\|. \tag{A.1}$$

Because this inequality also holds for $x = 0$, it holds for each $x \in \mathbf{R}^n$. Define $g(x) := y^T x - \|x\|$. Then $g(\lambda x) = \lambda g(x)$ for $\lambda \geq 0$. If $\|y\|_* > 1$, then there exists an x with $\|x\| \leq 1$ and $y^T x > 1$. Then $g(x) > 0$. Hence $g(\lambda x)$ goes to infinity if λ grows to infinity. We conclude from this that $y \in \operatorname{dom} f^*$ holds only if $\|y\|_* \leq 1$.

If $\|y\|_* \leq 1$, then (A.1) implies that $y^T x \leq \|y\|_* \|x\| \leq \|x\|$, whence $g(x) \leq 0$, for all $x \in \mathbf{R}^n$. Because $g(0) = 0$, we obtain

$$f^*(y) = 0, \quad \operatorname{dom} f^* = \{y \mid \|y\|_* \leq 1\}.$$

Let us recall that if $\|.\|$ represents the p-norm, then it is denoted as $\|.\|_p$, and

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}}, \, p \geq 1, \quad x \in \mathbf{R}^n.$$

In this case, the dual norm is given by $\|.\|_q$, with q such that $\frac{1}{p} + \frac{1}{q} = 1$.

A.4. Quadratic Function

In this section, we deal with line 13 in Table E.1. Let $f(x) = \frac{1}{2} x^T P x, x \in \mathbf{R}^n$, where P is a (symmetric) positive-semidefinite $n \times n$ matrix. We then have

$$f^*(y) = \sup_x \left\{y^T x - \frac{1}{2} x^T P x\right\}.$$

Because $y^T x - \frac{1}{2} x^T P x$ is concave in x, its maximal value occurs if $y = Px$. If P is nonsingular, then this implies that $x = P^{-1} y$, and hence,

$$f^*(y) = y^T P^{-1} y - \frac{1}{2} (P^{-1} y)^T P P^{-1} y = \frac{1}{2} y^T P^{-1} y.$$

If P is singular, then the equation $y = Px$ has a solution if and only if y belongs to the column space $L = \{Pz \mid z \in \mathbf{R}^n\}$. If $y \in L$, then the solution is given by $x = P^{\dagger} y$, where $P^{\dagger}$ is the generalized inverse of P, and we obtain

$$f^*(y) = y^T P^{\dagger} y - \frac{1}{2} \left(P^{\dagger} y\right)^T P P^{\dagger} y = \frac{1}{2} y^T P^{\dagger} y,$$

where the last equality is due to the fact that $P^\dagger P P^\dagger = P^\dagger$ (see Ben-Israel and Greville 2003).
If $y \notin L$, then we may write $y = y_1 + y_2$ with $y_1 \in L$ and $0 \neq y_2 \in L^\perp$. Take $\lambda > 0$ and $x = \lambda y_2$. Then $x^T y = x^T y_1 + x^T y_2 = x^T y_2 = \lambda y_2^T y_2 = \lambda \|y_2\|^2$. Because $x \in L^\perp$, we have $Px = 0$, whence $x^T P x = 0$. Hence, $y^T x - \frac{1}{2} x^T P x = \lambda \|y_2\|^2$, which goes to infinity if λ grows to infinity. This proves that $y \notin \operatorname{dom} f^*$ if $y \notin L$. Thus, we have shown that

$$f^*(y) = \frac{1}{2} y^T P^\dagger y, \quad \operatorname{dom} f^* = \{y \mid y = Pz,\ z \in \mathbf{R}^n\}.$$

A.5. Conjugate of $\frac{1}{p}x^p$, $x \geq 0$, $p > 1$

We prove the formula in line 12 of Table E.1. Let $f(x) = \frac{1}{p}x^p$, $x \geq 0$, $p > 1$. We have

$$f'(x) = x^{p-1},$$
$$f''(x) = (p-1)x^{p-2},$$

which shows that f is convex because $x \geq 0$ and $p > 1$. Then

$$f^*(y) = \sup_{x \geq 0}(yx - f(x)).$$

Because $yx - f(x) \leq 0$ if $y < 0$, the "sup" value occurs for $x = 0$, and then it equals $0 - f(0) = 0$. Otherwise, if $y \geq 0$, we have

$$f^*(y) = \sup_{x \geq 0}\left(yx - \frac{1}{p}x^p\right).$$

The "sup" value is then attained when $y = f'(x) = x^{p-1}$, which gives $x = y^{\frac{1}{p-1}}$, and

$$f^*(y) = y^{1+\frac{1}{p-1}} - \frac{1}{p}y^{\frac{p}{p-1}} = \left(1 - \frac{1}{p}\right)y^{\frac{p}{p-1}} = \frac{1}{q}y^q,$$

where q is such that $\frac{1}{p} + \frac{1}{q} = 1$. Summarizing,

$$f^*(y) = \begin{cases} 0, & y < 0, \\ \frac{1}{q}y^q, & y \geq 0. \end{cases}$$

Because $f^*(y)$ is monotonically increasing, this can be written as

$$f^*(y) = \inf_z\left\{\frac{1}{q}z^q \mid z \geq y,\ z \geq 0\right\}.$$

A.6. Conjugate of $\frac{1}{px^p}$, $x > 0$, $p > 0$

We prove the formula in line 11 of Table E.1. Let $f(x) = \frac{1}{px^p}$, $x > 0$, $p > 0$. We have

$$f'(x) = \frac{-1}{x^{p+1}},$$
$$f''(x) = \frac{p+1}{x^{p+2}},$$

which shows that f is convex because $x > 0$ and $p + 1 > 1 > 0$. Also note that $f(x)$ is monotonically decreasing to zero if x goes to ∞. We have

$$f^*(y) = \sup_{x > 0}(yx - f(x)).$$

If $y > 0$, then $yx - f(x)$ goes to ∞ if x goes to ∞. Hence, the "sup" value is ∞ if $y > 0$. Thus, we may assume that $y \leq 0$. Because then

$$f^*(y) = \sup_{x > 0}\left(yx - \frac{1}{px^p}\right),$$

the "sup" value is attained when $y = f'(x) = \frac{-1}{x^{p+1}}$, which gives $x = (-y)^{\frac{-1}{p+1}}$, and

$$f^*(y) = y(-y)^{\frac{-1}{p+1}} - \frac{1}{p}(-y)^{\frac{p}{p+1}} = \left(-1 - \frac{1}{p}\right)(-y)^{\frac{p}{p+1}} = -\left(\frac{p+1}{p}\right)(-y)^{\frac{p}{p+1}}.$$

Summarizing,

$$f^*(y) = -\left(\frac{p+1}{p}\right)(-y)^{\frac{p}{p+1}}, \qquad \operatorname{dom} f^* = \{y \mid y \leq 0\} = -\mathbf{R}_+.$$

Appendix B. Some Identities for Conjugate Functions

B.1. Conjugate of $f(x) = \max_i f_i(x)$

First, note that we may write

$$f(x) = \max_z \left\{ \sum_{i=1}^m z_i f_i(x) \,\middle|\, \sum_{i=1}^m z_i = 1,\, z \geq 0 \right\} = \max_{z \in S} \left\{ \sum_{i=1}^m z_i f_i(x) \right\},$$

where S denotes the unit simplex. By the definition of f^*, we have

$$\begin{aligned} f^*(y) &= \sup_x \left\{ y^T x - \max_{z \in S} \left\{ \sum_{i=1}^m z_i f_i(x) \right\} \right\} \\ &= \sup_x \min_{z \in S} \left\{ y^T x - \sum_{i=1}^m z_i f_i(x) \right\}. \end{aligned}$$

Because the argument is convex in z and concave in x and the domain of z is compact, we may interchange the "sup" and "min" operators, yielding

$$f^*(y) = \min_{z \in S} \sup_x \left\{ y^T x - \sum_{i=1}^m z_i f_i(x) \right\} = \min_{z \in S} \left(\sum_{i=1}^m z_i f_i \right)^* (y).$$

Using the sum rule for conjugates (Table E.2, line 6), we obtain

$$\begin{aligned} f^*(y) &= \min_{z \in S} \left\{ \sum_{i=1}^m (z_i f_i)^* (y^i) \,\middle|\, \sum_{i=1}^m y^i = y \right\} \\ &= \min_z \left\{ \sum_{i=1}^m (z_i f_i)^* (y^i) \,\middle|\, \sum_{i=1}^m y^i = y,\, \sum_{i=1}^m z_i = 1,\, z \geq 0 \right\}. \end{aligned}$$

B.2. Deriving Conjugates via the Adjoint, I

In this section, we prove the formula in line 10 of Table E.2. This formula enables us to write $f^*(y)$ as an "inf" expression in cases where no closed formula is available for $f^*(y)$, whereas such a formula exists for $(f^\diamond)^*(y)$. The proof that follows is an alternative to that in Gushchin (2008).

By the definition of $(f^\diamond)^*(y)$, we have

$$\inf_z \left\{ z \mid \left(f^\diamond\right)^*(-z) \leq -y \right\} = \inf_z \left\{ z \mid \sup_x \{ -zx - f^\diamond(x) \} \leq -y \right\}.$$

Because $\mathrm{dom} f^\diamond = \mathbf{R}_{++}$, we have $f^\diamond(x) = \infty$ if $x \leq 0$. Hence,

$$\inf_z \left\{ z \mid \left(f^\diamond\right)^*(-z) \leq -y \right\} = \inf_z \left\{ z \mid \sup_{x>0} \left\{ -zx - xf\left(\frac{1}{x}\right) \right\} \leq -y \right\}.$$

Because of the Lagrange duality, this implies that

$$\begin{aligned} \inf_z \left\{ z \mid \left(f^\diamond\right)^*(-z) \leq -y \right\} &= \sup_{\lambda \geq 0} \inf_z \left\{ z + \lambda \left[y + \sup_{x>0} \left\{ -zx - xf\left(\frac{1}{x}\right) \right\} \right] \right\} \\ &= \sup_{\lambda \geq 0} \inf_z \sup_{x>0} \left\{ \lambda y - \lambda x f\left(\frac{1}{x}\right) + z(1 - \lambda x) \right\}. \end{aligned} \tag{B.1}$$

Observe that the argument in this expression is the Lagrangian of the problem

$$\sup_{x>0} \left\{ \lambda y - \lambda x f\left(\frac{1}{x}\right) \mid 1 - \lambda x = 0 \right\}.$$

Because this problem is Slater regular, its Lagrangian has a saddle point. As a consequence, in (B.1) we may replace $\inf_z \sup_{x>0}$ with $\sup_{x>0} \inf_z$. We have

$$\sup_{x>0} \inf_z \left\{ \lambda y - \lambda x f\left(\frac{1}{x}\right) + z(1 - \lambda x) \right\} = \sup_{x>0} \left\{ \lambda y - \lambda x f\left(\frac{1}{x}\right) \mid \lambda x = 1 \right\} = \lambda y - f(\lambda).$$

Substitution into (B.1) yields

$$\inf_z\left\{z \mid \left(f^\lozenge\right)^*(-z) \le -y\right\} = \sup_{\lambda\ge 0}\{\lambda y - f(\lambda)\} = f^*(y).$$

This equality holds because the domain of f is $\mathbf{R}_+$.

B.3. Deriving Conjugates via the Adjoint, II

In this section, we prove formula 11 in Table E.2. This formula can be applied if f has an inverse, which is denoted as $h = f^{-1}$.

Because h is concave, we have, by definition, $h_*(y) = \inf_x\{yx - h(x)\}$. Thus, we may write

$$\begin{aligned}-(h_*)^\lozenge(y) &= -yh_*\left(\frac{1}{y}\right) = -y\inf_x\left\{\frac{1}{y}x - h(x)\right\} = y\sup_x\left\{h(x) - \frac{x}{y}\right\}\\ &= y\sup_{x,t}\left\{t - \frac{x}{y} \mid t \le h(x)\right\},\end{aligned}$$

where we used that h is (strictly) increasing. If $y < 0$, then $t - \frac{x}{y}$ goes to infinity if x goes to infinity. Hence, we may assume that $y \ge 0$. Also using that $t \le h(x)$ holds if and only if $f(t) \le x$, we may proceed as follows:

$$\begin{aligned}-(h_*)^\lozenge(y) &= \sup_{x,t}\{ty - x \mid t \le h(x)\} = \sup_{x,t}\{ty - x \mid f(t) \le x\}\\ &= \sup_t\{ty - f(t)\} = f^*(y).\end{aligned}$$

B.4. Linear Substitution Rule

In this section, we prove the formulas in lines 8a and 8b of Table E.2. Let f be defined by $f(x) = h(Ax + b)$, where A is an $m \times n$ matrix, $b \in \mathbf{R}^m$ and $h : \mathbf{R}^m \to \mathbf{R}$ convex. We then have

$$\begin{aligned}f^*(y) &= \sup_x\{y^Tx - h(Ax + b)\}\\ &= \sup_{x,v}\{y^Tx - h(v) \mid Ax + b = v\}.\end{aligned}$$

By writing the last problem as a minimization problem and then using Lagrange's duality theorem, we get

$$\begin{aligned}f^*(y) &= \min_z\sup_{x,v}\{y^Tx - h(v) + z^T(v - b - Ax)\}\\ &= \min_z\sup_{x,v}\{z^Tv - h(v) - z^Tb + x^T(y - A^Tz)\}.\end{aligned}$$

The contribution of the term containing x is finite if and only if $y - A^Tz = 0$. Thus, we obtain

$$\begin{aligned}f^*(y) &= \min_z\sup_v\{z^Tv - h(v) - z^Tb \mid A^Tz = y\}\\ &= \min_z\{h^*(z) - b^Tz \mid A^Tz = y\}.\end{aligned}$$

If A is square and nonsingular, then $A^Tz = y$ holds if and only if $z = A^{-T}y$. Hence, in that case, we have

$$f^*(y) = h^*\left(A^{-T}y\right) - b^TA^{-T}y.$$

B.5. Composite Function

In this section, we prove the formula in line 9 of Table E.2. Let $f(x) = g(h(x))$, with g and h convex and g nondecreasing. We may write

$$f(x) = \inf_z\{g(z) \mid h(x) \le z\}.$$

Hence,

$$\begin{aligned}f^*(y) &= \sup_x\left\{y^Tx - \inf_z\{g(z) \mid h(x) \le z\}\right\}\\ &= \sup_{x,z}\{y^Tx - g(z) \mid h(x) \le z\}.\end{aligned}$$

Using the Lagrange dual of the last problem, we get

$$\begin{aligned} f^*(y) &= \min_{u\geq 0} \sup_{x,z} \{y^T x - g(z) + u(z - h(x))\} \\ &= \min_{u\geq 0} \left\{ \sup_z \{uz - g(z)\} + \sup_x \{y^T x - uh(x)\} \right\} \\ &= \min_{u\geq 0} \{g^*(u) + (uh)^*(y)\}. \end{aligned}$$

B.6. Adjoint Function

In this section, we prove the formula in line 12 of Table E.2. With $g(x) = xf(\frac{a}{x})$, we may write

$$g^*(y) = \sup_{x\geq 0} \left\{ yx - xf\left(\frac{a}{x}\right) \right\} = \sup_{x\geq 0} \sup_w \{yx - xf(w) : xw = a\}.$$

Using Lagrange duality, we proceed as follows:

$$\begin{aligned} g^*(y) &= \sup_{x\geq 0} \sup_w \inf_z \{yx - xf(w) + z^T(a - xw)\} \\ &= \sup_{x\geq 0} \inf_z \left\{ yx + z^T a + x \sup_w \{-z^T w - f(w)\} \right\} \\ &= \inf_z \sup_{x\geq 0} \left\{ z^T a + x(y + f^*(-z)) \right\} \\ &= \inf_z \{z^T a \mid f^*(-z) \leq -y\}. \end{aligned}$$

It might be worth mentioning that if $a = 1$, then this formula implies the formula in line 9. This follows because the adjoint of the adjoint of a function yields the function itself.

B.7. Conjugates with "Branches"

In the examples so far, the conjugate of f was either a closed expression or the "inf" of a closed expression. This is not always the case, as we discuss in this section. Sometimes the natural domain of a function is restricted. Suppose that $f : \mathbf{R} \to \mathbf{R}$ is a convex function and that a problem contains the function $f_D : \mathbf{R} \to [-\infty, \infty]$, defined as $f_D(x) = f(x)$ if $x \in D$ and $f_D(x) = \infty$ otherwise.

By way of example, let $f(x) = x^2$. The natural domain is $\mathbf{R}$, and $f^*(y) = \frac{1}{4}y^2$, with domain $\mathbf{R}$. Now let $D = [1, \infty)$. Then $f_D(x) = x^2$ if $x \geq 1$; otherwise, $f_D(x) = \infty$. We then have

$$f_D^*(y) = \sup_x \{yx - f_D(x)\}.$$

Define $g(x) = yx - f_D(x)$. Then the "sup" is finite only if $x \geq 1$ and $g(x) = yx - x^2$. Then the largest value of $g(x)$ occurs for $x = \frac{1}{2}y$, which satisfies $x \geq 1$ only if $y \geq 2$. If $y < 2$, then the largest value occurs at $x = 1$, and then it equals $y - f(1) = y - 1$. Thus, we obain

$$f_D^*(y) = \begin{cases} y - 1, & \text{if } y \leq 2, \\ \frac{1}{4}y^2, & \text{if } y > 2. \end{cases}$$

Hence, f_D^* is now given by two closed formulas, one for $y \leq 2$ and one for $y \geq 2$. Also note that the resulting function is continuous in $y = 2$ but not differentiable. More examples of this phenomenon can be found in Ben-Tal et al. (2013, table 2). Many optimization algorithms cannot cope with nondifferentiable functions. A more tractable expression for $f_D^*(y)$ can be obtained by writing

$$f_D(x) = f(x) + \delta_D(x).$$

By applying the sum rule for conjugates (i.e., line 6 in Table E.2), we obtain

$$\begin{aligned} f_D^*(y) &= \inf_{y^1, y^2} \{f^*(y^1) + \delta_D^*(y^2) \mid y^1 + y^2 = y\} \\ &= \inf_{y^1} \{f^*(y^1) + \delta_D^*(y - y^1)\}. \end{aligned} \tag{B.2}$$

This expression can be used immediately when forming Fenchel's dual problem becauseI it is the "inf" of a tractable function. In the preceding example, we have

$$\delta_D^*(y) = \sup_{x\geq 1}\{yx\} = \begin{cases} y, & \text{if } y \leq 0 \\ \infty & \text{otherwise.} \end{cases}$$

Hence, when applying (B.2) to the example, we obtain

$$\begin{aligned} f_D^*(y) &= \inf_z\{f^*(z) + \delta_D^*(y-z)\} \\ &= \inf_z\left\{\frac{1}{4}z^2 + y - z \mid y - z \leq 0\right\} \\ &= \inf_z\left\{\left(\frac{1}{2}z - 1\right)^2 + y - 1 \mid y \leq z\right\}. \end{aligned}$$

When forming a dual problem, this expression is more convenient because the "branches" disappeared.

B.8. Support Function of a Cone

In this section, we compute the support function of a convex cone $\mathcal{K}$ in $\mathbf{R}^n$. We have

$$\delta_{\mathcal{K}}^*(y) = \sup_{x\in\mathcal{K}} y^T x.$$

Recall that the dual cone of $\mathcal{K}$ is defined by

$$\mathcal{K}_* = \{y \mid y^T x \geq 0,\ \forall x \in \mathcal{K}\}.$$

If $y^T x > 0$ for some $x \in \mathcal{K}$ and $\lambda > 0$, then $\lambda x \in \mathcal{K}$, whereas $y^T(\lambda x) = \lambda y^T x$ goes to infinity if λ goes to infinity. Hence, $\delta_K^*(y)$ is finite only if $y^T x \leq 0$ for all $x \in \mathcal{K}$, and then the maximal value is attained at $x = 0$. As a consequence, we have

$$\delta_{\mathcal{K}}^*(y) = 0, \quad \operatorname{dom}\delta_{\mathcal{K}}^* = -\mathcal{K}_*.$$

B.9. Discontinuity of the Closure of a Perspective Function

In this section, we deal with an example of a perspective function $(u_1 f_1)^*(y^1)$ that is finite on the closure of its domain despite the fact that, by definition (5), we have $(0 f_1)^*(y^1) = \infty$ if $y_1 \neq 0$. Notwithstanding this behavior, Fenchel's dual problem works out well.

Let $f_0(x) = x$ and $f_1(x) := \delta_{[-1,1]}(x) - 1$. We have

$$\min_x\{f_0(x) \mid f_1(x) \leq 0\} \equiv \min_x\{x \mid -1 \leq x \leq 1\} = -1.$$

Because $\delta_{[-1,1]}^*(y) = |y|$, we have $f_1^*(y) = |y| + 1$. Hence, Fenchel's dual problem is

$$\max_{y^0, y^1, u_1\geq 0}\left\{-\{0 \mid y^0 = 1\} - (u_1 f_1)^*(y^1) \mid y^0 + y^1 = 0\right\} = \max_{u_1\geq 0}\left\{-(u_1 f_1)^*(-1)\right\}.$$

Thus, we have

$$(u_1 f_1)^*(-1) = \begin{cases} u_1\left(\left|\frac{-1}{u_1}\right| + 1\right), & \text{if } u_1 > 0 \\ \infty, & \text{otherwise} \end{cases} = \begin{cases} 1 + u_1, & \text{if } u_1 > 0 \\ \infty, & \text{otherwise.} \end{cases}$$

Hence, the dual problem becomes

$$\sup_{u_1 > 0}\{-1 - u_1\}.$$

The "sup" value equals −1, as it should, but it is not attained! Note that at optimality, u_1 approaches zero, but $y^1 = -1$ stays away from zero.

Appendix C. Fenchel's Duality Theorem

In this appendix, we derive Fenchel's duality theorem from Lagrange's duality theorem. Let f and g be functions from $\mathbf{R}^n$ to $\mathbf{R} \cup \{-\infty, \infty\}$, with f proper convex and g proper concave. Then $f - g$ is convex. We consider the two problems:

$$\begin{aligned} &(P) \qquad \inf\{f(x) - g(x) \mid x \in \operatorname{dom} f \cap \operatorname{dom} g\}, \\ &(D) \qquad \sup\{g_*(y) - f^*(y) \mid y \in \operatorname{dom} f^* \cap \operatorname{dom} g_*\}. \end{aligned}$$

Theorem C.1 (Rockafellar 1970, theorem 31.1). *If ri.dom $f \cap$ ri.dom g is nonempty, then the optimal values of* (P) *and* (D) *are equal, and the maximal value of* (D) *is attained. If f is linear, then ri.dom f can be replaced by* $\operatorname{dom} f$. *Similarly, if g is linear, then ri.dom g can be replaced by* $\operatorname{dom} g$.

Proof. We may reformulate (P) as follows:

$$\inf_{u,v}\{f(u) - g(v) \mid u = v,\ u \in \operatorname{dom} f,\ v \in \operatorname{dom} g\}.$$

Defining

$$\mathscr{C} = \{(u, v) \mid u \in \operatorname{dom} f,\ v \in \operatorname{dom} g\},$$

the problem can be reformulated as

$$\inf_{u,v}\{f(u) - g(v) \mid u - v = 0,\ (u, v) \in \mathscr{C}\}.$$

The hypothesis in the theorem guarantees the existence of a point $x^* \in \text{ri.dom}(f) \cap \text{ri.dom}(g)$. As a consequence, $(x^*, x^*) \in \operatorname{ri}\mathscr{C}$. The constraints are linear, and obviously, (x^*, x^*) is feasible. Hence, we may apply Lagrange's duality theorem. Denoting the optimal value of (P) as p^*, we therefore have

$$\begin{aligned} p^* &= \max_y \inf_{u,v}\{f(u) - g(v) + (v-u)^T y \mid u \in \operatorname{dom}(f),\ v \in \operatorname{dom}(g)\} \\ &= \max_y \inf_{u,v}\{y^T v - g(v) - (u^T y - f(u)) \mid u \in \operatorname{dom}(f),\ v \in \operatorname{dom}(g)\}. \end{aligned}$$

Note that y is fixed in the last minimization problem and that u and v are independent variables. Hence, we obtain

$$\begin{aligned} p^* &= \max_y \left\{ \inf_{v \in \operatorname{dom}(g)} \{y^T v - g(v)\} - \sup_{u \in \operatorname{dom}(f)} \{u^T y - f(u)\} \right\} \\ &= \max_y \{g_*(y) - f^*(y)\}, \end{aligned}$$

where we used that $g(v) = -\infty$ if $v \notin \operatorname{dom}(g)$ and $f(v) = \infty$ if $v \notin \operatorname{dom}(f)$. For the proof of the remaining statements in the theorem, we refer to Rockafellar (1970). □

We just derived Fenchel's duality theorem from Lagrange's duality theorem. The converse, deriving Lagrange's duality theorem from Fenchel's duality theorem, is also possible. Thus, in essence, both theorems are equivalent. This has been worked out in Magnanti (1974).

Appendix D. Derivation of Fenchel's Dual Problem

In this appendix, we show how Fenchel's dual problem (D) of (P) in Section 3 can be obtained from Fenchel's duality theorem (Theorem C.1). Recall that (P) is the problem

$$(P) \qquad \inf_x\{f_0(x) \mid f_i(x) \le 0,\ i = 1, \ldots, m\},$$

where the functions $f_i : \mathbf{R}^n \to [-\infty, \infty]$ are proper convex functions for $i = 0, \ldots, m$. Let $\mathscr{S}_i := \{x \mid f_i(x) \le 0\}$, for $i = 1, \ldots, m$, and $\mathscr{S} = \cap_{i=1}^m \mathscr{S}_i$. Obviously, S is the feasible region of (P). The assumption that (P) is Slater regular implies that the set S is nonempty and that x ∈ ri.dom f for some $x \in S$. Moreover, if f_i is nonlinear, then $\operatorname{ri}\mathscr{S}_i$ is nonempty.
Given the definition of the indicator function $\delta_{\mathscr{S}}(x)$, we have

$$\inf_{x \in \mathscr{S}}\{f_0(x)\} = \inf_x\{f_0(x) + \delta_{\mathscr{S}}(x)\}.$$

Because f_0 is proper convex, $-\delta_{\mathscr{S}}$ proper concave, and

$$\text{ri.dom} f_0 \cap \text{ri.dom}\,\delta_S = \cap_{i=0}^m \text{ri.dom} f_i \neq \emptyset,$$

we may apply Fenchel's duality theorem. Hence, we may write

$$\begin{aligned} \inf_{x \in \mathscr{S}}\{f_0(x)\} &= \inf_x\{f_0(x) - (-\delta_{\mathscr{S}}(x))\} \\ &= \max_y\{(-\delta_{\mathscr{S}})_*(y) - f_0^*(y)\} \\ &= \max_y\{-f_0^*(y) - \delta_{\mathscr{S}}^*(-y)\}, \end{aligned} \tag{D.1}$$

where the last equality is due to (6).
At this stage, we need to compute $\delta^*_{\mathcal{S}}(-y)$. Because the sets $\mathcal{S}_i$ satisfy the Slater condition, we may apply corollary 16.4.1 from Rockafellar (1970), which gives (see also Table E.3, line 6)

$$\delta^*_{\mathcal{S}}(-y) = \min_{\{y^i\}_{i=1}^m}\left\{\sum_{i=1}^m \delta^*_{\mathcal{S}_i}(y^i) \;\middle|\; \sum_{i=1}^m y^i = -y\right\}.$$

Substitution into (D.1) yields

$$\begin{aligned}\inf_{x\in\mathcal{S}}\{f_0(x)\} &= \max_y\left\{-f_0^*(y) - \min_{\{y^i\}_{i=1}^m}\left\{\sum_{i=1}^m \delta^*_{\mathcal{S}_i}(y^i) \;\middle|\; \sum_{i=1}^m y^i = -y\right\}\right\}\\ &= \max_{y,\{y^i\}_{i=1}^m}\left\{-f_0^*(y) - \sum_{i=1}^m \delta^*_{\mathcal{S}_i}(y^i) \;\middle|\; \sum_{i=1}^m y^i = -y\right\},\end{aligned} \tag{D.2}$$

where the last expression follows because the "min" operator is absorbed by the "max" operator. Thus, we have shown that the maximization problem in (D.2) has the same optimal value as (P), and hence, it is a strong dual problem for (P). Replacing y by y^0, this dual problem simplifies to

$$\max_{\{y^i\}_{i=0}^m}\left\{-f_0^*(y^0) - \sum_{i=1}^m \delta^*_{\mathcal{S}_i}(y^i) \;\middle|\; \sum_{i=0}^m y^i = 0\right\}. \tag{D.3}$$

A nice feature of this dual problem is that its components are in one-to-one correspondence with the basic elements of the primal problem: its objective function and each of the constraints functions. By computing the conjugate of the objective function and the support function of each of the constraint sets, one obtains (D.3) in a structured way.
Usually, an explicit formula for the support function is in general not available. As already mentioned in Section 2, the support function of S_i can be expressed in the perspective of the conjugate f_i^* of the constraint function f_i. This is a consequence of the following lemma.

Lemma D.1. *We have*

$$\delta^*_{S_i}(y) = \min_{u\geq 0}\{(uf_i)^*(y)\}.$$

Proof. According to the definition of $\delta^*_{S_i}$, we have

$$\delta^*_{S_i}(y) = \sup_{x\in S_i}\{y^T x\} = \sup_x\left\{y^T x \mid f_i(x) \leq 0\right\}.$$

By writing the last problem as a minimimization problem and then applying Lagrange's duality theorem, we obtain

$$\delta^*_{S_i}(y) = \min_{u\geq 0}\sup_x\left\{y^T x - uf_i(x)\right\} = \min_{u\geq 0}\left\{(uf_i)^*(y)\right\},$$

which proves the lemma. □

By applying Lemma D.1 to (D.3), we obtain

$$\inf_{x\in\mathcal{S}}\{f_0(x)\} = \max_{\{y^i\}_{i=0}^m}\left\{-f_0^*(y^0) - \sum_{i=1}^m \min_{u_i\geq 0}\{(u_i f_i)^*(y^i)\} \;\middle|\; \sum_{i=0}^m y^i = 0\right\}.$$

Once again, the "min" operator is absorbed by the "max" operator, leaving us with the following dual problem of (P):

$$(D) \qquad \max_{\{y^i\}_{i=0}^m, u}\left\{-f_0^*(y^0) - \sum_{i=1}^m (u_i f_i)^*(y^i) \;\middle|\; \sum_{i=0}^m y^i = 0,\, u \geq 0\right\}.$$

Appendix E. Tables of Conjugates and Support Functions

Expressions for the convex (or concave) conjugates of many functions can be found in the existing optimization literature. Strange enough, in textbooks on optimization, tables of conjugates functions are rare. One positive example is in Borwein and Lewis (2000, p. 50), but the table in this book is not exploited in a structured way to derive dual problems.

An overview of conjugates is presented in Table E.1, with references to existing literature. As we made clear in Section 2, concave conjugates easily follow from these formulas by using (6).

In Table E.2, we present some transformation rules that are useful for the computation of conjugate functions. In line 9 of this table, the function g is assumed to be real valued. For the more general case with g vector valued, we refer to Hiriart-Urruty (2006).

Table E.3 gives a list of support functions, and Table E.4 provides a list of cones and their dual cones.

Table E.1. Examples of Convex Conjugate Functions Available from the Literature

Number	$f(x)$ (domain)	$f^*(y)$ (domain)	Reference
1	$a^T x + b$	$-b$ $(y = a)$	Boyd and Vandenberghe (2004, p. 91)
2	e^x	$y \log y - y$ $(y \geq 0, f^*(0) = 0)$	Rockafellar (1970, p. 105)
3	$-\log x$ $(x > 0)$	$-\log(-y) - 1$ $(y < 0)$	Rockafellar (1970, p. 106)
4	$x \log x$ $(x \geq 0, f(0) = 0)$	e^{y-1}	Boyd and Vandenberghe (2004, p. 92)
5	$\log(\sum_{i=1}^n e^{x_i})$	$\sum_{i=1}^n y_i \log y_i$ $(y \geq 0, \mathbf{1}^T y = 1)$	Rockafellar (1970, p. 148)
6	$\lVert x \rVert$	0 $(\lVert y \rVert_* \leq 1)$	Boyd and Vandenberghe (2004, p. 93)
7	$\frac{1}{2}\lVert x \rVert^2$	$\frac{1}{2}\lVert y \rVert_*^2$	Boyd and Vandenberghe (2004, p. 93)
8	$\frac{1}{p}x^p$ $(x \geq 0, p > 1)$	$\inf_z\{\frac{1}{q}z^q \mid z \geq y, z \geq 0\}$	Section A.5
9	$-\frac{1}{p}x^p$ $(x \geq 0, 0 < p < 1)$	$-\frac{1}{q}\lvert y \rvert^q$ $(y < 0)$	Rockafellar (1970, p. 106)
10	$\frac{1}{p}\lvert x \rvert^p$ $(p > 1)$	$\frac{1}{q}\lvert y \rvert^q$	Rockafellar (1970, p. 106)
11	$\frac{1}{px^p}$ $(x > 0, p > 0)$	$-\frac{p+1}{p}(-y)^{\frac{p}{p+1}}$ $(y \leq 0)$	Section A.6
12	$-(a^2 - x^2)^{1/2}$, $(\lvert x \rvert \leq a)$	$a(1 + y^2)^{1/2}$	Rockafellar (1970, p. 106)
13	$\frac{1}{2}x^T P x$, $P \succeq 0$	$\frac{1}{2}y^T P^{\dagger} y$ $(y = Pz)$	Rockafellar (1970, p. 108)
14	$-\log\det X$ $(X \in^n_{++})$	$-\log\det(-Y) - n$ $(-Y \in^n_{++})$	Boyd and Vandenberghe (2004, p. 92)
15	$-\sqrt{c^T X c}$, $X \succeq 0$	$\inf_{z<0}\{-\frac{1}{4z} \mid -Y \succeq zcc^T\}$	Gorissen and den Hertog (2015, p. 10)
16	$\cosh x$	$y \sinh^{-1} y - \sqrt{1 + y^2}$	Borwein and Lewis (2000, p. 50)
17	$-\log(\cos x)$ $(\lvert x \rvert < \frac{\pi}{2})$	$y \tan^{-1} y - \frac{1}{2}\log(1 + y^2)$	Borwein and Lewis (2000, p. 50)
18	$\log(\cosh x)$	$y \tanh^{-1} y + \frac{1}{2}\log(1 - y^2)$ $(\lvert y \rvert < 1)$	Borwein and Lewis (2000, p. 50)

Notes. In lines 6 and 7, $\lVert . \rVert$ denotes any norm, and $\lVert . \rVert_*$ the dual of this norm. In lines 8–10, the numbers p and q are related according to $\frac{1}{p} + \frac{1}{q} = 1$. In lines 13 and 14, $P^{\dagger}$ denotes the generalized inverse of matrix P. In lines 15 and 16, $\mathbf{S}_+^n$ and $\mathbf{S}_{++}^n$ denote the set of positive-semidefinite and positive-definite $n \times n$ matrices, respectively.

Table E.2. Rules for Calculating the Conjugate Function

Number	Function and assumptions	Conjugate function	Reference
1	$f(x) + a$	$f^*(y) - a$	Boyd and Vandenberghe (2004, p. 95)
2	$f(x + a)$	$f^*(y) - a^T y$	Rockafellar (1970, p. 107)
3	$f(x) + a^T x$	$f^*(y - a)$	Rockafellar (1970, p. 107)
4	$af(x)$, $a > 0$	$af^*\left(\frac{y}{a}\right)$	Rockafellar (1970, p. 140)
5	$f(ax)$, $a > 0$	$f^*\left(\frac{y}{a}\right)$	Rockafellar (1970, p. 107)
6	$\sum_{i=1}^m f_i(x)$	$\min_{y^i}\left\{\sum_{i=1}^m f_i^*(y^i) \,\middle\vert\, \sum_{i=1}^m y^i = y\right\}$	Rockafellar (1970, p. 145)
7	$\sum_{i=1}^m f_i(x^i)$	$\sum_{i=1}^m f_i^*(y^i)$	Boyd and Vandenberghe (2004, p. 95)
8	$f(Ax + b)$		
a	A nonsingular	$f^*(A^{-T}y) - b^T A^{-T} y$	Rockafellar (1970, p. 107)
b	otherwise	$\inf_z\{f^*(z) - b^T z \mid A^T z = y\}$	Section B.4
9	$f(g(x))$, f nondecreasing	$\inf_{z \geq 0}\{(zg)^*(y) + f^*(z)\}$	Hiriart-Urruty (2006, p. 39)
10	$\operatorname{dom} f = \mathbf{R}_{++}$	$\inf_z\{z \mid (f^{\diamond})^*(-z) \leq -y\}$	Gushchin (2008, p. 453)
11	$h : \mathbf{R} \to \mathbf{R}$ strictly increasing	$(h^{-1})^*(y) = -(h_*)^{\diamond}(y)$, $y > 0$	Ben-Tal et al. (1991, equation 3.3)
12	$xf(\frac{a}{x})$, $\operatorname{dom} f = \mathbf{R}^n$, $x \geq 0$	$\inf\{a^T z : f^*(-z) \leq -y\}$	Section B.6
13	$\max_i\{f_i(x) \mid i = 1, \ldots, m\}$	$\min_{z \in S}\left\{\sum_{i=1}^m (z_i f_i)^*(y^i) \,\middle\vert\, \sum_{i=1}^m y^i = y\right\}$	Rockafellar (1970, p. 149)

Notes: The functions $f(x)$, $f_i(x)$, and $h(x)$ in this table are real-valued functions, $f(x)$ and $f_i(x)$ are convex, and $h(x)$ is concave. In line 6, we assume that $\cap_{i=1}^m \operatorname{ri.dom} f_i \neq \emptyset$; if $f_i(x)$ is linear for some i, then the corresponding $\operatorname{ri.dom} f_i$ can be replaced by $\operatorname{dom} f_i$ in this condition. We call this the *sum rule for conjugate functions*. In line 7, $x^1 : x^m$ is a partition of x. We will refer to line 8 as the *linear substitution rule*. In lines 10 and 11, $f^{\diamond}$ denotes the *adjoint* of $f : \mathbf{R}_{++} \to \mathbf{R}$, which is defined as $f^{\diamond}(x) := xf(1/x)$. In line 13, S denotes the unit simplex.

Table E.3. Support Functions

Number	$S,\ S \neq \emptyset$	$\delta_S^*(y)$	Reference
1	$\{x \mid Ax = b\}$	$\min_z\{b^Tz \mid A^Tz = y\}$	Boyd and Vandenberghe (2004, p. 380)
2	$\{x \mid Ax \leq b\}$	$\min_z\{b^Tz \mid A^Tz = y, z \geq 0\}$	Boyd and Vandenberghe (2004, p. 380)
3	$\{x \mid b - Ax \in \mathcal{K}\}$	$\min_z\{b^Tz \mid A^Tz = y, z \in \mathcal{K}_*\}$	Ben-Tal et al. (2015, p. 272)
4	$\{x \mid \|x\|_p \leq \rho\}$	$\rho\|y\|_q,\ \frac{1}{p}+\frac{1}{q}=1$	Ben-Tal et al. (2015, p. 272)
5	$\{x \mid f_i(x) \leq 0, i = 1, \ldots, m\}$	$\min_{u\geq 0, y^i}\left\{\sum_{i=1}^m (u_if_i)^*(y^i) \,\middle\vert\, \sum_{i=1}^m y^i = y\right\}$	Rockafellar (1970, p. 146)
6	$S = \bigcap_{i=1}^m S_i$	$\min_{y^i}\left\{\sum_{i=1}^m \delta_{S_i}^*(y^i) \,\middle\vert\, \sum_{i=1}^m y^i = y\right\}$	Rockafellar (1970, p. 146)
7	$S = S_1 \times \cdots \times S_m$	$\sum_{i=1}^m \delta_{S_i}^*(y^i)$, $y^1 : y^m$ is a partition of y	Ben-Tal et al. (2015, p. 294)

Notes: In line 3, $\mathcal{K}$ is a pointed cone, and if $\mathcal{K}$ is nonlinear, then it is assumed that S contains a strictly feasible point. The functions f_i and the sets S_i in lines 5–7 are closed convex. In line 5, we assume that $\cap_{i=1}^m \mathrm{ri.dom} f_i \neq \emptyset$; if $f_i(x)$ is linear for some i, then the corresponding $\mathrm{ri.dom} f_i$ can be replaced by $\mathrm{dom} f_i$ in this condition. Similarly, in line 6, we assume that $\cap_{i=1}^m \mathrm{ri}\, S_i \neq \emptyset$, and if S_i is polyhedral, then the corresponding $ri\, S_i$ can be replaced by S_i.

Table E.4. An Overview of the Most Commonly Used Cones and Their Dual Cones

Name	Cone $\mathcal{K}$	Dual cone $\mathcal{K}_*$
Nonnegative orthant	$\mathbf{R}_+^n$	$\mathbf{R}_+^n$
Second-order cone	$\mathbf{L}^n = \{(x,t) \in \mathbf{R}^{n+1} \mid \|x\| \leq t\}$	$\mathbf{L}^n$
Positive-semidefinite cone	$\mathbf{S}_n^+ = \{X \in \mathbf{S}^n \mid X \succeq 0\}$	$\mathbf{S}_n^+$
Power cone	$\{x \in \mathbf{R}^n \mid \lvert x_n\rvert \leq \prod_{i=1}^{n-1} x_i^{\alpha_i}\}$	$\{y \in \mathbf{R}^n \mid \lvert y_n\rvert \leq \prod_{i=1}^{n-1}(y_i/\alpha_i)^{\alpha_i}\}$
Exponential cone	$\mathrm{cl}\,\{x \in \mathbf{R}^3 \mid x_2e^{x_1/x_2} \leq x_3, x_2 > 0\}$	$\mathrm{cl}\,\{y \in \mathbf{R}^3 \mid -y_1e^{y_2/y_1-1} \leq y_3, y_1 < 0\}$
Copositive cone	$\{X \in \mathbf{S}^n \mid v^TXv \geq 0\ \forall v \geq 0\}$	$\{Y \in \mathbf{S}^n \mid \exists A \in \mathbf{R}_+^{n\times m}\ Y = AA^T\}$

Notes: The first three cones are self-dual. The set $\mathbf{S}^n$ is the set of symmetric $n \times n$ matrices. The closure operator is denoted by cl.

Endnotes

[1] Here we adopt the calculation rules $0\infty = \infty 0 = 0$, as in Rockafellar (1970, p. 24).

[2] As in Boyd and Vandenberghe (2004), the notation $x \succeq 0$ defines a vector $x \in \mathbf{R}^n$ to be nonnegative; only if $n = 1$ do we use the notation $x \geq 0$.

[3] We use e to denote the base of the natural logarithm, according to ISO Standard 80000-2:2009.

References

Alber M, Nüsslin F (2001) A representation of an NTCP function for local complication mechanisms. *Physics Medicine Biol.* 46(2):439–447.
Ben-Israel A, Greville TNE (2003) *Generalized Inverses: Theory and Applications*, 2nd ed. (Springer, New York).
Ben-Tal A, Nemirovski A (2001) *Lectures on Modern Convex Optimization: Analysis, Algorithms, Engineering Applications* (SIAM, Philadelphia).
Ben-Tal A, Ben-Israel A, Teboulle M (1991) Certainty equivalents and information measures: duality and extremal principles. *J. Math. Anal. Appl.* 157(1):211–236.
Ben-Tal A, den Hertog D, Vial J-P (2015) Deriving robust counterparts of nonlinear uncertain inequalities. *Math. Programming* 149(1–2):265–299.
Ben-Tal A, den Hertog D, De Waegenaere A, Melenberg B, Rennen G (2013) Robust solutions of optimization problems affected by uncertain probabilities. *Management Sci.* 59(2):341–357.
Borwein JM, Lewis AS (2000) *Convex Analysis and Nonlinear Optimization: Theory and Examples* (Springer, New York).
Boyd S, Vandenberghe L (2004) *Convex Optimization* (Cambridge University Press, Cambridge, UK).
den Hertog D (1994) *Interior Point Approach to Linear, Quadratic and Convex Programming* (Kluwer, Dordrecht).
Duffin RJ, Peterson EL, Zener C (1967) *Geometric Programming: Theory and Application* (Wiley, New York).
Fiacco AV, McCormick GP (1990) *Nonlinear Programming: Sequential Unconstrained Minimization Techniques* (SIAM, Philadelphia).
Gorissen BL, den Hertog D (2015) Robust nonlinear optimization via the dual. Preprint, submitted April 29, http://www.optimization-online.org/DB_HTML/2015/04/4886.html.
Gushchin AA (2008) On an extension of the concept of *f*-divergence. *Theory Probab. Appl.* 52(3):439–455.
Hiriart-Urruty J-B (2006) A note on the Legendre-Fenchel transform of convex composite functions. Alart P, Maisonneuve O, Rockafellar RT, eds. *Nonsmooth Mechanics and Analysis* (Springer, New York), 35–46.
Jansen B, de Jong JJ, Roos C, Terlaky T (1997) Sensitivity analysis in linear programming: Just be careful! *Eur. J. Oper. Res.* 101(1):15–28.
Magnanti TL (1974) Fenchel and Lagrange duality are equivalent. *Math. Programming* 7(1):253–258.
Niemierko A (1999) A generalized concept of equivalent uniform dose (EUD). *Medical Physics* 26(6):1100.
Rockafellar RT (1970) *Convex Analysis* (Princeton University Press, Princeton, NJ).

Romeijn EH, Dempsey JF, Li JG (2004) A unifying framework for multi-criteria fluence map optimization models. *Physics Medicine Biol.* 49(10):1991–2013.

Slater ML (2014) Lagrange multipliers revisited. Giorgi G, Kjeldsen TH, eds. *Traces and Emergence of Nonlinear Programming* (Birkhäuser Verlag, Basel, Switzerland), 293–306.

Stavreva P, Hristov D, Warkentin B, Fallone B (2003) Investigating the effect of cell repopulation on the tumor response to fractionated external radiotherapy. *Medical Physics* 30(5):2948–2958.

Terlaky T (1985) On$\ell_p$programming. *Eur. J. Oper. Res.* 22(1):70–100.

Tomioka R (2010) Three strategies to derive a dual problem. Preprint, submitted May 8, https://www.scribd.com/document/31522215/Three-strategies-to-derive-a-dual-problem.