A universal and structured way to derive dual optimization problem formulations

Kees Roos\textsuperscript{1}, Marleen Balvert\textsuperscript{2}, Bram L. Gorissen\textsuperscript{3} and Dick den Hertog\textsuperscript{1}

\textsuperscript{1} Department of Electrical Engineering, Mathematics and Computer Science, Technical University Delft, 2628 CD Delft, The Netherlands
\textsuperscript{2} Department of Econometrics and Operations Research/Center for Economic Research (CentER), Tilburg University, PO Box 90153, 5000 LE Tilburg, The Netherlands
\textsuperscript{3} Faculty of Exact Sciences, Mathematics, Free University, 1081 HV Amsterdam, The Netherlands

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 in 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. We demonstrate the method by deriving dual problems for several classical problems and also for a practical model for radiotherapy treatment planning. Additional material is presented in appendices, including useful tables for finding conjugate functions of many functions.

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, 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 three 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, i.e., on how much their values may differ from the optimal value.

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.

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. We consider this the most structured way to teach deriving dual optimization problems. Using this structured way, one might even computerize the derivation of the dual optimization problem. We must emphasize that the ingredients in this approach are not new. About 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 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. On the other hand, 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.

The outline of the paper is as follows. In the next section 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 the Appendices C, D and E. 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 F are of crucial importance. In Appendix A we demonstrate how conjugates for some functions are computed.

The use of the scheme in Section 4 is illustrated in Section 5, 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, also problems with a variety of constraint functions can be handled systematically. In Section 5.7 we demonstrate this by applying our approach to a model for radiotherapy treatment.

## 2 Preliminaries

We recall in this section some notations and terminology that are common in the literature. The effective domain of a convex function \( f : \mathbb{R}^n \to \mathbb{R} \) is defined by

\[
\text{dom } f = \{ x \mid f(x) < \infty \}.
\]

We define \( f(x) = \infty \) if \( x \notin \text{dom } f \). A convex function \( f \) is proper if \( f(x) > -\infty \) for all \( x \in \mathbb{R}^n \) and \( f(x) < \infty \) for at least one \( x \in \mathbb{R}^n \).

The indicator function of a set \( S \subseteq \mathbb{R}^n \) 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 \( S \) is convex, and proper if \( S \) is nonempty. Obviously, we have

\[
x \in S \iff \delta_S(x) \leq 0.
\]

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

For any function \( f \) its convex conjugate \( f^* \) is defined by

\[
f^*(y) := \sup_{x \in \text{dom } f} \{ y^T x - f(x) \}.
\]

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

\[
\delta_S^*(y) = \sup_{x \in S} \{ y^T x \}.
\]

Since \( f^*(y) \) is the pointwise maximum of linear functions (linear in \( y \)), \( f^* \) is convex. Moreover, \( f^* \) is closed, \( (\text{cl } f)^* = f^* \) and \( f^{**} = \text{cl } f \), which implies \( f^{**} = f \) if \( f \) is closed (Rockafellar, 1970, Theorem 12.2). Our

---

2
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 \( \text{ri} S \) is nonempty if \( f \) is nonlinear, we have (cf. Lemma 1):

\[
\delta_S^*(y) = \min_{\lambda \geq 0} \{(\lambda f)^*(y)\}.
\]

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

\[
(\lambda f)^*(y) = \sup_x \left\{ y^T x - \lambda f(x) \right\} = \sup_x \left\{ \lambda \left( x^T \frac{y}{\lambda} - f(x) \right) \right\} = \lambda f^* \left( \frac{y}{\lambda} \right),
\]

and for \( \lambda = 0 \) we get\(^1\)

\[
(0 f)^*(y) = \sup_x \left\{ y^T x \right\} = \begin{cases} 0, & \text{if } y = 0 \\ \infty, & \text{otherwise} \end{cases} = \delta_{\{0\}}(y).
\]

The function \( (\lambda f)^* \) is called the perspective function of \( f^* \) (with \( \lambda \) as the new variable). Due to the first equality in (1), the perspective function of \( f^* \) is convex (more precisely: jointly convex in \( y \) and \( \lambda \)). It may be worth noting that \( (\lambda f)^*(y) \) may be well-defined even if \( \lambda \) approaches zero whereas \( y \) stays away from zero. Closing the perspective function can therefore result in the wrong dual. An example of this phenomenon can be found in Section B.8.

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

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

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

\[
g_*(y) = -f^*(-y) = -(g^*)^*(-y).
\]

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

### 3 Fenchel’s dual problem of a convex optimization problem

We consider the following (primal) convex optimization problem

\[
(P) \quad \inf_x \{ f_0(x) \mid f_i(x) \leq 0, \ i = 1, \ldots, m \},
\]

where the functions \( f_i : \mathbb{R}^n \to \mathbb{R} \) are proper convex functions, for \( i = 0, \ldots, m \). Fenchel’s dual problem of \((P)\) is given by\(^2\)

\[
(D) \quad \sup_{\{y^i\}_{i=0}^m, u} \left\{ -f_0^*(y^0) - \sum_{i=1}^m (u_i f_i)^*(y^i) \mid \sum_{i=0}^m y^i = 0, \ u \succeq 0 \right\}.
\]

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

It is clear that \((D)\) can be obtained straightforwardly by computing the convex conjugates \( f_i^* \) of the functions \( f_i \) in \((P)\). This important fact is the main motivation for this paper.

In the sequel we assume that \((P)\) is Slater regular (Slater 2014). In other words, we assume that there exists a feasible \( x \in \text{ri} \text{dom} f_0 \) such \( f_i(x) < 0 \) for each \( i = 1, \ldots, m \) for which \( f_i \) is nonlinear. In that case the optimal values of \((P)\) and \((D)\) are equal; for a proof we refer to Appendix E. We should add that if \( f_0 \) is linear then \( x \in \text{ri} \text{dom} f_0 \) can be replaced by \( x \in \text{dom} f_0 \) in the Slater condition.

In the sequel, 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

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

In the above expressions for \((D)\) and \((D')\) we used the sup operator to indicate that we are dealing with a maximization problem. Below 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 below, it is not excluded that the maximal value is attained, and similarly for the \( \inf \) operator.

\(^1\)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 \( x \in \mathbb{R}^n \) to be nonnegative; only if \( n = 1 \) we use the notation \( x \geq 0 \).
4 Recipe for setting up Fenchel’s dual problem

As became clear in the previous section, Fenchel’s dual gives us a straightforward method for finding the dual problem of \( (P) \). In this method the Tables 1-3 in Appendix F are useful. The method can be split into four steps as discussed below. We will give some demonstrations of this recipe in Section 5.

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, i.e., 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 that purpose one can use the three tables in Appendix F. Table 1 contains expressions for the conjugates of some basic functions and the domains of these functions, Table 2 some formulas for conjugates of function transformations, and Table 3 some support functions of common sets. Usually a conjugate function (or its perspective function) occurring in \( (D) \) or \( (D’) \) is written as \( \inf_x \{h(x,y)\} \), where \( h(x,y) \) is jointly convex in \( x \) and \( y \). A frequently used tool in our approach is the obvious equality

\[
\sup_y \left\{ -\inf_x h(x,y) \right\} = \sup_{x,y} \{-h(x,y)\}.
\]

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. If the supremum is attained we use max instead of sup (and in case of a minimization problem, min instead of inf). 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; in most cases it also enables to eliminate the variables \( y_i \). In some cases we get a conjugate function that consists of ‘branches’, i.e., one obtains different formulas for \( f^*(y) \) on several parts of the domain of \( f^* \). This phenomenon is demonstrated in Appendix B.6, where it is also shown how this can be prevented.

Step 3: Derive tractable convex expressions for perspectives of conjugates

Tables 1 and 2 can be used to obtain the conjugate for many functions \( f \). When substituting such a conjugate in \( (D) \) or \( (D’) \), its contribution to the dual problem is the term

\[
u f^* \left( \frac{y}{u} \right), \text{ with } \frac{y}{u} \in \text{dom } f^* ,
\]

where \( u > 0 \). One may be troubled by the fact that the quotient \( \frac{y}{u} \) is not jointly convex in \( y \) and \( u \). Indeed, sometimes the resulting formulation of the dual problem may be not convex. But a simple substitution yields a convex formulation. For example, when using line 7b in Table 2 one has

\[
f^*(y) = \inf_{z} \{h^*(z) - b^T z \mid A^T z = y\}.
\]

This gives rise to the term \( (uf)^*(y) \) in the dual problem, i.e.,

\[
(uf)^*(y) = \inf_{z} \left\{ uh^*(z) - ub^T z \mid A^T z = \frac{y}{u} \right\}, \quad u > 0
\]

\[
= \inf_{z} \left\{ uh^*(z) - b^T (uz) \mid A^T (uz) = y \right\}, \quad u > 0.
\]

Due to 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

\[
(uf)^*(y) = \inf_{\tilde{z}} \left\{ uh^* \left( \frac{z}{u} \right) - b^T \tilde{z} \mid A^T \tilde{z} = y \right\}
\]

\[
= \inf_{\tilde{z}} \left\{ (uh)^* (\tilde{z}) - b^T \tilde{z} \mid A^T \tilde{z} = y \right\}.
\]

Obviously, both the objective function and the constraint are now convex. In Section 5 one will find many substitutions of this type, that 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 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.

5.1 Linear optimization

We consider the standard linear optimization problem:

\[
\inf_x \{ c^T x \mid Ax = b, \ x \in \mathbb{R}^n_+ \},
\]

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

\[
f^*(y) = 0, \quad \text{dom } f^* = \{c\}.
\]

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

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

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

\[
g_2^*(y) = 0, \quad \text{dom } g_2^* = \{y \mid y \preceq 0\} = -\mathbb{R}^n_+.
\]

It follows that Fenchel’s dual problem is given by

\[
\sup_z \left\{ -f^*(y^0) - \sum_{i=1}^2 g_i^* (y^i) \mid \sum_{i=0}^2 y^i = 0 \right\}.
\]

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

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

We can eliminate the vectors \(y^1\), which gives

\[
\sup_z \left\{ -b^T z \mid c + A^T z \succeq 0 \right\}.
\]

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

\[
\inf_x \{ c^T x \mid Ax = b, \ x \succeq 0 \} = \sup_z \{ b^T z \mid A^T z \preceq c \}.
\]

5.2 Conic optimization

Let \(K\) be a convex cone in \(\mathbb{R}^n\). We consider the standard linear optimization problem over the cone \(K\):

\[
\inf_x \{ c^T x \mid Ax = b, \ x \in K \},
\]

with \(A, b\) and \(c\) as in Section 5.1. This problem differs from the linear optimization problem in Section 5.1 only in the set constraint \(x \in K\), so it suffices to find the support function of this constraint. Denoting this function as \(g_2^*\) we have (cf. Table 3, line 3)

\[
g_2^*(y) = 0, \quad \text{dom } g_2^* = -K_+.
\]

Just as in the previous section we obtain that Fenchel’s dual problem is given by

\[
\sup_z \left\{ -b^T z \mid A^T z + c \in K_+ \right\}.
\]

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

\[
\inf_x \{ c^T x \mid Ax = b, \ x \in K \} = \sup_z \{ b^T z \mid c - A^T z \in K_+ \}.
\]
5.3 Quadratically constrained quadratic optimization

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

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

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 (cf. Table 1, lines 1 and 13)

$$g_i^*(y) = \frac{1}{2} y^T P_i^T y, \quad \text{dom } g_i^* = \{ y \mid y = P_i z, z \in \mathbb{R}^n \},$$

$$h_i^*(x) = -r_i, \quad \text{dom } h_i^* = \{ q_i \}.$$

Hence, by using the sum rule in Table 2 (line 5), we get

$$f_i^*(y) = \min_{y^i, y^g} \left\{ g_i^*(y^i) + h_i^*(y^g) \mid y^i + y^g = y \right\}$$

$$= \min_{y^i, y^g} \left\{ \frac{1}{2} y^T P_i^T y^i \mid y^i = P_i z \right\} + \left\{ -r_i \mid y^g = q_i \right\} \mid y^i + y^g = y \right\}$$

$$= \min_{z} \left\{ \frac{1}{2} z^T P_i z - r_i \mid P_i z + q_i = y \right\}$$

$$= \min_{z} \left\{ \frac{1}{2} z^T P_i z - r_i \mid P_i z + q_i = y \right\},$$

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

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

$$= \sup_{u \geq 0, y^i} \left\{ -\sum_{i=0}^{m} u_i \min_{z^i} \left\{ \frac{1}{2} z^i^T P_i z^i - r_i \mid P_i z^i + q_i = \frac{y^i}{u_i} \right\} \mid \sum_{i=0}^{m} y^i = 0, u_0 = 1 \right\}$$

$$= \sup_{u, z^i} \left\{ \sum_{i=0}^{m} u_i (r_i - \frac{1}{2} z^i^T P_i z^i) \mid \sum_{i=0}^{m} u_i (P_i z^i + q_i) = 0, u \geq 0, u_0 = 1 \right\}.$$ 

Defining $z^i = u_i z^i$ this becomes

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

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

5.4 Second order cone optimization

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

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

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

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

In order to derive the conjugate of $g_i(x)$, first note that this is a function of a linear transformation of $x$, i.e., $g_i(x) = \sigma_i(A_i x - b_i)$, where $\sigma_i(x) = \| x \|_2$. Due to line 6 in Table 1 we have $\sigma_i^*(y) = 0$, with $\text{dom } \sigma^* = \{ y \mid \| y \|_2 \leq 1 \}$. Hence, by the linear substitution rule in Table 2 (line 7b), the conjugate of $g_i(x)$ is given by

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

Putting the above results together, we obtain:

$$f_i^*(y) = \inf_{y^i, y^g} \left\{ g_i^*(y^i) + h_i^*(y^g) \mid y^i + y^g = y \right\}$$

$$= \inf_{y^i, y^g} \left\{ \inf_{z^i} \{ b_i^T z^i \mid \| z^i \|_2 \leq 1, \ A_i^T z^i = y \} + \{ -q_i \mid y^g = -p_i \} \mid y^i + y^g = y \right\}$$

$$= \inf_{z^i} \{ b_i^T z^i - q_i \mid \| z^i \|_2 \leq 1, \ A_i^T z^i - p_i = y \}.$$
When plugging the above expression for \( f_1^*(y) \) into \((D')\) we obtain the following dual for the SOCO problem:

\[
\sup_{y', u \geq 0} \left\{ \sum_{i=0}^{m} u_i \inf_{z^i} \left( b_i^T z^i - q_i, \left\| z^i \right\|_2 \leq 1, A_i^T z^i - p_i = y^i / u_i \right) \right\} = \sum_{i=0}^{m} y^i = 0, u_0 = 1 .
\]

We eliminate the variables \( y^i \) and omit the inf-operator, which gives

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

Introducing \( \tilde{z}^i = u_i z^i \) we get

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

Finally, defining matrices \( A \) and \( P \) according to

\[
A^T = \begin{bmatrix} A_0^T & \cdots & A_m^T \end{bmatrix}, \quad P = \begin{bmatrix} p_0 & \cdots & p_m \end{bmatrix},
\]

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

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

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

### 5.5 Geometric optimization

Consider the geometric optimization problem (cf. Boyd and Vandenberghe, 2004, p. 254):\(^3\)

\[
\min_{\tilde{z}} \left\{ \log \sum_{k=1}^{K_0} e^{\tilde{z}^T x + b_{0k}} \left| \log \sum_{k=1}^{K_i} e^{\tilde{z}^T x + b_{ik}} \leq 0, \quad i = 1, \ldots, m \right\} ,
\]

where \( K_i \geq 1, a_{ik} \in \mathbb{R}^n, b_{ik} \in \mathbb{R} \) for \( i = 0, \ldots, m \). We define new variables \( z_i \in \mathbb{R}^{K_i} \), for \( i = 0, \ldots, m \) and reformulate the problems as

\[
\min_{\tilde{z}} \left\{ \log \sum_{k=1}^{K_0} e^{z_k} \left| \log \sum_{k=1}^{K_i} e^{z_k} \leq 0, \quad i = 1, \ldots, m, \quad z_i = A_i^T x + b_i, \quad i = 0, \ldots, m \right\} ,
\]

where \( A_i \) is the \( n \times K_i \) matrix with columns \( a_{i1}, \ldots, a_{iK_i} \) and \( b_i \) 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 \( i = 0, \ldots, m \),

\[
h_i(x) := \log \sum_{k=1}^{K_i} e^{z_k} , \quad f_i(x) := h_i(A_i^T x + b_i) = \log \sum_{k=1}^{K_i} e^{A_i^T x + b_{ik}} .
\]

Using the linear substitution rule (Table 2, line 7b) and the formula for the conjugate of \( h_i(x) \) (Table 1, line 5), the conjugate of \( f_i(x) \) can be written as:

\[
f_i^*(y) = \inf_{z} \left\{ h^*(z) - b_i^T z \left| A_i^T z = y \right\} = \inf_{z} \left\{ \sum_{k=1}^{K_i} z_k \log z_k - b_i^T z \left| z \geq 0, 1^T z = 1, A_i^T z = y \right\} .
\]

To simplify notation, let \( p^* \) be the optimal value of \((4)\). Using the dual problem as given by \((D')\) we obtain

\[
p^* = \sup_{y', u \geq 0} \left\{ \sum_{i=0}^{m} (u_i f_i)^* \left( y^i \right) \left| \sum_{i=0}^{m} y^i = 0, u_0 = 1 \right\} \right\} = \sup_{y', u \geq 0} \left\{ \sum_{i=0}^{m} u_i \inf_{z^i \geq 0} \left\{ \sum_{k=1}^{K_i} z_k \log z_k - b_i^T z^i \right\} \right\}
\]

s.t. \( \sum_{i=0}^{m} y^i = 0, 1^T z^i = 1, A_i^T z^i = y^i / u_i, \forall i, u_0 = 1 \).

\(^3\)We use \( e \) to denote the base of the natural logarithm, according to ISO Standard 80000-2:2009.
This is equivalent to

\[
p^* = \sup_{y^i, z^i \geq 0} \left\{ \sum_{i=0}^{m} u_i \left[ b_i^T z^i - \sum_{k=1}^{K_i} z_k^i \log z_k^i \right] \right\}
\]

subject to \( \sum_{i=0}^{m} y_i = 0, \quad 1^T z^i = 1, \quad A_i^T z^i = \frac{y_i}{u_i}, \quad \forall i, \quad u_0 = 1 \) ,

\[
= \sup_{u_0 \geq 0, z^i \geq 0} \left\{ \sum_{i=0}^{m} u_i \left[ b_i^T z^i - \sum_{k=1}^{K_i} z_k^i \log z_k^i \right] \right\} \left\{ \sum_{i=0}^{m} u_i A_i^T z^i = 0, \quad 1^T z^i = 1, \quad \forall i, \quad u_0 = 1 \right\} .
\]

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

\[
p^* = \sup_{\tilde{z}^i \geq 0} \left\{ \sum_{i=0}^{m} \left[ b_i^T \tilde{z}^i - \sum_{k=1}^{K_i} z_k^i \log z_k^i \right] \right\} \left\{ \sum_{i=0}^{m} A_i^T \tilde{z}^i = 0, \quad 1^T \tilde{z}^i = u_i, \quad \forall i, \quad u_0 = 1 \right\} .
\]

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 den Hertog (1994); Duffin et al. (1967); Fiacco and McCormick (1968).

5.6 \( \ell_p \)-norm optimization

We consider the following \( \ell_p \)-norm optimization problem den Hertog (1994); Terlaky (1985):

\[
\max_x \left\{ \eta^T x \left| \sum_{i \in I_k} \frac{1}{p_i} |a_i^T x - c_i|^p_i + b_k^T x - d_k \leq 0, \quad k = 1, \ldots, m \right. \right\} , \tag{6}
\]

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

\[
\min_x \left\{ -\eta^T x \left| \sum_{i \in I_k} \frac{1}{p_i} |a_i^T x - c_i|^p_i + b_k^T x - d_k \leq 0, \quad k = 1, \ldots, m \right. \right\} , \tag{7}
\]

whose optimal value is opposite to the optimal value of (6). We define

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

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 (7), we need the conjugates of \( f_0, \ldots, f_m \). The functions \( f_0 \) and \( h_k \) are linear, so one has (Table 1, line 1)

\[
f_0^*(y) = 0, \quad \text{dom } f^* = \{-\eta\}
\]

\[
h_k^*(y) = d_k, \quad \text{dom } h_k^* = \{b_k\}.
\]

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 2, line 5), the linear substitution rule (Table 2, line 7b) and the convex conjugate of \( \sigma_i(x) \) (Table 1, line 10), we obtain the following conjugate function of \( g_k(x) \):

\[
g_k^*(y) = \inf_{y^i} \left\{ \int_{i \in I_k} \int_{z_i \in \mathbb{R}} \left( \frac{1}{q_i} |z_i|^{q_i} + c_i z_i \right) \mid \sum_{i \in I_k} y^i = y \right\}
\]

\[
= \inf_{z_i} \left\{ \int_{i \in I_k} \left( \frac{1}{q_i} |z_i|^{q_i} + c_i z_i \right) \mid \sum_{i \in I_k} a_i z_i = y \right\}.
\]
with $q_i$ such that $1/p_i + 1/q_i = 1$. Now the convex conjugate of $f_k(x)$ can be written as:

$$f_k^*(y) = \inf_{y^1, y^2} \left\{ g_k^*(y^1) + h_k^*(y^2) \mid y^1 + y^2 = y \right\}$$

$$= \inf_{z_i} \left\{ \left( \sum_{i \in I_k} \frac{1}{q_i} |z_i|^{q_i} + c_i z_i \right) + d_k \mid \sum_{i \in I_k} a_i z_i + b_k = y \right\}.$$

Fenchel’s dual of (7) is now given by

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

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

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

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

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

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

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

By changing the sign of the optimal value we arrive at

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

which is the dual problem of (6) that we are looking for. Finally, defining matrices $A$ and $B$ according to

$$A = \left[ \begin{array}{c} a_1 \ldots a_n \end{array} \right], \quad B = \left[ \begin{array}{c} b_1 \ldots b_r \end{array} \right],$$

the last problem gets the form

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

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

### 5.7 Radiotherapy treatment planning

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 $i^{th}$ row of $D$ and $x$ 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, or biological metrics such as the tumor control probability (TCP). Treatment planning models may thus contain different types of constraints and do usually not fit a generic optimization format (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 Stavreva et al. (2003):

\[ TCP(x) = \prod_{i \in T} \exp(-N_0 v_i e^{-\alpha D_i x}), \]

where \( T \) 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 \( \alpha \) describes the radio resistance of cells in the target volume. Taking the log of TCP gives the concave function LTCP Romeijn et al. (2004):

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

that yields the same optimal solution for \( x \) as LTCP is increasing. Therefore, LTCP is optimized instead.

Our planning model contains four constraints on the dose in OARs. For simplicity, we assume 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 \( S \) Niemierko (1999):

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

where \( S \) denotes the set of voxels in structure \( S \), and \( p \geq 1 \) is a structure-dependent parameter for OARs that indicates if it is a serial or a parallel organ. Our model restricts gEUD from above.

The second constraint in our model concerns the normal tissue complication probability (NTCP). We use instead

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

where \( \Delta \) is a structure-dependent parameter that indicates the uniform dose level for which there is a 63% probability on complications and \( p > 1 \), since we restrict ourselves to a non-parallel organ. We would like to place a maximum on NTCP, however, NTCP is not a convex function. Therefore, following Romeijn et al. (2004), we use instead

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

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.

The maximum dose to structure \( S \) is bounded from above in the third constraint. As the number of variables may get large for radiotherapy treatment planning problems, in some cases it is beneficial to use the log-sum-exp function is instead, which is a convex approximation of the maximum function Fredriksson (2012). The fourth constraint restricts the mean dose to structure \( S \) from above. We can now formulate our treatment plan optimization model:

\[
\min_{x} \begin{cases}
N_0 \sum_{i \in T} v_i e^{-\alpha D_i x} & \geq \left( \sum_{i \in S} v_i (D_i x)^p \right)^{\frac{1}{p}} - \gamma_1 \leq 0 \\
\frac{1}{\Delta^p} \sum_{i \in S} v_i (D_i x)^p - \gamma_2 \leq 0 \\
\varepsilon \log \left( \sum_{i \in S} e^{D_i x/\varepsilon} \right) - \gamma_3 \leq 0 \\
\sum_{i \in S} v_i D_i x - \gamma_4 \leq 0 \\
x \geq 0
\end{cases}
\]

where \( \gamma_1, \gamma_2, \gamma_3 \) and \( \gamma_4 \) are predetermined upper bounds on \( gEUD, -\ln(1 - NTCP) \), the approximation of the maximum dose and the mean dose, respectively. The parameter \( \varepsilon \) determines the exactness of the approximation of a maximum function. The non-negativity constraint is included since a negative beam intensity is physically impossible.

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

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 1 (line 2) and the linear substitution rule (Table 2, line 7b), we obtain the convex conjugate of \( h_i(x) \):

\[ h_i^*(y) = \inf_{z_i} \left\{ z_i \log z_i - z_i | - \alpha z_i D_i^T = y, z_i \geq 0 \right\}. \]
Application of the sum rule for conjugates (Table 2, line 5) now enables us to write the conjugate of $f_0(x)$ as follows:

$$f_0^*(y) = N_0 \left( \sum_{i \in T} v_i h_i \right)^* \left( \frac{y}{N_0} \right) \quad \text{(Product with a scalar)}$$

$$= N_0 \inf_{(y') \in T} \left\{ \sum_{i \in T} (v_i h_i)^* \left( y' \right) \mid \sum_{i \in T} y^i = \frac{y}{N_0} \right\} \quad \text{(Sum rule)}$$

$$= N_0 \inf_{(y') \in T} \left\{ \sum_{i \in T} v_i h_i \left( \frac{y'}{v_i} \right) \mid \sum_{i \in T} y^i = \frac{y}{N_0} \right\} \quad \text{(Product with a scalar)}$$

$$= N_0 \inf_{(y') \in T} \left\{ \sum_{i \in T} v_i \inf_{z_i} \left\{ z_i \log z_i - z_i \mid - \alpha z_i D_i^T = \frac{y'_{v_i}, z_i \geq 0}{v_i} \right\} \mid \sum_{i \in T} y^i = \frac{y}{N_0} \right\}$$

$$= N_0 \inf_{w} \left\{ \sum_{i \in T} w_i \log \frac{w_i}{v_i} - w_i \mid \sum_{i \in T} - \alpha w_i D_i^T = \frac{y}{N_0}, w \succeq 0 \right\} \quad \text{(Product with a scalar)}$$

$$= N_0 \inf_{w} \left\{ \sum_{i \in T} w_i \log \frac{w_i}{v_i} - w_i \mid - \alpha N_0 D_i^T w = y, w \succeq 0 \right\},$$

where $D_T$ consists of the rows in $D$ related to the voxels in $T$.

gEUD constraint

First note that all constraint functions are a convex function minus a constant $\gamma_i$. The first constraint function is $f_1(x) := gEUD(x) - \gamma_1$. Hence we obtain $f_1^*(y) = gEUD^*(y) + \gamma_1$, by Table 2, line 1. Observe that $gEUD(x)$ can be written as a $p$-norm:

$$gEUD(x) = \left( \sum_{i \in S} v_i (D_i x)^p \right)^{\frac{1}{p}} = \left( \sum_{i \in S} (v_i^p D_i x)^p \right)^{\frac{1}{p}} = \left\| V_S^{\frac{1}{p}} D_S x \right\|_p,$$

where $D_S$ consists of the rows in $D$ related to the voxels in the OAR and $V_S = \text{Diag}(v_S)$.

Hence, denoting $g(x) := \|x\|_p$, we have $gEUD(x) = g(V_S^{\frac{1}{p}} D_S x)$. So the convex conjugate of $gEUD(x)$ can be obtained by using the linear substitution rule (Table 2, line 7b) and the convex conjugate of a norm (Table 1, line 6). Thus we get

$$gEUD^*(y) = \inf_{z} \left\{ 0 \mid D_S^T V_S^{\frac{1}{q}} z = y, \|z\|_q \leq 1 \right\},$$

with $q$ such that $\frac{1}{p} + \frac{1}{q} = 1$. Hence

$$f_1^*(y) = \inf_{z} \left\{ \gamma_1 \mid D_S^T V_S^{\frac{1}{q}} z = y, \|z\|_q \leq 1 \right\}.$$

NTCP constraint

The NTCP constraint function $f_2(x)$ is given by

$$f_2(x) = \frac{1}{\Delta p} \sum_{i \in S} v_i (D_i x)^p - \gamma_2 = \frac{1}{\Delta p} \sum_{i \in S} v_i g_i(x) - \gamma_2,$$

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 1, line 8 and Table 2, lines 3 and 7b),

$$h_i^*(y) = \inf_{t_i} \left\{ \frac{p}{q} t_i^q \mid t_i \geq 0, t_i \geq \frac{y}{p} \right\}$$

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

$$= \inf_{t_i, r_i} \left\{ \frac{p}{q} t_i^q \mid D_i^T r_i = y, t_i \geq 0, t_i \geq \frac{r_i}{p} \right\}.$$

By using the sum rule for conjugates, the multiply-with-a-constant rule and add-a-constant rule (Table 2, lines
5, 3 and 1, respectively) we can now compute $f_2^*(y)$:

$$f_2^*(y) = \frac{1}{\Delta p} \inf_{y'} \left\{ \sum_{i \in S} v_i g_i^* \left( \frac{y'}{v_i} \right) \bigg| \sum_{i \in S} y^i = \Delta p y \right\} + \gamma_2$$

$$= \frac{1}{\Delta p} \inf_{y'} \left\{ \sum_{i \in S} v_i \inf_{r \in T} \left[ \frac{p q_i^*}{q} r_i = \frac{y'}{v_i}, t_i \geq 0, t_i \geq \frac{r_i}{p} \right] \bigg| \sum_{i \in S} y^i = \Delta p y \right\} + \gamma_2$$

$$= \frac{p}{q \Delta p} \inf_{r, t} \left\{ \sum_{i \in S} v_i t_i^i \bigg| \sum_{i \in S} v_i D^T_1 r_i = \Delta p y, t_i \geq 0, p t_i \geq r_i \right\} + \gamma_2$$

$$= \inf_{r, t} \left\{ \gamma_2 + \frac{p}{q \Delta p} v^T_5 t^0 \big| D_5^T V_S r = \Delta p y, t \geq 0, p t \geq r \right\},$$

where $t^q$ is the vector with entries $t^q_i, i \in S$.

**Maximum dose constraint**

The third constraint function $f_3(x)$ is

$$f_3(x) = \varepsilon \log \left( \sum_{i \in S} e^{D_i x / \varepsilon} \right) - \gamma_3 = \varepsilon h(x) - \gamma_3,$$

where $h(x) = g(D_S x / \varepsilon)$, with $g(x) = \log \left( \sum_{i \in S} \exp(x_i) \right)$. The convex conjugate of $g(x)$ can be found in Table 1, line 5. Using the linear substitution rule from Table 2, line 7b, we obtain

$$g_i^*(y) = \sum_{i=1}^n y_i \log y_i, \quad \text{dom } g_i^* \{ y \mid y \geq 0, 1^T y = 1 \},$$

$$h^*(y) = \inf_s \{ g^*(s) \mid D_S^T s = \varepsilon y \}.$$

Hence we get the following convex conjugate of $f_3(x)$:

$$f_3^*(y) = (\varepsilon h)^*(y) + \gamma_3 = \varepsilon h^* \left( \frac{y}{\varepsilon} \right) + \gamma_3$$

$$= \varepsilon \inf_s \left\{ \sum_{i \in S} s_i \log s_i \bigg| D_S^T s = \varepsilon \frac{y}{\varepsilon}, s \geq 0, 1^T s = 1 \right\} + \gamma_3$$

$$= \inf_s \left\{ \gamma_3 + \varepsilon \sum_{i \in S} s_i \log s_i \bigg| D_S^T s = y, s \geq 0, 1^T s = 1 \right\}.$$

**Mean dose constraint**

The fourth constraint function is linear:

$$f_4(x) = \sum_{i \in S} v_i D_i x - \gamma_4 = v^T_4 D_S x - \gamma_4$$

Hence the convex conjugate of $f_4(x)$ is now straightforward (cf. Table 1, line 1):

$$f_4^*(y) = \gamma_4, \quad \text{dom } f_4^* = \{ y \mid y = D_4^T v_S \}.$$

**Non-negativity constraint**

We finally have to deal with the non-negativity constraint $x \succeq 0$. Denoting the indicator function of $R^+_n$ as $f_5(x)$, the corresponding support function is (cf. Table 3, line 2)

$$f_5^*(y) = 0, \quad \text{dom } f_5^* = \{ y \mid y \preceq 0 \}.$$

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

$$\sup_{(y')_{i=0}^u} \left\{ -f^*_5(y'^0) - \sum_{k=1}^4 u_k f^*_k \left( \frac{y'^k}{u_k} \right) - f^*_5(y'^0) \bigg| \sum_{k=0}^5 y'^k = 0, u \geq 0 \right\}.$$

We can already eliminate $y'^5$, yielding the equivalent problem

$$\sup_{(y')_{i=0}^u} \left\{ -f^*_5(y'^0) - \sum_{k=1}^4 u_k f^*_k \left( \frac{y'^k}{u_k} \right) \bigg| \sum_{k=0}^4 y'^k \geq 0, u \geq 0 \right\}.$$
Thus we obtain the following dual for the treatment planning optimization problem:

\[
\sup_{\{y^i\}_{i=0,u}} \left\{-N_0 \left\{ \inf_{w \in T} \left( \sum_{i \in T} w_i \log \frac{w_i}{v_i} - w_i \right) \right\} - \alpha N_0 D_T^2 w = y^0, w \succeq 0 \right\} - u_1 \inf_{z} \left\{ \gamma_1 \left| D_S^2 V_S^z = \frac{y^1}{u_1}, \|z\|_{q} \leq 1 \right\} - u_2 \inf_{r, t, s} \left\{ \gamma_2 + \frac{p}{q \Delta p} v_{r}^T t^q \left| D_S^2 V_S r = \Delta t \frac{y^2}{u_2}, t \succeq 0, p t \succeq r \right\} - u_3 \inf_{s} \left\{ \gamma_3 + \epsilon \sum_{i \in S} s_i \log s_i \left| D_S^2 s = \frac{y^3}{u_3}, s \succeq 0, 1^T s = 1 \right\} - u_4 \left\{ \gamma_4 \left| \frac{y^4}{u_4} = D_S^2 v_S \right\} \right\}.
\]

By omitting the \( \inf \) operators we get

\[
\sup_{\{y^i\}_{i=0,u}, w, z, r, s, u} \left\{-N_0 \left\{ \sum_{i \in T} w_i \log \frac{w_i}{v_i} - w_i \right\} - \alpha N_0 D_T^2 w = y^0, w \succeq 0 \right\} - u_1 \inf_{z} \left\{ \gamma_1 \left| D_S^2 V_S^z = \frac{y^1}{u_1}, \|z\|_{q} \leq 1 \right\} - u_2 \left\{ \gamma_2 + \frac{p}{q \Delta p} v_{r}^T t^q \left| D_S^2 V_S r = \Delta t \frac{y^2}{u_2}, t \succeq 0, p t \succeq r \right\} - u_3 \inf_{s} \left\{ \gamma_3 + \epsilon \sum_{i \in S} s_i \log s_i \left| D_S^2 s = \frac{y^3}{u_3}, s \succeq 0, 1^T s = 1 \right\} - u_4 \left\{ \gamma_4 \left| \frac{y^4}{u_4} = D_S^2 v_S \right\} \right\}.
\]

This is equivalent to

\[
\sup_{w, z, r, t, s, u} \left\{-N_0 \sum_{i \in T} \left( w_i \log \frac{w_i}{v_i} - w_i \right) - u^T \gamma - \frac{u_2 p v_{r}^T t^q}{q \Delta p} - \epsilon \sum_{i \in S} (u_3 s_i) \log s_i \right| w \succeq 0, \|z\|_{q} \leq 1, t \succeq 0, p t \succeq r, s \succeq 0, 1^T s = 1, u \succeq 0, \right\} - \alpha N_0 D_T^2 w + D_S^2 V_S^z (u z) + \frac{1}{\Delta p} D_S^2 S (u r) + D_S^2 (u s) + u_4 D_S^2 v_S \succeq 0\}.
\]

By redefining \( z, r, t \) and \( s \) according to \( z := u_1 z, r := u_2 r, t := u_2 t, s := u_3 s \) this becomes

\[
\sup_{w, z, r, s, u} \left\{-N_0 \sum_{i \in T} \left( w_i \log \frac{w_i}{v_i} - w_i \right) - u^T \gamma - \frac{u_2 p v_{r}^T t^q}{q \Delta p} - \epsilon \sum_{i \in S} (u_3 s_i) \log s_i \right| w \succeq 0, \|z\|_{q} \leq u_1, t \succeq 0, p t \succeq r, s \succeq 0, 1^T s = u_3, u \succeq 0, \right\} - \alpha N_0 D_T^2 w + D_T^2 V_S^z (u z) + \frac{1}{\Delta p} D_T^2 S (u r) + D_T^2 (u s) + u_4 D_T^2 v_S \succeq 0\}.
\]

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

6 Concluding remarks

Mathematical education usually includes the use of tables, e.g., tables of derivatives, primitive functions, Laplace transforms, Fourier transforms, etc. It may have become clear from this paper that the availability of tables as presented in in Appendix F 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.
Some examples of conjugate functions

By way of example we demonstrate how some of the results in Table 1 are obtained; this table also gives appropriate references to the existing literature whenever available. The examples in this section are used in Section 5 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 1. With \( f(x) = a^T x + b, \ a \in \mathbb{R}^n \) and \( b \in \mathbb{R} \), we have

\[
f^*(y) = \sup_x \left\{ y^T x - (a^T x + b) \right\} = \sup_x \left\{ (y - a)^T x - b \right\}.
\]

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

\[
f^*(y) = -b, \ \ \text{dom} \ f^* = \{a\}.
\]

A.2 Log-sum-exp function

We next deal with line 5 in Table 1. Let \( f(x) = \log \left( \sum_{i=1}^n e^{x_i} \right), \ x \in \mathbb{R}^n \). 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 \( i \), we take \( x = -\lambda e_i \), where \( \lambda > 0 \) and \( e_i \) is the \( i \)-th unit vector. Then \( g(x) = -\lambda y_i - \log(n - 1 + e^{-\lambda}) \), which goes to infinity if \( \lambda \) goes to infinity. If \( y \geq 0 \) but \( 1^T y \neq 1 \), take \( x = \lambda 1 \). Then \( g(x) = \lambda 1^T y - \log(ne^\lambda) = \lambda(1^T y - 1) - \log n \), which goes to infinity both if \( 1^T y > 1 \) (when \( \lambda \) goes to infinity) and if \( 1^T y < 1 \) (when \( \lambda \) goes to minus infinity). Hence \( y \in \text{dom} \ f^* \) holds only if \( 1^T y = 1 \) and \( y \geq 0 \). We proceed by computing the maximal value of \( g(x) \) under these conditions. One has

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

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

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

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

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

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

\[
f^*(y) = \sum_{i=1}^n y_i \log y_i, \ \ \text{dom} \ f^* = \{y | 1^T y = 1, \ y \geq 0\}.
\]

A.3 Arbitrary norm

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

\[
f^*(y) = \sup_x \left\{ y^T x - \|x\| \right\}.
\]

Recall that the dual norm is defined by

\[
\|y\|_* = \sup_x \left\{ y^T x | \|x\| \leq 1 \right\}.
\]

Since \( \|x\|/\|x\| = 1 \) if \( x \neq 0 \), we have for any nonzero \( x \in \mathbb{R}^n \)

\[
\|y\|_* \geq \frac{y^T x}{\|x\|} = \frac{y^T x}{\|x\|},
\]
which gives

\[ y^T x \leq \| y \|_p \| x \|. \]  

(8)

Since this inequality holds also for \( x = 0 \), it holds for each \( x \in \mathbb{R}^n \). Define \( g(x) := y^T x - \| x \| \). Then \( g(\lambda x) = \lambda g(x) \) for \( \lambda \geq 0 \).

If \( \| y \|_p > 1 \) 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 \( \lambda \) grows to infinity. We conclude from this that \( y \in \text{dom } f^* \) holds only if \( \| y \|_p \leq 1 \).

If \( \| y \|_p \leq 1 \) then (8) implies \( y^T x \leq \| y \|_p \| x \| \leq \| x \| \), whence \( g(x) \leq 0 \), for all \( x \in \mathbb{R}^n \). Since \( g(0) = 0 \), we obtain

\[ f^*(y) = 0, \quad \text{dom } f^* = \{ y \mid \| y \|_p \leq 1 \}. \]

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

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

In that 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 1. Let \( f(x) = \frac{1}{2} x^T P x, x \in \mathbb{R}^n \), where \( P \) is a (symmetric) positive semidefinite \( n \times n \) matrix. We then have

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

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

\[ f^*(y) = y^T P^{-1} y - \frac{1}{2} \left( P^{-1} y \right)^T P P^{-1} y = \frac{1}{2} y^T P^{-1} y. \]

If \( P \) is singular, the equation \( y = Px \) has a solution if and only if \( y \) belongs to the column space \( L = \{ Pz \mid z \in \mathbb{R}^n \} \). If \( y \in L \) then the solution is given by \( x = P^1 y \), where \( P^1 \) is the generalized inverse of \( P \), and we obtain

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

where the last equality is due to the fact that \( P^1 P P^1 = P^1 \) Ben-Israel and Greville (2003).

If \( y \notin L \) 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

\[ x^T y = x^T y_1 + x^T y_2 = x^T y_2 = \lambda y_2^T y_2 = \lambda \| y_2 \|^2. \]

Since \( x \in L^\perp \), 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 \( \lambda \) grows to infinity. This proves that \( y \notin \text{dom } f^* \) if \( y \notin L \).

Thus we have shown that

\[ f^*(y) = \frac{1}{2} y^T P^1 y, \quad \text{dom } f^* = \{ y \mid y = Pz, z \in \mathbb{R}^n \}. \]

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

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

\[ f'(x) = x^{p-1}, \]

\[ f''(x) = (p-1)x^{p-2}, \]

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

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

Since \( yx - f(x) \leq 0 \) if \( y < 0 \) the sup value the 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

\[ f^*(y) = y^{\frac{1}{p-1}} \frac{1}{p} y^{\frac{p-1}{p-1}} = \left( 1 - \frac{1}{p} \right) y^{\frac{p}{p-1}} = \frac{1}{q} y^q, \]
where \( q \) is such that \( \frac{1}{p} + \frac{1}{q} = 1 \). Summarizing,
\[
f^*(y) = \begin{cases} 
0, & y < 0, \\
\frac{1}{q}y^q, & y \geq 0.
\end{cases}
\]
Since \( f^*(y) \) 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 \( \frac{1}{px^r} \), \( x > 0, p > 0 \)

We prove the formula in line 11 of Table 1. Let \( f(x) = \frac{1}{px^r} \), \( x > 0, p > 0 \). One has
\[
f'(x) = -\frac{1}{x^{p+1}}, \quad f''(x) = \frac{p+1}{x^{p+2}},
\]
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 \( \infty \). We have
\[
f^*(y) = \sup_{x > 0} (yx - f(x)).
\]

If \( y > 0 \) then \( yx - f(x) \) goes to \( \infty \) if \( x \) goes to \( \infty \). Hence the sup value is \( \infty \) if \( y > 0 \). So we may assume that \( y \leq 0 \). Since then
\[
f^*(y) = \sup_{x > 0} (yx - \frac{1}{px^r}),
\]
the sup value is attained when \( y = f'(x) = \frac{1}{x^{p+1}} \), which gives \( x = (-y)^{\frac{r}{p}} \), and
\[
f^*(y) = y(-y)^{\frac{r}{p}} - \frac{1}{p}(-y)^{\frac{r}{p}+1} = \left(1 - \frac{1}{p}\right)(-y)^{\frac{r}{p}} = -\left(\frac{p+1}{p}\right)(-y)^{\frac{r}{p}}.
\]

Summarizing,
\[
f^*(y) = -\left(\frac{p+1}{p}\right)(-y)^{\frac{r}{p}}, \quad \text{dom} f^* = \{y \mid y \leq 0\} = -\mathbb{R}_+.
\]

### B Some identities for conjugate functions

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

First note that we may write
\[
f(x) = \max_z \left\{ \sum_{i=1}^m z_i f_i(x) \mid \sum_{i=1}^m z_i = 1, z \geq 0 \right\} = \max_{z \in S} \left\{ \sum_{i=1}^m z_i f_i(x) \right\},
\]
where \( S \) denotes the simplex. By the definition of \( f^* \) we have
\[
f^*(y) = \sup_x \left\{ y^T x - \max_{z \in S} \sum_{i=1}^m z_i f_i(x) \right\}
\]
\[
= \sup_{x \in S} \min_z \left\{ y^T x - \sum_{i=1}^m z_i f_i(x) \right\}.
\]

Since the argument is convex in \( z \) and concave in \( x \), and the domain of \( z \) is compact, we may interchange the sup and the 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 2, line 5) we obtain
\[
f^*(y) = \min_{z \in S} \left\{ \sum_{i=1}^m (z_i f_i)^*(y') \right\} \left\{ \sum_{i=1}^m y' = y \right\}
\]
\[
= \min_{z} \left\{ \sum_{i=1}^m (z_i f_i)^*(y') \right\} \left\{ \sum_{i=1}^m y' = y, \sum_{i=1}^m z_i = 1, z \geq 0 \right\}.
\]
B.2 Deriving conjugates via the adjoint I

In this section we prove the formula in line 9 of Table 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 \( (f^O)^*(y) \). The proof below is an alternative for that in Gushchin (2008).

By the definition of \( (f^O)^*(y) \) we have

\[
\inf_s \left\{ s \mid (f^O)^*(-s) \leq -y \right\} = \inf_s \left\{ s \mid \sup_x \left\{ -sx - f^O(x) \right\} \leq -y \right\}.
\]

Since \( \text{dom} f^O = R_+^+ \) we have \( f^O(x) = \infty \) if \( x \leq 0 \). Hence

\[
\inf_s \left\{ s \mid (f^O)^*(-s) \leq -y \right\} = \inf_x \left\{ s \mid \sup_{x>0} \left\{ -sx - xf \left( \frac{1}{x} \right) \right\} \leq -y \right\}.
\]

Due to Lagrange duality this implies

\[
\inf_s \left\{ s \mid (f^O)^*(-s) \leq -y \right\} = \sup_{\lambda \geq 0} \inf_s \left\{ s + \lambda \left( y + \sup_{x>0} \left\{ -sx - xf \left( \frac{1}{x} \right) \right\} \right) \right\} = \sup_{\lambda \geq 0} \inf_{s>0} \left\{ \lambda y - \lambda xf \left( \frac{1}{x} \right) + s(1 - \lambda x) \right\}.
\]  

(9)

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

\[
\sup_{x>0} \left\{ \lambda y - \lambda xf \left( \frac{1}{x} \right) \mid 1 - \lambda x = 0 \right\}.
\]

Since this problem is Slater regular, its Lagrangian has a saddle point. As a consequence, in (9) we may replace \( \inf_s \sup_{s>0} \) by \( \sup_{s>0} \inf_s \). One has

\[
\sup_{s>0} \inf_{\lambda \geq 0} \left\{ \lambda y - \lambda xf \left( \frac{1}{x} \right) + s(1 - \lambda x) \right\} = \sup_{s>0} \left\{ \lambda y - \lambda xf \left( \frac{1}{x} \right) \mid \lambda x = 1 \right\} = \lambda y - f(\lambda).
\]

Substitution into (9) yields

\[
\inf_s \left\{ s \mid (f^O)^*(-s) \leq -y \right\} = \sup_{\lambda \geq 0} \left\{ \lambda y - f(\lambda) \right\} = f^*(y).
\]

The last equality holds because the domain of \( f \) is \( R_+ \).

B.3 Deriving conjugates via the adjoint II

In this section we prove formula 10 in Table 2. This formula can be applied if \( f \) has an inverse, which is denoted as \( h = f^{-1} \).

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

\[
-(h_*)^O(y) = -y h_* \left( \frac{1}{y} \right) = -y \inf_x \left\{ \frac{1}{y} x - h(x) \right\} = y \sup_x \left\{ h(x) - \frac{x}{y} \right\} = y \sup_{x,t} \left\{ t - \frac{x}{y} \mid t \leq h(x) \right\},
\]

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 \( y \geq 0 \). Also using that \( t \leq h(x) \) holds if and only if \( f(t) \leq x \) we may proceed as follows:

\[
-(h_*)^O(y) = \sup_{x,t} \left\{ ty - x \mid t \leq h(x) \right\} = \sup_{x,t} \left\{ ty - x \mid f(t) \leq x \right\} = \sup_t \left\{ ty - f(t) \right\} = f^*(y).
\]

B.4 Linear substitution rule

In this section we prove the formulas in lines 7a and 7b of Table 2. Let \( f \) be defined by \( f(x) = h(Ax + b) \), where \( A \) is an \( m \times n \) matrix, \( b \in R^m \) and \( h : R^m \rightarrow R \) convex. We then have

\[
f^*(y) = \sup_x \left\{ y^T x - h(Ax + b) \right\} = \sup_{x,v} \left\{ y^T x - h(v) \mid Ax + b = v \right\}.
\]
By writing the last problem as a minimization problem and then using Lagrange’s duality theorem we get

\[
f^*(y) = \min_{z \geq 0} \sup_{x,v} \left\{ y^T x - h(v) + z^T (v - b - Ax) \right\}
\]

\[
= \min_{z \geq 0} \left\{ z^T v - h(v) - z^T b + x^T (y - A^T z) \right\}.
\]

The contribution of the term containing \( x \) is finite if and only if \( y - A^T z = 0 \). Thus we obtain

\[
f^*(y) = \min_{z \geq 0} \left\{ z^T v - h(v) - z^T b | A^T z = y \right\}
\]

\[
= \min_{z} \{ h^*(z) - b^T z | A^T z = y \}.
\]

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

\[
f^*(y) = h^*(A^{-T} y) - b^T A^{-T} y.
\]

### B.5 Composite function

In this section we prove the formula in line 8 of Table 2. Let \( f(x) = g(h(x)) \), with \( g \) and \( h \) convex and \( g \) nondecreasing. We may write

\[
f(x) = \inf_{z} \{ g(z) | h(x) \leq z \}.
\]

Hence

\[
f^*(y) = \sup_{x} \left\{ y^T x - \inf_{z} \{ g(z) | h(x) \leq z \} \right\}
\]

\[
= \sup_{x,z} \{ y^T x - g(z) | h(x) \leq z \}.
\]

Using the Lagrange dual of the last problem, we get

\[
f^*(y) = \min_{u \geq 0} \sup_{x,z} \{ y^T x - g(z) + u(z - h(x)) \}
\]

\[
= \min_{u \geq 0} \left\{ \sup_{z} \{ uz - g(z) \} + \sup_{x} \{ y^T x - uh(x) \} \right\}
\]

\[
= \min_{u \geq 0} \{ g^*(u) + (uh)^*(y) \}.
\]

### B.6 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. Suppose \( f : D \to \mathbb{R} \), with \( D \subseteq \mathbb{R} \). The ‘natural’ domain of \( f \) is the subset of \( \mathbb{R} \) where \( f(x) \) is finite. But the domain can also be a proper subset \( D \) of the natural domain. Let us denote the new function as \( f_D \). So we have \( f_D(x) = f(x) \) if \( x \in D \) and \( f_D(x) = \infty \) if \( x \notin D \).

By way of example, let \( f(x) = x^2 \). The natural domain is \( \mathbb{R} \), and \( f^*(y) = \frac{1}{2} y^2 \), with domain \( \mathbb{R} \). 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 \( x = \frac{y}{2} \), 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 obtain

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

Hence \( f_D^* \) is now given by two closed formulas, one for \( y \leq 2 \) and one for \( y > 2 \). 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).

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 5 in Table 2) we obtain

\[
  f_D^*(y) = \inf_{y^1, y^2} \left\{ f^*(y^1) + \delta_D^*(y^2) \mid y^1 + y^2 = y \right\}
\]

\[
  = \inf_y \left\{ f^*(y^1) + \delta_D^*(y - y^1) \right\}.
\]

(10)

This expression can be used immediately when forming Fenchel’s dual problem, as it is the inf of a tractable function. In the above example one has

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

Hence, when applying (10) to the example we obtain

\[
  f_D^*(y) = \inf_z \left\{ f^*(z) + \delta_D^*(y - z) \right\}
\]

\[
  = \inf_z \left\{ \frac{1}{2}z^2 + y - z \mid y - z \leq 0 \right\}
\]

\[
  = \inf_z \left\{ (\frac{1}{2}z - 1)^2 + y - 1 \mid y \leq z \right\}.
\]

When forming a dual problem this expression is more convenient, since the ‘branches’ disappeared.

### B.7 Support function of a cone

In this section we compute the support function of a convex cone \( K \) in \( \mathbb{R}^n \). One has

\[
  \delta_K^*(y) = \sup_{x \in K} y^T x.
\]

Recall that the dual cone of \( K \) is defined by

\[
  K^* = \{ y \mid y^T x \geq 0, \forall x \in K \}.
\]

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

\[
  \delta_K^*(y) = 0, \quad \text{dom} \delta_K^* = -K^*.
\]

### B.8 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, in spite of the fact that by definition (2) we have \((0 f_1)^* (y^1) = \infty\) if \( y_1 \neq 0 \). Notwithstanding this behaviour, Fenchel’s dual problem works out well.

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

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

Since \( \delta_{[-1,1]}^*(y) = |y| \), one has \( f_1^*(y) = |y| + 1 \). Hence Fenchel’s dual problem is

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

One has

\[
  (u_1 f_1)^* (-1) = \begin{cases} 
  u_1 \left( \left\lfloor \frac{-1}{u_1} \right\rfloor + 1 \right) & \text{if } u_1 > 0 \\
  1 + u_1 & \text{otherwise}.
\end{cases}
\]

Hence the dual problem becomes

\[
  \sup_{u_1 \geq 0} \left\{ -1 - u_1 \right\}.
\]

The sup value equals \(-1\), as it should, but is not attained! Note that at optimality \( u_1 \) approaches zero, but \( y^1 = -1 \) stays away from zero.
C Lagrange’s duality theorem for convex optimization

As in the main text, we consider the following minimization problem:

\[(P)\quad \inf \{ f_0(x) \mid f_i(x) \leq 0, \ i = 1, \ldots, m \}. \]

Then Lagrange’s dual problem is given by

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

It is well-known that if \((P)\) is Slater regular then \((P)\) and \((LD)\) have the same optimal value and the dual objective value attains the optimal value (see, e.g., (Boyd and Vandenberghe, 2004, Section 5.2.3)). So, we then have strong duality.

D Fenchel duality

In this section we derive Fenchel’s duality theorem from Lagrange’s duality theorem. Let \(f\) and \(g\) be functions from \(\mathbb{R}^n\) to \(\mathbb{R} \cup \{-\infty, \infty\}\), \(f\) proper convex and \(g\) proper concave. Then \(f - g\) is convex. We consider the two problems

\[(P)\quad \inf \{ f(x) - g(x) \mid x \in \text{dom } f \cap \text{dom } g \}, \]
\[(D)\quad \sup \{ g_u(y) - f^*(y) \mid y \in \text{dom } f^* \cap \text{dom } g_u \}. \]

**Theorem 1** (Theorem 31.1 in Rockafellar (1970)). If \(\text{ri.dom } f \cap \text{ri.dom } g\) 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 \(\text{ri.dom } f\) can be replaced by \(\text{dom } f\). Similarly, if \(g\) is linear, then \(\text{ri.dom } g\) can be replaced by \(\text{dom } g\).

**Proof.** We may reformulate \((P)\) as follows:

\[ \inf_{u,v} \{ f(u) - g(v) \mid u = v, u \in \text{dom } f, v \in \text{dom } g \} . \]

Defining

\[ C = \{(u, v) \mid u \in \text{dom } f, v \in \text{dom } g \}, \]

the problem can be reformulated as

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

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 \((x^*, x^*) \in \text{ri} C\). 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

\[ p^* = \max_y \inf_{u,v} \{ f(u) - g(v) + (v - u)^T y \mid u \in \text{dom } (f), v \in \text{dom } (g) \} \]
\[ = \max_y \inf_{u,v} \{ y^T v - g(v) - (u^T y - f(u)) \mid u \in \text{dom } (f), v \in \text{dom } (g) \} . \]

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

\[ p^* = \max_y \left\{ \inf_{v \in \text{dom } (g)} \{ y^T v - g(v) \} - \sup_{u \in \text{dom } (f)} \{ u^T y - f(u) \} \right\} \]
\[ = \max_y \{ g_u(y) - f^*(y) \} , \]

where we used that \(g(v) = -\infty\) if \(v \notin \text{dom } (g)\) and \(f(v) = \infty\) if \(v \notin \text{dom } (f)\). 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. So, in essence both theorems are equivalent. This has been worked out in Magnanti (1974).
E Derivation of Fenchel’s dual problem

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

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

where the functions $f_i : \mathbb{R}^n \to \mathbb{R}$ are proper convex functions for $i = 0, \ldots, m$. Let $S_i := \{x \mid f_i(x) \leq 0\}$, for $i = 1, \ldots, m$, and $S = \cap_{i=1}^m S_i$. Obviously, $S$ is the feasible region of (P). The assumption that (P) is Slater regular implies that the set $S$ is nonempty and $x \in \text{ri} \, \text{dom} \, f$ for some $x \in S$. Moreover, if $f_i$ is nonlinear then $\text{ri} \, S_i$ is nonempty.

Due to the definition of the indicator function $\delta_S(x)$, we have

$$\inf_{x \in S} \{f_0(x)\} = \inf_x \{f_0(x) + \delta_S(x)\}.$$  \hspace{1cm} \text{(10)}

Since $f_0$ is proper convex, $-\delta_S$ proper concave and

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

we may apply Fenchel’s duality theorem. Hence we may write

$$\inf_{x \in S} \{f_0(x)\} = \inf_x \{f_0(x) - (-\delta_S(x))\} = \max_y \{(-\delta_S)(y) - f_0^*(y)\} = \max_y \{-f_0^*(y) - \delta_S^*(-y)\},$$  \hspace{1cm} \text{(11)}

where the last equality is due to (3).

At this stage we need to compute $\delta_S^*(-y)$. Since the sets $S_i$ satisfy the Slater condition we may apply (Rockafellar, 1970, Corollary 16.4.1), which gives (see also Table 3, line 6)

$$\delta_S^*(-y) = \min_{(y^i)_{i=1}^m} \left\{ \sum_{i=1}^m \delta_{S_i}^*(y^i) \left| \sum_{i=1}^m y^i = -y \right. \right\}.$$  \hspace{1cm} \text{(12)}

Substitution into (11) yields

$$\inf_{x \in S} \{f_0(x)\} = \max_y \left\{ -f_0^*(y) - \min_{(y^i)_{i=1}^m} \left\{ \sum_{i=1}^m \delta_{S_i}^*(y^i) \left| \sum_{i=1}^m y^i = -y \right. \right\} \right\}.$$

where the last expression follows since the $\min$ operator is absorbed by the $\max$ operator. Thus we have shown that the maximization problem in (12) 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

$$\max_{(y^i)_{i=1}^m} \left\{ -f_0^*(y^0) - \sum_{i=1}^m \delta_{S_i}^*(y^i) \left| \sum_{i=1}^m y^i = 0 \right. \right\}.$$  \hspace{1cm} \text{(13)}

A nice feature of the above 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 (13) 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 1. One has

$$\delta_{S_i}(y) = \min_{u \geq 0} \{(uf_i)^*(y)\}.$$
Proof. According to the definition of $\delta^*_S$, we have

$$\delta^*_S(y) = \sup_{x \in S} \{ y^T x \} = \sup_x \{ y^T x \mid f_i(x) \leq 0 \}.$$ 

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

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

which proves the lemma.

By applying Lemma 1 to (13) we obtain

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

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

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

### F 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 text books on optimization tables of conjugates functions are rare. One positive example is (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 1, with references to existing literature. As we made clear in Section 2, concave conjugates easily follow from these formulas by using (3).

In Table 2 we present some transformation rules that are useful for the computation of conjugate functions. In line 8 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 3 gives a list of support functions.
Table 1: Examples of convex conjugate functions available from the literature. In lines 6 and 7 \( \| \cdot \| \) denotes any norm, and \( \| \cdot \|^* \) 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, \( S^n_+ \) and \( S^n_{++} \) denote the set of positive semidefinite and positive definite \( n \times n \) matrices, respectively.

| no. | \( f(x) \) (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   | \( \log (\sum_{i=1}^n e^{x_i}) \) | \( \sum_{i=1}^n y_i \log y_i \) (\( y \geq 0 \), \( 1^T y = 1 \)) | (Rockafellar, 1970, p.148) |
| 6   | \( \|x\| \)          | \( 0 \) (\( \|y\|^* \leq 1 \)) | (Boyd and Vandenberghe, 2004, p.93) |
| 7   | \( \frac{1}{2} \|x\|^2 \) | \( \frac{1}{2} \|y\|^2 \) | (Boyd and Vandenberghe, 2004, p.93) |
| 8   | \( \frac{1}{p} x^p \) (\( x \geq 0 \), \( p > 1 \)) | \( \inf_z \left\{ \frac{1}{q} z^q \mid z \geq y, z \geq 0 \right\} \) | (Rockafellar, 1970, p.106) |
| 9   | \( -\frac{1}{p} x^p \) (\( x \geq 0 \), \( 0 < p < 1 \)) | \( -\frac{1}{q} |y|^q \) (\( y < 0 \)) | (Rockafellar, 1970, p.106) |
| 10  | \( \frac{1}{p} |x|^p \) (\( p > 1 \)) | \( \frac{1}{q} |y|^q \) | (Rockafellar, 1970, p.106) |
| 11  | \( \frac{1}{p} x^p \) (\( x > 0 \), \( p > 0 \)) | \( -\frac{p+1}{p} (-y)^{\frac{p+1}{p}} \) (\( y \leq 0 \)) | (Rockafellar, 1970, p.106) |
| 12  | \(- (a^2 - x^2)^{1/2} \) (\( |x| \leq a \)) | \( a(1 + y^2)^{1/2} \) | (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  | \( \frac{1}{2} x^T P x + q^T x \) \( P \succeq 0 \) | \( \frac{1}{2} (y - q)^T P^\dagger (y - q) \) (\( y = q + Pz \)) | (Gorissen and den Hertog, p.7) |
| 15  | \(- \log \det X \) (\( X \in S^n_+ \)) | \(- \log \det(-Y) - n \) (\( -Y \in S^n_+ \)) | (Boyd and Vandenberghe, 2004, p.92) |
| 16  | \(- \sqrt{e^T X e} \) \( X \succeq 0 \) | \( \inf_{t < 0} \left\{ -\frac{1}{4t} \mid Y \succeq t e^T \right\} \) | (Gorissen and den Hertog, p.10) |
| 17  | \( \cosh x \) | \( y \sinh^{-1} y - \sqrt{1 + y^2} \) | (Borwein and Lewis, 2000, p.50) |
| 18  | \(- \log(\cos x) \) (\( |x| < \frac{\pi}{2} \)) | \( y \tan^{-1} y - \frac{1}{2} \log \left(1 + y^2\right) \) | (Borwein and Lewis, 2000, p.50) |
| 19  | \( \log(\cosh x) \) | \( y \tanh^{-1} y + \frac{1}{2} \log \left(1 - y^2\right) \) (\( |y| < 1 \)) | (Borwein and Lewis, 2000, p.50) |
Table 2: Rules for calculating the conjugate function. The functions $f(x)$, $f_i(x)$ and $h(x)$ in this table are real valued functions, $f(x)$ and $f_i(x)$ convex and $h(x)$ concave. In line 5 we assume $\cap_{i=1}^{m} \text{ri}\text{.dom} f_i \neq \emptyset$; if $f_i(x)$ is linear for some $i$ then the corresponding $\text{ri}\text{.dom} f_i$ can be replaced by $\text{dom} f_i$ in this condition. We call this the sum-rule for conjugate functions. In line 6, $x^1 : x_m$ is a partition of $x$. We will refer to line 7 as the linear substitution rule. In lines 9 and 10, $f^\diamond$ denotes the adjoint of $f : \mathbb{R}_{++} \to \mathbb{R}$, which is defined as $f^\diamond(x) := xf(1/x)$. In line 11, $S$ denotes the unit simplex.

| no. | Function and assumptions | Conjugate function | Reference |
|-----|-------------------------|-------------------|-----------|
| 1   | $f(x) + a$              | $f^*(y) - a$       | (Boyd and Vandenberghe, 2004, p.95) |
| 2   | $f(x + a)$              | $f^*(y) - a^Ty$    | (Rockafellar, 1970, p.107) |
| 3   | $af(x), a > 0$          | $af^*(\frac{y}{a})$ | (Rockafellar, 1970, p.140) |
| 4   | $f(ax), a > 0$          | $f^*(\frac{y}{a})$ | (Rockafellar, 1970, p.107) |
| 5   | $\sum_{i=1}^{m} f_i(x)$ | $\min_{y^*} \left\{ \sum_{i=1}^{m} f_i^*(y^i) \bigg| \sum_{i=1}^{m} y^i = y \right\}$ | (Rockafellar, 1970, p.145) |
| 6   | $\sum_{i=1}^{m} f_i(x')$ | $\sum_{i=1}^{m} f_i^*(y')$ | (Boyd and Vandenberghe, 2004, p.95) |
| 7   | $f(Ax + b)$             | $f^*(A^{-T}y) - b^TA^{-T}y$ | (Rockafellar, 1970, p.107) |
| 8   | $f(g(x)), f$ nondecreasing | $\inf_{u \geq 0} \{(ug)^*(y) + f^*(u)\}$ | (Hiriart-Urruty, 2006, p.39) |
| 9   | $\text{dom} f = \mathbb{R}_{++}$ | $\inf_{s} \{s | (f^\diamond)^*(-s) \leq -y\}$ | (Gushchin, 2008, p.453) |
| 10  | $h : \mathbb{R} \to \mathbb{R}$ strictly increasing | $(h^{-1})^*(y) = -(h_{-1})^\diamond(y), y > 0$ | (Ben-Tal et al., 1991, eq. (3.3)) |
| 11  | $\max_{i} \{f_i(x) | i = 1, \ldots, m\}$ | $\min_{z \in \mathcal{B}} \left\{ \sum_{i=1}^{m} (z_if_i)^*(y^i) \bigg| \sum_{i=1}^{m} y^i = y \right\}$ | (Rockafellar, 1970, p.149) |
Table 3: Support functions. 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 $\cap_{i=1}^m \text{ri} \text{ dom } f_i \neq \emptyset$; if $f_i(x)$ is linear for some $i$ then the corresponding $\text{ri} \text{ dom } f_i$ can be replaced by $\text{dom } f_i$ in this condition. Similarly, in line 6 we assume $\cap_{i=1}^m \text{ri } S_i \neq \emptyset$, and if $S_i$ is polyhedral then the corresponding $\text{ri } S_i$ can be replaced by $S_i$.

| no. | $S$, $S \neq \emptyset$ | $\delta^*_S(y)$ |
|-----|------------------------|------------------|
| 1   | $\{x \mid Ax = b\}$   | $\min_z \{b^Tz \mid A^Tz = y\}$ | (Boyd and Vandenberghe, 2004, p.380) |
| 2   | $\{x \mid Ax \preceq b\}$ | $\min_z \{b^Tz \mid A^Tz = y, z \succeq 0\}$ | (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   | $\{x \mid ||x||_p \leq \rho\}$ | $\rho ||y||_p, \frac{1}{p} + \frac{1}{q} = 1$ | (Ben-Tal et al., 2015, p.272) |
| 5   | $\{x \mid f_i(x) \leq 0, i = 1, \ldots, m\}$ | $\min_{u \geq 0, y'} \left\{ \sum_{i=1}^m (u_i f_i)^*(y') \mid \sum_{i=1}^m y^i = y \right\}$ | (Rockafellar, 1970, p.146) |
| 6   | $S = \bigcap_{i=1}^m S_i$ | $\min_{y'} \left\{ \sum_{i=1}^m \delta^*_{S_i}(y') \mid \sum_{i=1}^m y^i = y \right\}$ | (Rockafellar, 1970, p.146) |
| 7   | $S = S_1 \times \ldots \times S_m$ | $\sum_{i=1}^m \delta^*_{S_i}(y^i), y^1 : y^m$ is a partition of $y$ | (Ben-Tal et al., 2015, p.294) |
References

M. Alber and F. Nüsslin. A representation of an NTCP function for local complication mechanisms. *Physics in Medicine and Biology*, 46(2):439–447, 2001.

A. Ben-Israel and T. N. E. Greville. *Generalized inverses. Theory and applications.* Number 15 in CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC. Springer-Verlag, New York, 2003.

A. Ben-Tal and A. Nemirovski. *Lectures on modern convex optimization: analysis, algorithms, and engineering applications.* MPS-SIAM Series on Optimization, 2001.

A. Ben-Tal, A. Ben-Israel, and M. Teboulle. Certainty equivalents and information measures: duality and extremal principles. *Journal of Mathematical Analysis and Applications*, 157(1):211–236, 1991.

A. Ben-Tal, D. den Hertog, A. D. Waegenaere, B. Melenberg, and G. Rennen. Robust solutions of optimization problems affected by uncertain probabilities. *Management Science*, 59(2):341–357, 2013.

A. Ben-Tal, D. den Hertog, and J.-P. Vial. Deriving robust counterparts of nonlinear uncertain inequalities. *Mathematical Programming*, 149(1):265–299, 2015.

J. M. Borwein and A. S. Lewis. *Convex analysis and nonlinear optimization: theory and examples.* CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC. Springer-Verlag, New York, 2000.

S. Boyd and L. Vandenberghe. *Convex analysis and nonlinear optimization.* Cambridge University Press, 2004.

D. den Hertog. *Interior point approach to linear, quadratic and convex programming*, volume 277 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, 1994.

R. J. Duffin, E. L. Peterson, and C. Zener. *Geometric programming: theory and application.* John Whiley & Sons, 1967.

A. V. Fiacco and G. P. McCormick. *Nonlinear Programming. Sequential Unconstrained Minimization.* John Whiley & Sons, 1968.

A. Fredriksson. Automated improvement of radiation therapy treatment plans by optimization under reference dose constraints. *Physics in Medicine and Biology*, 57(23):7799–7811, 2012.

B. L. Gorissen and D. den Hertog. Robust nonlinear optimization via the dual.

A. A. Gushchin. On an extension of the concept of f-divergence. *Theory of Probability and its Applications*, 52(3):439–455, 2008.

J. B. Hiriart-Urruty. *Nonsmooth Mechanics and Analysis*, volume 12 of *Advances in Mechanics and Mathematics*, chapter 4 - A note on the Legendre-Fenchel transform of convex composite functions, pages 35–46. Springer, 2006.

T. L. Magnanti. Fenchel and Lagrange duality are equivalent. *Mathematical Programming*, 7(1):253–258, 1974.

A. Niemierko. A generalized concept of equivalent uniform dose (EUD). *Medical Physics*, 26(6):1100, 1999.

R. T. Rockafellar. *Convex Analysis*. Prentice-Hall, Inc., 1970.

E. Romeijn, J. Dempsey, and J. Li. A unifying framework for multi-criteria fluence map optimization models. *Physics in Medicine and Biology*, 49(10):1991–2013, 2004.

M. L. Slater. *Traces and emergence of nonlinear programming*, chapter Lagrange multipliers revisited., pages 293–306. Birkhäuser/Springer Basel AG, Basel, 2014.

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

T. Terlaky. On ℓ_p programming. *European Journal of Operations Research*, 22(1):70–100, 1985.

T. Tomioka. Three strategies to derive a dual problem.