STK4400 – Week 2

K. R. Dahl & A. B. Huseby

Department of Mathematics
University of Oslo, Norway

Optimization, Lagrange duality and convex duality

Optimization

Optimization is the mathematical theory of maximization and minimization problems.

Useful in many applications, for example in logistic problems, finding the best spot to set up a wind-farm, and in mathematical finance.

Example (finance): Consider an investor who wants to maximize her utility, given various constraints (for instance her salary).

How can we solve this problem?

Basic optimization problem

Let X be a vector space, $f : X \to \bar{\mathbb{R}}$, $g : X \to \mathbb{R}^n$ and $S \subseteq X$.

Consider an optimization problem of the form

$$\begin{array}{ll} \min & f(x) \\ \text{subject to} & \\ & g(x) \leq 0 \text{ (componentwise)} \\ & x \in S. \end{array} \tag{1}$$

In problem (1), f is called the *objective function*.

Furthermore, $g(x) \leq 0$, $x \in S$ are called the *constraints* of the problem.

Transforming an optimization problem

A useful technique when dealing with optimization problems is transforming the problem.

Example: A constraint of the form $h(x) \geq y$ (for $h : X \to \mathbb{R}^n$, $y \in \mathbb{R}^n$) is equivalent to $y - h(x) \leq 0$, which is of the form $g(x) \leq 0$ with $g(x) = y - h(x)$.

Similarly, any maximization problem can be turned into a minimization problem (and visa versa) by using that $\inf f(x) = -\sup(-f(x))$.

Any equality constraint can be transformed into two inequality constraints: $h(x) = 0$ is equivalent to $h(x) \leq 0$ *and* $h(x) \geq 0$.

Extreme value theorem

One of the most important theorems of optimization is the extreme value theorem (see Munkres [?]).

Theorem (The extreme value theorem)

If $f : X \to \mathbb{R}$ is a continuous function from a compact set into the real numbers, then there exist points $a, b \in X$ such that $f(a) \geq f(x) \geq f(b)$ for all $x \in X$. That is, f attains a maximum and a minimum.

The extreme value theorem gives the existence of a maximum and a minimum in a fairly general situation.

NOTE: These may not be unique.

Critical points

For convex (or concave) functions, any local minimum (maximum) is a global minimum (maximum) (from last lecture).

This makes convex functions useful in optimization.

For a function $f : \mathbb{R}^n \to \mathbb{R}$, the maximum and minimum are attained in *critical points*. Critical points are points x such that

- $f'(x) = 0$, where f is differentiable at x,
- the function f is not differentiable at x or
- x is on the boundary of the set one is optimizing over.

For a differentiable function (optimized without extra constraints): Can find maximum and minimum points by solving $f'(x) = 0$ and comparing the objective value in these points to those of the points on the boundary.

Constrained optimization

Constrained optimization can be more difficult.

Example of constrained optimization: Linear programming (LP); maximization of linear functions under linear constraints.

For LP: Strong theorems regarding the solution has been derived. Corresponding to each LP problem, there is a "dual" problem, and these two problems have the same optimal value.

This dual problem gives a second chance at solving an otherwise difficult problem.

There is also an effective numerical method for solving LP problems, called the simplex algorithm. See Vanderbei for more about linear programming.

Lagrange duality

K. R. Dahl & A. B. Huseby

The concept of deriving a "dual" problem to handle constraints is the idea of Lagrange duality as well.

Lagrange duality begins with a problem of the form (1) (or the corresponding maximization problem), and derives a dual problem which gives lower (upper) bounds on the optimal value of the problem.

Linear programming duality is a special case of Lagrange duality.

Since Lagrange duality is more general, one cannot get the strong theorems of linear programming.

The duality concept is generalized even more in convex duality theory.

Lagrange duality

Let X be a general inner product space with inner product $\langle \cdot, \cdot \rangle$.

Consider a problem of the following, very general, form

$$\text{maximize } f(x) \text{ subject to } g(x) \leq 0, \; x \in S \tag{2}$$

where g is a function such that $g : X \to \mathbb{R}^N$ and $S \neq \emptyset$ (to exclude a trivial case).

This is the the primal problem.

Lagrange duality: Transforming the problem to standard form

K. R. Dahl & A. B. Huseby

Equality constraints: Rewrite in the form of problem (2) by writing each equality as two inequalities.

$\geq$ can be turned into $\leq$ by multiplying with -1.

By basic algebra, one can always make sure there is 0 on one side of the inequality.

Note that there are no constraints on f or S and only one (weak) constraint on g. Hence, many problems can be written in the form (2).

Lagrange duality method

Let $\lambda \in \mathbb{R}^N$ be such that $\lambda \geq 0$ (componentwise), and assume that $g(x) \leq 0$ (componentwise) for all $x \in S$.

Then:

$$f(x) \leq f(x) - \lambda \cdot g(x) \tag{3}$$

because $\lambda \cdot g(x) \leq 0$ (where $\cdot$ denotes the Euclidean inner product).

This motivates the definition of the Lagrange function, $L(x, \lambda)$

$$L(x, \lambda) = f(x) - \lambda \cdot g(x).$$

Lagrange duality method, ctd.

Hence, $L(x, \lambda)$ is an upper bound on the objective function for each $\lambda \in \mathbb{R}^N$, $\lambda \geq 0$ and $x \in X$ such that $g(x) \leq 0$.

Taking supremum on each side of the inequality in (3), for each $\lambda \geq 0$,

$$\begin{aligned}
\sup\{f(x) : g(x) \leq 0,\ x \in S\} &\leq \sup\{f(x) - \lambda \cdot g(x) : \\
&\quad g(x) \leq 0, x \in S\} \\
&= \sup\{L(x, \lambda) : x \in S, \\
&\quad g(x) \leq 0\} \\
&\leq \sup_{x \in S} L(x, \lambda) \\
&:= L(\lambda)
\end{aligned} \tag{4}$$

where the second inequality follows because we are maximizing over a larger set, hence the optimal value cannot decrease.

Lagrange function is an upper bound

K. R. Dahl & A. B. Huseby

This implies that for all $\lambda \geq 0$, $L(\lambda)$ is an upper bound for the optimal value function. We want to find the smallest upper bound. This motivates the definition of the Lagrangian dual problem

$$\inf_{\lambda \geq 0} L(\lambda). \tag{5}$$

Therefore, the following theorem is proven (by taking the infimum on the right hand side of equation (4)).

Weak Lagrange duality

Theorem (Weak Lagrange duality)

In the setting above, the following inequality holds

$$\sup\{f(x) : g(x) \leq 0,\ x \in S\} \leq \inf\{L(\lambda) : \lambda \geq 0\}.$$

I.e., the Lagrangian dual problem is the smallest upper bound on the optimal value of problem (2) generated by the Lagrange function.

The Lagrangian dual problem has only one constraint: $\lambda \geq 0$.

This may mean that the dual problem is easier to solve than the original problem.

Strong duality

In some cases, we can show duality theorems:

$$\sup\{f(x) : g(x) \leq 0,\ x \in S\} = \inf_{\lambda \geq 0} L(\lambda)$$

If this is the case, we say that there is no duality gap.

Typically happens in convex optimization problems under certain assumptions.

Often, there is a duality gap, but the Lagrangian dual problem still gives us an upper bound.

Example of Lagrange duality with no duality gap: Linear programming (LP) duality.

Graphical illustration of Lagrange duality

Can illustrate Lagrange duality to see graphically whether there is a duality gap.

Consider problem (2) where $S = X$, and define the set $\mathcal{G} = \{(g(x), f(x)) \in \mathbb{R}^{N+1} : x \in X\}$.

The optimal value of problem (2), denoted p^*, can then be written as $p^* = \sup\{t : (u, t) \in \mathcal{G}, u \leq 0\}$ (from the definitions).

This can be illustrated for $g : X \to \mathbb{R}$ (i.e. for only one inequality) as in Figures 19 and 20.

Graphical illustration of Lagrange duality, ctd.

Figure 20 shows the set $\mathcal{G}$, the optimal primal value p^* and the Lagrange-function for two different Lagrange multipliers. The value of the function $L(l) = \sup_{x \in X}\{f(x) - lg(x)\}$ is given by the intersection of the line $t - lu$ and the t-axis.

The shaded part of $\mathcal{G}$ corresponds to the feasible solutions of problem (2).

To find the optimal primal solution p^* in the figure, find the point (u^*, t^*) in the shaded area of $\mathcal{G}$ such that t^* is as large as possible.

Illustration of Lagrange duality with duality gap

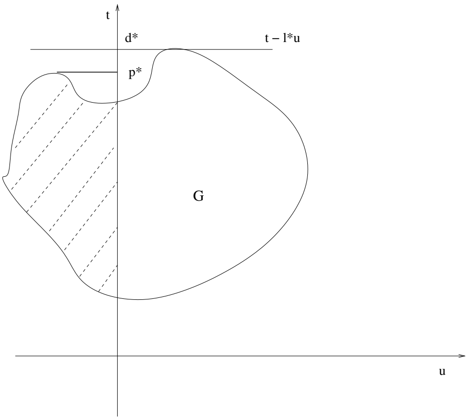
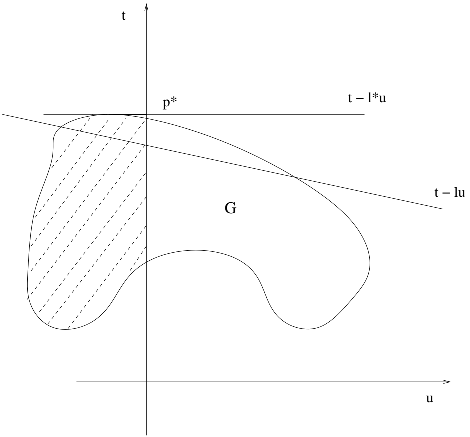


Illustration of Lagrange duality with no duality gap



Finding the optimal dual solution graphically

How can one find the optimal dual solution in Figure 20?

Fix $l \geq 0$, and draw the line $t - lu$.

Find the function $L(l)$ by parallel-adjusting the line so that the intersection of $t - lu$ is as large as possible, while making sure that the line still intersects $\mathcal{G}$.

Now, tilt the line such that l is still greater than or equal 0, but such that the intersection of the line and the t-axis becomes as big as possible. The final intersection is the optimal dual solution.

Actually, there is no duality gap in the problem of Figure 20, since the optimal primal value corresponds to the optimal dual value, given by the intersection of the line $t - l^* u$ and the t-axis.

Finding the optimal dual solution graphically, ctd.

In Figure 19 there is a duality gap, since the optimal dual value, denoted d^* is greater than the optimal primal value, denoted p^*.

What goes wrong?

By examining the two figures above, we note that the absence of a duality gap has something to do with the set $\mathcal{G}$ being "locally convex" near the t-axis.

Bertsekas formalizes this idea, and shows a condition for the absence of a duality gap (in the Lagrange duality case), called the Slater condition.

The Slater condition

The Slater condition for $X = \mathbb{R}^n$ (see Boyd and Vandenberghe): Assume there is a problem of the form (2). If f is concave, $S = X$, each component function of g is convex and there exists $x \in \mathrm{rint}(D)$ (relative interior), where D is defined as the set of $x \in X$ where both f and g are defined, such that $g(x) < 0$, then there is no duality gap.

This condition can be weakened in the case where the component-functions g are affine (and f is still concave) and $\mathrm{dom}(f)$ is open. In this case it is sufficient that there exists a feasible solution for the absence of a duality gap.

Note that for a minimization problem, the same condition holds as long as f is convex (since a maximization problem can be turned into a minimization problem by using that $\sup f = -\inf(-f)$).

Alternative formulation of the Slater condition

Alternative formulation of the Slater condition for $X = \mathbb{R}^n$, from Bertsekas et. al:

If the optimal value of the primal problem (2) is finite, S is a convex set, f and g are convex functions and there exists $x' \in S$ such that $g(x) < 0$, then there is no duality gap.

Convex duality and optimization

Next part is based on *Conjugate Duality and Optimization* by Rockafellar.

Convex functions are good for optimization because a local minimum is also a global minimum.

Also: One can exploit duality properties in order to solve problems.

As before, let X be a linear space, and let $f : X \to \mathbb{R}$ be a function.

Main idea of convex duality: View a given minimization problem $\min_{x \in X} f(x)$ as one half of a minimax problem where a saddle value exists.

Main idea of convex duality

Roughly: Do this by looking at an abstract optimization problem

$$\min_{x \in X} F(x, u) \tag{6}$$

where $F : X \times U \to \mathbb{R}$ is a function such that $F(x, 0) = f(x)$.

U is a linear space and $u \in U$ is a parameter one chooses depending on the particular problem at hand.

Example: u can represent time or some stochastic vector expressing uncertainty in the problem data.

Then, define an optimal value function for this problem

$$\varphi(u) = \inf_{x \in X} F(x, u), \quad u \in U. \tag{7}$$

Convexity of φ

Theorem

Let X, U be real vector spaces, and let $F : X \times U \to \overline{\mathbb{R}}$ be a convex function. Then φ is convex as well.

Proof: This follows from property 10 of the theorem on properties of convex functions from last week.

Convex duality method

Now: Present the dual optimization method in detail,

Let X and Y be general linear spaces, and let $K : X \times Y \to \overline{\mathbb{R}}$ be a function.

Define

$$f(x) = \sup_{y \in Y} K(x, y) \tag{8}$$

and

$$g(y) = \inf_{x \in X} K(x, y). \tag{9}$$

Then, consider two optimization problems

$$(P) \qquad \min_{x \in X} f(x)$$

and

$$(D) \qquad \max_{y \in Y} g(y).$$

Saddle value

From the definitions

$$g(y) \leq K(x,y) \leq f(x), \quad \forall\, x \in X, \forall\, y \in Y. \tag{10}$$

Taking the infimum over x and then the supremum over y in equation (10)

$$\inf_{x \in X} \sup_{y \in Y} K(x,y) = \inf_{x \in X} f(x) \geq \sup_{y \in Y} g(y) = \sup_{y \in Y} \inf_{x \in X} K(x,y). \tag{11}$$

If there is equality in equation (11), then the common value is called the saddle value of K.

Saddle points

The saddle value exists if K has a saddle point, i.e. there exists a point (x', y') such that

$$K(x', y) \leq K(x', y') \leq K(x, y') \tag{12}$$

for all $x \in X$ and for all $y \in Y$. If such a point exists, the saddle value of K is $K(x', y')$.

Existence of a saddle point

Can prove (from the definitions):

Theorem

A point (x', y') is a saddle point for K if and only if x' solves (P), y' solves (D), and the saddle value of K exists, i.e.

$$\inf_{x \in X} f(x) = \sup_{y \in Y} g(y)$$

The proof is left as an exercise.

Dual problems

(P) and (D) are called dual problems.

They can be viewed as half of the problem of finding a saddle point for K.

To prove that (P) and (D) have a solution, and actually find it, one can instead find a saddle point for K.

In convex optimization, we consider the opposite order: Starting with (P), where $f : X \to \mathbb{R}$, how can one choose a space Y and a function K on $X \times Y$ such that $f(x) = \sup_{y \in Y} K(x, y)$ holds?

Freedom: Choose Y and K in different ways, to (hopefully) achieve the properties we want of Y and K.

This idea is called the duality approach.

Examples of convex optimization via duality

We will now look at several examples of how this duality framework can be adapted to different optimization problems.

Example: Nonlinear programming

Let $f_0, f_1, \dots, f_m$ be real valued, convex functions on a nonempty, convex set C in the vector space X.

The duality approach consists of the following steps:

1. The given problem: $\min f_0(x)$ over $\{x \in C : f_i(x) \leq 0 \ \forall \ i = 1, \dots, m\}$.
2. Abstract representation: $\min f$ over X, where

$$f(x) = \begin{cases} f_0(x) & x \in C, \ f_i(x) \leq 0 \text{ for } i = 1, \dots, m \\ +\infty & \text{for all other } x \in X. \end{cases}$$

3. Parametrization: Define (for example) $F(x, u)$ for $u = (u_1, \dots, u_m) \in \mathbb{R}^m$ by $F(x, u) = f_0(x)$ if $x \in C, f_i(x) \leq u_i$ for i = 1, ..., m, and $F(x, u) = +\infty$ for all other x. Then, $F : X \times \mathbb{R}^m \to [-\infty, +\infty]$ is convex and $F(x, 0) = f(x)$

Example: Nonlinear programming with infinitely many constraints

Let $f_0 : C \to \mathbb{R}$ where $C \subset X$ is convex, and let $h : X \times S \to \bar{\mathbb{R}}$ be convex in the x-argument, where S is an arbitrary set.

1. The problem: $\min f_0(x)$ over $K = \{x \in C : h(x, s) \leq 0 \ \forall \ s \in S\}$.
2. Abstract representation: $\min f(x)$ over X, where $f(x) = f_0(x)$ if $x \in K$, and $f(x) = +\infty$ for all other x.
3. Parametrization: Choose u analoglously with Example **??**: Let U be the linear space of functions $u : S \to \mathbb{R}$ and let $F(x, u) = f_0(x)$ if $x \in C$, $h(x, s) \leq u(s) \ \forall \ s \in S$ and $F(x, u) = +\infty$ for all other x. As in the previous example, this makes F convex and satisfies $F(x, 0) = f(x)$.

Example: Stochastic optimization

Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $h : X \times \Omega \to \bar{\mathbb{R}}$ be convex in the x-argument, where X is a linear, topological space.

Let C be a closed, convex subset of X.

- The general problem: $\min h(x, \omega)$ over all $x \in C$, where ω is a stochastic element with a known distribution. The difficulty here is that x must be chosen before ω has been observed.
- Abstract representation:We therefore solve the following problem: Minimize the expectation $f(x) = \int_\Omega h(x, \omega) dP(\omega)$ over all $x \in X$. Here, it is assumed that h is measurable, so that f is well defined. Rockafellar then shows in [**?**], Theorem 3, that f actually is convex.

Example: Stochastic optimization, ctd.

- Parametrization: Let $F(x, u) = \int_\Omega h(x - u(\omega), \omega) dP(\omega) + \delta_C(u)$ for $u \in U$, where U is a linear space of measurable functions and δ_C is the indicator function of C, as defined in Definition **??**. Then F is (by the same argument as for f) well defined and convex, with $F(x, 0) = f(x)$.

Conjugate functions in paired spaces

The material in this section is based on Rockafellar and Rockafellar and Wets.

Definition (Pairing of spaces)

A pairing of two linear spaces X and V is a real valued bilinear form $\langle \cdot, \cdot \rangle$ on $X \times V$.

The pairing associates for each $v \in V$ a linear function $\langle \cdot, v \rangle : x \mapsto \langle x, v \rangle$ on X, and similarly for X.

Compatible topology

K. R. Dahl & A. B. Huseby

Definition (Compatible topology)

Assume there is a pairing between the spaces X and V. A topology on X is compatible with the pairing if it is a locally convex topology such that the linear function $\langle \cdot, v \rangle$ is continuous, and any continuous linear function on X can be written in this form for some $v \in V$. A compatible topology on V is defined similarly.

Definition (Paired spaces)

X and V are paired spaces if one has chosen a pairing between X and V, and the two spaces have compatible topologies with respect to the pairing.

Examples of paired spaces

Example

Let $X = \mathbb{R}^n$ and $V = \mathbb{R}^n$. Then, the standard Euclidean inner product is a bilinear form, so X and V become paired spaces.

Example

Let $X = L^1(\Omega, \mathcal{F}, P)$ and $V = L^\infty(\Omega, \mathcal{F}, P)$. Then X and V are paired via the bilinear form $\langle x, v \rangle = \int_\Omega x(s)v(s)dP(s)$. Similarly, the spaces $X = L^p(\Omega, F, P)$ and $V = L^q(\Omega, F, P)$, where $\frac{1}{p} + \frac{1}{q} = 1$, are paired.

Convex conjugate of a function

Now: A central notion of convex duality, the conjugate of a function.

Definition (Convex conjugate of a function, f^*)

Let X and V be paired spaces. For a function $f : X \to \bar{\mathbb{R}}$, define the conjugate of f, denoted by $f^* : V \to \bar{\mathbb{R}}$, by

$$f^*(v) = \sup\{\langle x, v\rangle - f(x) : x \in X\}. \tag{13}$$

Finding f^* is called taking the conjugate of f in the convex sense.

One may also define the conjugate g^* of a function $g : V \to \bar{\mathbb{R}}$ similarly.

Biconjugate of a function and the Fenchel transform

K. R. Dahl & A. B. Huseby

Similarly, define

Definition (Biconjugate of a function, f^{**})

Let X and V be paired spaces. For a function $f : X \to \bar{\mathbb{R}}$, define *the biconjugate* of f, f^{**}, to be the conjugate of f^*, so $f^{**}(x) = \sup\{\langle x, v\rangle - f^*(v) : v \in V\}$.

Definition

The operation $f \mapsto f^*$ is called *the Fenchel transform*.

If $f : \mathbb{R}^n \to \bar{\mathbb{R}}$, then the operation $f \mapsto f^*$ is sometimes called the Legendre-Fenchel transform.

Why is the conjugate function important?

K. R. Dahl & A. B. Huseby

To understand why, we consider f^* via the epigraph.

Most easily done in $\mathbb{R}^n$, so let $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ and consider $X = \mathbb{R}^n = V$.

From equation (13), it is not difficult to show that (try this!)

$$(v, b) \in \text{epi}(f^*) \iff b \geq \langle v, x \rangle - a \text{ for all } (x, a) \in \text{epi}(f). \quad (14)$$

This can also be expressed as (check!)

$$(v, b) \in \text{epi}(f^*) \iff l_{v,b} \leq f \quad (15)$$

where $l_{v,b}(x) := \langle v, x \rangle - b$.

Why is the conjugate function important? Ctd.

Specifying a function on $\mathbb{R}^n$ is equivalent to specifying its epigraph.

Hence, f^* describes the family of all affine functions that are majorized by f (since all affine functions on $\mathbb{R}^n$ are of the form $\langle v, x\rangle - b$ for fixed v, b).

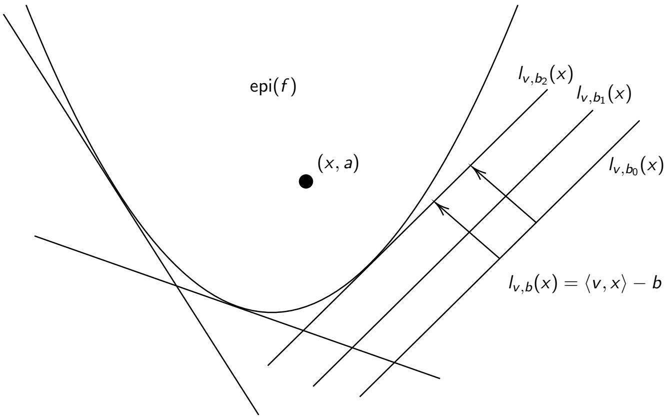
But from equation (14)

$$b \geq f^*(v) \iff b \geq l_{x,a}(v) \text{ for all } (x, a) \in \mathrm{epi}(f).$$

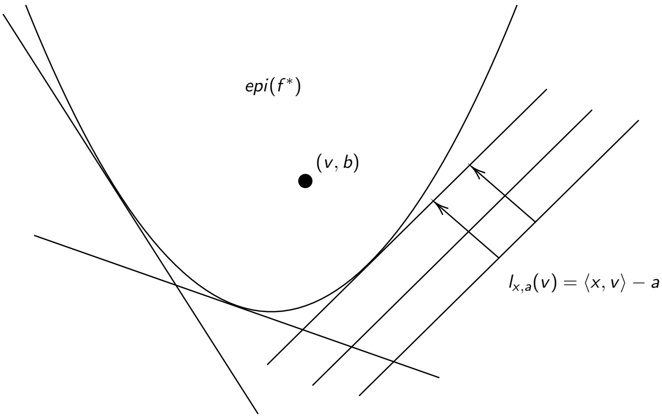
This means that f^* is the pointwise supremum of all affine functions $l_{x,a}$ for $(x, a) \in \mathrm{epi}(f)$.

This is illustrated in the following figures.

Affine functions majorized by f



Affine functions majorized by f^*



$f^{**} = \mathrm{cl}(\mathrm{co}(f))$

We then have the following very central theorem on duality:

Theorem

Let $f : X \to \bar{\mathbb{R}}$ be arbitrary. Then the conjugate f^* is a closed, convex function on V and $f^{**} = \mathrm{cl}(\mathrm{co}(f))$. Similarly if one starts with a function in V.

In particular, the Fenchel transform induces a one-to-one correspondence $f \mapsto h, h = f^*$ between the closed, convex functions on X and the closed, convex functions on V.

Proof

Proof: By definition f^* is the pointwise supremum of the continuous, affine functions $V \mapsto \langle x, v\rangle - \alpha$, where $(x, \alpha) \in \mathrm{epi}(f)$. Therefore, f^* is convex and lsc, hence it is closed. $(v, \beta) \in \mathrm{epi}(f^*)$ if and only if the continuous affine function $x \mapsto \langle x, v\rangle - \beta$ satisfies $f(x) \geq \langle x, v\rangle - \beta$ for all $x \in X$, that is if the epigraph of this affine function contains the epigraph of f. Thus, $\mathrm{epi}(f^{**})$ is the intersection of all the nonvertical, closed halfspaces in $X \times \mathbb{R}$ containing $\mathrm{epi}(f)$. This implies, using what a closed, convex set is, that $f^{**} = \mathrm{cl}(\mathrm{co}(f))$.

Theorem 13 implies that if f is convex and closed, then $f = f^{**}$. This gives a one-to-one correspondence between the closed convex functions on X, and the same type of functions on V. Hence, all properties and operations on such functions must have conjugate counterparts (see Rockafellar and Wets).

Example: The indicator function

Let X and V be paired spaces, and let $f = \delta_L$ where $L \subseteq X$ is a subspace (so in particular, L is convex) and δ_L is the indicator function of L, as defined last week.

From last week, we know that $f = \delta_L$ is convex. Then

$$\begin{aligned} \delta_L^*(v) &= \sup\{\langle x, v\rangle - \delta_L(x) : x \in X\} \\ &= \sup\{\langle x, v\rangle ; x \in L\} \end{aligned}$$

since $\langle x, v\rangle - \delta_L(x) = -\infty$ if $x \notin L$. This function δ_L^* is called the *support function* of L (and is often denoted by ψ_L).

Note also that

$$\delta_L^*(v) = \delta_{L^\perp}(v)$$

because if $v \in L^\perp$, then $\langle x, v\rangle = 0$ for all $x \in L$, but if $v \notin L^\perp$ then $\langle x', v\rangle \neq 0$ for some $x' \in L$.

Example: The indicator function, ctd.

Hence, since L is a subspace, $\langle x', v \rangle$ can be made arbitrarily large by multiplying x' by either $+t$ or $-t$ (in order to make $\langle x', v \rangle$ positive), and letting $t \to +\infty$.

By a similar argument

$$\delta_L^{**} = \delta_{(L^\perp)^\perp}. \tag{16}$$

Conjugate in the concave sense

For a concave function $g : X \to \bar{\mathbb{R}}$ one can define the conjugate as:

$$g^*(v) = \inf\{\langle x, v\rangle - g(x) : x \in X\} \tag{17}$$

This is called taking the conjugate of g in the concave sense.

Status quo: Dual problems and Lagrangians

We have an abstract minimization problem:

$$(P) \qquad \min_{x \in X} f(x)$$

which is assumed to have the representation:

$$f(x) = F(x, 0), \quad F : X \times U \to \bar{\mathbb{R}} \tag{18}$$

(where U is some linear space).

Everything depends on the choice of U and F!

Want to exploit duality: Let X be paired with V, and U paired with Y, where U and Y are linear spaces (the choice of pairings may also be important in applications).

Preferably, we want to choose (F, U) such that F is a closed, jointly convex function of x and u.

The Lagrange function

K. R. Dahl & A. B. Huseby

Definition (The Lagrange function, $K(x, y)$)

Define the Lagrange function $K : X \times Y \to \bar{\mathbb{R}}$ to be

$$K(x, y) = \inf\{F(x, u) + \langle u, y \rangle : u \in U\}. \tag{19}$$

K is closed and convex

K. R. Dahl & A. B. Huseby

Theorem

The Lagrange function K is closed, concave in $y \in Y$ for each $x \in X$, and if $F(x, u)$ is closed and convex in u

$$f(x) = \sup_{y \in Y} K(x, y). \tag{20}$$

Conversely, if K is an arbitrary extended-real valued function on $X \times Y$ such that (20) holds, and if K is closed and concave in y, then K is the Lagrange function associated with a unique representation $f(x) = F(x, 0), F : X \times U \rightarrow \bar{\mathbb{R}}$ where F is closed and convex in u. This means that

$$F(x, u) = \sup\{K(x, y) - \langle u, y \rangle : y \in Y\}.$$

Further, if F is closed and convex in u, K is convex in x if and only if $F(x, u)$ is jointly convex in x and u.

Proof of the theorem

Proof: Everything in the theorem, apart from the last statement, follows from Theorem 13. For the last statement, assume that F and K respectively are convex, use the definitions of F and K and that the supremum and infimum of convex functions are convex (see properties of convex functions from last week).

The dual problem gives a lower bound on the primal

We now define, motivated by equation (20), the dual problem of (P),

$$(D) \qquad \max_{y \in Y} g(y)$$

where $g(y) = \inf_{x \in X} K(x, y)$.
Note that this dual problem gives a lower bound on the primal problem, from (20) since

$$K(x, y) \geq \inf_{x \in X} K(x, y) = g(y).$$

But then

$$\sup_{y \in Y} K(x, y) \geq \sup_{y \in Y} g(y).$$

Weak duality

So from equation (20), $f(x) \geq \sup_{y \in Y} g(y)$.

Therefore, taking the infimum with respect to $x \in X$ on the left hand side implies $(D) \leq (P)$.

This is called weak duality.

Sometimes, one can prove that the dual and primal problems have the same optimal value. If this is the case, there is no duality gap and strong duality holds.

If φ is convex and l.s.c., there is no duality gap

Theorem

The function g in (D) is closed and concave. By taking the conjugate in concave sense, $g = -\varphi^*$, hence $-g^* = \mathrm{cl}(\mathrm{co}(\varphi))$, so

$$\sup_{y \in Y} g(y) = \mathrm{cl}(\mathrm{co}(\varphi))(0)$$

while

$$\inf_{x \in X} f(x) = \varphi(0)$$

In particular, if $F(x, u)$ is convex in (x, u), then $-g^* = \mathrm{cl}(\varphi)$ and $\sup_{y \in Y} g(y) = \liminf_{u \to 0} \varphi(u)$ (except if $0 \notin \mathrm{cl}(\mathrm{dom}(\varphi)) \neq \emptyset$, and $\mathrm{lsc}(\varphi)$ is nowhere finite valued).

For the proof, see Rockafellar.

Summary of the convex duality method

- To begin, there is a minimization problem $\min_{x \in X} f(x)$ which cannot be solved directly.
- Find a function $F : X \times U \to \bar{\mathbb{R}}$, where U is a vector space, such that $f(x) = F(x, 0)$.
- Introduce the linear space Y, paired with U, and define the Lagrange function $K : X \times Y \to \bar{\mathbb{R}}$ by $K(x, y) = \inf_{u \in U}\{F(x, u) + \langle u, y \rangle\}$.
- Try to find a saddle point for K. If this succeeds, Theorem 4 tells us that this gives the solution of (P) and (D).
- Theorem 16 tells us that K has a saddle point if and only if $\varphi(0) = (\mathrm{cl}(\mathrm{co}(\varphi)))(0)$. Hence, if the value function φ is convex, the lower semi-continuity of φ is a sufficient condition for the absence of a duality gap.

Example: Nonlinear programming

The Lagrange function takes the form

$$
\begin{aligned}
K(x,y) &= \inf\{F(x,u) + \langle u, y\rangle : u \in U\} \\
&= \inf\{\begin{cases} f_0(x) + \langle u, y\rangle ; x \in C, f_i(x) \leq u_i \\ +\infty + \langle u, y\rangle ; \forall \text{ other } x \end{cases} : u \in U\} \\
&= \begin{cases} f_0(x) + \inf\{\langle u, y\rangle : u \in U, f_i(x) \leq u_i\}, x \in C \\ +\infty, \ \text{otherwise.} \end{cases} \\
&= \begin{cases} \inf\{f_0(x) + f_1(x)y_1 + \ldots + f_m(x)y_m\}, u \in U, x \in C \\ -\infty, x \in C, y \notin \mathbb{R}^m_+ \\ +\infty, \ \text{otherwise.} \end{cases}
\end{aligned}
$$

where the last equality follows because if there is at least one negative y_j, one can choose u_j arbitrarily large and make the above expression arbitrarily small.

STK
W

K. R
A. B

Therefore, the dual function is

$$
\begin{aligned}
g(y) &= \inf_{x\in X} K(x,y) \\
&= \inf_{x\in X} \begin{cases} f_0(x) + f_1(x)y_1 + \ldots + f_m(x)y_m \text{ if } x \in C, y \in \mathbb{R}^m_+ \\ -\infty, x \in C, y \notin \mathbb{R}^m_+ \\ +\infty, \text{ otherwise.} \end{cases} \\
&= \begin{cases} \inf_{x\in C}\{f_0(x) + f_1(x)y_1 + \ldots + f_m(x)y_m\} \text{ if } y \in \mathbb{R}^m_+ \\ -\infty, y \notin \mathbb{R}^m_+. \end{cases}
\end{aligned}
$$

Convex duality generalizes LP

By making some small alterations to the approach above, Rockafellar shows that by beginning with the standard primal linear programming problem

$$\max\{\langle c, x\rangle : Ax \leq b, \ x \geq 0\}$$

where c and b are given vectors and A is a given matrix, and finding its dual problem (in the above sense), one gets the standard dual LP-problem back. That is

$$\min\{\langle b, y\rangle : A^T y \geq c, \ y \geq 0\}$$

(see Vanderbei).