MA427 Lecture 9
Lagrangian duality and Karush-Kuhn-Tucker conditions

Giacomo Zambelli

Department of Mathematics

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

11 March, 2019

Today's lecture

- Lagrangian duality for convex programs
- Complementarity: Karush-Kuhn-Tucker conditions

Convex Optimisation Problems

$$\begin{aligned} \min \quad & f_0(x) \\ & f_i(x) \leq 0 \quad i = 1, \ldots, m \\ & h_i(x) = 0 \quad i = 1, \ldots, k \end{aligned}$$

- Domain of problem: $\mathcal{D} = \left(\bigcap_{i=0}^{m} \textbf{dom}\, f_i \right) \bigcap \left(\bigcap_{i=1}^{k} \textbf{dom}\, h_i \right)$.
- Feasible region: set X of all points in $\mathcal{D}$ satisfying the constraints.

Convex problem if:

- $f_0, \ldots, f_m$ are convex,
- $h_1, \ldots, h_k$ are affine (i.e. there exist $a_i \in \mathbb{R}^n$, $b_i \in \mathbb{R}$ such that $h_i(x) = a_i^\top x - b_i$)

Important facts

1. If f_i convex, then $\{x \in \mathbb{R}^n : f_i(x) \leq 0\}$ is a convex set.
2. The feasible region X is convex, because it is the intersection of convex sets.
3. Every local optimum for f_0 in X is also a global optimum.

Example: Markowitz portfolio optimisation

- Invest a budget B in n assets. Return of asset i is random variable p_i, with expected value $\bar{p}_i$.
- Covariance matrix Σ.
- Target expected return at least $r_{\min}$.
- Among such portfolios, minimise risk, which is the variance $\mathbb{E}[(p^\top x - \bar{p}^\top x)^2] = x^\top \Sigma x$.

Example: Markowitz portfolio optimisation

- Invest a budget B in n assets. Return of asset i is random variable p_i, with expected value $\bar{p}_i$.
- Covariance matrix Σ.
- Target expected return at least r_{min}.
- Among such portfolios, minimise risk, which is the variance $\mathbb{E}[(p^\top x - \bar{p}^\top x)^2] = x^\top \Sigma x$.

$$
\begin{aligned}
\min\ & x^\top \Sigma x \\
\text{s.t. } \bar{p}^\top x & \geq r_{\text{min}} \\
\textstyle\sum_{i=1}^n x_i & = B \\
x & \geq 0
\end{aligned}
$$

- Convex program: Σ is *positive semidefinite*.

Concave functions

- A function f is *concave* if $-f$ is convex.
- The previous theorems concerning minima of convex functions are true if we replace "convex" with "concave" and "minimum" with "maximum".
  - Local maximum $=$ Global maximum.
  - Local maximum if and only if gradient is zero.

Concave maximisation

- A maximisation problem of the form

$$\begin{aligned} \max \quad & f_0(x) \\ & f_i(x) \leq 0 \quad i = 1, \ldots, m \\ & h_i(x) = 0 \quad i = 1, \ldots, k \end{aligned}$$

  is a convex optimisation problem if f_0 is concave, while $f_1, \ldots, f_m$ are convex and $h_1, \ldots, h_k$ affine,
- Indeed, the equivalent problem obtained replacing "$\max f_0(x)$" with "$\min -f_0(x)$" is a convex optimisation problem.

Lagrangian duality

Lagrangian

To get lower-bounds on the optimal value p^* of (P), we assign multipliers to the constraints:

- λ_i: Lagrange multiplier of $f_i(x) \leq 0$, $i = 1, \ldots, m$;
- ν_i: Lagrange multiplier of $h_i(x) = 0$, $i = 1, \ldots, k$

The *Lagrangian* of (P) is the function

$$L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^k \to \mathbb{R}$$

defined by

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{k} \nu_i h_i(x),$$

where $\textbf{dom}\, L = \mathcal{D} \times \mathbb{R}^m \times \mathbb{R}^k$.

We do not assume convexity for now!

Lagrangian

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{k} \nu_i h_i(x),$$

Lemma

For every feasible point $\bar{x}$ for (P) and every λ, ν with $\lambda \geq 0$, we have that

$$L(\bar{x}, \lambda, \nu) \leq f_0(\bar{x}).$$

Example

$$p^* = \min \quad x^2 + 1$$
$$(x-2)(x-4) \le 0$$

- The domain of the problem is $\mathbb{R}$,
- The feasible region is the interval $[2,4]$,
- Minimum is attained at $x^* = 2$, $p^* = 5$.
- The Lagrangian of the above problem is the function of two variables

$$\begin{aligned} L(x,\lambda) &= x^2 + 1 + \lambda(x-2)(x-4) \\ &= (1+\lambda)x^2 - 6\lambda x + 8\lambda + 1. \end{aligned}$$

Example

$$L(x, \lambda) = x^2 + 1 + \lambda(x-2)(x-4)$$

Lagrangian dual function

Lagrange dual function $g : \mathbb{R}^m \times \mathbb{R}^k \to \mathbb{R}$

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu).$$

Lagrangian dual function

Lagrange dual function $g : \mathbb{R}^m \times \mathbb{R}^k \to \mathbb{R}$

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu).$$

- Computing $g(\lambda, \nu)$ is an unconstrained problem in x.

Lagrangian dual function

Lagrange dual function $g : \mathbb{R}^m \times \mathbb{R}^k \to \mathbb{R}$

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu).$$

- Computing $g(\lambda, \nu)$ is an unconstrained problem in x.
- For every λ, ν with $\lambda \geq 0$,

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) \leq \inf_{x \in X} L(x, \lambda, \nu) \leq \inf_{x \in X} f_0(x) = p^*.$$

$\Rightarrow$ For every λ, ν with $\lambda \geq 0$, $g(\lambda, \nu)$ is a lower bound on p^*.

Lagrangian dual function

Lagrange dual function $g : \mathbb{R}^m \times \mathbb{R}^k \to \mathbb{R}$

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu).$$

- Computing $g(\lambda, \nu)$ is an unconstrained problem in x.
- For every λ, ν with $\lambda \geq 0$,

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) \leq \inf_{x \in X} L(x, \lambda, \nu) \leq \inf_{x \in X} f_0(x) = p^*.$$

$\Rightarrow$ For every λ, ν with $\lambda \geq 0$, $g(\lambda, \nu)$ is a lower bound on p^*.

Lemma
The Lagrangian dual $g(\lambda, \nu)$ is a concave function (even if (P) is not a convex problem).

Example (continued)

- Problem

$$p^* = \min \quad x^2 + 1 \\ \quad (x-2)(x-4) \leq 0$$

- Lagrangian

$$L(x, \lambda) = (1+\lambda)x^2 - 6\lambda x + 8\lambda + 1.$$

- The Lagrangian dual is

$$g(\lambda) = \frac{-\lambda^2 + 9\lambda + 1}{1+\lambda}.$$

$\mathbf{dom}\, g = \{\lambda \in \mathbb{R} \,:\, \lambda > -1\}$

Example

$$L(x, \lambda) = x^2 + 1 + \lambda(x - 2)(x - 4)$$

Lagrangian dual problem

- For every λ, ν with $\lambda \geq 0$, $g(\lambda, \nu)$ is a lower bound on p^*.
- The best possible lower bound is therefore

$$\begin{aligned} d^* = \max \quad & g(\lambda, \nu) \\ \text{s.t.} \quad & \lambda \geq 0 \\ & \lambda \in \mathbb{R}^m,\ \nu \in \mathbb{R}^k \end{aligned}$$

- The above is the *Lagrangian dual problem*.
- It is a convex optimization problem because g is a concave function.

Lagrangian dual problem

- For every λ, ν with $\lambda \geq 0$, $g(\lambda, \nu)$ is a lower bound on p^*.
- The best possible lower bound is therefore

$$\begin{aligned} d^* = \max \quad & g(\lambda, \nu) \\ \text{s.t.} \quad & \lambda \geq 0 \\ & \lambda \in \mathbb{R}^m,\ \nu \in \mathbb{R}^k \end{aligned}$$

- The above is the *Lagrangian dual problem*.
- It is a convex optimization problem because g is a concave function.

Theorem (Week Lagrangian Duality)

$$d^* \leq p^*.$$

Example (continued)

- Problem

$$
\begin{aligned}
p^* = \min \quad & x^2 + 1 \\
& (x-2)(x-4) \leq 0
\end{aligned}
$$

- Lagrangian Dual problem

$$
\begin{aligned}
d^* = \max \quad & \frac{-\lambda^2 + 9\lambda + 1}{1+\lambda} \\
\text{s.t.} \quad & \lambda \geq 0
\end{aligned}
$$

Example (continued)

- Problem

$$p^* = \min \quad x^2 + 1$$
$$(x-2)(x-4) \leq 0$$

- Lagrangian Dual problem

$$d^* = \max \quad \frac{-\lambda^2 + 9\lambda + 1}{1+\lambda}$$
$$\text{s.t.} \quad \lambda \geq 0$$

- Solving: $d^* = 5$ achieved for $\lambda = 2$.

Duality gap and strong duality

- Duality gap: $p^* - d^*$.
- Weak duality says that the duality gap is ≥ 0.
- If the duality gap is 0, we say that strong duality holds.
- Strong duality might not hold, even if (P) is convex.

Duality gap and strong duality

- Duality gap: $p^* - d^*$.
- Weak duality says that the duality gap is ≥ 0.
- If the duality gap is 0, we say that strong duality holds.
- Strong duality might not hold, even if (P) is convex.
- Example:

$$\begin{aligned} p^* = \min \quad & e^{-x_1} \\ \text{s.t.} \quad & \frac{x_1^2}{x_2} \leq 0 \end{aligned}$$

- Domain: $\mathcal{D} = \{x \in \mathbb{R}^2 : x_2 > 0\}$,
- Feasible region: $X = \{x \in \mathbb{R}^2 : x_1 = 0, x_2 > 0\}$.

Duality gap and strong duality

- Duality gap: $p^* - d^*$.
- Weak duality says that the duality gap is ≥ 0.
- If the duality gap is 0, we say that strong duality holds.
- Strong duality might not hold, even if (P) is convex.
- Example:

$$p^* = \min \quad e^{-x_1} \\ \text{s.t.} \quad \frac{x_1^2}{x_2} \leq 0$$

- Domain: $\mathcal{D} = \{x \in \mathbb{R}^2 : x_2 > 0\}$,
- Feasible region: $X = \{x \in \mathbb{R}^2 : x_1 = 0, x_2 > 0\}$.
- All feasible points are optimal, with value $p^* = 1$.
- The Lagrangian dual has optimum $d^* = 0$.
- Duality gap is 1.

Duality gap and strong duality

The previous situation above is "pathological": the inequality constraint is always satisfied as "$=$".

Definition (Slater conditions)

(P) satisfies the Slater conditions if there exists a feasible solution x^* in the interior of $\mathcal{D}$ such that $f_i(x) < 0$ for $i = 1, \ldots, m$.

Theorem (Strong duality under Slater conditions)

Strong duality holds for every convex optimization problem satisfying Slater condition.

Lagrangian dual of LP problems

What is the Lagrangian dual of

$$\begin{aligned} \min \quad & c^\top x \\ \text{s.t.} \quad & Ax \geq b \end{aligned}$$

?

Lagrangian dual of LP problems

What is the Lagrangian dual of

$$\begin{aligned} \min \quad & c^\top x \\ \text{s.t.} \quad & Ax \geq b \end{aligned}$$

?

$$\begin{aligned} \max \quad & b^\top \lambda \\ \text{s.t.} \quad & A^\top \lambda = c \\ & \lambda \geq 0. \end{aligned}$$

The usual LP dual!

Karush-Kuhn-Tucker conditions

Karush-Kuhn-Tucker conditions

Lemma

Assume the following:

- $f_0, f_1, \ldots, f_m,\ h_1, \ldots, h_k : \mathbb{R}^n \to \mathbb{R}$ *are differentiable.*
- *strong duality holds.*
- *Primal and Dual both admit optima:* x^* *and* (λ^*, ν^*)*.*

Then x^* *and* (λ^*, ν^*) *satisfy the following:*

$$\begin{array}{rcll} f_i(x^*) & \leq & 0 & (i = 1, \ldots, m) \\ h_i(x^*) & = & 0 & (i = 1, \ldots, k) \\ \lambda_i^* & \geq & 0 & (i = 1, \ldots, m) \\ \lambda_i^* f_i(x^*) & = & 0 & (i = 1, \ldots, m) \\ \multicolumn{4}{l}{\nabla f_0(x^*) + \sum_{i=1}^m \lambda_i^* \nabla f_i(x^*) + \sum_{i=1}^k \nu_i^* \nabla h_i(x^*) = 0.} \end{array} \qquad (KKT)$$

KKT conditions

Theorem

Let $f_0, f_1, \ldots, f_m$ be convex differentiable functions and $h_1, \ldots, h_k$ be affine functions. If the KKT conditions have a solution (x^*, λ^*, ν^*), then x^* is optimal for the primal, (λ^*, ν^*) is optimal for the dual, and strong duality holds.

KKT conditions

Theorem

Let $f_0, f_1, \ldots, f_m$ be convex differentiable functions and $h_1, \ldots, h_k$ be affine functions. If the KKT conditions have a solution (x^*, λ^*, ν^*), then x^* is optimal for the primal, (λ^*, ν^*) is optimal for the dual, and strong duality holds.

Proof.

$$
\begin{aligned}
g(\lambda^*, \nu^*) &= \inf_{x \in \mathcal{D}} \left\{ f_0(x) + \sum_{i=1}^{m} \lambda_i^* f_i(x) + \sum_{i=1}^{k} \nu_i^* h_i(x) \right\} \\
&= f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* f_i(x^*) + \sum_{i=1}^{k} \nu_i^* h_i(x^*) \\
&= f_0(x^*).
\end{aligned}
$$

KKT conditions

Note that the previous statement does not hold for non-convex problems. That is, if (x^*, λ^*, ν^*) satisfies the KKT conditions, it does not mean that x^* is optimal for the primal.

Example:

- $p^* = \min\{x^3 : x^2 \leq 1\}$ (not convex)
- Optimal solution is $x = -1$, $p^* = -1$.
- KKT conditions

$$\begin{aligned} x^2 - 1 &\leq 0 \\ \lambda &\geq 0 \\ \lambda(x^2 - 1) &= 0 \\ 3x^2 + 2\lambda x &= 0. \end{aligned}$$

- $(x^*, \lambda^*) = (0, 0)$ is a solution for the KKT conditions, but $x^* = 0$ is not a primal optimum.

Example

Consider the problem

$$\begin{array}{lll} \min & \frac{1}{2}x^\top Px + q^\top x + r & \\ s.t. & \quad -1 \leq x_i \leq 1 & \quad i = 1,2,3 \end{array}$$

where

$$P = \begin{pmatrix} 13 & 12 & -2 \\ 12 & 17 & 6 \\ -2 & 6 & 12 \end{pmatrix}, \quad q = \begin{pmatrix} -22 \\ -29/2 \\ 13 \end{pmatrix}, \quad r = 1.$$

We will show that the point $x^* = (1, 1/2, -1)$ is a global optimum.