Lecture 11: Convex Optimization II

Isabel Valera

Machine Learning Group
Department of Mathematics and Computer Science
Saarland University, Saarbrücken, Germany

10.06.2024

Outline

Main references

- Bishop: Chapters 1–5
- Summary from Zico Kolter

Motivation of the Lagrange dual problem:

The general form of an optimization problem (OP) is given by

$$
\begin{aligned}
p^* = \min_{\mathbf{x} \in Domain} & \ f(\mathbf{x}), \\
\text{subject to: } & \ g_i(\mathbf{x}) \leq 0, \ i = 1, \ldots, r \\
& \ h_j(\mathbf{x}) = 0, \ j = 1, \ldots, s.
\end{aligned}
$$

From now on, we will refer to the above OP as the **primal problem**.

Idea behind the Lagrange dual problem:

Map a hard constrained optimization problem (primal problem) into an easier (e.g., unconstrained) one (dual problem)!

Outline

The Lagrange function I

Definition

The **Lagrangian** or **Lagrange function** $L : \mathbb{R}^n \times \mathbb{R}^r_+ \times \mathbb{R}^s \to \mathbb{R}$ associated with the optimization problem is defined as

$$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = f(\mathbf{x}) + \sum_{j=0}^{r} \lambda_j \, g_j(\mathbf{x}) + \sum_{i=0}^{s} \nu_i \, h_i(\mathbf{x}),$$

with $\operatorname{dom} L = D \times \mathbb{R}^r_+ \times \mathbb{R}^s$ where D is the domain of the OP. The variables λ_j and ν_i are called **Lagrange multipliers** associated with the inequality and equality constraints.

Note: The Lagrange multipliers $\{\lambda_j\}_{j=1}^{r}$ of inequality constraints are non-negative!

The Lagrange function II

For all feasible $\mathbf{x}$ (*reminder:* $\mathbf{x}$ is feasible if the inequality and equality constraints hold at x), **the Lagrange function** $L(\mathrm{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$ **is a lower-bound of the objective function** $f(\mathrm{x})$, since

$$\sum_{j=0}^{r} \lambda_j \, g_j(\mathbf{x}) + \sum_{i=0}^{s} \nu_i \, h_i(\mathbf{x}) \leq 0 \quad \Longrightarrow \quad \boxed{L(\mathbf{x}, \lambda, \nu) \leq f(\mathbf{x})}$$

Outline

The dual function I

Definition

The **Lagrange dual function** $q : \mathbb{R}^r_+ \times \mathbb{R}^s \to \mathbb{R}$ associated with the OP is defined as

$$q(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \inf_{\mathbf{x} \in D} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = \inf_{\mathbf{x} \in D} \left(f(\mathbf{x}) + \sum_{j=0}^{r} \lambda_j \, g_j(\mathbf{x}) + \sum_{i=0}^{s} \nu_i \, h_i(\mathbf{x}) \right),$$

$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$ is unbounded from below in $\mathbf{x}$.

The dual function II

Properties:

- **The dual function** $q(\boldsymbol{\lambda}, \boldsymbol{\nu})$ **is** a pointwise infimum of a family of concave functions (in $\boldsymbol{\lambda}$ and $\boldsymbol{\nu}$) and therefore **concave**. This holds irrespectively of the character of the primal problem (even if non-convex). Moroever, $q(\boldsymbol{\lambda}, \boldsymbol{\nu})$ can be $-\infty$ for some values of $\boldsymbol{\lambda}, \boldsymbol{\nu}$. *Note:* The pointwise supremum of a family of convex functions is convex. The pointwise infimum of family of concave functions is concave. Refer to Boyd for operations that preserve convexity.
- For any $\boldsymbol{\lambda} \succeq \mathbf{0}$ and $\boldsymbol{\nu} \in \mathbb{R}^s$, any feasible $\mathbf{x}'$ satisfies that

$$\inf_{\mathbf{x} \in D} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \leq f(\mathbf{x}') \quad \Longrightarrow \quad \boxed{q(\boldsymbol{\lambda}, \boldsymbol{\nu}) \leq p^*}.$$

In other words, **the dual function lower-bounds the optimal value of the primal problem.**

What is the best possible lower bound of p^*?

The dual problem

Definition

The **Lagrange dual problem** is defined as

$$d^* = \max_{\boldsymbol{\lambda},\boldsymbol{\nu}} q(\boldsymbol{\lambda}, \boldsymbol{\nu}),$$

$$\text{subject to: } \lambda_i \geq 0, \; i = 1, \ldots, r.$$

Properties:

- For each OP the dual problem is **convex**, as the objective to be maximized is concave and the constraint is convex.
- The dual problem finds the best lower bound on p^*, obtained from Lagrange dual function. Dual problems are often be easier to solve than their corresponding primal problems.
- $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ is **dual feasible** if $q(\boldsymbol{\lambda}, \boldsymbol{\nu}) > -\infty$.

Weak and strong duality

Definition

Let d^* and p^* be the optimal values of the dual and primal problem. Then, weak duality holds (independently of the nature of the primal problem), i.e,

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

Definition

We say that **strong duality** holds if $d^* = p^*$.

- The difference $p^* - d^*$ is the **optimal duality gap**.
- Strong duality does not hold in general. But for convex problems strong duality holds most of the times.
- **Constraint qualifications** are conditions under which strong duality holds.

Slaters' constraint qualification

Theorem

Assume a convex primal problem, then if there exists a primal feasible $\mathbf{x}$ *such that*

$$g_i(\mathbf{x}) < 0, \quad i = 1, \ldots, r,$$

(i.e., all the inequality constraints are strictly satisfied) then Slater's condition holds, which implies that strong duality holds, i.e., $p^* = d^*$*.*

Note: Strict inequality is not necessary if $g_i(\mathbf{x})$ is an affine constraint.
Observation: There exist many other types of constraint qualifications.

Optimality conditions

Theorem: Given that:

- f, g_i and h_j differentiable, and
- strong duality holds.

Then, the necessary conditions for primal and dual optimal points $\mathbf{x}^*$ and $(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$ are the **Karush-Kuhn-Tucker(KKT) conditions:**

1. primal constraints:
   $g_i(\mathbf{x}^*) \leq 0,\ i = 1, \dots, r, \quad h_j(\mathbf{x}^*) = 0,\ j = 1, \dots, s,$
2. dual constraints: $\lambda_i^* \geq 0,\ i = 1, \dots, r$
3. complementary slackness: $\lambda_i^* g_i(\mathbf{x}^*) = 0,\ i = 1, \dots, r$
4. gradient of Lagrangian with respect to $\mathbf{x}^*$ vanishes:
   $\nabla f(\mathbf{x}^*) + \sum_{i=1}^{r} \lambda_i^* \nabla g_i(\mathbf{x}^*) + \sum_{j=1}^{s} \nu_j^* \nabla h_j(\mathbf{x}^*) = 0$

Observation: If the primal problem is convex, then the KKT conditions are **necessary and sufficient** for primal and dual optimality with zero duality gap (i.e., strong duality holds).

KKT conditions for Convex Problems

- If $\mathbf{x}', \boldsymbol{\lambda}', \boldsymbol{\nu}'$ satisfy KKT for a convex problem, then they are optimal:
  - from complementary slackness: $f(\mathbf{x}') = L(\mathbf{x}', \boldsymbol{\lambda}', \boldsymbol{\nu}')$
  - from 4th condition and convexity: $g(\boldsymbol{\lambda}', \boldsymbol{\nu}') = L(\mathbf{x}', \boldsymbol{\lambda}', \boldsymbol{\nu}')$

  Hence, $f(\mathbf{x}') = g(\boldsymbol{\lambda}', \boldsymbol{\nu}')$ (i.e., zero duality gap).
- If Slater's condition is satisfied: $\mathbf{x}'$ is optimal if and only if there exist $\boldsymbol{\lambda}', \boldsymbol{\nu}'$ that satisfy the KKT conditions. Remarks:
  - Recall that Slater's qualification implies strong duality and dual optimum is attained.
  - Thus, it generalizes optimality condition

Outline

Convex optimization problem

Question: *what is the minimum distance of a hyperplane* $\{\mathbf{x}|\,\langle\mathbf{w},\mathbf{x}\rangle + b = 0\}$ *to the origin?*

Primal problem: $\min_{\mathbf{x}\in R^d} \|\mathbf{x}\|^2$ s.t. $\langle\mathbf{w},\mathbf{x}\rangle + b = 0$

Primal Lagrangian formulation: $L(\mathbf{x},\lambda) = \frac{1}{2}\|\mathbf{x}\|^2 - \lambda(\langle\mathbf{w},\mathbf{x}\rangle + b)$

Dual Lagrangian fomulation: Using that $\nabla_{\mathbf{x}} L(\mathbf{x},\lambda) = \mathbf{x} - \lambda\mathbf{w} \overset{!}{=} 0$, we can write:

$$q(\lambda) = \inf_{\mathbf{x}\in R^d} L(\mathbf{x},\lambda) = -\frac{\lambda^2}{2}\|\mathbf{w}\|^2 - \lambda b,$$

which is convex w.r.t λ. Thus, we can find a stationary point of the dual problem by $\frac{dq(\lambda)}{d\lambda} = -\lambda\|\mathbf{w}\|^2 - b \overset{!}{=} 0$, to obtain

$$\lambda^* = \arg_\lambda \max q(\lambda) = -\frac{b}{\|\mathbf{w}\|^2} \text{and, thus } \mathbf{x}^* = \lambda^*\mathbf{w} = -\frac{b\mathbf{w}}{\|\mathbf{w}\|^2}$$

.

Then, the solution to the original question is $\|\mathbf{x}^*\|^2 = \frac{b}{\|\mathbf{w}\|}$.

Outline

Why is the dual problem useful?

- Every dual feasible point $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ satisfies that $p^* \geq q(\boldsymbol{\lambda}, \boldsymbol{\nu})$,
- Every feasible point $\mathbf{x}$ satisfies that $d^* \leq f(\mathbf{x})$,
- Any primal/dual feasible pair $\mathbf{x}$ and $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ provides an upper bound on the duality gap: $f(\mathbf{x}) - q(\boldsymbol{\lambda}, \boldsymbol{\nu})$, or

$$p^* \in [q(\boldsymbol{\lambda}, \boldsymbol{\nu}), f(\mathbf{x})], \qquad d^* \in [q(\boldsymbol{\lambda}, \boldsymbol{\nu}), f(\mathbf{x})].$$

- Duality gap is zero $\Longrightarrow$ $\mathbf{x}$ and $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ is primal/dual optimal.

Practical considerations

- Equivalent formulations of a problem can lead to very different duals
- Reformulating the primal problem can be useful when the dual is difficult to derive, or uninteresting
- Common reformulations include:
  - introduce new variables and equality constraints
  - make explicit constraints implicit or vice-versa
  - transform objective or constraint functions, e.g., replace $f(x)$ by $\phi(f(x))$ with ϕ convex, increasing
- Refer to supplementary material for more examples.