Convex Duality

1 The Lagrange Dual Problem

Recall the standard form of a convex program:

$$\begin{aligned} \min \;\; & f_0(\boldsymbol{x}) \\ \text{s.t.} \;\; & f_i(\boldsymbol{x}) \leq 0 \\ & h_j(\boldsymbol{x}) = 0 \end{aligned}$$

Where the f_i are convex and the h_j are affine. Let $\mathcal{D}$ be the domain of all the functions.

The basic idea of Lagrangian duality is to relax the constraints by replacing each constraint with a linear penalty term in the objective. So, we obtain a new optimization problem with no constraints and the **Lagrange function** as the objective:

$$L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = f_0(\boldsymbol{x}) + \boldsymbol{\lambda}^\intercal \boldsymbol{f}(\boldsymbol{x}) + \boldsymbol{\nu}^\intercal \boldsymbol{h}(\boldsymbol{x})$$

We restrict $\boldsymbol{\lambda} \succeq \mathbf{0}$. We call each $\boldsymbol{\lambda}_i$ the Lagrange multiplier for the i^{th} inequality constraint, and $\boldsymbol{\nu}_j$ the **Lagrange multiplier** for the j^{th} equality constraint. Now, the **Lagrange dual function** gives the optimal value for each parameter subject to the soft constraints we imposed:

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

Note that g is a concave function of the Lagrange multipliers. It is common for the dual function to be unbounded for some choice of multipliers; so, we can restrict the domain to g to be $\{(\boldsymbol{\lambda}, \boldsymbol{\nu}) : g(\boldsymbol{\lambda}, \boldsymbol{\nu}) > -\infty\}$. Then we have the **Lagrange dual problem**:

$$\begin{aligned} \max \;\; & g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \\ \text{s.t.} \;\; & \boldsymbol{\lambda} \succeq \mathbf{0} \end{aligned}$$

1.1 Convex Duality

Since the Lagrange dual problem solves a relaxed version of the primal, we know that $OPT(dual) \leq OPT(primal)$ under the primal constraints – this is exactly weak duality. Strong duality is more subtle, and requires **Slater's condition** (i.e., strict feasiblity). If this condition holds, along with convexity of the problem, then $\exists \boldsymbol{\lambda}, \boldsymbol{\nu}$ such that $OPT(dual) = OPT(primal)$.

Because of strong duality, dual solutions can serve as a certificate of optimality for the primal solution. In particular, if $f_0(\boldsymbol{x}) = g(\boldsymbol{\lambda}, \boldsymbol{\nu})$ and both are feasible, then both are optimal. Furthermore, primal-dual algorithms use dual certificates to bound the suboptimality of a solution. If $f_0(\boldsymbol{x}) - g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \leq \epsilon$, then both solutions are within ϵ of optimality (i.e., $OPT(primal)$ and $OPT(dual)$ lie in $[g(\boldsymbol{\lambda}, \boldsymbol{\nu}), f_0(\boldsymbol{x})]$).

If strong duality holds and $\boldsymbol{x}^*$ and $(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$ are optimal, then:

- $\boldsymbol{x}^*$ minimizes $L(\boldsymbol{x}, \boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$ over all $\boldsymbol{x}$.
- $\boldsymbol{\lambda}_i^* \boldsymbol{f}_i(\boldsymbol{x}^*) = 0 \; \forall i$ (complementary slackness). Specifically, $\boldsymbol{\lambda}_i^* = 0$ for loose constraints and $\boldsymbol{f}_i(\boldsymbol{x}^*) = 0$ for tight constraints.

1.2 KKT Optimality Conditions

Suppose the primal is convex and defined on an open domain, the objective and constraints are differentiable everywhere, and strong duality holds. Then $\boldsymbol{x}^*$ and $(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$ are optimal if:

- $\boldsymbol{x}^*$ and $(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$ are feasible.
- $\boldsymbol{\lambda}_i^* \boldsymbol{f}_i(\boldsymbol{x}^*) = 0 \ \forall i$.
- $\nabla_{\boldsymbol{x}} L(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\nu}^*) = 0$.

These are called the KKT conditions, and are often useful for reframing a problem to make it easier to solve. For example, given an equality-constrainted quadratic program of the form:

$$\begin{aligned} \min \ & \frac{1}{2}\boldsymbol{x}^\intercal \mathbf{P}\boldsymbol{x} + \boldsymbol{q}^\intercal \boldsymbol{x} + r \\ \text{s.t.} \ & \mathbf{A}\boldsymbol{x} = \boldsymbol{b} \end{aligned}$$

Then the KKT conditions state that $\mathbf{A}\boldsymbol{x}^* = \boldsymbol{b}$ and $\mathbf{P}\boldsymbol{x}^* + \boldsymbol{q} + \mathbf{A}^\intercal \boldsymbol{\nu}^*$, and we can solve this system of equations to find the optimal solution. This is a much easier problem because it is simply a linear system with $m+n$ constraints and $m+n$ variables.