EECS 127/227AT Discussion 6 Slides

Druv Pai

October 7, 2020

Lagrangians

- Motivation: want to solve constrained optimization problem
$$\min_{x \in \mathcal{F}} f(x),$$
especially when $\mathcal{F}$, f are convex.
- We have tools to solve (convex) unconstrained optimization problems, so we want to convert constrained optimization problems into unconstrained problems.
- First idea: indicator functions, if K is a set then
$$I_K(x) = \begin{cases} \infty, & x \notin K \\ 0, & x \in K \end{cases}.$$
Then
$$\min_{x \in \mathcal{F}} f(x) = \min_{x} (f(x) + I_{\mathcal{F}}(x)).$$
- Theoretically – great! Solutions to constrained problem are exactly solutions to unconstrained problem.

- Practically – terrible! How do we optimize over indicator functions? First derivatives not obvious, iterative optimization is a huge failure. 2/10 would not recommend.
- What we want: smooth version of $I_{\mathcal{F}}(x)$ which fulfills this property: optimal solutions are the same for constrained/unconstrained. Lets us optimize using regular methods.

Notation: constrained **primal** problem $\mathcal{P}_c$ is

$$\begin{aligned} p^* = \min_x \quad & f(x) \\ \text{s.t.} \quad & g(x) \leq 0 \\ & h(x) = 0 \end{aligned}$$

Recall: f scalar valued, g, h vector valued.

Definition (Lagrangian)

Lagrangian $\mathcal{L}(x, \lambda, \nu)$ defined as

$$\mathcal{L}(x, \lambda, \nu) = f(x) + \lambda^\mathsf{T} g(x) + \nu^\mathsf{T} h(x).$$

Primal problem:

$$p^* = \min_x \max_{\substack{\lambda \geq 0 \\ \nu}} \mathcal{L}(x, \lambda, \nu).$$

Dual problem:

$$d^* = \max_{\substack{\lambda \geq 0 \\ \nu}} \min_x \mathcal{L}(x, \lambda, \nu).$$

Switching min / max converts from primal to dual problem.

Fact: $d^* \leq p^*$, consequence of Minimax Theorem. (Simple algebra, but not proved here.)

Claim

Our p^* equations are both valid, i.e.

$$\min_{\substack{g(x)\leq 0 \\ h(x)=0}} f(x) = \min_x \max_{\lambda\geq 0,\nu} \mathcal{L}(x,\lambda,\nu).$$

Proof.

Write out

$$\mathcal{L}(x,\lambda,v) = f(x) + \lambda^\mathsf{T} g(x) + \nu^\mathsf{T} h(x).$$

If even one $g(x)_i > 0$, then $\lambda_i \to \infty$ implies $\mathcal{L}(x,\lambda,\nu) \to \infty$, so we pick x for which $g(x) \leq 0$ (note because $\lambda_i \geq 0$ we can't pick $\lambda_i \to -\infty$ for $g(x)_i < 0$). If even one $h(x)_i \neq 0$, $\nu_i \to \operatorname{sign}(h(x)_i) \cdot \infty$ implies $\mathcal{L}(x,\lambda,\nu) \to \infty$, so we only pick x for which $h(x) = 0$. $\square$

- Our Lagrangian is exactly what we want for converting constrained problems to unconstrained problems!
- Single most important innovation in convex optimization.
- Sadly, not always easy to find optimal solution to Lagrangian.

Dual Problem

Dual problem is sometimes easier to solve than primal problem.

$$d^* = \max_{\substack{\lambda \geq 0 \\ \nu}} \min_x \mathcal{L}(x, \lambda, nu).$$

- Weak duality: $d^* \leq p^*$ – intuition: any solution to the dual problem is a lower bound on the primal solution, and if we find one dual solution and primal solution with the same value then they're both optimal.
- Strong duality: $d^* = p^*$. If we can solve the dual problem then we're done.

Theorem (Slater's Condition)

If $g(x)_i$ is convex and feasible region contains an open set (there is an x for which $g(x) < 0$) then strong duality holds.

KKT

How to actually solve primal/dual problems?

- In some cases we can just take the derivative of $\mathcal{L}$ and set to 0 (if $\mathcal{L}$ is convex in x, since $\mathcal{L}$ is linear in λ, ν).
- In other cases, we have to use KKT conditions to find necessary/sufficient conditions for solution.

Theorem (KKT Necessary Conditions)

Any feasible solution (x^*, λ^*, ν^*) *must obey*

- *Primal feasibility:* $g(x^*) \leq 0$, $h(x^*) = 0$
- *Dual feasibility:* $\lambda^* \geq 0$
- *Stationarity:*

$$\nabla_x \mathcal{L}(x^*, \lambda^*, \nu^*) = \nabla_x f(x^*) + \underbrace{(\nabla_x g(x^*))}_{matrix} \lambda^* + \underbrace{(\nabla_x h(x^*))}_{matrix} \nu^* = 0$$

- *Complementary slackness:*

$$\lambda^{*\mathsf{T}} g(x^*) = 0 \textit{ or equivalently } \lambda_i^* g(x^*)_i = 0.$$

Another (more useful) way to write stationarity is

$$\nabla_x f(x^*) + \sum_i \lambda_i^* \left(\nabla_x g_i(x^*)\right) + \sum_i \nu_i^* \left(\nabla_x h_i(x^*)\right) = 0$$

Obviously first three conditions are necessary. What about complementary slackness? Solution is by exchange argument; if $g(x^*)_i \neq 0$ then $\lambda_i^* = 0$ improves objective value.