B9120-001 Dynamic Programming Winter 2020

Prof. Daniel Russo

Homework Assignment 9: Due Monday April 20 at 2:30PM

This problem explores a precise connection between policy iteration and the conditional gradient algorithm (a.k.a Frank Wolfe) applied to the policy gradient loss.

<u>*MDP setup:*</u> Let $\mathcal{S} = \{1, \cdots, n\}$ and $U = \{u \in \mathbb{R}^k : \mathbf{1}^\top u = 1\,,\, u \succeq 0\}$ be the set of probability distributions over k base actions. Due to the linearity of expectations, expected costs and transition probabilities are linear in the stochastic action vector, with

$$g(s,u) = \sum_{i=1}^{k} g(s, e_i) u_i \qquad P_{ss'}(u) = \sum_{i=1}^{k} P_{ss'}(u_i)$$

where e_i is the i-th standard basis vector. The set of stochastic stationary policies $M = \{\mu \in \mathbb{R}^{n\times k} : \mu_s \in U \ \forall s\}$ is the set of matrices whose rows are probability distributions. Define

$$\ell(\mu) = \nu J_\mu = \sum_{s=1}^{n} \nu(s) J_\mu(s) \qquad \mu \in M$$

for given state-relevance weights ν where $\nu(s) > 0$ and $\sum_{s=1}^{n} \nu(s) = 1$.

<u>*Conditional gradient algorithm:*</u> Consider the conditional gradient (CG) algorithm applied to $\min_{\mu\in M} \ell(\mu)$. Beginning with some initial iterate μ_0, CG produces a sequence of iterates $\{\mu_{k+1}\}_{k\in\mathbb{N}}$ where

$$\mu_{k+1} = (1-\gamma_k)\mu_k + \gamma_k y_k \tag{1}$$

$$y_k \in \operatorname*{arg\,min}_{\mu\in M} \langle \nabla \ell(\mu_k)\,,\, \mu - \mu_k \rangle \tag{2}$$

We approximate ℓ by linearizion around μ_k, minimize that globally (here solving an LP), and then take a small step in that direction (reflecting that the linearization is not globally accurate).

You may assume a formula for the gradients of ℓ given in class,

$$\nabla \ell(\mu) = \sum_{s=1}^{n} \nu_\mu(s) \nabla_{\bar{\mu}}\, Q_\mu(s, \bar{\mu}_s)\Big|_{\bar{\mu}=\mu} \in \mathbb{R}^{n\times k}$$

where $\nu_\mu = \sum_{t=0}^{\infty} \alpha^t \nu P_\mu^t$. (The notation on the right side indicates we evaluate the partial derivative with respect to $\bar{\mu}$ around the point where $\bar{\mu} = \mu$)

If you can stuck on a subproblem, you may solve the remaining subproblems assuming its claim.

Problem (a) Recognize that $Q_\mu(s,u)$ is linear in u.

Problem (b) Calculate $\frac{\partial}{\partial \mu_{si}} \ell(\mu)$.

Problem (c) Show that y_k in (2) is a policy iteration update to μ_k.

Problem (d) Consider a fixed stepsize $\gamma_k = \gamma \in (0, 1)$. Show that $\|J_{\mu_k} - J^*\|_\infty \leq (1 - \gamma(1 - \alpha))^k \|J_{\mu_0} - J^*\|_\infty$. Provide the same convergence result for $\ell(\cdot)$. (Hint, if you're stuck, re-read Class 9, slide 7 and class 10, slide 21)

Problem (e) What stepsize choice does your analysis suggest?

You don't need to write anything, but think about how to reconcile (e) with the usual stepsizes in smooth optimization.