Mehryar Mohri
Advanced Machine Learning 2015
Courant Institute of Mathematical Sciences
Homework assignment 2
April 13, 2015
Due: April 27, 2015

A. RWM and FPL

Let RWM(β) denote the RWM algorithm described in class run with parameter $\beta > 0$. Consider the version of the FPL algorithm FPL(β) defined using the perturbation:

$$\mathsf{p}_1 = \left[\frac{\log(-\log(u_1))}{\beta}, \ldots, \frac{\log(-\log(u_N))}{\beta}\right]^\top.$$

where, for $j \in [1, N]$, u_j is drawn from the uniform distribution over $[0, 1]$. At round $t \in [1, T]$, $\mathbf{w}_t$ is found via $\mathbf{w}_t = M(\mathbf{x}_{1:t-1} + \mathsf{p}_1) = \operatorname{argmin}_{\mathbf{w} \in \mathcal{W}} \mathbf{w} \cdot \mathbf{x}_{1:t-1} + \mathsf{p}_1$ using the notation adopted in the class lecture for FPL, with $\mathcal{W}$ the set of coordinate unit vectors. Show that FPL(β) coincides with RWM(β).

B. Zero-sum games

For all the questions that follow, we consider a zero-sum game with payoffs in $[0, 1]$.

1. Show that the time complexity of the RWM algorithm to determine an ϵ-approximation of the value of the game is in $O(\log N/\epsilon^2)$.

2. Use the proof given in class for von Neumann's theorem to show that both players can come up with a strategy achieving and ϵ-approximation of the value of the game (or Nash equilibrium) that are sparse: the support of each mixed strategy is in $O(\log N/\epsilon^2)$. What fraction of the payoff matrix does it suffice to consider to compute these strategies?

C. Bregman divergence

1. Given an open convex set C, provide necessary and sufficient conditions for a differentiable function $G\colon C \to \mathbb{R}$ to be a Bregman divergence. That is, give conditions for the existence of a convex function

$F\colon C \to \mathbb{R}$ such that $G(x,y) = F(x) - F(y) - \nabla F(y)(x-y)$.
Hint: Show that a Bregman divergence satisfies the following identity

$$B_F(y||x) + B_F(x||z) = B_F(y||z) + (y-x)(\nabla F(z) - \nabla F(x)).$$

2. Using the results of the previous exercise, decide whether or not the following functions are a Bregman divergence.

   - The KL-divergence: the function $G\colon \mathbb{R}^n_+ \to \mathbb{R}$ defined for $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ by $G(x,y) = \sum_{i=1}^n x_i \log\left(\frac{x_i}{y_i}\right)$.
   - The function $G\colon \mathbb{R}_+ \to \mathbb{R}_+$ given by $G(x,y) = x(e^x - e^y) - ye^y(x - y)$.