CSE599s Spring 2012 - Online Learning
Homework Exercise 3 - due 5/22/12

1. **Strong convexity** Strongly convex functions have a number of important properties that make them particularly nice to use as regularizers. We investigate some of them here. For all of these, assume the functions are from $\mathbb{R}^n \to \mathbb{R}$, and strong convexity is with respect to an arbitrary norm $\|\cdot\|$. Unless otherwise specified, assume the strong convexity holds on some convex set $\mathcal{W}$.

   (a) Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is σ-strongly convex on a convex set $\mathcal{W}$ (possibly $\mathbb{R}^n$). Show that f is strongly convex on any convex $\mathcal{W}' \subseteq \mathcal{W}$.

   (b) Let f be σ-strongly convex, and let h be α-strongly-convex. Show that $c(x) = f(x) + h(x)$ is $(\sigma + \alpha)$-strongly-convex. An important corollary is that if f is σ-strong-convex and h is an arbitrary convex function, then their sum is also σ-strongly-convex.

   (c) Suppose f is 1-strongly-convex. Show that $h(x) = \alpha f(x)$ is α-strongly-convex for $\alpha \in [0, \infty)$.

   (d) Let f be σ-strongly-convex on a convex set $\mathcal{W}$. Show that f has a unique minimizer $w^* \in \mathcal{W}$.

2. **Online Gradient Descent with Strongly Convex Loss Functions** Recall the analysis of the Online Gradient Descent algorithm (see notes for lecture 5)

   (a) Prove that if the loss functions f_t are all σ-strongly convex then regret is upper bounded by

   $$\frac{1}{2}\|w^\star\|^2 \left(\frac{1}{\eta_t} - \sigma\right) \;+\; \frac{1}{2}\sum_{t=2}^{T} \|w_t - w^\star\|^2 \left(\frac{1}{\eta_t} - \frac{1}{\eta_{t-1}} - \sigma\right) \;+\; \frac{G^2}{2}\sum_{t=1}^{T} \eta_{t+1} \quad .$$

   (b) Set $\eta_t = \frac{1}{\sigma t}$ and conclude that

   $$\sum_{t=1}^{T} f_t(w_t) - \sum_{t=1}^{T} f_t(w^\star) \;\leq\; \frac{G^2}{H}(1 + \log T) \quad .$$

   Note that we obtain a *logarithmic* bound on regret, which is much smaller than a *square-root* bound.

3. **Implementing FTRL with Proximal and L1 Regularization** We consider the FTRL algorithm with adaptive proximal regularization and an L_1 penalty to indroduce sparsity. We consider the unconstrained problem, so the update is

$$w_{t+1} = \arg\min_{w \in \mathbb{R}^n} g_{1:t} \cdot w + t\lambda \|w\|_1 + \sum_{s=1}^{t} \frac{\sigma_s}{2} \|w - w_s\|_2^2$$

Here, the σ_s in $\mathbb{R}^+$ give the strengh of each incremental regularization function, and $\lambda \geq 0$ gives the strength of the per-round L_1 penalty.

(a) Consider the $1D$ optimization problem

$$w^* = \arg\min_{w \in \mathbb{R}} \frac{a}{2} w^2 + bw + c\|w\|_1,$$

where $a, b, c \in \mathbb{R}$ are constants and $a, c \geq 0$. Derive a closed-form solution for w^*. Hint: Consider the subdifferential of this objective, and recall that if $0 \in \partial f(w^*)$ then w^* is a minimizer. Your closed-form solution may still contain cases.

(b) Suppose that the σ_t are chosen only as a function of t, for example so $\sigma_{1:t} = \sqrt{t}$, corresponding to a learning rate of $\frac{1}{\sqrt{t}}$. Write pseudocode for the algorithm, using an implementation that only requires storing a single vector in $\mathbb{R}^n$. For simplicity, structure your code like this:

```
/*TODO: Define variables for the state of the algorithm*/
for round t = 1, 2, ... do
    Observe gradient g_t
    /*TODO: Implement the update.*/
    /*TODO: Compute and output w_{t+1} */
end for
```