CS591 E2: Optimization Methods and their Applications – Fall 2016

Homework 2

Due date: November 1, 2016

Homework Policy and Guidelines

You are encouraged to collaborate on the solution of the homeworks and to consult any materials, but you must write up your own answers and you must acknowledge all of your collaborators and sources.

The problems marked with a (∗) are more challenging. You may not be able to completely solve some of the more challenging problems, that is completely normal!

Some of the problems ask you to fill in a proof that we did not cover in class; the readings will often have these proofs, you are free to consult them but you must write up a complete proof in your own words. In general it may be good to keep in mind that some of the proofs in the textbooks may leave some steps to the reader, and it is very important to make sure that you know how to fill in those missing steps. Also, thinking about those proofs on your own will help you understand the material better.

Problem 1 *Describe and analyze a projected gradient descent algorithm for the constrained minimization* $\min_{x \in \mathcal{X}} f(x)$*, where* f *is a* L*-smooth convex function and* $\mathcal{X}$ *is a convex set.*

Problem 2 *In this problem, we extend the Frank-Wolfe algorithm that we saw in class to take advantage of smoothness in any norm. Consider some norm* $\|\cdot\|$ *over* $\mathbb{R}^n$*, and let* $\|\cdot\|_*$ *denote its dual norm:* $\|x\|_* = \sup_{y \in \mathbb{R}^n : \|y\| \leq 1} \langle x, y \rangle$*. A differentiable function* f *is* L*-smooth in the norm* $\|\cdot\|$ *if, for all* $x, y \in \mathrm{R}^n$*,*

$$\|\nabla f(x) - \nabla f(y)\|_* \leq L\|x - y\|.$$

- *Show that the dual norm of* ℓ_2 *is* ℓ_2 *(and thus smoothness in the* ℓ_2 *norm is precisely the definition of smoothness we saw in class).*
- *Show that, for any* p*, the dual norm of* ℓ_p *is* ℓ_q*, where* $\frac{1}{p} + \frac{1}{q} = 1$*.*
- *Generalize the Frank-Wolfe algorithm and its analysis for minimizing a function* f *that is* L*-smooth in an arbitrary norm* $\|\cdot\|$*.*

Problem 3 *Consider the analysis of the Frank-Wolfe algorithm that we saw in class. Based on that analysis, what is the optimal setting for the step sizes? How does this setting compare to the simple setting that we saw in class, namely setting* $\gamma_t = \frac{2}{t+1}$*? Does the optimal setting improve the convergence guarantee?*

Problem 4 *Let* f *be a twice-differentiable function. Show that, if* f *is* L*-smooth then* $\nabla^2 f \leq LI_n$ *and, if* f *is* ℓ*-strongly convex then* $\nabla^2 f \geq \ell I_n$*.*

Problem 5 *In this problem, we explore the convergence of the gradient descent and accelerated gradient descent methods we saw in class on some very simple examples, namely unconstrained quadratic minimization problems in two dimensions* $\min_{x \in \mathbb{R}^2} \frac{1}{2}x^T Ax - bx$*, where* $A \in \mathbb{R}^{2\times 2}$ *is a matrix and* $b \in \mathbb{R}^2$ *is a vector. Using Python, Matlab, Mathematica, or any other programming language or tool, we can visualize the contour plot of the function* $f(x) = \frac{1}{2}x^T Ax - bx$ *and the iterates of the gradient descent and accelerated gradient descent schemes. Create these plots for several examples of quadratic functions with varying condition numbers to contrast the behavior of the algorithms for instances with small and large condition number (for example, condition number 2 versus 100). In class, we saw similar plots for the example* $A = \begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix}$ *and* $b = \begin{bmatrix} 2 \\ -2 \end{bmatrix}$.
