Homework 1 (a and b)
Convex Sets and Convex Functions

CMU 10-725/36-725: Convex Optimization (Fall 2017)

OUT: Sep 1

DUE: Prob 1-3 **Sep 11, 5:00 PM**; Prob 4 **Sep 15, 5:00 PM**

START HERE: Instructions

- **Collaboration policy:** Collaboration on solving the homework is allowed, after you have thought about the problems on your own. It is also OK to get clarification (but not solutions) from books or online resources, again after you have thought about the problems on your own. There are two requirements: first, cite your collaborators fully and completely (e.g., "Jane explained to me what is asked in Question 3.4"). Second, write your solution independently: close the book and all of your notes, and send collaborators out of the room, so that the solution comes from you only.

- **Submitting your work:** Assignments should be submitted as PDFs using Gradescope unless explicitly stated otherwise. Each derivation/proof should be completed on a separate page. Submissions can be handwritten, but should be labeled and clearly legible. Else, submissions can be written in LaTeX. Upon submission, label each question using the template provided by Gradescope. Please refer to Piazza for detailed instruction for joining Gradescope and submitting your homework.

- **Programming**: All programming portions of the assignments should be submitted to Gradescope as well. We will not be using this for autograding, meaning you may use any language which you like to submit.

Problem 1: Convex Set (Hongyang - 20 pts)

1. **[5pts]** (Definition of convexity) Denote by $C \subseteq \mathbb{R}^n$ a convex set. Let $x_1, ..., x_k \in C$, and $\theta_1, ..., \theta_k \in \mathbb{R}$ satisfy $\theta_i \geq 0$ and $\sum_{i=1}^k \theta_i = 1$. Show that $\theta_1 x_1 + ... + \theta_k x_k \in C$ for arbitrary k (Recall that in the class, we define the convexity by the case of $k = 2$.)

2. (Example of convex set)
(a) **[5pts]** Show that if $a, b \geq 0$ and $0 \leq \theta \leq 1$, then $a^\theta b^{1-\theta} \leq \theta a + (1-\theta) b$. Hint: Use concavity of log functions.
(b) **[5pts]** Show that $S_n = \{x \in \mathbb{R}_+^n \mid \prod_{i=1}^n x_i \geq 1\}$ is convex.

3. **[5pts]** (Operations that preserve convexity) Suppose that S_1 and S_2 are convex sets in $\mathbb{R}^{m+n}$. Show that their partial sum

$$S = \{(x, y_1 + y_2) \mid x \in \mathbb{R}^m,\ y_1, y_2 \in \mathbb{R}^n,\ (x, y_1) \in S_1,\ (x, y_2) \in S_2\}$$

is a convex set.

Problem 2: Convex Functions (Devendra - 20 pts)

1. **[5pts]** Suppose that a non-convex function $f(x)$ is given as a sum of terms of the form $g : R^n_{++} \rightarrow R, g(x) = \beta x_1^{\alpha_1}..x_n^{\alpha_n}, \beta > 0, \alpha_i \in R$. Show that the substitution $y_i = \log x_i$ can transform $f(x)$, into a into a convex function in y.
   (Hint: Use a simple example such as $f(x_1, x_2) = x_1^{-2} + (x_1 x_2)^{\frac{1}{3}} + 2x_2^{-4}, x_1 > 0, x_2 > 0$ to prove the claim and then generalize it)

2. Show that the following functions are convex
   - **[5pts]** $f : R^n_{++} \rightarrow R$, $f(x) = \sum_{i=1}^n x_i \log x_i$, such that $\sum_{i=1}^n x_i = 1$. This is also called negative entropy function.
   - **[5pts]** $f : R^n_{++} \rightarrow R_-$, $f(x) = -(\sum_{i=1}^n x_i^a)^{\frac{1}{a}}$, given that $a < 1, a \neq 0$.
   - **[5pts]** $f : R_+ \rightarrow R$
   $$f(x) = \begin{cases} \frac{1}{x}\int_0^x g(y)dy & \text{if } x > 0 \\ 0 & \text{if } x = 0 \end{cases}$$
   given that $g : R_+ \rightarrow R$ is a convex, non-negative function and $g(0) = 0$.

Problem 3: Lipschitz gradients and strong convexity (Yifeng - 30 pts)

Let f be convex and twice differentiable.

1. **[6pts]** Prove the *monotonicity of gradient* ∇f, i.e.,

   f is convex if and only if $(\nabla f(x) - \nabla f(y))^T(x - y) \geq 0$, for all x, y.

   (Note: Feel free to use this property in the proof of problem 2 or 3 if necessary.)

2. **[12pts]** (Smoothness of f) Show that the following statements are equivalent.

   i. ∇f is Lipschitz with constant L;

   ii. $(\nabla f(x) - \nabla f(y))^T(x - y) \leq L\|x - y\|_2^2$ for all x, y;

   iii. $\nabla^2 f(x) \preceq LI$ for all x;

   iv. $f(y) \leq f(x) + \nabla f(x)^T(y - x) + \frac{L}{2}\|y - x\|_2^2$ for all x, y.

3. **[12pts]**(Curvature of f) Show that the following statements are equivalent.

   i. f is strongly convex with constant m;

   ii. $(\nabla f(x) - \nabla f(y))^T(x - y) \geq m\|x - y\|_2^2$ for all x, y;

   iii. $\nabla^2 f(x) \succeq mI$ for all x;

   iv. $f(y) \geq f(x) + \nabla f(x)^T(y - x) + \frac{m}{2}\|y - x\|_2^2$ for all x, y.

Programming Problem: Solving optimization problems with CVX (Hao - 30 pts)

CVX is a fantastic framework for disciplined convex programming - its rarely the fastest tool for the job, but its widely applicable, and so its a great tool to be comfortable with.

In this exercise, we will set up the CVX environment and solve a convex optimization problem. In this class, your solution to coding problems should include plots and whatever explanation necessary to answer the questions asked. In addition, full code should be submitted as an appendix to the homework document.

CVX variants are available for each of the major numerical programming languages. There are some minor syntactic and functional differences between the variants but all provide essentially the same functionality. The Matlab version (and by extension, the R version which calls Matlab under the covers) is the most mature but all should be sufficient for the purposes of this class.

Download the CVX variant of your choosing:

- Matlab - http://cvxr.com/cvx/
- Python - http://www.cvxpy.org/en/latest/
- Julia - https://github.com/JuliaOpt/Convex.jl
- R - http://faculty.bscb.cornell.edu/ bien/cvxfromr.html

and consult the documentation to understand the basic functionality. To feel comfortable with CVX, make sure that you can solve the least squares problem for a vector y and matrix X. Check your answer by comparing with the analytic least squares solution. You do not need to submit answers for the least squares part. The following will be graded.

1. Using CVX, we can solve the SVM model discussed in Lecture 1:

$$\min_{w,b} \|w\|^2 + C \sum_{(x_i,y_i)\in E} \left(1 - (w^T x_i + b) y_i\right)_+ .$$

   The set E is the set of data points in the training set.

   a) **[9pts]** Load the training set from train.csv and solve the SVM model with $C = 1$. Report the objective value obtained at the solution. Load the test set from test.csv and use the trained SVM model to predict the class of each data point. Report the error rate of SVM on the train and test set.

   b) **[6pts]** Next, we consider how the solution changes as we vary C. Use the same data set as (1) and solve the SVM model for this training set for $C \in \{10^{-k/2} : k = -8, ..., -3, -2, -1, 0, 1, 2, 3, ..., 8\}$. Plot the curve of train and test error rate with the varying C. What changes in test error you can observe with varying C?

2. **[15pts]** Disciplined convex programming or DCP is a system for composing functions while ensuring their convexity. It is the language that underlies CVX. Essentially, each node in the parse tree for a convex expression is tagged with attributes for curvature (convex, concave, affine, constant) and sign (positive, negative) allowing for reasoning about the convexity of entire expressions. The website `http://dcp.stanford.edu/` provides visualization and analysis of simple expressions.

   Typically, writing problems in the DCP form is natural, but in some cases manipulation is required to construct expressions that satisfy the rules. For each set of mathematical expressions below, **give equivalent DCP expressions and specify whether this equation or inequality defines a convex set**. DCP expressions should be given in a form that passes analysis at `http://dcp.stanford.edu/analyzer`.

   (a) $\|(x, y)\|_1^3 \leq 5x + 7y$

   (b) $\frac{2}{x} + \frac{9}{z-y} \leq 3, x > 0, y < z$

(c) $\max(x, y, z) < \frac{2}{x}, x > 0$

(d) $\log\left(e^{-3\sqrt{x}} + e^{2z}\right) \leq -e^{5y}, x \geq 0$

(e) $\sqrt{\|(2x - 3y, y + x)\|_1} = 0$