Homework 1

Convex Optimization 10-725

Due Friday September 13 at 11:59pm

Submit your work as a single PDF on Gradescope. Make sure to prepare your solution to each problem on a separate page. (Gradescope will ask you select the pages which contain the solution to each problem.)

Total: 66 points (+ 10 bonus points)

1 Convex sets (16 points)

(a, 12 pts) Closed sets and convex sets.

i. Show that a polyhedron $\{x \in \mathbb{R}^n : Ax \leq b\}$, for some $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$, is both convex and closed.

ii. Show that if $S_i \subseteq \mathbb{R}^n$, $i \in I$ is a collection of convex sets, then their intersection $\cap_{i \in I} S_i$ is also convex. Show that the same statement holds if we replace "convex" with "closed".

iii. Given an example of a closed set in $\mathbb{R}^2$ whose convex hull is not closed.

iv. Let $A \in \mathbb{R}^{m \times n}$. Show that if $S \subseteq \mathbb{R}^m$ is convex then so is $A^{-1}(S) = \{x \in \mathbb{R}^n : Ax \in S\}$, which is called the preimage of S under the map $A : \mathbb{R}^n \to \mathbb{R}^m$. Show that the same statement holds if we replace "convex" with "closed".

v. Let $A \in \mathbb{R}^{m \times n}$. Show that if $S \subseteq \mathbb{R}^n$ is convex then so is $A(S) = \{Ax : x \in S\}$, called the image of S under A.

vi. Give an example of a matrix $A \in \mathbb{R}^{m \times n}$ and a set $S \subseteq \mathbb{R}^n$ that is closed and convex but such that $A(S)$ is not closed.

(b, 4 pts) Polyhedra.

i. Show that if $P \subseteq \mathbb{R}^n$ is a polyhedron, and $A \in \mathbb{R}^{m \times n}$, then $A(P)$ is a polyhedron. Hint: you may use the fact that

$$P \subseteq \mathbb{R}^{m+n} \text{ is a polyhedron } \Rightarrow \{x \in \mathbb{R}^n : (x, y) \in P \text{ for some } y \in \mathbb{R}^m\} \text{ is a polyhedron.}$$

ii. Show that if $Q \subseteq \mathbb{R}^m$ is a polyhedron, and $A \in \mathbb{R}^{m \times n}$, then $A^{-1}(Q)$ is a polyhedron.

2   Convex functions (14 points)

(a, 2 pts) Prove that the *entropy function*, defined as

$$f(x) = -\sum_{i=1}^{n} x_i \log(x_i),$$

with $\text{dom}(f) = \{x \in \mathbb{R}_{++}^n : \sum_{i=1}^n x_i = 1\}$, is strictly concave.

(b, 4 pts) Let f be twice differentiable, with $\text{dom}(f)$ convex. Prove that f is convex if and only if

$$(\nabla f(x) - \nabla f(y))^T (x - y) \geq 0,$$

for all x, y. This property is called *monotonicity* of the gradient ∇f.

(c, 2 pts) Give an example of a strictly convex function that does not attain its infimum.

(d, 3 pts) A function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be *coercive* provided that $f(x) \to \infty$ as $\|x\|_2 \to \infty$. A key fact about coercive functions is that they attain their infimums. Prove that a twice differentiable, strongly convex function is coercive and hence attains its infimum. Hint: use Q3 part (b.iv).

(e, 3 pts) Prove that the maximum of a convex function over a bounded polyhedron must occur at one of the vertices. Hint: you may use the fact that a bounded polyhedron can be represented as the convex hull of its vertices.

3   Partial optimization with ℓ_2 penalties (10 bonus points)

Consider the problem

$$\min_{\beta, \, \sigma \geq 0} \ f(\beta) + \frac{\lambda}{2} \sum_{i=1}^{n} g(\beta_i, \sigma_i), \tag{1}$$

for some convex f with domain $\mathbb{R}^n$, $\lambda \geq 0$, and

$$g(x, y) = \begin{cases} x^2/y + y & \text{if } y > 0 \\ 0 & \text{if } x = 0, \, y = 0 \\ \infty & \text{else.} \end{cases}$$

In other words, the problem (1) is just the weighted ℓ_2 penalized problem

$$\min_{\beta, \, \sigma \geq 0} \ f(\beta) + \frac{\lambda}{2} \sum_{i=1}^{n} \left(\frac{\beta_i^2}{\sigma_i} + \sigma_i \right),$$

but being careful to treat the ith term in the sum as zero when $\beta_i = \sigma_i = 0$.

(a, 5 pts) Prove that g is convex. Hence argue that (1) is a convex problem. Note that this means we can perform partial optimization in (1) and expect it to return another convex problem. Hint: use the definition of convexity.

(b, 2 pts) Argue that $\min_{y \geq 0} g(x, y) = 2|x|$.

(c, 3 pts) Argue that minimizing over $\sigma \geq 0$ in (1) gives the ℓ_1 penalized problem

$$\min_{\beta} \ f(\beta) + \lambda \|\beta\|_1.$$

4 Lipschitz gradients and strong convexity (18 points)

Let f be convex and twice continuously differentiable.

(a, 10 pts) 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.

Your solution should have 5 parts, where you prove i $\Rightarrow$ ii, ii $\Rightarrow$ iii, iii $\Rightarrow$ iv, iv $\Rightarrow$ ii, and iii $\Rightarrow$ i.

(b, 8 pts) 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.

Your solution should have 4 parts, where you prove i $\Rightarrow$ ii, ii $\Rightarrow$ iii, iii $\Rightarrow$ iv, and iv $\Rightarrow$ i.

5 Solving optimization problems with CVX (18 points)

CVX is a fantastic framework for disciplined convex programming: it's rarely the fastest tool for the job, but it's widely applicable, and so it's a great tool to be comfortable with. In this exercise we will set up the CVX environment and solve a convex optimization problem.

Generally speaking, for homeworks in this class, your solution to programming-based problems should include plots and whatever explanation necessary to answer the questions asked. In addition, you 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. Download the CVX variant of your choosing:

- Matlab: `http://cvxr.com/cvx/`

- Python: `http://www.cvxpy.org/`

- R: `https://cvxr.rbind.io`

- Julia: `https://github.com/JuliaOpt/Convex.jl`

and consult the documentation to understand the basic functionality. Make sure that you can solve the least squares problem $\min_\beta \|y - X\beta\|_2^2$ for an arbitrary vector y and matrix X. Check your answer by comparing with the closed-form solution $(X^T X)^{-1}X^T y$.

(a, 10 pts) Given labels $y \in \{-1, 1\}^n$, and a feature matrix $X \in \mathbb{R}^{n \times p}$ with rows $x_1, \ldots x_n$, recall the support vector machine (SVM) problem

$$\min_{\beta, \beta_0, \xi} \quad \frac{1}{2}\|\beta\|_2^2 + C\sum_{i=1}^n \xi_i$$
$$\text{subject to} \quad \xi_i \geq 0, \ i = 1, \ldots n$$
$$y_i(x_i^T\beta + \beta_0) \geq 1 - \xi_i, \ i = 1, \ldots n.$$

i. Load the training data in `xy_train.csv`. This is a matrix of $n = 200$ row and 3 columns. The first two columns give the first $p = 2$ features, and the third column gives the labels. Using CVX, solve the SVM problem with $C = 1$. Report the optimal crtierion value, and the optimal coefficients $\beta \in \mathbb{R}^2$ and intercept $\beta_0 \in \mathbb{R}$.

ii. Recall that the SVM solution defines a hyperplane

$$\beta_0 + \beta^T x = 0,$$

which serves as the decision boundary for the SVM classifier. Plot the training data and color the points from the two classes differently. Draw the decision boundary on top.

iii. Now define $\tilde{X} \in \mathbb{R}^{n \times p}$ to have rows $\tilde{x}_i = y_i x_i$, $i = 1, \ldots, n$, and solve using CVX the problem

$$\max_w \quad -\frac{1}{2} w^T \tilde{X} \tilde{X}^T w + 1^T w$$
$$\text{subject to} \quad 0 \le w \le C1, \ w^T y = 0,$$

(Above, we use 1 to denote the vector of all 1s.) Report the optimal criterion value; it should match that from part i. Also report $\tilde{X}^T w$ at the optimal w; this should mach the optimal β from part i. Note: this is not a coincidence, and is an example of *duality*, as we will study in detail later in the course.

iv. Investigate many values of the cost parameter $C = 2^a$, as a varies from -5 to 5. For each one, solve the SVM problem, form the decision boundary, and calculate the misclassification error on the test data in `xy_test.csv`. Make a plot of misclassification error (y-axis) versus C (x-axis, which you will probably want to put a log scale).

(b, 8 pts) Disciplined convex programming (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, first briefly explain why each defines a convex set. Then, give an equivalent DCP expression along with a brief explanation of why the DCP expression is equivalent to the original for each set. DCP expressions should be given in a form that passes analysis (a green tick on the left of the expression box) at `http://dcp.stanford.edu/analyzer`. Note: this question is really about developing a better understanding of the various composition rules for convex functions.

i. $\|(x, y, z)\|_2^2 \le 1$

ii. $\sqrt{x^2 + 1} \le 3x + y$

iii. $1/x + 2/y \le 5, x > 0, y > 0$

iv. $(x + y)^2 / \sqrt{y} \le x - y + 5, y > 0$

v. $(x + z)y \ge 1, x + z \ge 0, y \ge 0$

vi. $\|(x + 2y, x - y)\|_2 = 0$

vii. $x\sqrt{y} \ge 1, \ x \ge 0, y \ge 0$

viii. $\log(e^{y-1} + e^{x/2}) \le -e^x$