Homework 1

Convex Optimization 10-725

Due Friday September 14 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: 75 points

1 Convex sets (18 points)

(a, 8 pts) Closed and convex sets.

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

ii. Show that if $S \subseteq \mathbb{R}^m$ is convex, and $A \in \mathbb{R}^{m\times n}$, then $A^{-1}(S) = \{x : Ax \in S\}$, called the preimage of S under A, is convex.

iii. Show that (ii) also holds if we replace "convex" by "closed".

iv. Show that (i) does not hold if we replace "convex" by "closed", i.e., give a counterexample.

(b, 4 pts) Polyhedra.

i. Show that any polyhedron is both closed and convex.

ii. 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

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

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

(c, 2 pts) Given some integer $k \geq 0$, the *Fantope* of order k is

$$\{Z \in \mathbb{S}^n : 0 \preceq Z \preceq I,\ \mathrm{tr}(Z) = k\}.$$

You can think of this like a "polyhedron in matrix land". Prove that it is convex, in one of two ways: (i) straight from the definition of convexity; or (ii) by recognizing that it is the convex hull of another set.

(d, 4 pts) The following is a "strict" variant of the Separating Hyperplane Theorem: if $C, D \subseteq \mathbb{R}^n$ are disjoint, closed and convex, and (say) D is bounded, then there exists $a \in \mathbb{R}^n, b \in \mathbb{R}$ with $a \neq 0$ such that $a^T x > b$ for all $x \in C$ and $a^T x < b$ for all $x \in D$ (i.e., the hyperplane $\{x \in \mathbb{R}^n : a^T x = b\}$ strictly separates C, D). Use this to prove *Farkas' Lemma*: given $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, exactly one of the following is true:

- there exists $x \in \mathbb{R}^n$ such that $Ax = b$, $x \geq 0$;
- there exists $y \in \mathbb{R}^m$ such that $A^T y \geq 0$, $y^T b < 0$.

Hint: it will help you to use part (b.ii), to deduce that the set $\{Ax : x \geq 0\}$ is a polyhedron, and hence closed and convex by part (b.i).

2 Convex functions (14 points)

(a, 2 pts) Prove that the function $f : \mathbb{R}^n \to \mathbb{R}$, defined by

$$f(x) = \min_{\sigma} \sum_{i=1}^{n-1} |x_{\sigma(i)} - x_{\sigma(i+1)}|,$$

where the minimum above is over all permutations σ of the set $\{1, \ldots, n\}$, is convex. Hint: you can rewrite this function in a much simpler form.

(b, 2 pts) Prove that the *log barrier function* $f : \mathbb{R}^n_{++} \to \mathbb{R}$, defined as

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

is strictly convex.

(c, 4 pts) Let f be twice differentiable, with $\mathrm{dom}(f)$ 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.

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

(e, 2 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).

(f, 2 pts) Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function. Its *perspective transform* $g : \mathbb{R}^{n+1} \to \mathbb{R}$ is defined by

$$g(x, t) = t f(x/t),$$

with domain $\mathrm{dom}(g) = \{(x, t) \in \mathbb{R}^{n+1} : x \in \mathrm{dom}(f),\ t > 0\}$. Prove that if f is convex, then so is its perspective transform g, in one of two ways: (i) straight from the definition of convexity; or (ii) by following the proof in Section 3.2.6 in Boyd and Vandenberghe, explaining each step appropriately.

3 Lipschitz gradients and strong convexity (16 points)

Let f be convex and twice continuously differentiable.

(a, 8 pts) Show that the following statements are equivalent.

i. ∇f is Lipschitz with constant L;

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

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

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

(b, 8 pts) Show that the following statements are equivalent.

i. f is strongly convex with constant m;

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

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

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

4 Solving optimization problems with CVX (27 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 $\min_\theta \; \|y - X\theta\|_2^2$ for a vector y and matrix X. Check your answer by comparing with the closed-form solution $(X^TX)^{-1}X^Ty$.

(a) Using CVX, we will solve the 2d lasso problem and its variants:

$$\min_{\theta \in \mathbb{R}^{mn}} \; \frac{1}{2}\sum_{i=1}^{mn}(y_i - \theta_i)^2 + \lambda \sum_{(i,j)\in E} |\theta_i - \theta_j| \,.$$

The set E is the set of all undirected edges connecting horizontally or vertically neighboring pixels in the image. More specifically, $(i, j) \in E$ if and only if pixel i is the immediate neighbor of pixel j on the left, right, above or below.

1. (5 pts) Load the basic test data from `toy.csv` and solve the 2d lasso problems with $\lambda = 1$. Report the objective value obtained at the solution and plot the solution and original data as images. Why does the shape change its form?

2. (5 pts) Another way to formulate the 2d lasso problem is as follows:

$$\min_{\theta \in \mathbb{R}^{m \times n}} \ \frac{1}{2} \sum_{a=1}^{m} \sum_{b=1}^{n} (y_{a,b} - \theta_{a,b})^2 + \lambda \sum_{a=1}^{m-1} \sum_{b=1}^{n-1} \left\| \begin{pmatrix} \theta_{a,b} - \theta_{a+1,b} \\ \theta_{a,b} - \theta_{a,b+1} \end{pmatrix} \right\|_p .$$

Note that the index a, b here refers to the coordinates of pixel i. When taking a 1-norm ($p = 1$), the formulation reduces to the 2d fused lasso mentioned above, and the latter term is called an "anisotropic" total variation penalty. When taking a 2-norm ($p = 2$), the term is called an "isotropic" total variation penalty.

Solve the "isotropic" 2d lasso problems with $\lambda = 1$ on `toy.csv`. Report the objective value obtained at the solution and plot the solution and original data as images. Informally speaking, why is the output different from the "anisotropic" penalty, and what's the difference?

Hint: For `cvxpy` users, the `diff` function, the `hstack` function, and the axis option in the `norm` function would be useful. For Matlab `CVX` users, there is a `norms(x,p,dim)` function that can compute the norm along different dimensions.

3. (7 pts) Next, we consider how the solution changes as we vary λ. Load a grayscale 64×64 pixel image from `baboon.csv` and solve the isotropic and anisotropic 2d lasso problem for this image for $\lambda \in \{10^{-k/4} : k = 0, 1, \dots, 8\}$. For each λ, report the value of the optimal objective value, plot the optimal image and show a histogram of the pixel values (100 bins between values 0 and 1). What change in the histograms can you observe with varying λ for the isotropic and anisotropic penalties?

(b, 10 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.

1. $\frac{1}{x-y} + \frac{1}{(x+y)^2} \leq z$, $x > |y|$
2. $x^2 + \frac{2}{\log^2 y} \leq 5z^{\frac{1}{4}}$, $y > 1$, $z \geq 0$
3. $2x^2 - 2xy + 5y^2 = 0$
4. $(x^2 + 4y^2 + 1)^{3/2} \leq 7x + y$
5. $2xyz \geq yz + 8$, $x \geq \frac{1}{2}$, $y \geq 0$, $z \geq 0$
6. $x \exp\left(\frac{y}{x}\right) \leq z$, $x > 0$, $z > 0$
7. $z^2 \leq 4xy$
8. $\frac{(x+y)^4}{\sqrt{z}} \leq 2x + y$, $z > 0$

9. $\max^2(x, 2) + |y|^2 - z^2 \leq 0$

10. $\log\left(e^{-\sqrt{x}} + \left(1 + \frac{y}{z}\right)^y\right) \leq -\exp(-z),\ x \geq 0,\ y > 0,\ z > 0$

Bonus. Is the function $f(x) = \log\left(\alpha + \sum_{i=1}^n \frac{\beta_i}{x_i}\right) + \lambda\|x\|_2$, where $\alpha \geq 0$, $\beta > 0$, $\lambda > 0$, convex on $\mathbb{R}^n_{++}$? Is it possible to express this into an equivalent DCP expression? Now, consider the following optimization problem:

$$\begin{array}{ll} \text{minimize} & \log\left(\alpha + \sum_{i=1}^n \frac{\beta_i}{x_i}\right) + \lambda\|x\|_2 \\ \text{subject to} & \sum_{i=1}^n x_i \leq 1 \end{array}$$

where $x \in \mathbb{R}^n_{++}$ is the optimization variable, and $\alpha \geq 0$, $\beta > 0$, and $\lambda > 0$ are fixed problem parameters. Can you use CVX to solve the above problem for given α, β, and λ?