Homework 1

Convex Optimization 10-725/36-725

Due Wednesday September 16 at 4:00pm
submitted to Mallory Deptola in GHC 8001
(Remember to submit each problem on a separate sheet of paper, with your name on at the top)

1 Convex sets

(a) This is Boyd & Vandenberghe's Exercise 2.12, copied here for your convenience. Identify which of the following sets are convex, and provide a brief justification for each.

1. A *slab*, i.e., a set of the form $\{x \in \mathbb{R}^n \mid \alpha \leq a^T x \leq \beta\}$.
2. A *rectangle*, i.e., a set of the form $\{x \in \mathbb{R}^n \mid \alpha_i \leq x_i \leq \beta_i, i = 1, ..., n\}$.
3. A *wedge*, i.e., a set of the form $\{x \in \mathbb{R}^n \mid {a_1}^T x \leq b_1, {a_2}^T x \leq b_2\}$.
4. The set of points closer to a given point than a given set, i.e., $\{x \mid \|x - x_0\|_2 \leq \|x - y\|_2, \forall y \in S\}$, where $S \subseteq \mathbb{R}^n$.
5. The set of points closer to a set than another, i.e., $\{x \mid \mathrm{dist}(x, S) \leq \mathrm{dist}(x, T)\}$, where $S, T \subseteq \mathbb{R}^n$.
6. The set $\{x \mid x + S_2 \subseteq S_1\}$, where $S_1, S_2 \subseteq \mathbb{R}^n$ with S_1 convex.
7. The set of points whose distance to a does not exceed a fixed fraction θ of the distance to b, i.e., the set $\{x \mid \|x - a\|_2 \leq \theta \|x - b\|_2\}$. You can assume that $a \neq b$, and $0 \leq \theta \leq 1$.

(b) The Separating Hyperplane theorem ensures that two disjoint convex sets C and D can always be separated by a hyperplane such that $a^T x \leq b$ $\forall x \in C$, and $a^T x \geq b$ $\forall x \in D$. However, strict separability might not be always possible. Give an example of two *closed* disjoint convex sets that are not strictly separable by a hyperplane. i.e., $\nexists\, a \in \mathbb{R}^n$ such that $a^T x < b$ $\forall x \in C$ and $a^T x > b$ $\forall x \in D$

(c) Prove Farkas' Lemma, which states that for $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$, exactly one of the following is true:

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

Hint: Consider the conic hull of the columns of A, and use the Separating Hyperplane theorem.

2 Convex functions

(a) This is from B&V Additional Exercises 2.9 and 2.20(a) (separate from the main textbook). Prove that the following functions are convex:

1. $\frac{\|Ax-b\|^2}{1-x^T x}$ on $\{x \mid \|x\|_2 < 1\}$
2. The difference between the maximum and minimum value of a polynomial on a fixed interval as function of its coefficients,

$$f(x) = \sup_{t \in [a,b]} p(t) - \inf_{t \in [a,b]} p(t), \text{ where } p(t) = x_1 + x_2 t + x_3 t^2 + \cdots x_n t^{n1}$$

   Here a and b are real constants, with $a < b$.

(b) Specify whether the function is strongly convex, strictly convex, convex, or nonconvex, and give a brief justification for each.

1. $f(x) = x \log x$ for $x > 0$
2. $f(x) = x^4$
3. $f(x) = \log(1 + e^x)$
4. $f(x) = \frac{1}{2} x^T Q x$ for $x \in \mathbb{R}^n$

(c) Using properties of convex functions, argue that the maximum value of a convex function over a closed and bounded polyhedron $\{x \mid Ax \leq b\}$ is achieved at one of its vertices.

3 Lipschitz gradients and strong convexity

Let f be convex and twice differentiable.

(a) Show that the following statements are equivalent.

- ∇f is Lipschitz with constant L;
- $\nabla^2 f(x) \preceq LI$ for all x;
- $f(y) \leq f(x) + \nabla f(x)^T (y - x) + \frac{L}{2} \|y - x\|_2^2$ for all x, y.

(b) Show that the following statements are equivalent.

- f is strongly convex with constant m;
- $\|\nabla f(x) - \nabla f(y)\|_2 \geq m \|x - y\|_2$ for all x, y;
- $\nabla^2 f(x) \succeq mI$ for all x;
- $f(y) \geq f(x) + \nabla f(x)^T (y - x) + \frac{m}{2} \|y - x\|_2^2$ for all x, y.

4 Solving optimization problems with CVX

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.

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.

(a) 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. Make sure that you can solve the least squares problem $\min_\beta \|y - X\beta\|_2^2$ for a vector y and matrix X. Check your answer by comparing with the analytic least squares solution.

(b) Using CVX, we will solve the 1d fused lasso problem discussed in Lecture 1:

$$\min_{\beta \in \mathbb{R}^n} \frac{1}{2} \sum_{i=1}^{n} (y_i - \beta_i)^2 + \lambda \sum_{i=1}^{n-1} |\beta_i - \beta_{i+1}|.$$

The data for this problem, `y.txt` and `beta0.txt`, are available on the class website.

1. Load the data `y.txt` and solve the 1d fused lasso problem with $\lambda = 1$. Report the objective value obtained at the solution.
2. Next, we consider how the solution changes as we vary λ. Solve the optimization problem for 100 logarithmically spaced values from 10^1 to 10^{-2} (in Matlab, `logspace(1, -2, 100)`). For each λ, compute the mean squared error (MSE) of the solution $\hat{\beta}$ and the true β_0 (from `beta0.txt`) as well as the number of changepoints in $\hat{\beta}$. For numerical purposes, define a changepoint to be absolute difference greater than 10^{-8}. Plot MSE and number of changepoints as a function of λ.
3. Find $\hat{\beta}$ that minimizes the MSE in part (ii) and plot $\hat{\beta}$, β_0 and the data y. Is the MSE minimized at the estimate $\hat{\beta}$ that has the same number of jumps (changepoints) as the truth β_0? If not, is the estimate number of jumps at the MSE-optimal solution too small, or too big? (Extra credit: can you explain what you see, statistically?)

(c) 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 (all define a convex set), give equivalent DCP expressions along with a brief explanation of why the DCP expressions are equivalent to the original. DCP expressions should be given in a form that passes analysis 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. $\|(x, y, z)\|_2^2 \leq 1$
2. $\sqrt{x^2 + 1} \leq 3x + y$
3. $1/x + 2/y \leq 5, x > 0, y > 0$
4. $(x + y)^2/\sqrt{y} \leq x - y + 5, y > 0$
5. $(x + z)y \geq 1, x + z \geq 0, y \geq 0$
6. $\|(x + 2y, x - y)\|_2 = 0$
7. $x\sqrt{y} \geq 1,\ x \geq 0, y \geq 0$
8. $\log(e^{y-1} + e^{x/2}) \leq -e^x$