University of Toronto
Department of Electrical and Computer Engineering

ECE367 MATRIX ALGEBRA AND OPTIMIZATION

Problem Set #6

Autumn 2018

Prof. S. C. Draper **Due:** In class on Tuesday, 04 Dec. 2018

Homework policy: Problem sets must be turned in in class on the due date. Late problem sets will not be accepted. See information sheet for full discussion of problem set policy. Some problems are drawn from the course text "Optimization Models" by Giuseppe Calafiore and Laurent El Ghaoui, abbreviated as "OptM", others from the text "Introduction to Applied Linear Algebra", abbreciated "IALA", by Stephen Boyd and Lieven Vandenberghe. An electronic version of the latter is available on the authors' websites. In both cases we indicate the appropriate problem number and name.

Grading: Grading is not uniform. Typically the numerical/design questions are expected to take longer and, correspondingly, are more heavily weighted in the final grade.

Optional problems: Optional problems (if there are any) are located at the end of the problem set and are marked "optional". Although you are not required to do these problems, you are expected to review and understand the solutions to them.

Problem Set #6 problem categories: A quick categorization by topic of the problems in this problem set is as follows:

- Basic problems about convexity: 6.1-6.4
- Related optimization problems: 6.5
- Sparse coding: problem 6.6
- Optional problem (convexity and LPs): 6.7

Problem 6.1 (Convex, affine, and conic hulls)

(a) Consider the set $\mathcal{S} = \{[1\ 1]^T, [1\ 2]^T\} \subseteq \mathbb{R}^2$. Sketch conv($\mathcal{S}$), affine($\mathcal{S}$) and conic($\mathcal{S}$), respectively the convex, affine, and conic hulls of the set $\mathcal{S}$. Each is the union of all combinations of the respective type (convex, affine or conic).

(b) Repeat part (a) for the set $\mathcal{S} = \{[1\ 1]^T, [1\ 2]^T, [0.5\ 0.25]^T\}$.

(c) Consider a set $\mathcal{S}$. What are the respective inclusion relations between the convex hull, the affine hull, and the conic hull of $\mathcal{S}$. I.e., which of these three sets are *always* subsets of the other, regardless of the original $\mathcal{S}$?

Problem 6.2 (Proving convexity-preserving operations)

(a) Prove that the set $\mathcal{S}$ resulting from taking the intersection of a set of convex sets $\mathcal{S}_\alpha$ is itself a convex set. I.e., $\mathcal{S} = \cap_\alpha \mathcal{S}_\alpha$ is a convex set when all the $\mathcal{S}_\alpha$ are convex sets.

(b) Consider any affine function $f : \mathbb{R}^n \to \mathbb{R}^m$ and convex set $\mathcal{S} \subseteq \mathbb{R}^n$. Prove that the image of $\mathcal{S}$ under f, i.e., $f(\mathcal{S}) = \{f(x) | x \in \mathcal{S}\}$, is a convex set.

(c) Consider any affine function $f : \mathbb{R}^n \to \mathbb{R}^m$ and convex set $\mathcal{S} \subseteq \mathbb{R}^m$. Prove that the *inverse* (or *pre-*) image of $\mathcal{S}$ under f, i.e., $f^{-1}(\mathcal{S}) = \{x | f(x) \in \mathcal{S}\}$, is a convex set.

Problem 6.3 (Identifying convexity)
For each of the functions listed in parts (a)-(c) below identify whether the function is (i) convex, is (ii) quasi-convex, is (iii) concave, is (iv) quasi-concave.

(a) $f(x) = e^x - 1$ where $\text{dom} f = \mathbb{R}$.

(b) $f(x_1, x_2) = x_1 x_2$ where $\text{dom} f = \mathbb{R}^2_{++}$.

(c) $f(x) = 1/(x_1 x_2)$ where $\text{dom} f = \mathbb{R}^2_{++}$.

Recall that a function f is "quasiconvex" if all its sublevel sets $\mathcal{S}_\alpha = \{x \in \text{dom} f | f(x) \leq \alpha\}$ are convex sets, i.e., are convex sets for all $\alpha \in \mathbb{R}$. (Note the empty set is a convex set.) A function f is "concave" if the function $-f$ is convex. A function is "quasiconcave" if $-f$ is quasiconvex; equivalently, a function is quasiconcave if every "superlevel" set $\{x \in \text{dom} f | f(x) \geq \alpha\}$ is a convex set.

Problem 6.4 (Quadratic Inequalities)
OptM Problem 8.1.

Problem 6.5 (Monotonicity and locality)
OptM Problem 8.6.

Problem 6.6 (Sparse coding of images)
In this problem you will experiment with an image compression method that uses ℓ_1 regularization. To set the context for this problem, we recommend you to read Section 9.6.2 and Example 9.19 in OptM. For solving the numerical parts you may use `CVX`, `MATLAB` or any other computational tool of your choice. In Example 9.19 of OptM, a vectorized image is considered. In this problem we will consider an image without vectorizing it. To avoid possible confusion with the notation for vectorized version in OptM, we use a different notation in this problem. You will note that the two versions are quite similar.

Let $M \in \{0, \ldots, 255\}^{n\times n}$ be a matrix that represents a grayscale image of dimensions $n \times n$. Similar to matrix A in Example 9.19, we use an orthonormal matrix $H \in \mathbb{R}^{n\times n}$, i.e., $H^T H = HH^T = I_n$, to perform a wavelet transform. Analogous to the relation $\tilde{y} = A^T y$ in Example 9.19, in our case the wavelet transform coefficients of M are obtained as $\tilde{M} = HMH^T$. This is called the "2-D wavelet transform" where we note that $\tilde{M}$ is a matrix, $\tilde{M} \in \mathbb{R}^{n\times n}$. The transform matrix H is designed so that, when applied to natural images. the entries in the $\tilde{M}$ matrix are typically quite "concentrated", i.e., most are of very small magnitude clustered about zero with a much smaller number of large coefficients. The particular H we use in this problem corresponds to the family of "Haar" wavelets.

As in OptM Ex. 9.19 we use ℓ_1 regularization to search for a sparse encoded $X \in \mathbb{R}^{n\times n}$ to represent M in the (Haar) wavelet domain. In this problem the matrix X play the role of the vector x in the OptM example. As the natural equivalence of the LASSO problem discussed in OptM, but now using our matrix notation, we define the optimization problem

$$f(\lambda) = \min_{X\in\mathbb{R}^{n\times n}} \frac{1}{2}\|H^T XH - M\|_F^2 + \lambda\|X\|_1 \tag{1}$$

and solve for best choice of the matrix X. Given the optimizer X^* the optimal (regularized) approximation to M is $\hat{M} = H^T X^* H$. In these expressions, $\|X\|_1 = \sum_{i\in[n]}\sum_{j\in[n]} |X_{i,j}|$ and $\|\cdot\|_F$ denotes the Frobenius norm defined as $\|V\|_F^2 = \sum_{i\in[n]}\sum_{j\in[n]} |V_{i,j}|^2$ for any real matrix V.

(a) Find an expression for the inverse wavelet transform, i.e., use H to express M in terms of $\tilde{M}$. From the point of view of the optimization problem stated in (1), what is the importance of selecting a specially designed transform matrix H, opposed to using a random orthonormal matrix?

(b) Show that $f(\lambda)$ is a *separable* problem. A separable problem reduces to a set of single variable problems. See OptM Example 9.19 for an example of a separable problem. After comparing your result to that in Example 9.19 you should be able to write down the solution to $f(\lambda)$. Do so. You may find following matrix properties useful: $\|V\|_F^2 = \mathrm{trace}(VV^T)$ and $\mathrm{trace}(ABC) = \mathrm{trace}(CAB) = \mathrm{trace}(BCA)$.

We now move to the numerical part of the problem. Download the file `sparseCoding.mat` from the course website. You can load the content of this file onto a `MATLAB` workspace by executing `load ‘sparseCoding’`. You will see two variables `M` and `H` which correspond to the matrices M and H. In this case $n = 256$. You can view the image by executing `imshow(M, [])`. One measure of the goodness of an approximation $\hat{M}$ to M is the “mean-squared-error” (MSE). The MSE is defined as

$$\mathrm{MSE}(M, \hat{M}) = \frac{1}{n^2} \sum_{i\in[n]} \sum_{j\in[n]} |M_{i,j} - \hat{M}_{i,j}|^2.$$

(c) Produce three histograms. The first two are histograms of the values of M and $\tilde{M}$, and should be similar to Figure 9.19 and 9.20 in OptM. For the third histogram, zoom in on your histogram of $\tilde{M}$ around the origin of the x-axis. You should observe that most coefficients are clustered around zero. This indicates the Haar wavelet transform is doing what it was designed to do. Include all three plots in your solutions. Also indicate the number of non-zero coefficients in M and the number of non-zero coefficients in $\tilde{M}$.

(d) Compute the optimal X^* for the problem of (1) when $\lambda = 30$. Also compute the corresponding compression factor and MSE. As is discussed in OptM Problem 9.19, the compression factor is the ratio between the number of non-zero components in X^* and the size of M (which is n^2). For the X^* you found, perform the inverse wavelet transform and plot the resulting approximation $\hat{M}$ alongside the original image. As in part (c), produce a histogram of non-zero components of X^*. Zoom in along the x-axis to see what is happening close to the origin, include a plot of the close-up of the histogram in that regime and comment on what you see.

(e) Repeat part (d) for $\lambda = 10$. What are the differences that you see between the results for two λ values in terms of the respective $\hat{M}$ and the histograms?

Optional problem:

Problem 6.7 (Convexity and concavity of optimal value of an LP)
OptM Problem 9.8.

Hints: (i) For the first part think about re-expressing $\min_x c^T x = \min_x f_x(c)$ where $f_x : \mathbb{R}^n \to \mathbb{R}$ as $f_x(c) = c^T x$. Next, recall that f is concave if $-f$ is convex and apply the equivalent (for concave functions) of the “pointwise supremum or maximum” property of Sec. 8.2.2.4. For the second part one can use the epigraph property characterization of convexity. For the third think about scalar examples.)