Convex and Nonsmooth Optimization
HW9: The Nuclear Norm and Matrix Completion

Michael Overton

Spring 2022

Each question is worth 15 points except #3 which is 25 points

1. The *singular value decomposition* of a p by q matrix A can be defined in several ways which can lead to confusion. Let's assume that $p \geq q$, and, for the moment, that the columns are linearly independent, so the rank is q. Then the definition in the text on p. 648 is consistent with the "reduced" or "economy-sized" SVD in Matlab, `[U,S,V]=svd(A,0)`. When the matrix has rank $r < q$, the definition in the book defines only nonzero singular values, but in the more standard definition, there are $q - r$ zero singular values (and additional columns in U and V). Then, the singular values of A are the square roots of the eigenvalues of the symmetric positive semidefinite q by q matrix $A^T A$. (This is shown on p. 648; see also p. 646.) For more information, see my notes on eigenvalues and singular values.

   Assuming to avoid confusion that the rank is q, define the $p+q$ by $p+q$ symmetric indefinite matrix

   $$B = \begin{bmatrix} 0 & A^T \\ A & 0 \end{bmatrix}$$

   Show that B has $2q$ nonzero eigenvalues which are plus and minus the singular values of A, and $p - q$ eigenvalues which are zero, by coming up with eigenvectors for B that establish this result, constructed from the singular vectors of A.

2. In class we showed that the dual norm of the 2-norm $\|\cdot\|_2$ (the largest singular value) is the *nuclear norm* (the sum of all of the singular values) by writing down an SDP characterization for the 2-norm (p. MN3 of notes on the nuclear norm, which are based on this annotated article

by Recht et al., and then, since this SDP (D), equivalently (D') or (D"), is in standard dual form, looking at the corresponding primal SDP (P), equivalently (P'). At the top of p. MN4, a matrix W which is feasible for (P) is exhibited.

(a) Explain why, as a consequence of the argument on the rest of p. MN4, this feasible W is actually optimal for (P).

(b) Give a formula for a matrix Y, in terms of the SVD of X, which is both feasible and optimal for (D), establishing this by verifying both that it satisfies the inequality in (D) and that the corresponding dual objective value equals the primal objective value for the primal optimal W mentioned above.

(c) Also check complementary: verify that the matrix product of the primal matrix W and the dual slack matrix associated with (D"), constructed from the Y given in (b), is zero.

3. (25 pts) *Low rank matrix completion* is a problem where a $p \times q$ matrix, which is supposed to have low rank, needs to be completed from knowledge of a relatively small number of its entries. The most famous example is the Netflix Prize. It is now well known that just as convex L_1 optimization problems such as LASSO typically result in *sparse* solutions, Nuclear Norm convex optimization problems typically result in *low rank* matrix solutions. The formulation is: minimize the nuclear norm of the matrix subject to the constraint that the matrix has the correct known entries. So, as explained on p. MN5 of the notes on the nuclear norm, we can address the matrix completion problem using the SDP characterization (P) with additional constraints imposing values for known entries in X.

   Write a MATLAB function that uses CVX to minimize the nuclear norm, given a set of k row and column index pairs $(i_1, j_1), \ldots, (i_k, j_k)$ and the associated data X_{ij}. Using the three low-rank examples in Xdata.mat, measure how good the answer is by computing the Frobenius norm $\|X - X^*\|_F$ (`norm(X-Xstar,'fro')`) where X is the data matrix and X^* is the best approximation returned by CVX, starting with $k = 1$ and a randomly chosen index pair, and then successively increasing k, choosing a new (and different!) random index pair each time k is increased. Continue this process until you successfully recover the matrix within a tolerance, not too small, maybe 10^{-5}. Does the norm suddenly drop when k reaches a threshold value? For each of the three data matrices, plot $\|X - X^*\|_F$ as a function of k using `semilogy`, overlaying your plot with more plots using different randomly chosen index sequences (but for the same data matrix X) if this doesn't take too much computation time. (Use `semilogy(...,...,'.-')`; when you call this several times, each time a different color will be used. Use `ceil(n*rand)` to generate a random integer between 1 and n.)

4. Prove the inequality (2.1) on p. 477 of the annotated article by Recht et al. It may be easier to start by proving the analogous inequality for vector norms, with cardinality (number of nonzero entries in the vector) replacing the matrix rank r.

5. Consider the nuclear norm relaxation of the generalization of the matrix completion problem stated on the left side of the bottom of p. 480 of the annotated article by Recht et al. Show that the program written on the right side is indeed its dual. Here, if $A(X) = b$ means $\langle A_k, X\rangle = b_k$, $k = 1, ..., p$, then $A^*(z)$ means the adjoint operation, $\sum_k z_k A_k$.

6. (a) Complete the proof of claims (3) and (4) in the hand-written notes on p. 482 of the annotated article by Recht et al, assuming, although it doesn't say so, that $m = 2$. In particular, for claim (3), find a matrix X which proves that there is no $\delta_r < 1$ that works, i.e., satisfies (3.2), for all X, and, for claim (4), complete the proof that $\delta_r = 1 - \frac{1}{\sqrt{2}}$ works, and find an example X that shows that no smaller δ_r works.

   (b) Complete the proof of Lemma 3.4 on p. 482 of the same paper, showing that the matrices B_1 and B_2 on p. 483 indeed satisfy conditions (1)–(4).