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

Michael Overton

Spring 2020

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 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.

To see how to answer these questions it may be helpful to try some examples in CVX, for example with a random 3 by 2 matrix X, and comparing its SVD with the primal and dual solutions computed by CVX.

3. *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. 5 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. Using the cooked-up low-rank examples in Xdata.mat, use CVX to minimize the nuclear norm specifying only a relatively small number of the entries, where you generate the row and column indexes for these specified entries randomly. Run successive SDPs with increasing numbers of entries specified until you successfully recover the matrix. For each of the 3 matrices in the data file, approximately how many of the entries do you need to specify to be able to recover the matrix almost exactly? Average your results over repeated runs with different row and column index pairs. (Use `ceil(k*rand)` to generate a random integer between 1 and k.)

4. For randomly generated square matrices X of order q and (i) fixed rank $= 3$; (ii) fixed rank $= 10$, do a more systematic investigation of how the norm of Frobenius norm of $\|X - X^*\|$, where X^* is the computed approximation, varies with the number of specified entries. Do you find that the norm drops dramatically when the number of specified entries reaches a certain threshold? Plot your results using `semilogy`. Make the matrices as large as you can while still being able to solve the problem with CVX in a reasonable amount of time (meaning, you don't have to wait too long for the program to run).

5. Investigate the practicality of solving larger problems by using ADMM to solve the SDP. See the bottom of p. 36 of the ADMM paper for how to solve an LP using ADMM, using projection onto the nonnegative orthant (see p. 26). For an SDP, you would have to project onto the cone of positive semidefinite matrices, which can be done by computing the eigenvalues of a symmetric matrix using MATLAB's `eig` and replacing the negative eigenvalues in the factorization $Q\Lambda Q^T$ by zeros. Can you solve larger problems than with CVX at least approximately? How much larger? How good is the accuracy?

6. 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, the sum of $z_k A_k$.

7. 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.

8. Verify claims (3) and (4) in the hand-written notes on p. 482 of the annotated article by Recht et al.