EE263, Summer 2013-14 Instructor: Alex Lemon

Final Review for EE263

TA: Paris Syminelakis

1 Introduction

These are some brief notes to supplement our last problem session that was interrupted. The material we had time to cover during the problem session, included some fundamental properties of:

- "Orthonormal", Projection, Orthogonal and Symmetric matrices.
- SVD and Full SVD.
- Applications of SVD.

As part of the final review I would like to present to you a problem solving strategy that provides a high level overview of the steps involved in solving the kind of optimization problems we have been seeing throughout the course. At this point, I would like to stress that this strategy is simply a rough "road-map" to getting to or near the solution of the problem. *In no way should it be used as a substitute of actual understanding of the concepts in this course.*

2 Problem Solving Strategy

In many cases you might be aware of all the individual elements involved in the solution of a given problem, but you might have a hard time getting started or retracing you steps once you are stuck. The following steps might help you in that respect.

1. *Identify Knowns and Unknowns.* We start by writing down what the known quantities (data) and unknown quantities (variables) are. Then we identify any constraints on our unknowns and properties of the knowns. The constraints might be given explicitly, e.g. $\text{rank}(A) = k$, or implicitly, e.g. the constraint $A = uv^T$ also implies that $\text{rank}(A) = 1$.
2. *Algebraic Expression for Objective Function.* In our problems usually you are either given an explicit objective function, e.g. $J(x) = ||Ax - y||^2$, $J(X) = \|A - X\|_F$, or a description of the objective function in words. In both cases, to come a step closer to solving the problem you should get some (albeit possibly messy) algebraic expression of the objective function.
3. *Simplify Objective Function.* Having obtained an algebraic expression, our next step is to clean up that expression. That could involve grouping terms together where the same unknown quantity appears, e.g. factoring out a matrix or a vector. It could also involve grouping unknowns into a vector or matrix. Further, to improve the compactness of your expression for the objective function, the following two tricks are useful:

(a) Vectorize: $\sum_{i=1}^{N} y_i^2 = \left\| \begin{bmatrix} y_1 \\ \dots \\ y_N \end{bmatrix} \right\|^2$

(b) Stack Columns: $\sum_{k=1}^{m} \|y^{(k)}\|^2 = \| \begin{bmatrix} y^{(1)} & y^{(2)} & \dots & y^{(m)} \end{bmatrix} \|_F^2$

We have also seen throughout the homework problems that in many cases in order to get a clean expression for the objective, we might need to utilize *properties of the known* quantities, e.g. "centered data" $\sum x_i = 0$, and the *constraints on the unknown* quantities, e.g. $XX^T = I$.

4. *Formulate the Optimization Problem.* After we have cleaned our expressions for the objective function and the constraints. We formulate the optimization problem we are asked to solve:

$$\begin{aligned} \min \ & J(x) \\ s.t \ & \text{constraints} \end{aligned}$$

5. *Partial Optimization.* Sometimes our optimization problem might involve more than one unknown, e.g. $J(A, X, b) = ||Y - (AX + b)||_F$. If we fix a subset of the variables, e.g. fix (A_0, X_0), and are able to find the optimal solution to the problem with those variables fixed, e.g. solve $J(A_0, X_0, b) = \tilde{J}(b)$ to get $b^*(A, X)$, we can plug the partial solution back in the original objective to get a problem involving only the variables that we originally fixed, i.e. $J(A_0, X_0, b^*(A_0, X_0)) = \hat{J}(A_0, X_0)$. This process is called partial minimization and is mostly helpful when one can get simple analytic solution to the partial problem, e.g. involving b.

6. *Identify Convenient Relaxation.* Having, simplified the problem as much as possible, your next step would be to find out if the problem falls into one of the category of problems that you know how to solve. If it does then all is well. If not, then you should try to find out a variation of the problem that would. Typically, this is achieved by relaxing some constraints (omitting them temporarily, or slightly weakening them). Hopefully, after the relaxation you would be able to solve the problem exactly.

7. *Translate the Solution of the Relaxation to a Solution of the Original Problem.* Having solved the relaxation does not directly give the solution to our original problem. However, most of the times we can use the properties of the solution of the relaxed problem, to satisfy the constraints that we omitted from the original problem. Even, when this is not directly possible the solution of the relaxation provides as with a good idea of what the solution of the original problem might look like. The last three homework sets included many problems of this kind.

3 Conclusion

In this course, you have been exposed in a variety of theoretical and practical aspects of linear algebra. The power of linear algebra lies in the ability to represent spaces (range, nullspace, orthogonal complement) in terms of their basis, and perform all actions and manipulations through that basis. Further, an important fact is that various optimality properties are expressed as orthogonality conditions. For instance, least-square problems ($\min ||y - Ax||$) involve basically finding an estimate $\hat{x}$ such that the residual $y - A\hat{x}$ is orthogonal to the range of A, i.e. $(y - A\hat{x})^T A = 0 \Leftrightarrow A^T y = A^T A\hat{x}$ (normal equations). Orthogonality also plays a role into the variational characterization (we have not seen in this course) of left and right singular-spaces (spaces spanned by singular vectors) and ultimately accounts for the remarkable properties of the SVD.

The first part of the course treated almost exclusively least-squares problems, whereas the second part involved mostly maximum/minimum norm, extremal trace and low-rank matrix approximation problems. Both parts are unified through the Singular Value Decomposition

$$A = U_1 \Sigma V_1^T = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix}^T$$

that gives a complete picture of the properties of matrices. It reveals the range $R(A) = R(U_1)$ and the nullspace $N(A) = R(V_2) = R(V_1)^\perp$ of the matrix. It gives a formula for the pseudo-inverse $A^\dagger = V_1 \Sigma^{-1} U_1^T$ and as such the solution to least-square problems involving A. In conclusion, the SVD is a potent tool that can be used both for theoretical investigation as well as applications. Good luck with the Final and have a great rest of the summer!