Lecture 11 - Unit 11,12A Notes

11.1 Minimal norm solution

- There is an assumption is SVD of A exists

- If $\sigma_i$ is not 0, then $\hat{c}_i = \hat{b}_i / \sigma_i$. If $\sigma_i$ is 0, then we cannot do the division. Be careful of what is 0 in computers, eg: $e^{-16}$.

- Solve $Ac = b$, get $c = A^{-1}b$. In this case, $A$ is not full rank, we have $c = A^+b$. This is the minus norm solution $c$.

Figure 1: Write $A = U\Sigma V^T$, which is SVD decomposition. Then we can substitute $\hat{c} = v^Tc$, $\hat{b} = u^Tb$ and get $\Sigma \hat{c} = \hat{b}$. $c = V\hat{c}$ because $V^TV = I$.

And we can calculate $\hat{c}_i = \frac{\hat{b}_i}{\sigma_i}$ if $\sigma_i \neq 0$. Be careful what is 0 numerically in computers.
11.2 Sum of Rank One Matrices

\[ A = U \Sigma V^*, \text{where } \Sigma = \begin{bmatrix} \sigma_1 & 0 & 0 & \ldots \\ 0 & \sigma_2 & 0 & \ldots \\ 0 & 0 & \sigma_3 & \ldots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \]

- If we expand \( U, V \) in vectors, then we have \( A = \sum_i \sigma_i u_i v_i^T \). \( u_i v_i^T \) is the outer product of \( u, v \) and the result is a matrix of the same size as \( A \). Here \( u_i v_i^T \) are matrices of rank 1.
- \( A = \sum_{\sigma_i \neq 0} \sigma_i u_i v_i^T \). We can drop the \( u, v \) if their corresponding \( \sigma = 0 \).

11.3 Approximating a Matrix

- Numerically, we can drop terms of very small \( \sigma_i \), so \( A = \sum_{\sigma_i > \epsilon} \sigma_i u_i v_i^T \).
- If keep \( r \) largest singular values, then \( \sum_{i=1}^{r} \sigma_i u_i v_i^T \) is the rank-\( r \) approximation.

11.4 PCA

- We can use a finite number of the largest \( \sigma_i \)’s to reconstruct \( A \).

11.5 Finding Low Rank Approximation

- \( A^T A = V \Sigma^T \Sigma V^T \). \( A^T A \) and \( A A^T \) is symmetric but not positive definite.
  If \( A \) is not full rank (there exists \( \sigma_i = 0 \)), then \( A^T A \) and \( A A^T \) will not be SPD.
- We choose to use the smaller one from \( A^T A \) and \( A A^T \), and the dimension of it means the number of \( \sigma \) you can have.

11.6 Computing Eigenvalues

- In practice, we never solve polynomials to compute eigenvalues, because it is too expensive.
- Similarity transform preserves eigenvalues. \( T^{-1} A T \) and \( A \) has the same eigenvalues.
- If \( A \) is real and symmetric, then \( A \) will have real eigenvalues.
- Triangular matrices’ eigenvalues are the diagonal entries. We want to use similarity transform to make \( T^{-1} A T \) triangular matrices.
- Schur form is \( A = Q U Q^{-1} \), where \( U \) is an upper triangular matrix.

11.7 Condition Number

- The condition number is \( \frac{\| \lambda \|}{\| \lambda \|} \) where \( \lambda, \mu \) are normalized right and left eigenvectors.
- Symmetric matrices have the same right and left eigenvectors, so they have condition number 1.
11.8 QR Iteration

• We compute QR decomposition of $A_k$: $A_k = Q_k R_k$. Then $A_{k+1} = R_k Q_k = Q_k^T Q_k R_k Q_k = Q_k^T A_k Q_k$. All $A_k$ are similar to $A$.

• We can use power method to find largest eigenvalue of $A^{-1}$, then we can find the smallest eigenvalue of $A$.

• If $A$ has distinct eigenvalues, then $A_k$ converge to Schur form (a triangular matrix) of $A$.

11.9 Power Method

• Power method computes the largest eigenvalue and its corresponding eigenvector.

• If we can take out the largest eigenvalue and eigenvector, we can keep using power method to compute the second largest eigenvalue and eigenvector.

11.10 Unit 12A - Regularization

• As we solve least square problems, we can add regularization term. We form a minimization problem: $\min ||Ax - b||_2^2$ and add a regularization term $\lambda ||x||_2^2$. We also observe that $\min(||Ax - b||_2^2 + \lambda ||x||_2^2)$ is equal to $\min(||A^T \sqrt{\lambda} I - A^T \sqrt{\lambda} I x - b||_2^2$.

• If $\lambda = 1$, we have the equation $A^T I x = b$.

• Go back to the minimal norm equation solution, if one $\sigma_i$ is very small, we need to be careful of $\hat{c}_i = \hat{b}_i / \sigma_i$. Adding regularization would fix this problem.

• In diagonal entries of $\Sigma^T \Sigma$, we add 1 to each $\sigma_i$. This changes more if $\sigma_i$ is small.

11.11 Unit 12A - Full Rank Scenario

• After adding regularization, $\hat{c}_i = \hat{b}_i * \frac{\sigma_i}{\sigma_i^2 + 1}$. This will stabilize the small singular values by pushing $\hat{c}_i$ to 0.

• You can use $\lambda = \epsilon$ and this will affect a lot on cases where $\sigma_i < \epsilon$. 
