Modified on Monday 23rd November, 2020, 11:39:4711:39

Nov 23, 2020

- In the last lecture, for solving a linear system $A\mathbf{x} = \mathbf{b}$, we assumed that A is an invertible $m \times m$ matrix and $\mathbf{b}, A\mathbf{b}, \ldots, A^{n-1}\mathbf{b}$ are linearly independent.
  - We defined Krylov subspace $\mathcal{K}_n = \mathrm{span}\{\mathbf{b}, A\mathbf{b}, \ldots, A^{n-1}\mathbf{b}\}$. What happens if $\{\mathbf{b}, A\mathbf{b}, \ldots, A^{n-1}\mathbf{b}\}$ is not linearly independent?
  - What happens if A is not invertible? Do we have to know in advance if A is invertible or not?
- Least squares: Four ways
- Now we are interested in $A\mathbf{x} = \mathbf{b}$ when the system doesn't have a solution, i.e., $\mathbf{b}$ is not in the column space of A. The system has a least squares solution $\hat{\mathbf{x}}$ which minimizes $\|\mathbf{b} - A\hat{\mathbf{x}}\|$. We already introduced the normal equation $A^TA\mathbf{x} = A^T\mathbf{b}$, and showed that $\hat{\mathbf{x}}$ satisfies $A^TA\hat{\mathbf{x}} = A^T\mathbf{b}$. If A^TA is invertible, i.e., A has full column rank, then the normal equation has a unique solution.
- Four ways of computing a least squares solution to $A\mathbf{x} = \mathbf{b}$.
  1. Pseudoinverse A^+ and $\hat{\mathbf{x}} = A^+\mathbf{b}$.
  2. $A^TA\hat{\mathbf{x}} = A^T\mathbf{b}$ when A has independent columns.
  3. The Gram-Schmidt process. $A = QR$, QR-decomposition.
  4. Minimize $\|\mathbf{b} - A\mathbf{x}\|^2 + \delta^2\|\mathbf{x}\|^2$. $(A^TA + \delta^2 I)\mathbf{x}_\delta = A^T\mathbf{b}$. Let $\delta \to 0$.
- Pseudoinverse A^+ of A.
  - If A is invertible then $A^+ = A^{-1}$.
  - Let $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_k, 0, \ldots, 0)$ with $\sigma_i \neq 0$. Then, $\Sigma^+ = \mathrm{diag}(1/\sigma_1, \ldots, 1/\sigma_k, 0, \ldots, 0)$.
  - Let $A = U\Sigma V^T$. Then $A^+ = V\Sigma^+U^T$. If A is an $m \times n$ matrix, A^+ is an $n \times m$ matrix.
  - If A has independent columns, then $A^+ = (A^TA)^{-1}A^T$ and $A^+A = I$.
  - If A has independent rows, then $A^+ = A(AA^T)^{-1}$ and $AA^+ = I$.
- $\mathbf{x}^+ = A^+\mathbf{b}$ is a least squares solution to $A\mathbf{x} = \mathbf{b}$.
  - $\mathbf{x}^+$ solves $\min \|\mathbf{b} - A\mathbf{x}\|$, i.e., $\mathbf{x}^+$ is a least squares solution.
  - $\|\mathbf{x}^+\| < \|\hat{\mathbf{x}}\|$ for any least squares solution $\hat{\mathbf{x}} \neq \mathbf{x}^+$, i.e., $\mathbf{x}^+$ is the minimum norm least squares solution.
  - If $A = U\Sigma V^T$, then $\mathbf{x}^+ = V\Sigma^+U^T\mathbf{b} = A^+\mathbf{b}$. But it costs to get SVD.
  $$\|\mathbf{b} - A\mathbf{x}\| = \|\mathbf{b} - U\Sigma V^T\mathbf{x}\| = \|U^T\mathbf{b} - \Sigma V^T\mathbf{x}\|$$
  Solve $\min \|U^T\mathbf{b} - \Sigma\mathbf{w}\|$ to get $\mathbf{w}^+ = \Sigma^+U^T\mathbf{b}$. $V^T\mathbf{x}^+ = \mathbf{w}^+$. We obtain $\mathbf{x}^+ = V\Sigma^+U^T\mathbf{b} = A^+\mathbf{b}$.
- The least squares solution to $A\mathbf{x} = \mathbf{b}$ is a solution to $A^TA\mathbf{x} = A^T\mathbf{b}$.
  - Projection to the column space of A.
  - $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to the column space of A, i.e., $A^T(\mathbf{b} - A\hat{\mathbf{x}}) = \mathbf{0}$.
  - If A has independent columns, then A^TA is invertible and $\hat{\mathbf{x}} = (A^TA)^{-1}A^T\mathbf{b}$.
- The third way to compute $\hat{\mathbf{x}}$: Gram-Schmidt, $A = QR$
  - Assume that A has full column rank. $A = \begin{bmatrix}\mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n\end{bmatrix}$. Do the Gram-Schmidt orthogonalization to obtain a QR-decomposition. Q has orthonormal columns, $m \times n$, and R is upper triangular, $n \times n$.

  - The least squares solution to $A\mathbf{x} = b$ is $\hat{\mathbf{x}} = R^{-1}Q^T\mathbf{b}$.

$$A^T A\mathbf{x} = A^T b, \quad (QR)^T QR\mathbf{x} = (QR)^T b, \quad , R^T Q^T QR\mathbf{x} = R^T Q^T \mathbf{b}, \quad R\mathbf{x} = Q^T\mathbf{b}$$

  - In computation, it may be better if we rearrange the columns of A. Use the largest column among the remaining columns. This is called a column pivoting. Then we obtain $AP = QR$ for some permutation matrix P.

- The fourth way to compute $\hat{\mathbf{x}}$: Least squares with a penalty term
  - Assume that A does not have full column rank.
  - Minimize $\|\mathbf{b} - A\mathbf{x}\|^2 + \delta^2\|\mathbf{x}\|^2$. Solve $(A^T A + \delta^2 I)\mathbf{x}_\delta = A^T\mathbf{b}$. Let $\delta \to 0$.
  - Recall that if A has independent columns, then $A^+ = (A^T A)^{-1}A^T$ and $A^+A = I$.
  - Let $A = U\Sigma V^T$ be an SVD of A.

$$\lim_{\delta\to 0}(A^T A + \delta^2 I)^{-1}A^T = \lim_{\delta\to 0} V[(\Sigma^T\Sigma + \delta^2 I)^{-1}\Sigma^T]U^T = V\Sigma^+U^T = A^+$$