Math 381 - Fall 2022

Jay Newby

University of Alberta

Week 10

Last Time

1. Stability and accuracy for solving linear systems with QR
2. Conditioning of the least squares problem

Today

1. Rank deficient least squares problem
2. Low rank approximations

Rank deficient least squares

$Ap = 0,\ p \neq 0,\quad x$ solves $\min_x \|Ax-b\|_2^2$

Then so does $x + \alpha p,\ \forall \alpha \in \mathbb{R}$

The idea is to minimize $\|b - Ax\|_2^2$ and $\|x\|_2$. SVD is backward stable for solving the problem. The general SVD is regain unique soln

$$A = \begin{bmatrix} \hat{U} & U_0 \end{bmatrix} \begin{bmatrix} \hat{\Sigma} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \hat{V}^T \\ V_0^T \end{bmatrix}. \quad N(A)$$

The above can be use to derive a general solution to the rank deficient least squares problem.

Rank deficient least squares

Homework Problem:

Let $A \in \mathbb{R}^{m \times n}$ be rank $r \leq \min\{n, m\}$. The singular values of A are typically ordered so that

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{n-1} \geq \sigma_n.$$

The number of positive (i.e., nonzero) singular values is the rank of A. In practice, some singular values might be nonzero but very small. One can solve the rank deficient least squares problem (see previous slide) using a low rank approximation of A. This can be done by defininig a threshold $\delta > 0$ and using only those singular values that are above the threshold. In other words, we can find $k \leq r$ such that

$$k = \max\{1 \leq j \leq r \mid \sigma_j > \delta\},$$

and solve the rank-deficient least squares problem using only the first k singular values.

New Topic: Low Rank Approximations

Low rank approximation

Let $A \in \mathbb{R}^{m\times n}$ have rank $r \leq \min\{n, m\}$. We want to approximate A with a matrix A_k that is rank $k < r$. The error is

$$\mathcal{E}_k = \|E_k\| = \|A_k - A\|.$$

Rank one matrices

Claim:

Let $x, y \in \mathbb{R}^n$, with $x \neq 0$ and $y \neq 0$. The $n \times n$ matrix,

$$W = xy^T,$$

has rank $r = 1$.

$$xy^T = \begin{bmatrix} x \end{bmatrix} \begin{bmatrix} y^T \end{bmatrix} = \begin{bmatrix} y_1 x & y_2 x & \cdots & y_n x \end{bmatrix}$$

SVD

Let $A \in \mathbb{R}^{m \times n}$ and let u_j and v_j be the orthonormal columns of the SVD of A. Let σ_j be the singular values in non increasing order so that

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$$

Then, we can represent the matrix A as the sum of rank one matrices as follows,

$$A = \sum_{j=1}^{n} \sigma_j u_j v_j^T.$$

What if some singular values are equal to zero?

SVD: rank k approximation

Let $A \in \mathbb{R}^{m \times n}$ be rank $r \leq \min\{n, m\}$. Define the rank $k \leq r$ approximation as

$$A_k = \sum_{j=1}^{k} \sigma_j u_j v_j^T.$$

Theorem (Schmidt 1907, Eckart and Young 1936)

Let $A \in \mathbb{R}^{m\times n}$. Let $\|\cdot\|_F$ denote the Frobenius norm

$$\|A\|_F = \left[\sum_{i,j} |a_{ij}|^2\right]^{1/2}.$$

For each k with $1 \leq k \leq n-1$, A_k is the best rank k approximation to A with respect to the Frobenius norm, with corresponding error $E_k = A - A_k$ of magnitude

$$\|E_k\|_F = \left[\sum_{j=k+1}^{n} \sigma_j^2\right]^{1/2}$$

A starting point for connecting rank one expansions to several common matrix factorizations

Theorem: Wedderburn rank-one reduction formula (Wedderburn 1934; Chu, Funderlic, and Golub 1995)

Let $A \in \mathbb{R}^{m\times n}$. If $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$ are vectors such that $\omega = y^T Ax \neq 0$ then the matrix

$$B = A - \frac{1}{\omega}Axy^T A = A\overbrace{\left[I - \frac{1}{\omega}xy^T A\right]}^{M}$$

has rank exactly one less than the rank of A.

Connected result: for $n=m$ with A nonsingular

If $u \in \mathbb{R}^n$, $v \in \mathbb{R}^n$ and $A+uv^T$ is nonsingular then $[A+uv^T]^{-1} = \left[I - \frac{1}{1+v^TA^{-1}u}A^{-1}uv^T\right]A^{-1}$.

Lemma: $\det(I+A^{-1}uv^T) = 1+v^TA^{-1}u$

This is useful to update an inverse matrix given a "rank 1 change" to A

SVD as an iterative procedure

- Define $E_0 = A$
- Find $\sigma_1 \geq 0$ and unit vectors u_1, v_1 such that $\sigma_1 u_1 v_1^T$ is the best rank 1 approximation to E_0.
- Define $E_1 = E_0 - \sigma_1 u_1 v_1^T$
- Find $\sigma_2 \geq 0$ and unit vectors u_2, v_2 such that $\sigma_2 u_2 v_2^T$ is the best rank 1 approximation to E_1
- Repeat

There is no direct algorithm that yields the SVD in a finite number of operations.

QR

Let $A \in \mathbb{R}^{m \times n}$ be full rank, and let Q and R be the QR factorization such that $A = QR$. Then, a rank 1 expansion is given by

$$A = \sum_{j=1}^{n} q_j r_j^T.$$

LU

Let $A \in \mathbb{R}^{m \times n}$ be full rank. Assuming that an LU decomposition of A exists, we have the rank 1 expansion,

$$A = \sum_{j=1}^{n} l_j u_j^T.$$

Cholesky

Let $A \in \mathbb{R}^{n \times n}$ be symmetric and positive definite. Let $A = RR^T$ be the Cholesky factorization of A. A rank 1 expansion is given by,

$$A = \sum_{j=1}^{n} r_j r_j^T.$$

Gaussian elimination with pivoting as an iterative procedure

The idea is similar to the SVD procedure, but we do not find the best rank 1 approximation at each step.

- Define $E_0 = A$
- Find pivot $i_1, j_1 = \text{argmax}\{|[E_0]_{ij}|\}$
- Let u_1 be the j_1 column of E_0 and v_1^T be the i_1 row of E_0
- Define $A_1 = \frac{1}{[E_0]_{i_1,j_1}} u_1 v_1^T$
- Define $E_1 = E_0 - A_1$
- Repeat

QR as an iterative procedure

Exercise: You can encode Gram-Schmidt this way.

[Examples in Jupyter]