Fall-23 5304 LecN7 Notes

Wan

November 13, 2023

Topics: The Gram-Schmidt algo and the QR Factorization; Least-squares problems; Applications; Data fitting.

Math tools: Eigenvalues and Eigencevectors

1 Least-Squares Systems and Data fitting

Problem BG

1, overdetermined system, more euqations than unknowns, no sol

如果一个方程组无解，那么这个方程组被称为不一致。例如下面的方程组：

$x_1 + x_2 = 2$
$x_1 - x_2 = 1$
$x_1 + x_2 = 3$

根据线性代数的知识，m个方程n个未知量 $m > n$ 时通常无解，但是虽然不能求出 $Ax = b$ 的解，那何不退而求其次，去寻找与解近似的向量 x 。

那么如何定义与解相似，一般使用欧氏距离来进行度量，即两点间的距离，这其实很好理解，越相似，欧氏距离越近，这样求出的 x 被称为最小二乘解。

2, Function approximation

3, Goal: Find the best approximation to the system of equations.

Geometric Interpretation

Good illustration: Data fitting

2 QR Factorization

2.1 Concept

Definition 8.6 QR-factorization

*Let A be an $m \times n$ matrix with independent columns. A **QR-factorization** of A expresses it as $A = QR$ where Q is $m \times n$ with orthonormal columns and R is an invertible and upper triangular matrix with positive diagonal entries.*

The importance of the factorization lies in the fact that there are computer algorithms that accomplish it with good control over round-off error, making it particularly useful in matrix calculations. The factorization is a matrix version of the Gram-Schmidt process.

A = QR where Q is an orthogonal matrix and R is an invertible upper triangular matrix (diagonal entry elements are all positive). If A is non-singular, then this factorization is unique.

QR Visulization (c_j can be expressed as a linear combination of q_j):

Using these equations, express each $\mathbf{c}_k$ as a linear combination of the $\mathbf{q}_i$:

$$
\begin{aligned}
\mathbf{c}_1 &= \|\mathbf{f}_1\|\mathbf{q}_1 \\
\mathbf{c}_2 &= (\mathbf{c}_2 \cdot \mathbf{q}_1)\mathbf{q}_1 + \|\mathbf{f}_2\|\mathbf{q}_2 \\
\mathbf{c}_3 &= (\mathbf{c}_3 \cdot \mathbf{q}_1)\mathbf{q}_1 + (\mathbf{c}_3 \cdot \mathbf{q}_2)\mathbf{q}_2 + \|\mathbf{f}_3\|\mathbf{q}_3 \\
\vdots \quad & \quad \vdots \\
\mathbf{c}_n &= (\mathbf{c}_n \cdot \mathbf{q}_1)\mathbf{q}_1 + (\mathbf{c}_n \cdot \mathbf{q}_2)\mathbf{q}_2 + (\mathbf{c}_n \cdot \mathbf{q}_3)\mathbf{q}_3 + \cdots + \|\mathbf{f}_n\|\mathbf{q}_n
\end{aligned}
$$

These equations have a matrix form that gives the required factorization:

$$
\begin{aligned}
A &= \begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \mathbf{c}_3 & \cdots & \mathbf{c}_n \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \mathbf{q}_3 & \cdots & \mathbf{q}_n \end{bmatrix}
\begin{bmatrix}
\|\mathbf{f}_1\| & \mathbf{c}_2 \cdot \mathbf{q}_1 & \mathbf{c}_3 \cdot \mathbf{q}_1 & \cdots & \mathbf{c}_n \cdot \mathbf{q}_1 \\
0 & \|\mathbf{f}_2\| & \mathbf{c}_3 \cdot \mathbf{q}_2 & \cdots & \mathbf{c}_n \cdot \mathbf{q}_2 \\
0 & 0 & \|\mathbf{f}_3\| & \cdots & \mathbf{c}_n \cdot \mathbf{q}_3 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & \|\mathbf{f}_n\|
\end{bmatrix}
\end{aligned}
\tag{8.5}
$$

Here the first factor $Q = \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \mathbf{q}_3 & \cdots & \mathbf{q}_n \end{bmatrix}$ has orthonormal columns, and the second factor is an $n \times n$ upper triangular matrix R with positive diagonal entries (and so is invertible). We record this in the following theorem.

3 Using classical Gram-Schmidt to get QR

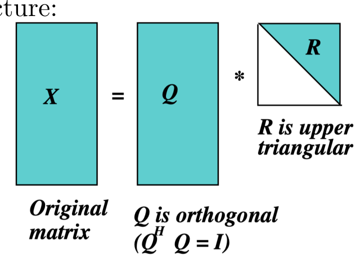
Classical Gram-Schmidt

Input and Ouput Description

Input: $X = [x_1, x_2 ... x_n]$, should be full rank (columns are linearly independent), but not necessarily sqare.

Output: Will get Q and R where Q is orthogonal ($Q^T Q = I$), and R is upper triangular.

Picture:



Process

Input: $X = [x_1, x_2 \ldots x_n]$ X full rank

Goal: Find $Q = [q_1, q_2 \ldots q_n]$, orthonormal cols and

$\text{span}(Q) = \text{span}(X)$.

即找到X的标准正交基向量组.

Note: Each round 主要分2步: ① 正交化 ② 归一化.

update $\hat{q}$

逐列处理

Step ① (Initialization):

$\hat{q} = x_1$ (此时没有已知的投影分量需要减去).

$q_1 = \frac{\hat{q}}{\|\hat{q}\|}$ (normalize)

Step② : Make $\hat{q}$ orthogonal to q_1

$\hat{q} = x_2 - \underbrace{(x_2, q_1)}_{\text{inner product, scalar.}} q_1$

$q_2 = \frac{\hat{q}}{\|\hat{q}\|}$.

step③ : Make $\hat{q}$ 同时正交于 q_1, q_2

$\hat{q} = x_3 - (x_3, q_1) q_1 - (x_3, q_2) q_2$

$q_3 = \frac{\hat{q}}{\|\hat{q}\|}$.

⋮

step ⓙ : Conclusion.

$\hat{q} = x_j - (x_j, q_1) q_1 - (x_j, q_2) q_2 \ldots - (x_j, q_{j-1}) q_{j-1}$.

$q_j = \frac{\hat{q}}{\|\hat{q}\|}$.

x_j is the linear combination of q_j

Using Classical Gram-Schmidt to get QR, start from step j

QR factorization. Start from Classical Gram-Scmidt

j-th step

$$\hat{q} = x_j - \overbrace{\boxed{(x_j, q_1)}}^{\text{scalar}} q_1 - \dots - (x_j, q_{j-1})\, q_{j-1}$$

$$q_j = \hat{q} / \|\hat{q}\|$$

set $r_{ij} = (x_j, q_i)$ and $r_{jj} = \|\hat{q}\|$ (拐到/存储)

$$r_{jj} q_j = x_j - \sum_{1:j-1} r_{ij} q_i$$

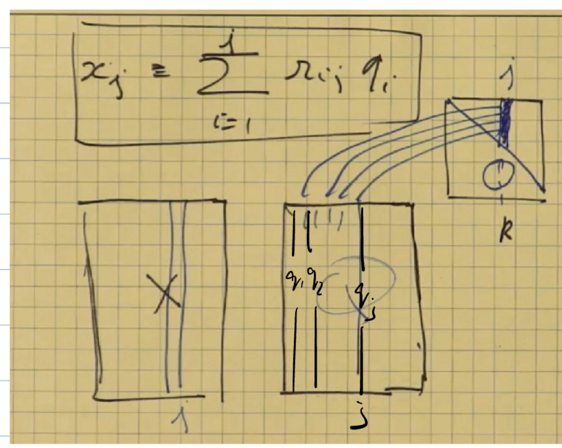
$$x_j - r_{1j} q_1 - \dots - r_{j-1\,j} q_{j-1} = r_{jj} q_j \quad \Longrightarrow$$

$$x_j = r_{1j} q_1 \dots + r_{j-1\,j} q_{j-1} + r_{jj} q_j$$

$$x_j = \sum_{i=1}^{j} r_{ij} q_i \quad \Longrightarrow \quad X = Q\,R$$

R = upper triangular matrix n x n

The j-th col of X is the linear combination of r_{ij} and q_i



$$X = [x_1, x_2 \ldots x_n]$$

inner product.

$$= [q_1, q_2 \ldots q_n] \begin{bmatrix} \|\hat{q}_1\| & x_2 \cdot q_1 & x_2 \cdot q_1 & x_n \cdot q_1 \\ 0 & \|\hat{q}_2\| & x_3 \cdot q_2 & x_n \cdot q_2 \\ 0 & 0 & \|\hat{q}_3\| & x_n \cdot q_3 \\ \vdots & \vdots & 0 & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \|\hat{q}_n\| \end{bmatrix}$$

Cost

The cost of classical Gram-Schmidt is $2mn^2 - \frac{2}{3}n^3$ flops.

4 Using modified Gram-Schmidt to get QR (better)

Verification of Algo:
1, A = QR. by checking whether norm(A - Q*R) is a small number.
2, Orthogonality: Q is orthogonal. by checking whether $Q^T Q = I$, norm(Q'* Q - eye(n)) should be small.

4.1 Modified Gram-Schmidt

Cost of QR by using classical Gram-Schmidt:

Cost Analysis

Examples

See topic Notes.

Cost

lalala

Application

QR分解主要有三个用途：

- （1）首先它可以用于求解n个方程n个未知变量的系统 $Ax = b$，如果A非奇异，那么就可以将A分解为QR，回代则求解出x，但是QR分解线性方程组比LU分解方法的计算代价大三倍还多，用的不多。
- （2）用于求解最小二乘问题
- （3）用于特征值计算，这个之后的帖子再写

Application1: Use QR to solve least-squares problems

11-09

unitary matrix perserves the length of a vector.

rank-1 update

3 steps:

1, transform problem 2, gradient