5374 Fall '22

Numerical Linear Algebra Lecture 14

QR decomposition continued

Least square problems.

Last time we reviewed the Gram-Schmidt process and saw how given a basis

$$a_1, \dots, a_n$$

we can produce an orthonormal basis

$$q_1, \dots, q_n$$

by

$$(r_{11} = \|a_1\|_2)$$

$$r_{11} q_1 = a_1$$

$$r_{22} q_2 = a_2 - (a_2, q_1) q_1$$

$$\vdots$$

$$r_{kk} q_k = a_k - \sum_{j=1}^{k-1} (a_k, q_j) q_j$$

$$r_{kk} = \| a_k - \sum_{j=1}^{k-1} (a_k, q_j) q_j \|_2$$

Clearly this is an iterative procedure where q_k is built from a_k and $q_1, \dots, q_{k-1}$.

From the construction it follows that

$$\text{span}[a_1, \dots, a_k] = \text{span}[q_1, \dots, q_k]$$
$$k = 1, 2, \dots, n$$

and $(q_i, q_j) = \delta_{ij}$. Moreover,

$$a_1 = r_{11} q_1$$
$$a_2 = (a_2, q_1) q_1 + r_{22} q_2$$
$$\vdots$$
$$a_k = (a_k, q_1) q_1 + \dots + (a_k, q_{k-1}) q_{k-1} + r_{kk} q_k$$

Let's define r_{ij} (for $i \neq j$) by

$$r_{ij} = \begin{cases} (a_j, q_i) & \text{if} \quad i < j \\ 0 & \text{if} \quad i > j \end{cases}$$

In this notation

$$a_1 = r_{11} q_1$$
$$a_2 = r_{12} q_1 + r_{22} q_2$$
$$\vdots$$
$$a_k = r_{1k} q_1 + \dots + r_{kk} q_k$$
$$\vdots$$
$$a_n = r_{1n} q_1 + \dots + r_{nn} q_n$$

This n vector equations can be expressed as a single matrix equation

$$\left(a_1 \middle| \cdots \middle| a_n \right) = \left(q_1 \middle| \cdots \middle| q_n \right) \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ 0 & r_{22} & & r_{2n} \\ \vdots & 0 & \ddots & \vdots \\ 0 & 0 \cdots & 0 & r_{nn} \end{pmatrix}$$

If A is an arbitrary $n \times n$ invertible matrix, we can apply the above process to the columns of A $(a_1, \dots, a_n)$ and obtain n orthonormal vectors $(q_1, \dots, q_n)$ and coefficients r_{ij} such that if we define Q and R by

$$Q = \left(q_1 \middle| \cdots \middle| q_n \right)$$

$$R = \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ 0 & \ddots & & \\ \vdots & & \ddots & \\ 0 & \cdots & 0 & r_{nn} \end{pmatrix}$$

Then

$$A = QR$$

This proves any invertible matrix admits a QR decomposition.

A detour on projections

A $n \times n$ matrix P is said to be idempotent if

$$P^2 = P$$

A projector or projection operator is a symmetric idempotent operator, i.e. P such that

$$P^2 = P \quad , \quad P^t = P$$

Examples :

① $A = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$

② Let $q \in \mathbb{R}^n$ be a unit vector, and consider

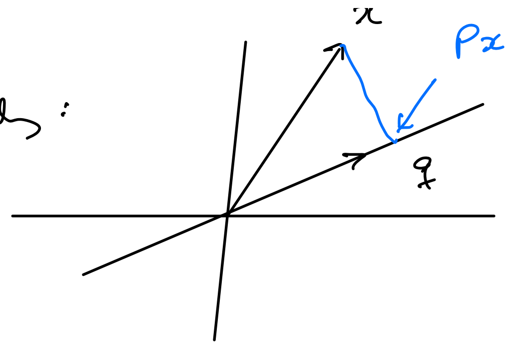
$$\underline{P} = q \otimes q$$

$$(\ \underline{P}_{ij} = q_i q_j \ )$$

$$\underline{P} x = (q, x)\, q \quad (*)$$

Is easy to check from $(*)$ that $P^2 = P$

Geometrically:



$$
\begin{aligned}
\underline{P}(Px) &= \underline{P}\left((q,x)\,q\right) \\
&= (q,x)\,\underline{P}q \\
&= (q,x)\,q \qquad (\underline{P}q = q)
\end{aligned}
$$

The example in (1) corresponds to

$$\begin{pmatrix}0\\1\end{pmatrix} \otimes \begin{pmatrix}0\\1\end{pmatrix}$$

(3) If $q_1, \dots, q_k$ are an orthonormal family of vectors $(1 \le k \le n)$, let

$$\underline{P} = q_1 \otimes q_1 + q_2 \otimes q_2 + \dots + q_k \otimes q_k$$

Clearly $\underline{P}^t = \underline{P}$

$$
\left(\begin{aligned}
\underline{P}^t &= (q_1 \otimes q_1)^t + \dots + (q_k \otimes q_k)^t \\
&= q_1 \otimes q_1 + \dots + q_k \otimes q_k
\end{aligned} \right)
$$

$$\underline{P}x = (q_1, x)\,q_1 + \dots + (q_k, x)\,q_k$$

In particular,

$$\underline{P}\, q_i = (q_1, q_i)\, q_1 + \cdots + (q_k, q_i)\, q_k$$

Since $(q_i, q_j) = \delta_{ij}$, the above reduces to

$$\underline{P}\, q_i = (q_i, q_i)\, q_i = q_i$$

$\Longrightarrow$ if x is a linear combn of $q_1, \dots, q_k$ then

$$\underline{P}\, x = x.$$

$\Longrightarrow$ $P^2 = P$.

So $\underline{P}$ is the orthogonal projection onto the space spanned by $q_1, \dots, q_k$.

As it turns out, these are all the projectors

(P is $n \times n$)

Lemma: Let P be a projector, then $\forall\ x \in \mathbb{R}^n$ we can write x uniquely as a sum

$$x = x_p + x_0$$

where

$$x_p \in \mathrm{Im}(P), \quad x_0 \in \mathrm{Ker}(P)$$

Proof: Let $x_p = Px$,

$$x_0 = x - Px$$

Clearly $x_p \in \mathrm{Im}(P)$, $x = x_p + x_0$,

lastly,

$$\begin{aligned} P x_0 &= Px - P^2 x \\ &= Px - Px = 0 \end{aligned}$$

so $x_0 \in \mathrm{Ker}(P)$. ⊠

In particular, if P is a projector, we can select an orthonormal basis of P, $q_1, \dots, q_k$, and we will have

$$P = q_1 \otimes q_1 + \cdots + q_k \otimes q_k$$

This defines a 1-1 correspondence between subspaces of $\mathbb{R}^n$ and projectors; and here $\operatorname{tr}(P)$ is equal to the dimension of the corresponding subspace.

Projectors and more general QR decompositions

A remark about Gram-Schmidt.

If $a_1, \dots, a_n$ are some linearly independent family of vectors in $\mathbb{R}^m$ (so $m \geq n$) we can apply the Gram-Schmidt process and obtain n orthonormal vectors

$$q_1, \dots, q_n$$

and coefficients r_{ij} $(i, j = 1, \dots n)$ such that

$$a_1 = r_{11} q_1$$

$$a_2 = r_{12} q_1 + r_{22} q_2$$

$$\vdots$$

$$a_n = r_{1n} q_1 + \dots + r_{nn} q_n$$

Define

$$Q = \left(q_1 \,|\, \cdots \,|\, q_n \right)$$

this is a $m\times n$ matrix, and define

$$R = \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ 0 & r_{22} & & \\ \vdots & & \ddots & \\ 0 & \cdots & & r_{nn} \end{pmatrix}$$

Then

$$\underbrace{A}_{m\times n} = \underbrace{Q}_{m\times n}\;\underbrace{R}_{n\times n}$$

Sometimes this is also called "the reduced QR decomposition" (see Trefethen and Bau)

When $m>n$, Q cannot be orthogonal, since it is a rectangular matrix.

But, what do we get from

$$\underbrace{QQ^T}_{m\times m} \quad \text{and} \quad \underbrace{Q^TQ}_{n\times n}$$

$$\begin{pmatrix} 1 & 0 & \cdots & & 0 \\ \vdots & \ddots & & & \\ & & 1 & & \\ \vdots & & & 0 & \\ 0 & & & \cdots & 0 \end{pmatrix} \qquad I_{n\times n}$$

n (upper diagonal block), $m-n$ (lower diagonal block)

Observation: Let $q_1, \dots, q_k$ be orthonormal vectors in $\mathbb{R}^n$ ($k \leq n$) and set P as the projector onto the subspace spanned by them. Let

$$Q = (q_1 | \dots | q_k) \quad (n \times k)$$

Then $P = QQ^t$, why?

If $x \in \mathbb{R}^n$, then

$$Px = (q_1, x) q_1 + \dots + (q_k, x) q_k$$

(thanks to the orthonormality of the q_i's)

but

$$(q_1, x) q_1 + \dots + (q_k, x) q_k$$

$$= \begin{pmatrix} q_1 \dots q_k \end{pmatrix} \begin{pmatrix} (q_1, x) \\ \vdots \\ (q_k, x) \end{pmatrix}$$

$$= (q_1 \dots q_k) \begin{pmatrix} q_1^t \\ \vdots \\ q_k^t \end{pmatrix} x = QQ^t x$$

Least Square Problems

If $n < m$, the equation

$$Ax = b \quad \left(\begin{array}{ll} A & \text{is } m \times n \\ b & \text{is in } \mathbb{R}^m \end{array} \right)$$

hardly ever has a solution, since it only covers by definition those b's with

$$b \in \mathrm{Im}(A)$$

and this is a small part of $\mathbb{R}^m$.

A problem that always has some solution is "find x so that Ax is as close as possible to b"

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad \|Ax - b\|$$

The nature of this minimization problem depends drastically on the choice of norm we use. The most popular choices are $\|\cdot\|_1$, $\|\cdot\|_\infty$, and $\|\cdot\|_2$, the latter being the absolutely most studied and used one.

This is a linear algebra class, and we will only study $\|\cdot\| = \|\cdot\|_2$.

In this case the problem is written as

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad \|Ax-b\|_2^2$$

and this is a sum of squares:

$$\|Ax-b\|_2^2 = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} a_{ij} x_j - b_i \right)^2$$

so we call it the least squares problem.

Let us use the following notation

$$f(x) = \tfrac{1}{2}\|Ax-b\|_2^2$$

next class we will study $\nabla f(x)$, $D^2 f(x)$

$$\nabla f(x) = A^t(Ax-b), \quad D^2 f(x) = A^t A$$

(what about 1-d ! $f(x) = \frac{1}{2}(ax-b)^2$, $a, b \in \mathbb{R}$)