5374 Fall '22

Numerical Linear Algebra Lecture 16

Today:
* More on QR decomposition
* Regression / function approximation
* Positive matrices and their properties

The last thing we learned is how the QR decomposition aids in solving least squares problems:

If A is $m\times n$ and

$$A = Q\,R \qquad (m \geq n)$$

$$\begin{pmatrix} \\ A \\ \\ \end{pmatrix}_{m\times n} = \begin{pmatrix} \\ \quad\quad \\ \\ \end{pmatrix}_{m\times n} \begin{pmatrix} \quad \end{pmatrix}_{n\times n}$$

(this picture emphasizes there are more rows than columns in Q and A)

This decomposition, which we obtained from the Gram-Schmidt process applied to the column of A, is sometimes known ($m>n$) as the reduced

QR decomposition.

This is in contrast to the "generalized QR decomposition" which has the form ($m \geq n$)

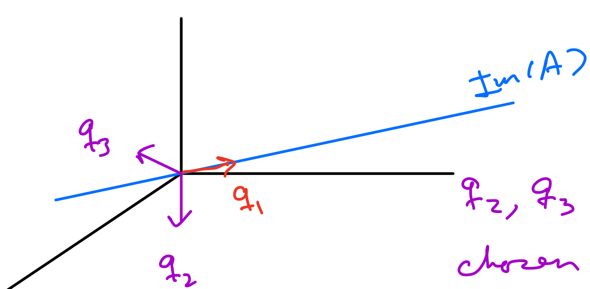
$$\begin{pmatrix} A \end{pmatrix}_{m\times n} = \begin{pmatrix} \tilde{Q} \end{pmatrix}_{m\times m} \begin{pmatrix} \tilde{R} \end{pmatrix}_{m\times n}$$

where observe Q above was $m\times n$ and $\tilde{Q}$ here is $m\times m$ (so, a square matrix), and R above was $n\times n$ (so, a square matrix), and $\tilde{R}$ is $m\times n$. How are $\tilde{Q}$ and $\tilde{R}$ defined?

$\tilde{Q}$: Start from $q_1, \dots, q_n$ the column of Q, they form a basis of $\mathrm{Im}(A)$ ($\neq \mathbb{R}^n$), what we do is we select $m-n$ additional vectors

$$q_{n+1}, \dots, q_m$$

such that $q_1, \dots, q_m$ form an orthonormal basis of $\mathbb{R}^m$



q_2, q_3 could be chosen out of infinitely many choices that produce an orthonormal basis ($n<m-1$)

Then the matrix $\tilde{Q}$ is:

$$\tilde{Q} = \begin{pmatrix} q_1 & \cdots & q_m \end{pmatrix} = \begin{pmatrix} \underbrace{q_1 \cdots q_n}_{Q} & q_{n+1} \cdots q_m \end{pmatrix}$$

For $\tilde{R}$, we simply take R and add $m-n$ rows all equal to zero.

Then

$$A = \tilde{Q}\tilde{R} = \begin{pmatrix} q_1 & \cdots & q_n & q_{n+1} & \cdots & q_m \end{pmatrix} \begin{array}{c} \begin{pmatrix} r_{11} & \cdots & r_{1n} \\ 0 & r_{22} & \vdots \\ \vdots & \ddots & \ddots \\ 0 & 0 \cdots 0 & r_{nn} \\ 0 & 0 \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{pmatrix} \end{array} \Big\} m-n$$

<u>Remark</u>: Observe that for a square matrix A, the QR decomposition provides a way to compute $\det(A)$. Since

$$\det(A) = \det(QR) = \det(Q)\det(R)$$
$$= \det(R)$$
$$= \det \begin{pmatrix} r_{11} & x & x & x \\ 0 & r_{22} & & \vdots \\ \vdots & 0 & \ddots & x \\ 0 & \cdots & 0 & r_{nn} \end{pmatrix} = r_{11} r_{22} \cdots r_{nn}$$

This shows how to compute $\det(A)$ in no more than $O(n^4)$ FLOPs (thanks to the Gram-Schmidt algorithm). Compare this with our discussion about $\det(A)$ and Cramer's rule

(i.e. compare for say $n = 100$:

$100^4 = 10^8$

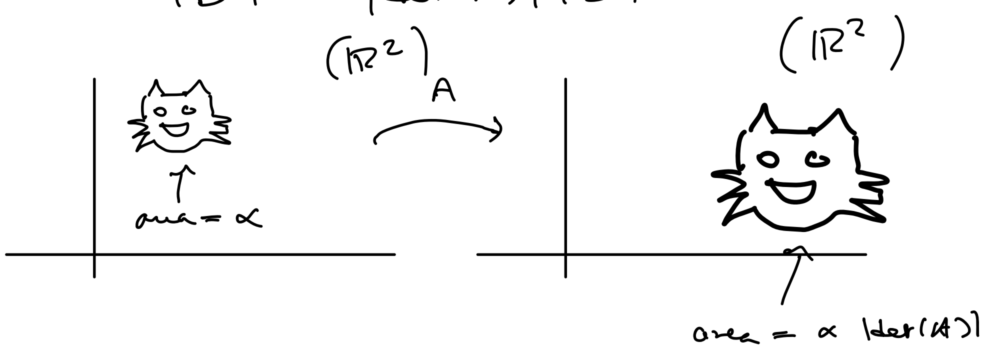
$100! = \dots$ )

Remark: How many ways are there to see $\det(Q) = \pm 1$ for any orthogonal matrix?

① The determinant (well, its absolute value) represents the change in area/volume/n-dimensional Lebesgue measure caused by transforming a set via A:

If $D \subset \mathbb{R}^n$ is Lebesgue measurable and $D' = \{ y \mid y = Ax \text{ for } x \in D \}$ then

$$|D'| = |\det(A)|\,|D|$$



In particular, if A ($n \times n$) preserves distances, Then it will preserves n-dimensional Lebesgue measure, so

$$|\det(A)| = 1$$

(Camille Jordan)

(2) The Jordan decomposition theorem says that if Q is an orthogonal matrix then there exists a change of basis R such that

$$RQR^{-1} = \begin{pmatrix} \cos(\alpha_1) & -\sin(\alpha_1) & 0 & 0 & \cdots & & 0 & 0 \\ \sin(\alpha_1) & \cos(\alpha_1) & 0 & 0 & & & 0 & 0 \\ 0 & 0 & \cos(\alpha_2) & -\sin(\alpha_2) & 0 & 0 \cdots & 0 & 0 \\ 0 & 0 & \sin(\alpha_2) & \cos(\alpha_2) & 0 & 0 \cdots & 0 & 0 \\ \vdots & & 0 & 0 & \ddots & & & \\ 0 & 0 & \vdots & & & \lambda_1 & 0 \cdots & 0 \\ 0 & 0 & & & & \vdots & \lambda_2 \cdots & \\ & & & & & 0 & & \lambda_m \end{pmatrix}$$

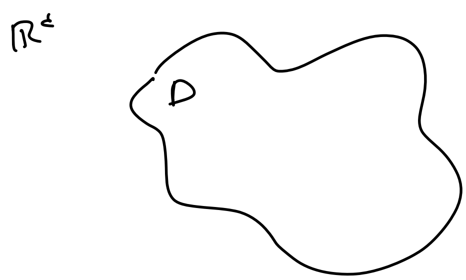
when $\lambda_1, \ldots, \lambda_m$ are all ± 1's.

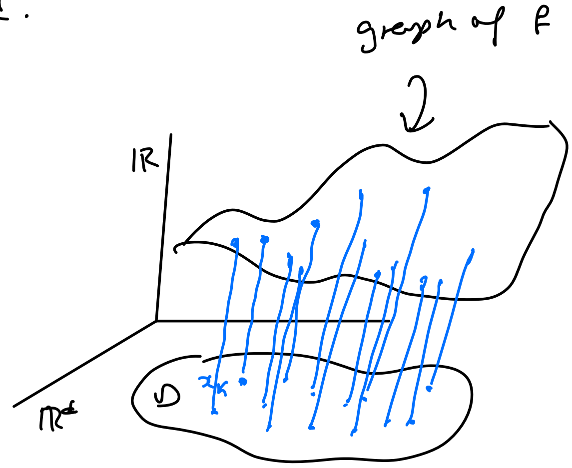
The determinant of the matrix on the right will be ± 1 (according to the # of -1s on the diagonal)

A word about QR decomposition and function approximation / parametric regression.

The context:

* We want to study / approximate a function $f(x)$ defined for $x \in D \subset \mathbb{R}^d$ for some dimension d.





* As data you are given a large but finite sample of values:

  $x_1, \ldots, x_m$ all in D,

  $y_1, \ldots, y_m$ respective values

i.e. you are told $y_k = f(x_k)$ for all k.

* Problem: Determine / estimate $f(x)$ for other values of x from this data, under the hypothesis that f can be

well approximated by a function of the form:

$$\tilde{f}(x) = c_1 f_1(x) + \cdots + c_n f_n(x)$$

where $f_1(x), \ldots, f_n(x)$ are a given family of linearly independent functions and $c_1, \ldots, c_n$ are to be determined.

* The most popular criterium to select the coefficients $c_1, \ldots, c_n$ is the mean square error (i.e. the Euclidean norm):

$$\text{minimize} \quad \sum_{i=1}^{m} \left(y_i - \sum_{j=1}^{n} c_j f(x_i) \right)^2$$

(In machine learning / statistical inference)

We see how parametric regression + with quadratic loss function is a least squares problem:

The matrix A is given by

$$A = \left(f_j(x_i) \right)_{ij} \qquad \begin{array}{ll} i = 1, \ldots, m & \text{(rows)} \\ j = 1, \ldots, n & \text{(column)} \end{array}$$

In all practical application m will be many orders of magnitude larger than n.

and $b = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix}$, so we are looking for a vector

$$c = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}$$

minimizing

$$\| Ac - b \|_2^2$$

The A encodes: the points $x_1, \ldots, x_m$
the functions $f_1, \ldots, f_n$

If we are in a situation where we will solve many problems with the same A but different values $y_1, \ldots, y_m$ then it becomes practical to compute the QR decomposition of A, and then the solution c, as we now know, will be given by solving

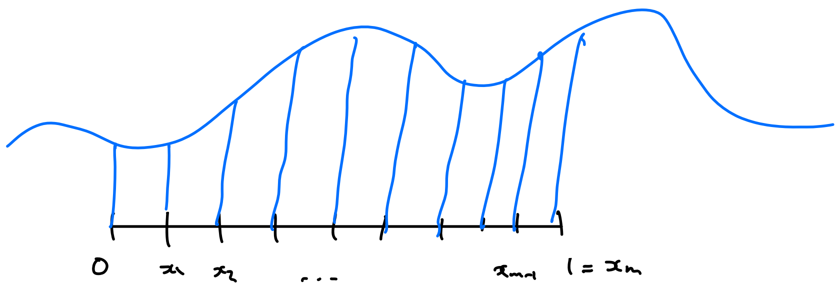
$$Rc = Q^t b$$

(when $A = QR$ is the reduced QR decomposition)

A special and not uncommon situation arises when you choose your functions $f_1, \dots, f_n$ to have some orthogonality from the start; in such a case the columns of A may be close to or exactly orthonormal, and computing the QR decomposition is "easy" or trivial.

Example : Fix m and let

$$x_k = \frac{k}{m} \qquad (k = 1, \dots, m)$$



For each $\ell = 1, \dots, m$, define

$$e_\ell(x) = \frac{1}{\sqrt{m}} e^{2\pi i \ell x}, \quad \text{for } x \in [0,1].$$

These functions are defined in $[0,1]$, but by restricting them to $\{x_k\}_{k=1}^{m}$ we get elements of $\mathbb{C}^m$.

Given two functions $f, \tilde{f} : \{x_k\}_{k=1}^{m} \longrightarrow \mathbb{C}$, their inner product is defined as

$$(f, \tilde{f}) = \sum_{k=1}^{m} f(x_k) \overline{\tilde{f}(x_k)}$$

Lemma: For every fixed m:

$$(e_\ell, e_{\ell'}) = \begin{cases} 1 & \text{if } \ell = \ell' \\ 0 & \text{if } \ell \neq \ell' \end{cases}$$

The proof of this discrete identity uses an important polynomial identity:

$$1 + z + z^2 + \cdots + z^m = \frac{z^{m+1} - 1}{z - 1}$$

Apply this with

$$z = e^{2\pi i(\ell - \ell')\frac{1}{m}}$$

$$1 + z + z^2 + \cdots + z^m = \sum_{k=0}^{m} e^{2\pi i(\ell-\ell')\frac{k}{m}}$$

$$= e^{2\pi i(\ell-\ell')\frac{1}{m}} \underbrace{\sum_{k=1}^{m-1} e^{2\pi i(\ell-\ell')\frac{k}{m}}}_{m(e_\ell, e_{\ell'})}$$

(by defin)

$$(k \in \mathbb{Z}) \quad e^{2\pi i k} = 1$$

Then,

$$m\, e^{2\pi i(\ell-\ell')\frac{1}{m}}(e_\ell, e_{\ell'}) = \frac{\overset{=1}{e^{2\pi i(\ell-\ell')\frac{m}{m}}} - 1}{e^{2\pi i(\ell-\ell')\frac{1}{m}} - 1}$$

since $\ell-\ell' \in \mathbb{Z}$, $\frac{m}{m} = 1$

(provided $\ell-\ell' \neq 0$) so

$$\ell \neq \ell' \implies (e_\ell, e_{\ell'}) = 0$$

Fr $\ell = \ell'$, $\quad (e_\ell, e_\ell) = \sum_{k=1}^{m} \frac{1}{\sqrt{m}} \cdot \frac{1}{\sqrt{m}} = 1$