§3.3. Least squares

Least squares soln to $ax=b$ is $\hat{x} = \frac{a^Tb}{a^Ta}$.

$e = x - \hat{x}a$ is $\perp$ to b.

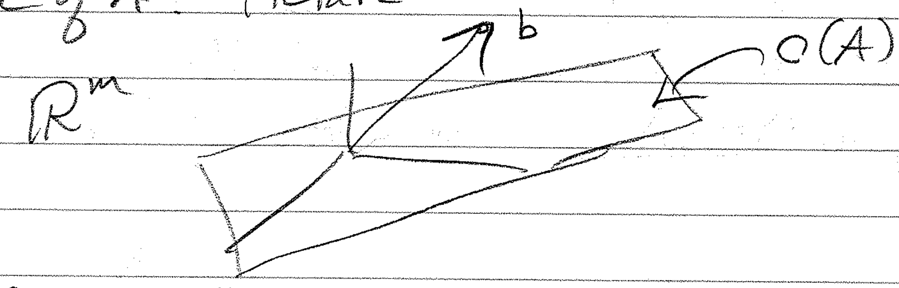
Let's do in several variables now.

Let A be an $m \times n$ matrix — this really is especially interesting when $m > n$, since

$$A : \mathbb{R}^n \to \mathbb{R}^m.$$

So the <u>range</u> of $A = \{y ; y = Ax \text{ for some } x\} =$ is a subspace of $\mathbb{R}^m$, and the range is at most n-dimensional.

We use the notation <u>column space</u> $C(A)$ for the range of A. Picture



So if $b \in \mathbb{R}^m$, the closest point in $C(A)$ is $A\hat{x}$ where $\hat{x}$ (there may be more than one) is chosen so that

$$\min_x \|Ax - b\| = \|A\hat{x} - b\|$$

Again,

$$e = (b - A\hat{x}) \text{ is } \perp C(A)$$

Formula

$$Pb = p = A(A^TA)^{-1}A^Tb.$$

(yesterday, when projecting on the line determined by a, we had

$$Pb = p = \frac{aa^T}{a^Ta} b.$$

Notice we write it differently here).

Let's review geometry.

a) If v is $\perp$ to $C(A)$ then v is in the left nullspace, or $N(A^T)$. So if

$$b = A\hat{x} \qquad (\hat{x} \in \mathbb{R}^n)$$

is best solution, then

$$A^T(b - A\hat{x}) = 0 \ ;$$

$$\boxed{A^T b = A^T A \hat{x}}$$

Then $\hat{x} = (A^T A)^{-1} A^T b$

$$p = A\hat{x} = A(A^T A)^{-1} A^T b$$

Note $(A^T A)^{-1}$ will usually exist even if A^{-1} does not. What we need is that the row space of A have dimension n, which is usually less than m; A is to have maximum rank (see p 163). In other words, if $Ax = 0$ only when $x = 0$, the

$$(A^T A)^{-1}$$

exists.

$A^T A$ is symmetric (and square). And if $A^T A x = 0$, then

$$x^T A^T A x = 0 \quad \text{or} \quad \|Ax\|^2 = 0 \ !$$

Properties of P: $\quad P = P^T, \ P^2 = P$

$$(P = A(A^T A)^{-1} A^T$$

Note that if P satisfies $P^2 = P$, $P = P^T$, then Pb is in the column space of P: if c is any vector, then

$$(b - Pb)^T Pc \qquad = 0$$

since

3.3-3

$$(b-Pb)^T Pc = b^T (I-P)^T Pc \quad \text{(symm.)}$$

$$= b^T \underbrace{(P - P^2)}_{=0} c = 0$$

So $b - Pb$ is $\perp$ to $\mathcal{C}(A)$

P 165 - Simplifies if A is invertible: $P = I$

Fitting data to a line

Suppose we have at times $t_1, \ldots, t_m$ data $b_1, \ldots, b_m$, and we want to choose C and D so that the line

$$Ct + D$$

is the best fit. This gives the matrix "equation"

$$\underset{A}{\overset{m\times 2}{}} \rightarrow \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ 1 & \vdots \\ 1 & t_m \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ \vdots \\ b_m \end{bmatrix} \; ; \quad Ax = b$$

If $m > 2$, we can only hope for best fit. So $x = \begin{bmatrix} C \\ D \end{bmatrix}$ and best value is to take

$$p = \underset{(m\times 2)}{A} \underset{(2\times 2)}{(A^TA)^{-1}} \underset{(2\times m)}{A^T} \underset{m\times 1}{b} \rightarrow 2\times 1 \text{ matrix}$$

P170 #6 $A = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ -2 & 4 \end{bmatrix}$ $\quad A^TA = \begin{bmatrix} 6 & -8 \\ -8 & 18 \end{bmatrix}$

So $p = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ -2 & 4 \end{bmatrix} \begin{bmatrix} 6 & -8 \\ -8 & 18 \end{bmatrix} \begin{bmatrix} 1 & 1 & -2 \\ 1 & -1 & 4 \end{bmatrix} \begin{pmatrix} 1 \\ 2 \\ 7 \end{pmatrix}$ $\qquad$ $b - p$ is in left nullspace

Orthogonal bases, Gram/Schmidt

3.4

Def of orthonormal. (ON) (Generalization of $e_1, e_2, \ldots, e_n$)
Can convert any basis to an ON basis. (Gram-Schmidt)

Def An orthogonal matrix Q is a square matrix with orthonormal columns (so they form a basis). Key property:

$$\boxed{Q^T = Q^{-1}}$$

(serves as a definition too)

Ex rotation, permutation matrix (think!)) Ex: rotation

Check: $\|Qx\| = \|x\|$ (for projection, $\|Px\| \le \|x\|$)

Why convenient? Let $q_1, \ldots, q_n$ be an ON basis. Then if x is a vector, $x = \sum c_j q_j$. And

$$c_j = (x, q_j)$$

(so it depends only on q_j, not the other basis vectors!

$$c_j = q_j^T q_j \left(= \frac{q_j^T q_j}{q_j^T q_j}\right) \quad \text{(Pythagoras)}$$

$$Q^TQ = I, \ QQ^T \text{ is not.}$$

Least squares (so now Q is $m \times n$ with $m > n$)

$$p = A\hat{x} = A(A^TA)^{-1}A^Tb,$$

we get

$$p = \underbrace{QQ^T}_{\text{Proj matrix}}b, \quad \underbrace{Q^Tb}_{\hat{x}} \text{ not}$$

G-S process: Let $b_1, \ldots, b_n$ be a basis.
Wlog, assume $|b_j| = 1$ for all j. Then let

$$q_1 = b_1,$$

(A low operation) Strang: leave q_2, q_3' etc.

$$q_2' = b_2 - (q_1, b_2)q_1, \quad q_2 = q_2'/|q_2'|,$$

$$q_3' = b_3 - (q_1, b_3)q_1 - (q_2, b_3)q_2; \quad q_3 = q_3'/|q_3'| \ldots$$

Factorization

$$\begin{array}{ccc} & \text{col vectors} & Q \quad R \text{ invertible} \end{array}$$

$$A = [a_1 \ a_2 \ a_3 \cdots a_n] = [q_1 \ q_2 \cdots q_n]\begin{bmatrix} q_1^Ta_1 & q_1^Ta_2 & \cdots \\ & q_2^Ta_2 & \cdots \\ & & q_n^Ta_n \end{bmatrix}$$

Earlier, $A = LU$, but here the columns of Q are orthogonal. Also, the determinant is the product of the diagonal entries of R. — orthog expansion —

$$a_j = (q_1^Ta_j)q_1 + (q_2^Ta_j)q_2 + \cdots + (q_j^Ta_j)q_j \; ;$$
(this is $Q \times$ (col j of R))

$$(q_j^Ta_j)q_j = a_j - (q_1^Ta_j)q_1 - \cdots - (q_{j-1}^Ta_j)q_{j-1}$$

Least squares simplifies a lot. We had

$$P = A(A^TA)^{-1}A^T$$

But now we have Q instead of A, Q is $m \times n$, $m > n$, so we had before

$$A^TA = R^TQ^TQR = R^TR$$

and $A^TA\hat{x} = b$ becomes

$$R^TR\hat{x} = A^Tb = R^TQ^Tb \quad \text{or}$$

$$R\hat{x} = Q^Tb$$

triangular so easier!

P 186 #9 q_1, q_2, q_3 ON, what linear combination of q_1, q_2 closest to q_3? : $0q_1 + 0q_2$

#12 $\left(\begin{pmatrix}4\\0\end{pmatrix} - c\begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}1\\1\end{pmatrix}\right) = 0$

$0 = 4 - c - c \cdot 1 = 4 - 2c$, so $c = 2$.

Check $\begin{pmatrix}2\\-2\end{pmatrix} \perp \begin{pmatrix}1\\1\end{pmatrix}$ ✓

3.4-3

3.4' Application: Fourier Series

Two infinite-dimensional vector spaces, one of which is of interest

1. Hilbert space: $(x_1, x_2, x_3, \ldots)$ (infinitely many) such that $\sum x_n^2 < \infty$

Ex $(1, \frac{1}{2}, \frac{1}{3}, \ldots)$

$$\|x\| = \left(\sum x_j^2\right)^{1/2}$$

$$x \cdot y = \sum x_j y_j$$

2. $L^2[a,b]$. Say $a = 0$, $b = 2\pi$

$$L^2[a,b] = f(x) : \int_a^b (f(x))^2 < \infty.$$

~~So~~ Take $a = 0$, $b = 2\pi$. Then any continuous function is in L^2. But $1/x$ or $1/(2\pi - x)$ is NOT in L^2. $(f,g) = \int f(x) g(x)\, dx$

$$\|f\| = \left(\int f(x)^2 dx\right)^{1/2} \quad \text{(like } R^n\text{!)}$$

Ex $\sin x$ is $\perp$ 1 and $\cos x$ on $(0, 2\pi)$.

$$\|\sin x\|^2 = \int_0^{2\pi} \sin^2 x\, dx = \int_0^{2\pi} \frac{1 - \cos 2x}{2}\, dx$$

$$= \pi$$

So $\|\sin x\| = \sqrt{\pi}$

$\sin mx \cos nx = \frac{1}{2}(\sin(m+n)x + \sin(m-n)x)$

Remark $\cos nx$, $\sin mx$ are all orthogonal.

$\cos mx \cos nx = \frac{1}{2}[\cos(m+n)x + \cos(m-n)x]$, $\sin mx \sin nx = \frac{1}{2}[\cos(m-n)x - \cos(m+n)x]$

3. Let's do polynomials, on $[-1, 1]$ (formulas better there)

Try $1, x, x^2$ (we could go on),

1 is $\perp$ x: $\int_{-1}^{1} 1 \cdot x\, dx = 0$

1 is NOT $\perp$ x^2: $\int_{-1}^{1} 1 \cdot x^2 dx > 0$ $(= \frac{2}{3})$

So we have

$$A^T A = \begin{bmatrix} (1,1) & (1,x) & (1,x^2) \\ (x,1) & (x,x) & (x,x^2) \\ (x^2,1) & (x^2,x) & (x^2,x^2) \end{bmatrix}$$

So 1, and x are $\perp$

Now for x^2 replace by

$$q_3 = x^2 - \frac{(x^2,1)}{(1,1)} \cdot 1 - \frac{(x^2,x)}{(x,x)} x$$

Check that $q_3 \perp 1, x$.

Study Item 5 p 185. (for hw 21, 25

P 187 #22. Find best ftn $b \sin x$ to approximate $\cos x$. Ans has to be zero (explain), but we see from calculus!

$$\int_0^{2\pi} (b \sin x - \cos x)^2 dx$$

$$= b^2 \int_0^{2\pi} \sin^2 x \, dx + 2b \int \sin x \cos x \, dx + \int \cos^2 x \, dx$$

$$= \pi b^2 - 0 + \pi \quad \text{Min when } b = 0.$$

#25 Min

$$F = \int (x^2 - (C + Dx))^2 dx \quad (\text{fn of } C, D).$$

MA 261 - set $\partial F / \partial C = \partial F / \partial D = 0$

End of Ch 3

Ch 4 Determinant (of a square matrix)

I consider this review.

Always awkward to introduce — We follow book.

3 Key properties — it is a # assigned to a square matrix

$|A|$ or $\det(A)$

(1) $\det I = 1$

(2) det changes sign with row exchange

(3) det is linear w.r.t. 1st row (2nd row too)

Ex Two equal rows gives $|A| = 0$; row of 0's also

Ex Add a multiple of one row to another doesn't change $|A|$.

$(a \neq 0)$ Ex $\begin{vmatrix} a & b \\ c & d \end{vmatrix} = \begin{bmatrix} a & b \\ c - \frac{c}{a}a & d - \frac{c}{a}b \end{bmatrix} = \begin{bmatrix} a & \\ 0 & d - \frac{c}{a}b \end{bmatrix}$

$\rightarrow \begin{bmatrix} a & 0 \\ c & b - \frac{c}{a} \end{bmatrix} \rightarrow a \begin{bmatrix} 1 & 0 \\ 0 & d - \frac{c}{a} \end{bmatrix}$

$\rightarrow a\left(d - \frac{cb}{a}\right) \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$!

Ex $|A| = 0 \Leftrightarrow A$ is sing; if A is $\triangle$ det A is product of diagonal elements. This shows $|A| = |A^T|$

Key nontrivial fact (property 9):
$|AB| = |A||B|$

Proof (1) p 205 is too slick! We'll go over both