MA-106 Linear Algebra

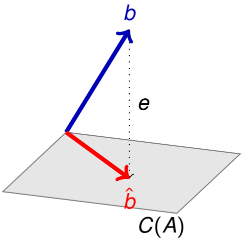
H. Ananthnarayan

Department of Mathematics
Indian Institute of Technology Bombay
Powai, Mumbai - 76

22nd February 2018
D1 - Lecture 22

Linear Least Squares and Projections

Suppose system $Ax = b$ is inconsistent, i.e. $b \notin C(A)$.
The error $E = \|Ax - b\|$ is the distance from b to $Ax \in C(A)$.



We want the least square solution $\hat{x}$ of $Ax = b$, which minimizes E,
i.e., we want to find $\hat{b}$ closest to b such that $A\hat{x} = \hat{b}$ is a consistent system.
Therefore, $\hat{b} = \text{proj}_{C(A)}(b)$ and $A\hat{x} = \hat{b}$.

The error vector $e = b - A\hat{x}$ must be perpendicular to $C(A)$.
So $e \in C(A)^{\perp}$ = left null space of A, $N(A^T)$,
i.e.,
$$A^T(b - A\hat{x}) = 0 \text{ or } \boxed{A^T A\,\hat{x} = A^T b}$$

Therefore, to find $\hat{x}$, we need to solve $A^T A\,\hat{x} = A^T b$.

Linear Least Squares and Projections

Let A be $m \times n$. Then $A^T A$ is a symmetric $n \times n$ matrix.

• $\boxed{N(A^T A) = N(A)}$.

Proof. $Ax = 0 \Rightarrow A^T Ax = 0$. So, $N(A) \subset N(A^T A)$.
For the converse, take $x \in N(A^T A)$. $A^T Ax = 0 \Rightarrow$

$$x^T(A^T Ax) = (Ax)^T(Ax) = \|Ax\|^2 = 0$$
$$\Rightarrow Ax = 0 \quad \Rightarrow x \in N(A).$$

• Since $N(A) = N(A^T A)$, by rank-nullity theorem, rank$(A) = n - \dim\,(N(A)) =$ rank$(A^T A)$.

• A has linearly independent columns $\Leftrightarrow$ rank$(A) = n$ $\Leftrightarrow$ rank$(A^T A) = n \Leftrightarrow A^T A$ is invertible.

• If rank$(A) = n$, then the least square solution of $Ax = b$ is given by $A^T A\,\hat{x} = A^T b \Rightarrow \boxed{\hat{x} = (A^T A)^{-1} A^T b}$ and

the orthogonal projection of b on $C(A)$ is $\boxed{\hat{b} = A\hat{x} = Pb}$, where $\boxed{P = A(A^T A)^{-1} A^T}$ is the projection matrix. **Q** Is $P^2 = P$?

Linear Least Squares: Example

Example: Find the least square solution to the system

$$\begin{pmatrix} -1 & 2 \\ 2 & -3 \\ -1 & 3 \end{pmatrix} x = \begin{pmatrix} 4 \\ 1 \\ 2 \end{pmatrix} \quad (Ax = b)$$

We need to solve $A^T A\,\hat{x} = A^T b$. Now $A^T b = \begin{pmatrix} -4 \\ 11 \end{pmatrix}$ and

$$A^T A = \begin{pmatrix} -1 & 2 & -1 \\ 2 & -3 & 3 \end{pmatrix} \begin{pmatrix} -1 & 2 \\ 2 & -3 \\ -1 & 3 \end{pmatrix} = \begin{pmatrix} 6 & -11 \\ -11 & 22 \end{pmatrix}.$$

$$[A^T A \mid A^T b] = \left(\begin{array}{cc|c} 6 & -11 & -4 \\ -11 & 22 & 11 \end{array}\right) \to \left(\begin{array}{cc|c} 6 & -11 & -4 \\ 0 & 11/6 & 11/3 \end{array}\right)$$

Therefore $\hat{x}_2 = 2$, and $\hat{x}_1 = 3$.

Ex: Find the projection matrix P, and check that $Pb = A\hat{x}$.

Exercise: Line of Best Fit

Q: We want to find the best line $y = C + Dx$ which fits the given data and gives least square error.
Data: (x, y) = $(-2, 4)$, $(-1, 3)$, $(0, 1)$, and $(2, 0)$.

The system $\begin{pmatrix} 1 & -2 \\ 1 & -1 \\ 1 & 0 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} C \\ D \end{pmatrix} = \begin{pmatrix} 4 \\ 3 \\ 1 \\ 0 \end{pmatrix} \quad (Ax = b)$
is inconsistent.

Find the least square solution by solving $A^T A \hat{x} = A^T b$.

Q: Find the best quadratic curve $y = C + Dx + Ex^2$ which fits the above data and gives least square error.

Hint. The first row of the matrix A in this case will be $[1 \quad -2 \quad 4]$.

Extra Reading: Inner Product spaces : Definition

We can define inner product spaces in more generality.
An **inner product** on a vector space V is which gives a real number $\langle u\,,v\rangle$ for every pair $u,v\in V$ such that it satisfies the following axioms for all u,v,w in V and c in $\mathbb{R}$. (or $\mathbb{C}$).

- $\langle u\,,v\rangle = \langle u\,,v\rangle$
- $\langle u\,,u\rangle \geq 0$ and $\langle u\,,u\rangle = 0 \iff u = 0$.
- $\langle u+v\,,w\rangle = \langle u\,,w\rangle + \langle v\,,w\rangle$
- $\langle cu\,,v\rangle = c\langle u\,,v\rangle = \langle u\,,cv\rangle$.

This helps us define length of a vector in such spaces
Define **length(norm)** of v as $\|v\| = \sqrt{\langle v\,,v\rangle}$.

Extra Reading: Inner Product spaces : Properties

Let V be a vector space with an inner product.

Then it satisfies the Cauchy-Schwartz inequality, that is, for all $u, v \in V$,

$$|\langle u\,, v\rangle| \leq \|u\|\|v\|,$$

and the Pythagores theorem, that is, for $u, v \in V$,

$$\langle u\,, v\rangle = 0 \iff \|u - v\|^2 = \|u\|^2 + \|v\|^2.$$

We define two vectors $u, v \in V$ to be **orthogonal** to each other if $\langle u\,, v\rangle = 0$.

Why do we want to talk about these generalities?

Extra Reading : Inner Product spaces : Example

Consider the space $C[0, 2\pi]$ of all continuous functions on $[0, 2\pi]$. We already saw that $C[0, 2\pi]$ is a vector space. We now define an inner product on this space as follows; f, $g \in C[0, 2\pi]$

$$\langle f, g \rangle = \int_0^{2\pi} f(t)g(t)\, dt.$$

This satisfies all the axioms of an inner product.
Some of the well understood functions are the cosine and sine functions on $[0, 2\pi]$ and it is often useful to write periodic functions in terms of sines and cosines.

This idea is a simple application of taking orthogonal projections in the vector space $C[0, 2\pi]$ onto the subspace
$T = \mathrm{Span}\{\cos nx, \sin mx \mid m, n \in \mathbb{Z}, m, n \geq 0\}$.

In order to do take orthogonal projections we need an orthogonal basis of T.

Extra Reading: Fourier Series Expansion

The set $S = \{\cos nx, \sin mx \mid m, n \geq 0, m, n \in \mathbb{Z}\}$ is orthogonal Why?

For example, $\langle \cos x\,, \sin x \rangle = \int_0^{2\pi} \cos t \sin t \; dt = \frac{1}{2} \int_0^{2\pi} \sin 2t \; dt = 0.$

$\langle \cos x\,, \cos x \rangle = \int_0^{2\pi} \cos^2 t \; dt = \frac{1}{2} \int_0^{2\pi} (1 - \cos 2t) \; dt = 2\pi.$

Then S is an orthogonal basis of T. Let $f \in C[0, 2\pi]$. Then

$$\mathrm{proj}_T f(x) = a_0 + a_1 \cos x + b_1 \sin x + \ldots + a_k \cos kx + b_k \sin kx + \ldots,$$

where $a_0 = \dfrac{\langle 1\,, f \rangle}{2\pi}$, $a_k = \dfrac{\langle \cos kx\,, f \rangle}{2\pi}$, $b_k = \dfrac{\langle \sin kx\,, f \rangle}{2\pi}$.

This is exactly the **Fourier series expansion of** f.

Diagonalizing Symmetric matrices :Example

Ex: Let $A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$. Then $A - \lambda I = \begin{bmatrix} 1-\lambda & 1 & 1 \\ 1 & 1-\lambda & 1 \\ 1 & 1 & 1-\lambda \end{bmatrix}$ and

$$\det(A - \lambda I) = \\ (1-\lambda)[(1-\lambda)^2 - 1] - 1[1 - \lambda - 1] + 1[1 - (1-\lambda)] \\ = (3-\lambda)\lambda^2 \qquad \text{Eigenvalues: } \lambda_1 = 3, \lambda_2 = 0, \lambda_3 = 0.$$

To find $N(A - 3I)$, solve $A - 3I = \begin{bmatrix} -2 & 1 & 1 \\ 1 & -2 & 1 \\ 1 & 1 & -2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$.

$N(A)$ is the plane $x + y + z = 0$. Hence, the associated eigenvectors are $v_1 = (1, 1, 1)^T$, $v_2 = (-1, 0, 1)^T$ and $v_3 = (0, -1, 1)^T$.

Example: $A = Q\Lambda Q^T$

A has eigenvalues $\lambda_1 = 3, \lambda_2 = 0, \lambda_3 = 0$ with associated eigenvectors $v_1 = (1, 1, 1)^T$, $v_2 = (-1, 0, 1)^T$ and $v_3 = (0, -1, 1)^T$. Note that v_2 and v_3 are linearly independent in $N(A)$. Observe $v_1^T v_2 = 0 = v_1^T v_3$.

How do we get an orthogonal Q such that $A = Q\Lambda Q^T$, where Λ is diagonal with entries 3, 0, 0 on the diagonal?

Steps: 1. Let $u_1 = v_1/\|v_1\|$.

2. Start with the basis $\{v_2, v_3\}$ of $N(A)$, and apply the Gram-Schimdt process to get an orthonormal basis $\{u_2, u_3\}$ for $N(A)$,

Note that u_2 and u_3 are eigenvectors of A associated to $\lambda = 0$, and are linearly independent since they are non-zero orthogonal vectors.

3. Then $Q = [u_1 \;\; u_2 \;\; u_3]$ is orthogonal, and $Q^{-1}AQ = \Lambda$.

4. Since $Q^{-1} = Q^T$, $A = Q\Lambda Q^T$.

Real Symmetric Matrices: Properties

• Let A be a symmetric matrix. Then all its eigenvalues are real.

Let $\lambda \in \mathbb{C}$ and $v \in \mathbb{C}^n$ be eigenvalue and eigenvector of A.
Then $Av = \lambda v \Rightarrow (\bar{A}\bar{v})^T = \bar{\lambda}\bar{v}^T$.

$$\text{Now } \bar{v}^T A v = \lambda \bar{v}^T v$$

Alternately, $A = A^T$ and $\bar{A}^T = A^T$ implies

$$\bar{v}^T A v = \bar{v}^T \bar{A}^T v = \bar{\lambda}\bar{v}^T v.$$

Since $\bar{v}^T v > 0$ for a v non-zero, we have,

$$\lambda = \bar{\lambda} \Rightarrow \lambda \text{ is real.}$$

Real Symmetric Matrices: Properties

• λ has an associated eigenvector with real entries.
If $x \notin \mathbb{R}^n$, write $x = u + iv$, where u, $v \in \mathbb{R}^n$.Then $\lambda \in \mathbb{R} \Rightarrow u$ and v are also eigenvectors associated to λ.

• Let A be a symmetric $n \times n$ matrix with real entries, λ_1 and λ_2 are distinct eigenvalues of A, with associated eigenvectors v_1 and $v_2 \in \mathbb{R}^n$. Then v_1 and v_2 are orthogonal.

Want to prove: $v_1^T v_2 = 0$. Now $\lambda_1(v_1^T v_2) = (\lambda_1 v_1)^T v_2$
$= (Av_1)^T v_2 = (v_1^T A^T)v_2 = v_1^T(Av_2) = v_1^T(\lambda_2 v_2) = \lambda_2(v_1^T v_2)$.
Since $\lambda_1 \neq \lambda_2$, $v_1^T v_2 = 0$.

Real Symmetric Matrices

• If A is a real symmetric matrix with n distinct eigenvalues, then there is an orthogonal matrix Q and a diagonal matrix Λ such that $A = Q\Lambda Q^T$.

If A has n distinct eigenvalues, then A is diagonalizable, i.e., there is an invertible P and a diagonal Λ such that $P^{-1}AP = \Lambda$.

The matrix P has eigenvectors of A as its columns. Choose $v_1, \ldots, v_n \in \mathbb{R}^n$, respective eigenvectors associated to the n distinct eigenvalues.

Let $Q = \left[\frac{v_1}{\|v_1\|} \ \cdots \ \frac{v_n}{\|v_n\|} \right]$. Then Q is orthogonal, i.e., $Q^{-1} = Q^T$ and $Q^{-1}AQ = \Lambda \Rightarrow A = Q\Lambda Q^T$.

Spectral Theorem for a Real Symmetric Matrix

Theorem Every real symmetric matrix A can be diagonalised. In particular, there is an orthogonal matrix Q and a diagonal matrix Λ such that $A = Q\Lambda Q^T$.

Given that A can be diagonalised, then $A = Q\Lambda Q^T$ follows from the previous slide (+ Gram-Schmidt process applied to each eigenspace of A).

Extra Reading - Application: SVD

Singular Value Decomposition: Given an $m \times n$ matrix A, there exists an $m \times n$ "diagonal" matrix Σ, orthogonal matrices U ($m \times m$) and V ($n \times n$), such that $A = U\Sigma V^T$.

Note: If U, V and Σ are as above, they satisfy:
$AA^T = (U\Sigma V^T)(V\Sigma^T U^T) = U\Sigma^2 U^T$, and $A^T A = V\Sigma^2 V^T$. Thus:
1. The non-zero diagonal entries in Σ, called the singular values of A, are the square roots of the common eigenvalues of AA^T and $A^T A$, .
2. Columns of U are eigenvectors of AA^T, and those of V are eigenvectors of $A^T A$.
3. If the first r columns of Σ are non-zero, then $\{U_{*1}, \ldots, U_{*r}\}$ and $\{V_{*1}, \ldots, V_{*r}\}$ are orthonormal bases for the column space of A, $C(A)$ and the row space of A, $C(A^T)$, respectively.
Furthermore, $AV = U\Sigma \Rightarrow AV_{*j} = \sigma_j U_{*j}$ for each $j \leq r$.

Importance: Used in image compression, e.g.,
If an image is represented by A, find U, Σ, V as in SVD. Replace "small" singular values in Σ by 0 to $\widetilde{\Sigma}$. Then $\widetilde{A} = U\widetilde{\Sigma} V^T$ represents the compressed image.