MA232 Linear Algebra

Alex Myasnikov

Stevens Institute of Technology

October 11, 2011

Orthogonal Vectors.

Two vectors $\bar{\mathbf{v}}$ and $\bar{\mathbf{w}}$ are orthogonal if $\bar{\mathbf{v}} \cdot \bar{\mathbf{w}} = 0$ or $\bar{\mathbf{v}}^T \bar{\mathbf{w}} = 0$

If $\bar{\mathbf{v}}$ and $\bar{\mathbf{w}}$ are orthogonal then

$$||\bar{\mathbf{v}}||^2 + ||\bar{\mathbf{w}}||^2 = ||\bar{\mathbf{v}} + \bar{\mathbf{w}}||^2$$

Orthogonal Subspaces.

A set $S \subseteq V$ is orthogonal if each pair $\bar{\mathbf{v}}, \bar{\mathbf{w}} \in S$ are orthogonal: $\bar{\mathbf{v}} \cdot \bar{\mathbf{w}} = 0$

Two subspaces V and W are orthogonal if every vector in V is orthogonal to every vector in W:

$$\bar{\mathbf{v}}^T \bar{\mathbf{w}} = 0, \text{ for all } \bar{\mathbf{v}} \in V, \bar{\mathbf{w}} \in W.$$

If S is an orthogonal set of nonzero vectors, then S is linearly independent

Orthogonal Subspaces.

Let $\bar{\mathbf{u}}_1, \dots \bar{\mathbf{u}}_n$ be an orthogonal basis of V, then for any $\bar{\mathbf{v}} \in V$

$$\bar{\mathbf{v}} = \frac{\bar{\mathbf{v}} \cdot \bar{\mathbf{u}}_1}{\bar{\mathbf{u}}_1 \cdot \bar{\mathbf{u}}_1}\bar{\mathbf{u}}_1 + \frac{\bar{\mathbf{v}} \cdot \bar{\mathbf{u}}_2}{\bar{\mathbf{u}}_2 \cdot \bar{\mathbf{u}}_2}\bar{\mathbf{u}}_2 + \cdots + \frac{\bar{\mathbf{v}} \cdot \bar{\mathbf{u}}_n}{\bar{\mathbf{u}}_n \cdot \bar{\mathbf{u}}_n}\bar{\mathbf{u}}_n$$

Orthogonal Subspaces.

The nullspace $N(A)$ and the row space of A are orthogonal subspace.

From equations:

- $\bar{\mathbf{x}} \in N(A) \Rightarrow A\bar{\mathbf{x}} = \bar{\mathbf{0}} \Rightarrow A_{row}(i) \cdots \bar{\mathbf{x}} = 0$ for all i
- $\bar{\mathbf{y}} \in C(A^T)$ - row space then
  $\bar{\mathbf{y}} = c_1 A_{row}(1) + c_2 A_{row}(2) + \cdots + c_n A_{row}(n)$
- Multiplying both sides by $\bar{\mathbf{x}}$ we have:

$$\bar{\mathbf{y}} \cdot \bar{\mathbf{x}} = c_1 A_{row}(1) \cdot \bar{\mathbf{x}} + c_2 A_{row}(2) \cdot \bar{\mathbf{x}} + \cdots + c_n A_{row}(n) \cdot \bar{\mathbf{x}} = 0$$

Orthogonal Subspaces.

The nullspace $N(A)$ and the row space of A are orthogonal subspace.

Another view:

- $\bar{\mathbf{x}} \in N(A) \Rightarrow A\bar{\mathbf{x}} = \bar{\mathbf{0}}$ and $\bar{\mathbf{y}} \in C(A^T) \Rightarrow \bar{\mathbf{y}} = A^T\bar{\mathbf{v}}$
- $\bar{\mathbf{x}}^T\bar{\mathbf{y}} = \bar{\mathbf{x}}^T(A^T\bar{\mathbf{v}}) = (A\bar{\mathbf{x}})^T\bar{\mathbf{v}} = \bar{\mathbf{0}}\bar{\mathbf{v}} = 0$

Orthogonal Subspaces.

The left nullspace $N(A^T)$ and the column space $C(A)$ are orthogonal subspace.

Another view:

- $\bar{\mathbf{x}} \in N(A^T) \Rightarrow A^T\bar{\mathbf{x}} = \bar{\mathbf{0}}$ and $\bar{\mathbf{y}} \in C(A) \Rightarrow \bar{\mathbf{y}} = A\bar{\mathbf{v}}$
- $\bar{\mathbf{x}}^T\bar{\mathbf{y}} = \bar{\mathbf{x}}^T(A\bar{\mathbf{v}}) = (A^T\bar{\mathbf{x}})^T\bar{\mathbf{v}} = \bar{\mathbf{0}}\bar{\mathbf{v}} = 0$

Orthogonal Complement.

Orthogonal complement $W^{\perp}$ of a subspace $W \subseteq V$ of vector space V contains every vector that is perpendicular to W:

$$W^{\perp} = \{\bar{\mathbf{v}} \in V \mid \bar{\mathbf{v}} \cdot \bar{\mathbf{w}} = 0 \text{ for every } \bar{\mathbf{w}} \in W\}$$

If W is a subspace then $W^{\perp}$ is a subspace.

Orthogonal Complement.

- Nullspace is the orthogonal complement of the row space: $N(A)^{\perp} = C(A^T)$
- The left nullspace is the orthogonal complement of the column space: $N(A^T)^{\perp} = C(A)$

Orthogonal Complement.

- If V and W are orthogonal then any $\bar{\mathbf{v}} \in V$ and $\bar{\mathbf{w}} \in W$ are linearly independent (except for $\bar{\mathbf{0}}$)
- Let A be of rank r, B_N be the basis of $N(A)$ and B_R be the basis of $C(A^T)$.
- We know that $|B_N| = n - r$ and $|B_R| = r$
- Since $N(A)$ and $C(A^T)$ are orthogonal then all vectors in B_N are linearly independent of vectors in B_R, therefore

$$|B_N \cup B_R| = n - r + r = n \;\Rightarrow\; span(B_N \cup B_R) = \mathbb{R}^n$$

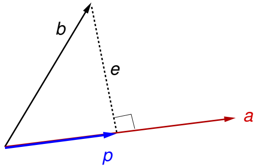
Every $\bar{\mathbf{x}} \in \mathbb{R}^n$ is a sum of row space vector and a nullspace vector:
$\bar{\mathbf{x}} = \bar{\mathbf{x}}_r + \bar{\mathbf{x}}_n$

Projections.

Intuitively: projection of $\bar{\mathbf{b}}$ onto a subspace V is a vector in V which is closest to $\bar{\mathbf{b}}$.

Projection onto a line

1. $\bar{\mathbf{p}} = c\bar{\mathbf{a}}$
2. $\bar{\mathbf{e}} \perp \bar{\mathbf{a}}$ and $\bar{\mathbf{e}} = \bar{\mathbf{b}} - \bar{\mathbf{p}}$
3. Combining: $\bar{\mathbf{a}} \cdot (\bar{\mathbf{b}} - c\bar{\mathbf{a}}) = 0$



- Solve for c:

$$\bar{\mathbf{a}} \cdot (\bar{\mathbf{b}} - c\bar{\mathbf{a}}) = 0 \;\Rightarrow\; \bar{\mathbf{a}} \cdot \bar{\mathbf{b}} - c\bar{\mathbf{a}}\bar{\mathbf{a}} = 0 \;\Rightarrow\; c = \frac{\bar{\mathbf{a}} \cdot \bar{\mathbf{b}}}{\bar{\mathbf{a}} \cdot \bar{\mathbf{a}}} = \frac{\bar{\mathbf{a}}^T \bar{\mathbf{b}}}{\bar{\mathbf{a}}^T \bar{\mathbf{a}}}$$

Projection onto a line.

So we have the projection of $\bar{\mathbf{b}}$ onto the line through $\bar{\mathbf{a}}$:

$$\bar{\mathbf{p}} = c\bar{\mathbf{a}} = \frac{\bar{\mathbf{a}}^T\bar{\mathbf{b}}}{\bar{\mathbf{a}}^T\bar{\mathbf{a}}}\bar{\mathbf{a}}$$

Compare to linear combination of vectors in orthogonal basis!

Projection onto a line.

Projection of $\bar{\mathbf{b}}$ onto the line through $\bar{\mathbf{a}}$:

$$\bar{\mathbf{p}} = \frac{\bar{\mathbf{a}}^T \bar{\mathbf{b}}}{\bar{\mathbf{a}}^T \bar{\mathbf{a}}} \bar{\mathbf{a}}$$

Denote matrix

$$P = \frac{\bar{\mathbf{a}} \bar{\mathbf{a}}^T}{\bar{\mathbf{a}}^T \bar{\mathbf{a}}}$$

Then

$$\bar{\mathbf{p}} = P \bar{\mathbf{b}}$$

P is a projection matrix onto the line through $\bar{\mathbf{a}}$.

Projections.

If P is a projection matrix then

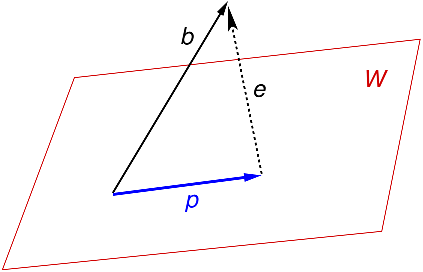
- $P^2 = P$
- $(I - P)$ is a projection onto a perpendicular subspace.

Projection onto a subspace.

Let W be a subspace with basis $\bar{\mathbf{a}}_1, \ldots, \bar{\mathbf{a}}_n$. Projection $\bar{\mathbf{p}}$ of $\bar{\mathbf{b}} \in \mathbb{R}^M$ onto W is a vector

$$\bar{\mathbf{p}} = c_1\bar{\mathbf{a}}_1 + c_2\bar{\mathbf{a}}_2 + \cdots + c_n\bar{\mathbf{a}}_n$$

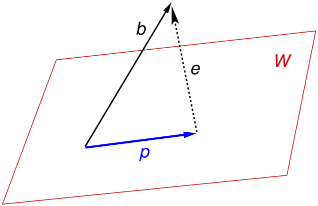
closets to $\bar{\mathbf{b}}$.



Projection onto a subspace.

Let A be the matrix with columns $\bar{\mathbf{a}}_1, \ldots, \bar{\mathbf{a}}_n$ and $\bar{\mathbf{c}} = [c_1 \ c_2 \ \ldots \ c_n]^T$

1. $\bar{\mathbf{p}} = A\bar{\mathbf{c}}$
2. $\bar{\mathbf{e}} \perp \bar{\mathbf{a}}_i, i = 1, \ldots, n$ and $\bar{\mathbf{e}} = \bar{\mathbf{b}} - A\bar{\mathbf{c}}$



Combining:

$$\left.\begin{aligned} \bar{\mathbf{a}}_1^T(\bar{\mathbf{b}} - A\bar{\mathbf{c}}) &= 0 \\ \cdots \\ \bar{\mathbf{a}}_n^T(\bar{\mathbf{b}} - A\bar{\mathbf{c}}) &= 0 \end{aligned}\right. \Rightarrow A^T(\bar{\mathbf{b}} - A\bar{\mathbf{c}}) = \bar{\mathbf{0}} \Rightarrow A^T A\bar{\mathbf{c}} = A^T\bar{\mathbf{b}}$$

Projection onto a subspace.

$A^T A\bar{\mathbf{c}} = A^T\bar{\mathbf{b}}$: To solve for $\bar{\mathbf{c}}$ need inverse of $A^T A$

Note: $A^T A$ is a square symmetric matrix

Let A be $m \times n$ matrix. $A^T A$ is invertible if and only if A has linearly independent columns

- $N(A^T A) = N(A)$:
  1. Let $\bar{\mathbf{x}} \in N(A)$: $A\bar{\mathbf{x}} = \bar{\mathbf{0}} \Rightarrow A^T A\bar{\mathbf{x}} = \bar{\mathbf{0}}$
  2. Let $\bar{\mathbf{x}} \in N(A^T A)$: $A^T A\bar{\mathbf{x}} = \bar{\mathbf{0}} \Rightarrow (\bar{\mathbf{x}}^T)A^T A\bar{\mathbf{x}} = 0 \Rightarrow$
     $(A\bar{\mathbf{x}})^T(A\bar{\mathbf{x}}) = 0 \Rightarrow ||A\bar{\mathbf{x}}||^2 = 0 \Rightarrow$
     $A\bar{\mathbf{x}} = \bar{\mathbf{0}}$
- If A has independent columns then

$$N(A) = \{\bar{\mathbf{0}}\} = N(A^T A) \Rightarrow A^T A \text{ is invertible}$$

Projection onto a subspace.

$A^T A\bar{\mathbf{c}} = A^T\bar{\mathbf{b}}$:

- $A^T A$ is a symmetric $n \times n$ matrix
- $A^T A$ is invertible because $\bar{\mathbf{a}}_i$ are linearly independent
- $\bar{\mathbf{c}} = (A^T A)^{-1} A^T \bar{\mathbf{b}}$

The projection of $\bar{\mathbf{b}} \in \mathbb{R}^M$ onto subspace W

$$\bar{\mathbf{p}} = A\bar{\mathbf{c}} = A(A^T A)^{-1} A^T \bar{\mathbf{b}}$$

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

$$\bar{\mathbf{p}} = P\bar{\mathbf{b}}$$

Projection onto a subspace.

- A is $m \times n$ matrix so we cannot use A^{-1} in general
- If A^{-1} exists then $m = n$ and there are m independent columns, therefore, $W = \mathbb{R}^M$
- Projection of $\bar{\mathbf{b}} \in \mathbb{R}^M$ onto $\mathbb{R}^m$ is the vector $\bar{\mathbf{b}}$ itself:

$$\bar{\mathbf{p}} = A(A^T A)^{-1} A^T \bar{\mathbf{b}} = AA^{-1}(A^T)^{-1} A^T \bar{\mathbf{b}} = \bar{\mathbf{b}}$$

And projection matrix

$$P = I$$

Least squares approximation.

Let A be $m \times n$ and $m >> n$ - number of rows significantly greater then number of columns

- Solution to $A\bar{\mathbf{x}} = \bar{\mathbf{b}}$ exists if $\bar{\mathbf{b}} \in C(A)$
- Note that $\bar{\mathbf{b}} \in \mathbb{R}^m$ and $dim(C(A))$ at most n
- There are many vectors in $\mathbb{R}^m$ which are not in $C(A)$

It is quite possible in practice to have systems with no solutions

Least squares approximation.

System of linear equations as optimization problem:

Given a system $A\bar{\mathbf{x}} = \bar{\mathbf{b}}$ we would like to find the solution $\bar{\mathbf{x}}$ such that the error

$$||\bar{\mathbf{e}}|| = ||A\bar{\mathbf{x}} - \bar{\mathbf{b}}||$$

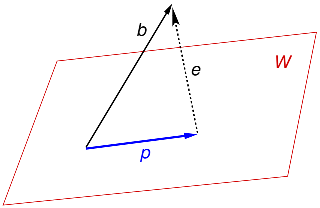
is minimal.

- I.e. the distance from vector obtained using our solution to $\bar{\mathbf{b}}$ is the minimal possible.
- This is equivalent to minimizing $||e||^2 = ||A\bar{\mathbf{x}} - \bar{\mathbf{b}}||^2 = (A\bar{\mathbf{x}} - \bar{\mathbf{b}})^T(A\bar{\mathbf{x}} - \bar{\mathbf{b}}) = \sum_{i=1}^{m}(A\bar{\mathbf{x}} - \bar{\mathbf{b}})_i^2$ therefore method of least squares

Least squares approximation.

Minimize $||\bar{\mathbf{e}}|| = ||A\bar{\mathbf{x}} - \bar{\mathbf{b}}||$:

1 W spanned by columns of A
2 Any vector $\bar{\mathbf{y}}$ for which solution to $A\bar{\mathbf{x}} = \bar{\mathbf{y}}$ exists is in W
3 The closest vector to $\bar{\mathbf{b}}$ in W is the projection $\bar{\mathbf{p}}$



Least squares approximation.

The least squares solution to $A\bar{\mathbf{x}} = \bar{\mathbf{b}}$ is the solution to

$$A\bar{\mathbf{x}} = \bar{\mathbf{p}} = A(A^TA)^{-1}A^T\bar{\mathbf{b}} \Rightarrow \bar{\mathbf{x}} = (A^TA)^{-1}A^T\bar{\mathbf{b}}$$

- To obtain least squares solution to $A\bar{\mathbf{x}} = \bar{\mathbf{b}}$ we solve

$$A^TA\bar{\mathbf{x}} = A^T\bar{\mathbf{b}}$$

- A must have independent columns - important requirement in practical applications

Least squares approximation: curve fitting.

Fitting a straight line:

- Suppose we are given n points:

$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$$

- Out goal is to find an equation of a line closest to the n points

Least squares approximation: curve fitting.

Suppose our line has equation $\beta + \alpha x$ our goal is to find unknowns α, β.

- In the best case scenario all points will lie alone a line
- In this case the best fitted line will pass through all points:

$$\begin{aligned} \beta + \alpha x_1 &= y_1 \\ \beta + \alpha x_2 &= y_2 \\ &\dots \\ \beta + \alpha x_n &= y_n \end{aligned}$$

- α and β are found by solving this system of linear equations (in fact wee need just two)

Least squares approximation: curve fitting.

- In the general points will not lie on the same line
- The system does not have a solution
- The idea is to use least squares to find the parameters of the best fit

Least squares approximation: curve fitting.

Denote

$$A = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \ \bar{\mathbf{y}} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}; \ \bar{\mathbf{d}} = \begin{bmatrix} \beta \\ \alpha \end{bmatrix}$$

- The system for least squares:

$$A\bar{\mathbf{d}} = \bar{\mathbf{y}}$$

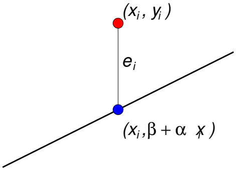
- We find $\bar{\mathbf{d}} = [\beta \ \alpha]^T$ by solving

$$A^T A\bar{\mathbf{d}} = A^T \bar{\mathbf{y}}$$

Least squares approximation: curve fitting.

What is the error $||A\bar{\mathbf{d}} - \bar{\mathbf{y}}||$?

1. $e_i = y_i - (\beta + \alpha x_i)$
2. $e_i^2 = (\beta + \alpha x_i - y_i)^2$
3. $||A\bar{\mathbf{d}} - \bar{\mathbf{y}}|| = \sum_{i=1}^{n} e_i^2$



$$||A\bar{\mathbf{d}} - \bar{\mathbf{y}}|| = \sum_{i=1}^{n} (\beta + \alpha x_i - y_i)^2$$

Least squares approximation: curve fitting.

Let's fit a parabola $\beta + \alpha x + \gamma x^2$ to n points. We need to find a β, α, γ s.t.

$$\begin{aligned} \beta + \alpha x_1 + \gamma x_1^2 &= y_1 \\ \beta + \alpha x_2 + \gamma x_2^2 &= y_2 \\ &\cdots \\ \beta + \alpha x_n + \gamma x_n^2 &= y_n \end{aligned}$$

Fitting a nonlinear curve is still a linear problem!

Least squares approximation: curve fitting.

$$A = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \\ 1 & x_n & x_n^2 \end{bmatrix}, \; \bar{\mathbf{y}} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}; \; \bar{\mathbf{d}} = \begin{bmatrix} \beta \\ \alpha \\ \gamma \end{bmatrix}$$

To find β, α, γ solve using least squares

$$A^T A \bar{\mathbf{d}} = A^T \bar{\mathbf{y}}$$