Linear Algebra

2.1 understanding of Ax = b

In **Row space**, the equation we are trying to solve is two lines in the xy plane. The solution is to find what is the intersection of the two lines.

Click here to expand...

Two equations / **Two unknowns**

$$\begin{aligned} x - 2y &= 1 \\ 3x + 2y &= 11 \end{aligned} \qquad (1)$$

We begin *a row at a time*. The first equation $\mathbf{x - 2y = 1}$ produces a straight line in the xy plane. The point $x = 1, y = 0$ is on the line because it solves that equation. The point $x = 3, y = 1$ is also on the line because $3 - 2 = 1$. If we choose $x = 101$ we find $y = 50$.

The slope of this particular line is $\frac{1}{2}$, because y increases by 1 when x changes by 2. But slopes are important in calculus and this is linear algebra!

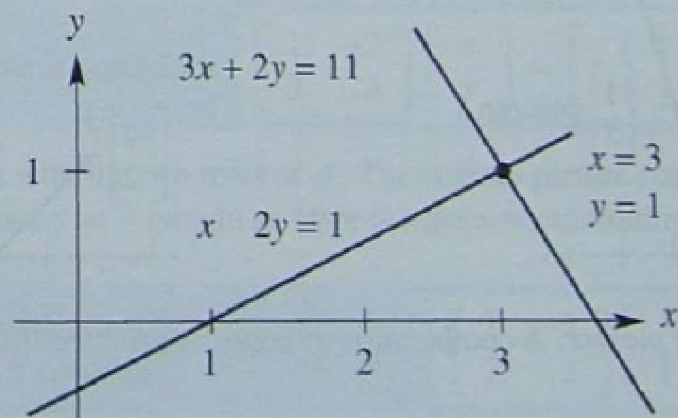


Figure 2.1: *Row picture*: The point (3, 1) where the lines meet is the solution.

In **column space**, the equations are two vectors in column space, the solution is trying to find the scale that combines two vectors into b vector.

Click here to expand...

Combination equals b $$x\begin{bmatrix}1\\3\end{bmatrix}+y\begin{bmatrix}-2\\2\end{bmatrix}=\begin{bmatrix}1\\11\end{bmatrix}=b. \qquad (2)$$

This has two column vectors on the left side. The problem is *to find the combination of those vectors that equals the vector on the right.* We are multiplying the first column by x and the second column by y, and adding. With the right choices $x = 3$ and $y = 1$ (the same numbers as before), this produces $3(\textit{column 1}) + 1(\textit{column 2}) = b$.

COLUMNS ***The column picture combines the column vectors on the left side to produce the vector b on the right side.***

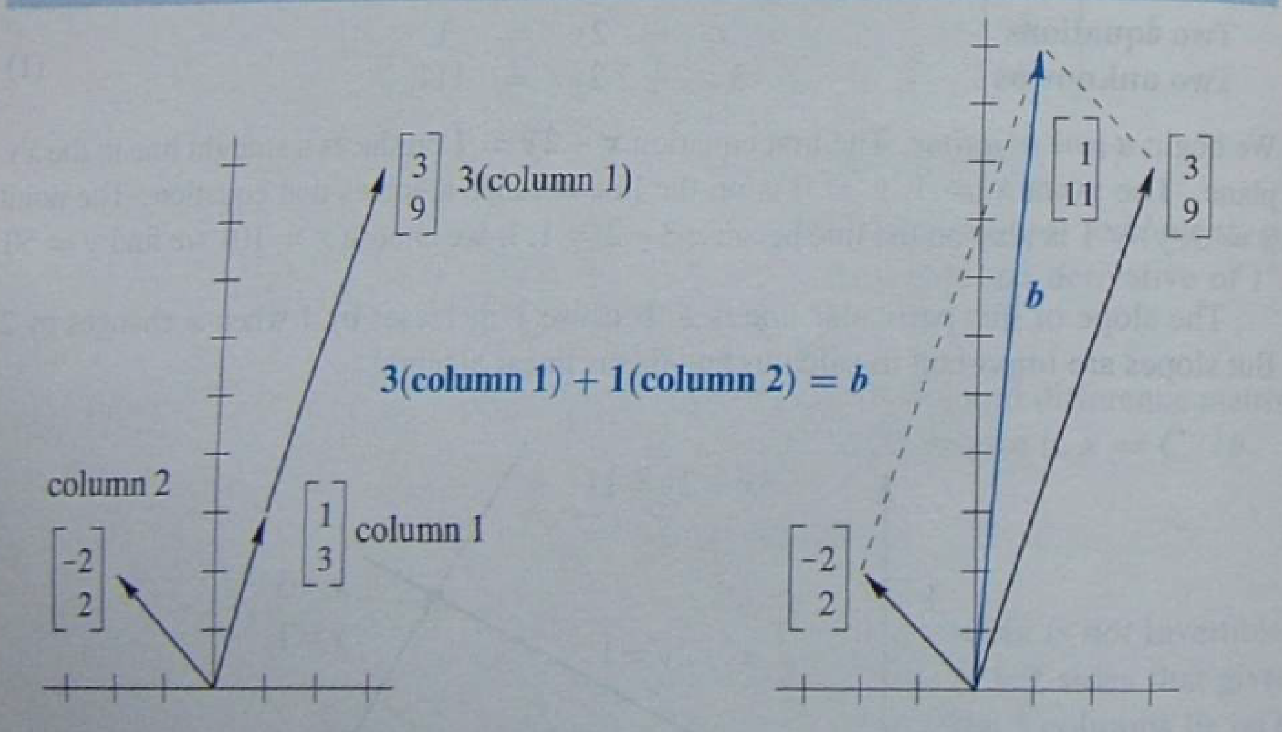


Figure 2.2: *Column picture*: A combination of columns produces the right side (1,11).

3.4 The Complete Solution to Ax=b

Ax = b Rx = d with the same x to solve the question by augmented matrix

Example

Now b is not zero. Row operations on the left side must act also on the right side. $Ax = b$ is reduced to a simpler system $Rx = d$. One way to organize that is to ***add b as an extra column of the matrix***. I will "*augment*" A with the right side $(b_1, b_2, b_3) = (1, 6, 7)$ and reduce the bigger matrix $\begin{bmatrix} A & b \end{bmatrix}$:

$$\begin{bmatrix} 1 & 3 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ 1 & 3 & 1 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 6 \\ 7 \end{bmatrix} \quad \begin{matrix} \text{has the} \\ \text{augmented} \\ \text{matrix} \end{matrix} \quad \begin{bmatrix} 1 & 3 & 0 & 2 & \mathbf{1} \\ 0 & 0 & 1 & 4 & \mathbf{6} \\ 1 & 3 & 1 & 6 & \mathbf{7} \end{bmatrix} = \begin{bmatrix} A & b \end{bmatrix}.$$

The augmented matrix is just $\begin{bmatrix} A & b \end{bmatrix}$. When we apply the usual elimination steps to A, we also apply them to b. That keeps all the equations correct.

In this example we subtract row 1 from row 3 and then subtract row 2 from row 3. This produces a *complete row of zeros* in R, and it changes b to a new right side $d = (1, 6, 0)$:

$$\begin{bmatrix} 1 & 3 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 6 \\ 0 \end{bmatrix} \quad \begin{matrix} \text{has the} \\ \text{augmented} \\ \text{matrix} \end{matrix} \quad \begin{bmatrix} 1 & 3 & 0 & 2 & \mathbf{1} \\ 0 & 0 & 1 & 4 & \mathbf{6} \\ 0 & 0 & 0 & 0 & \mathbf{0} \end{bmatrix} = \begin{bmatrix} R & d \end{bmatrix}.$$

That very last zero is crucial. The third equation has become $0 = 0$ and the equations can be solved. In the original matrix A, the first row plus the second row equals the third row. If the equations are consistent, this must be true on the right side of the equations also! The all-important property on the right side was $1 + 6 = 7$.

Here are the same augmented matrices for a general $b = (b_1, b_2, b_3)$:

$$\begin{bmatrix} A & b \end{bmatrix} = \begin{bmatrix} 1 & 3 & 0 & 2 & b_1 \\ 0 & 0 & 1 & 4 & b_2 \\ 1 & 3 & 1 & 6 & b_3 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 3 & 0 & 2 & b_1 \\ 0 & 0 & 1 & 4 & b_2 \\ 0 & 0 & 0 & 0 & b_3 - b_1 - b_2 \end{bmatrix} = \begin{bmatrix} R & d \end{bmatrix}$$

Now we get $0 = 0$ in the third equation provided $b_3 - b_1 - b_2 = 0$. This is $b_1 + b_2 = b_3$.

- First: X particular : one vector
  - check zero rows in R results in 0 elements in b (to guarantee the exist of the solutions)
  - set all free variables s into 0
  - solve pivot variables Rx = d

Click here to expand...

For a solution to exist, zero rows in R must also be zero in d. Since I is in the pivot rows and pivot columns of R, the pivot variables in $x_{\text{particular}}$ come from d:

$$Rx_p = \begin{bmatrix} \mathbf{1} & 3 & \mathbf{0} & 2 \\ \mathbf{0} & 0 & \mathbf{1} & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{1} \\ 0 \\ \mathbf{6} \\ 0 \end{bmatrix} = \begin{bmatrix} \mathbf{1} \\ \mathbf{6} \\ 0 \end{bmatrix} \qquad \begin{matrix} \textbf{Pivot variables } 1, 6 \\ \textbf{Free variables } 0, 0 \end{matrix}$$

X complete = Xp + X null

Reason: A(Xp + Xnull) = b + 0 whole null space

Click here to expand...

$x_{\text{particular}}$	*The particular solution solves*	$Ax_p = b$
$x_{\text{nullspace}}$	*The $n-r$ special solutions solve*	$Ax_n = \mathbf{0}$.

That particular solution is $(1, 0, 6, 0)$. The two special (nullspace) solutions to $Rx = \mathbf{0}$ come from the two free columns of R, by reversing signs of 3, 2, and 4. ***Please notice how I write the complete solution $x_p + x_n$ to $Ax = b$:***

Complete solution
one x_p
many x_n

$$x = x_p + x_n = \begin{bmatrix} 1 \\ 0 \\ 6 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} -3 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -2 \\ 0 \\ -4 \\ 1 \end{bmatrix}.$$

Note, if Ax=b has multiple solutions, should think x as a batch of cols in a subspace, instead of simply a vector.

Click here to expand...

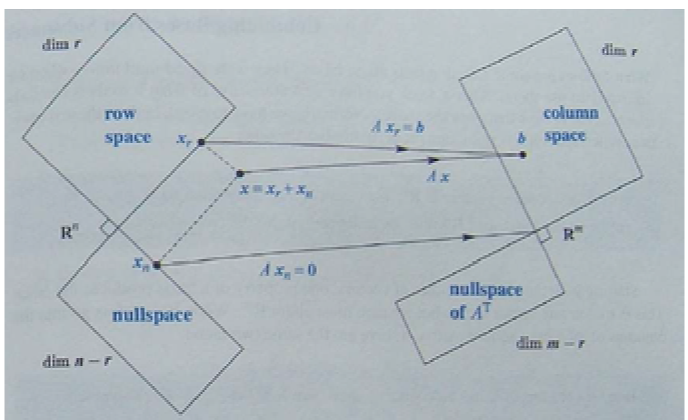


Figure 4.3: This update of Figure 4.2 shows the true action of A on $x = x_r + x_n$. Row space vector x_r to column space, nullspace vector x_n to zero.

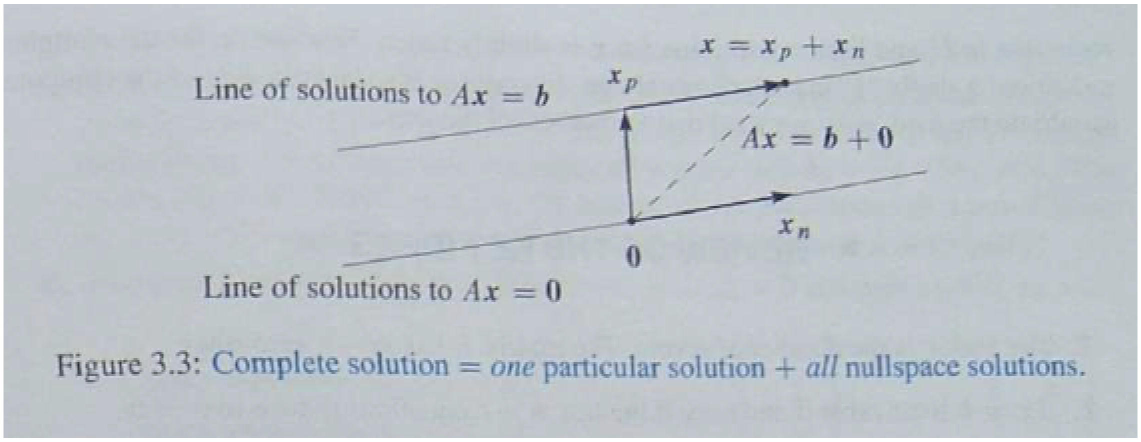


Figure 3.3: Complete solution = *one* particular solution + *all* nullspace solutions.

Note: the Xn in the both images is the null subspace, meaning a batch of vectors there (if r < n). Any Xn in subspace leads to Ax = b (X = xr+ xn). So x is infinite in this case.

3. 6 Four subspaces

1. In the context of Ax = b: Four subspaces is a property of the matrix, it could be used to infer the existence and number of solutions in Ax = b. (one of the usage of four subspaces, might be more…)
2. another one: everything can be vectors, group of them is matrix. But there is the limitation of the matrix, when you use it (to multiple other matrix or vector), those matrix or vector should be in its space. So subspace is like setting the

Click here to expand...

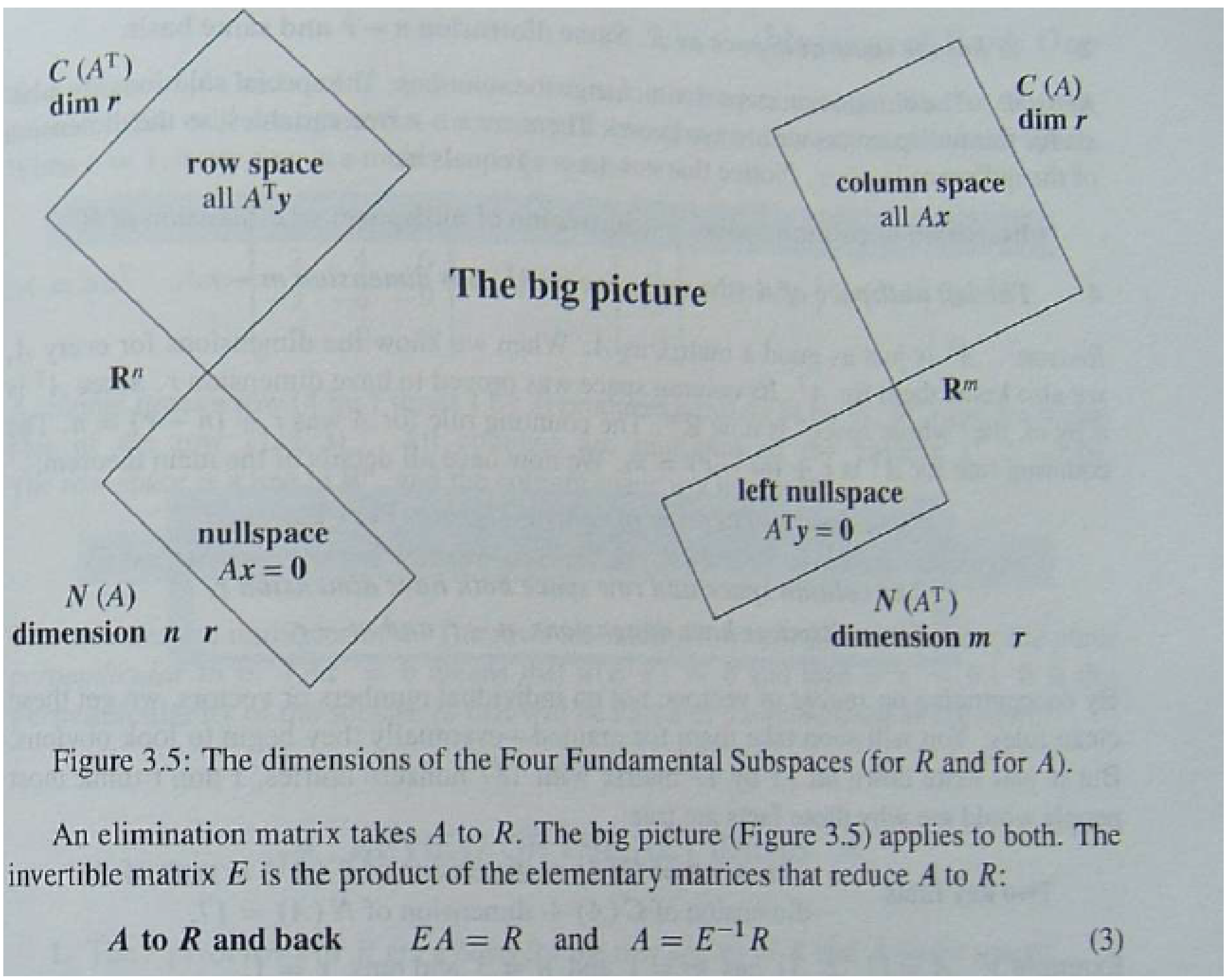


Figure 3.5: The dimensions of the Four Fundamental Subspaces (for R and for A).

An elimination matrix takes A to R. The big picture (Figure 3.5) applies to both. The invertible matrix E is the product of the elementary matrices that reduce A to R:

$$\textbf{A to R and back} \qquad EA = R \quad \text{and} \quad A = E^{-1}R \qquad (3)$$

Rank and Dimension and Basis

- A **basis** consists of linearly independent vectors that span the space
- The **rank** of a **matrix** is the number of pivots.
- The **dimension** of a **subspace** is the number of vectors in a basis.
- Fundamental Theorem of Lineart Algebra1.

  The column space and row space both have dimensions r

  The nullspace has dimensions n-r and m-r

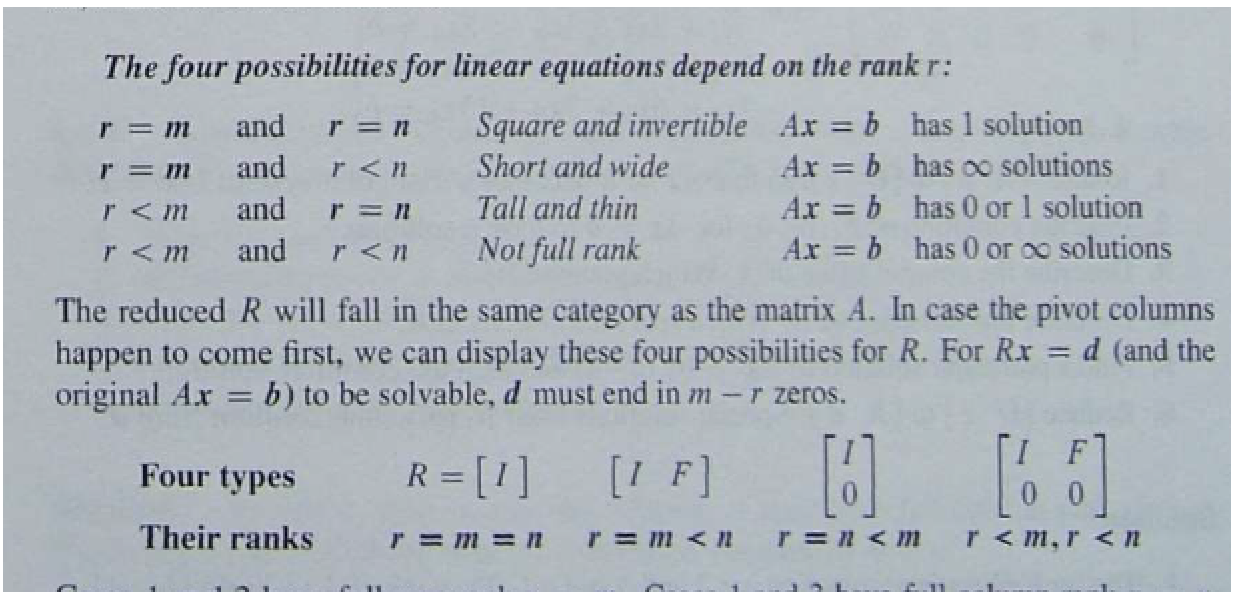
Basis of subspaces:

- C(A) pivot cols dim:r
- N(A) special solutions dim: n-r
- C(AT) rows in the R dim: r
- N(AT): The row in E (EA = R), which leads to 0 rows in R (normally last several rows)
  - To find E : Do the same elimination step for I dim: m-r

Note for Rand A, different columns space yet the same two spaces. because A row exchanges to R, which is a linear combination of rows. the linear combination will not change space. Yet for columns, it is not a linear combination.

Solution Summary How r, n, m related to a number of solutions:

Click here to expand...



The four possibilities for linear equations depend on the rank r:

$r = m$	and	$r = n$	*Square and invertible*	$Ax = b$	has 1 solution
$r = m$	and	$r < n$	*Short and wide*	$Ax = b$	has ∞ solutions
$r < m$	and	$r = n$	*Tall and thin*	$Ax = b$	has 0 or 1 solution
$r < m$	and	$r < n$	*Not full rank*	$Ax = b$	has 0 or ∞ solutions

The reduced R will fall in the same category as the matrix A. In case the pivot columns happen to come first, we can display these four possibilities for R. For $Rx = d$ (and the original $Ax = b$) to be solvable, d must end in $m - r$ zeros.

Four types	$R = [\,I\,]$	$[\,I \;\; F\,]$	$\begin{bmatrix} I \\ 0 \end{bmatrix}$	$\begin{bmatrix} I & F \\ 0 & 0 \end{bmatrix}$
Their ranks	$r = m = n$	$r = m < n$	$r = n < m$	$r < m, r < n$

1. **No solution case (r with m)**

   when r < m and the 0 rows in R are not matched to zero elements in b, there is no solution

2. Have a solution case (r and n to null space)

Whether there are (infinite) solutions in null space.

if r = n null space only has 0, there must be a solution, which is x particular (0 or one up to m)

whereas if r < n null space if not 0, has infinite solutions. This infinite mainly comes from null pace.

4. Orthogonality

Def: The orthogonal complement of a subspace V contains every vector that is perpendicular to V. This orthogonal subspace is denoted by V perp.

Fundamental Theorem of Lineart Algebra 2

N(A) is the orthogonal complement of the row space C(AT) in Rn

N(AT) is the orthogonal complement of the column space C(A) in Rm

4.3 Least Squares Approximations

Problem: when Ax = b has no solution (b is out of col space of A), we need to find x that minimizes b-Ax.

How to find x^ as min b-Ax?

By geometry: The nearest point in col space of A is the projection of B into that space, Which give vector p - **The nearest point is the projection p**

Now convert sovling **Ax = b to Ax^ = p**

Has proved that ATAx^ = ATb. Then we can get x^.

We call this x^ the least squares solution.

But why “least squares solution“ this name for x ^ as projection of b?

\

The origin of the name of x^

Example 1 A crucial application of least squares is fitting a straight line to m points. Start with three points: *Find the closest line to the points* $(0,6)$, $(1,0)$, *and* $(2,0)$.

No straight line $b = C + Dt$ goes through those three points. We are asking for two numbers C and D that satisfy three equations. Here are the equations at $t = 0, 1, 2$ to match the given values $b = 6, 0, 0$:

$t = 0$	The first point is on the line $b = C + Dt$ if	$C + D \cdot 0 = 6$
$t = 1$	The second point is on the line $b = C + Dt$ if	$C + D \cdot 1 = 0$
$t = 2$	The third point is on the line $b = C + Dt$ if	$C + D \cdot 2 = 0.$

This 3 by 2 system has *no solution*: $\boldsymbol{b} = (6, 0, 0)$ is not a combination of the columns $(1, 1, 1)$ and $(0, 1, 2)$. Read off $A, \boldsymbol{x}$, and $\boldsymbol{b}$ from those equations:

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \quad \boldsymbol{x} = \begin{bmatrix} C \\ D \end{bmatrix} \quad \boldsymbol{b} = \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix} \quad A\boldsymbol{x} = \boldsymbol{b} \text{ is } \textit{not} \text{ solvable.}$$

Example: To find the fitting line to m points.

The least squares solution $\widehat{x}$ makes $E = \|Ax - b\|^2$ as small as possible.

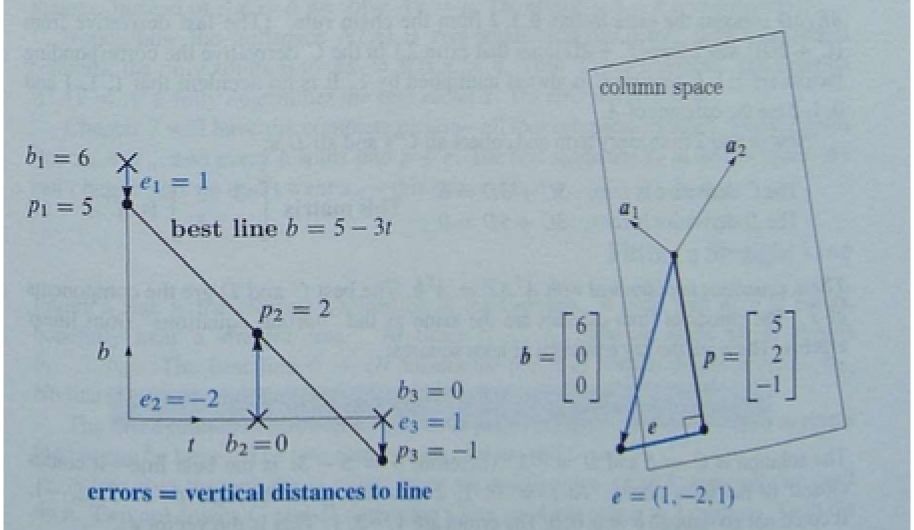


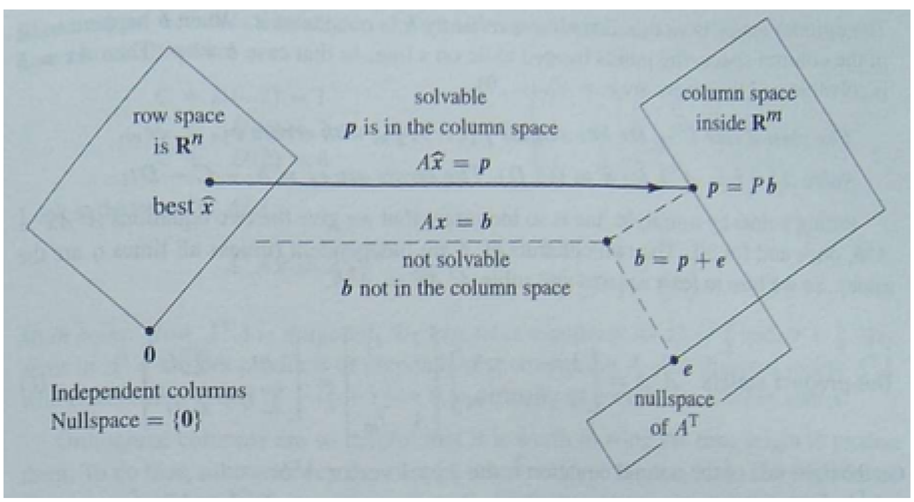
Figure 4.6: **Best line and projection: Two pictures, same problem.** The line has heights $p = (5, 2, -1)$ with errors $e = (1, -2, 1)$. The equations $A^TA\widehat{x} = A^Tb$ give $\widehat{x} = (5, -3)$. The best line is $b = 5 - 3t$ and the projection is $p = 5a_1 - 3a_2$.

left: row picture: find the fitting line

right: column picture: find a projection p in the col space of A

understanding of “least squares solution“:

b could split into p and e. e is the part of the b out off the space. so it is totally unsolvable. The only is to solve p, in the sense that ||Ax^ - p|| = 0



So, x^ is least squares solution because the relationship of Ax-b and Ax-p and e meet the right triangle formula.

The solution to $A\widehat{x} = p$ leaves the least possible error (which is e):

Squared length for any x $$\|Ax - b\|^2 = \|Ax - p\|^2 + \|e\|^2. \quad (2)$$

This is the law $c^2 = a^2 + b^2$ for a right triangle. The vector $Ax - p$ in the column space is perpendicular to e in the left nullspace. We reduce $Ax - p$ to zero by choosing x to be $\widehat{x}$. That leaves the smallest possible error $e = (e_1, e_2, e_3)$.

So, x^ is the least squares solution because the relationship of Ax-b and Ax-p and e meet the right triangle formula.

Eigenvalue and Eigenvector

All in all:

The basic equation is **Ax =x**

eigenvalue first:

Ax = x --> (A - I)x = 0.

If exis **,** there is the solution (The eigenvector x is the zero vector. The eigenvectors make up the null space of A-I) A-I is singular --> To find **:** solving **det(A-I) = 0**

if **= 0** Ax = 0 is solvable **-->**A is singular

Given to find eigenvetor:

By solving (A-I)x = 0.

if all **of A are distinct eigenvectors are independent A is diagonalizable** (Any matrix with no repeat eigenvalues is diagonalizable this is **sufficient** proof : no repeat dia but dia !--> no repeat)

Important : There is no connection between invertibility and diagonalizability

- Invertibility is concerned with the eigenvalues (= 0 or != 0).
- Diagonalizability is concerned with the eigenvectors (too few or enough for S)
- Independent x from different

Proof: Eigenvector of distinct eigenvalues are linearly independent

Eigenvector of distinct eigenvalues are linearly indepent

$T \quad \{V_1, V_2, V_3 \cdots V_n\}$. P_i ~~?~~ if distinct eigenvalues $[\lambda_1 \lambda_2 \cdots \lambda_n]$ Then $[V_1 \cdots V_n]$ independe

To proof V is indepent ; if $\sum_{i=1}^{n} a_i V_i = 0$ show $\forall i \le n \quad a_i = 0$. Then V is indep

logic : $P(0)$ is True $P(1)$ is True Assume $P(k)$ is T show $P(k+1)$ is T

① : $P(k)$ is T ② $\lambda_{k+1} \ne \lambda_k$

$\therefore \sum_{i=1}^{k} a_i V_i \Rightarrow \forall i \le k \quad a_i = 0$

Show $\sum_{i=1}^{k+1} a_i V_i = 0$ & $\forall i \le k+1 \quad a_i = 0$

$k=0$. S_0 is empty indepent ✓.

$k=1 \quad S_1 = \{V_1\}$ independent ✓.

Let $\sum_{i=1}^{k+1} a_i V_i = 0$

$\sum_{i=1}^{k+1} a_i V_i \lambda_i = 0$? :

$\because TV = \lambda V$

$T\left(\sum_{i=1}^{k+1} a_i V_i\right) / T(0) = 0$?

$T \cdot 0 = T \cdot (0+0) = T \cdot 0 + T \cdot 0$

$T \cdot 0 = T \cdot 0 + T \cdot 0$

~~$T(0)$~~ $T \cdot 0 = 0$.

$T\left(\sum_{i=1}^{k+1} a_i V_i\right) = 0 = \lambda \sum_{i=1}^{k+1} a_i V_i = \sum_{i=1}^{k+1} a_i \lambda_i V_i$

$0 = \sum_{i=1}^{k+1} a_i \lambda_i V_i = \sum_{i=1}^{k} a_i \lambda_i V_i + a_{k+1} \lambda_{k+1} V_{k+1}$ - - - - - ①

λ_{k+1} ?

$0 = 0 \lambda_{k+1} = \left[\sum_{i=1}^{k+1} a_i V_i\right] \cdot \lambda_{k+1} = \sum_{i=1}^{k} a_i \lambda_{k+1} V_i + a_{k+1} \lambda_{k+1} V_{k+1}$ - - ②

② - ① : $\sum_{i=1}^{k} a_i \lambda_{k+1} V_i - \sum_{i=1}^{k} a_i \lambda_i V_i$

$= \sum_{i=1}^{k} a_i V_i (\lambda_{k+1} - \lambda_i)$

$\because S_k$ is independent $\therefore \forall i \le k$ ~~$\sum_{i=1}^{k} a_i V_i$~~ $a_i(\lambda_{k+1} - \lambda_i) = 0$

$\because \lambda_{k+1} - \lambda_i \ne 0$

$\therefore \quad a_i = 0$

$\sum_{i=0}^{k+1} a_i V_i = 0 = \underbrace{\sum_{i=0}^{k} a_i V_i}_{= 0} + a_{k+1} V_{k+1} = 0$

$\Rightarrow a_{k+1} = 0$ 得证

others

- The product of the n eigenvalues equals the determinant.
- The sum of the n eigenvalues equals the sum of the n diagonal

Symmetrix matrix

For symmetric matrix: p = p^T, For orthonormal matrix : A^T A = I

Spectral theorem: every symmetric matrix has real 's an orthonormal(orthogonal and unit vectors) x's.

Spectral theorem

1. A symmetric matrix has only real eigenvalues.
2. The eigenvectors can be chosen orthonormal.

Every symmetric matrix can be diagonalized

A is a combination of projection matrice

A is a combination of projection matrice

This formula shows the n by n ***projection matrix*** that produces $p = Pb$:

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

when A= A^T

$$A = Q\Lambda Q^{\mathrm{T}} = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} \lambda_1 & \\ & \lambda_2 \end{bmatrix} \begin{bmatrix} x_1^{\mathrm{T}} \\ x_2^{\mathrm{T}} \end{bmatrix}.$$

The columns x_1 and x_2 multiply the rows $\lambda_1 x_1^{\mathrm{T}}$ and $\lambda_2 x_2^{\mathrm{T}}$ to produce A:

Sum of rank-one matrices $\quad A = \lambda_1 x_1 x_1^{\mathrm{T}} + \lambda_2 x_2 x_2^{\mathrm{T}}.$

$A = \lambda_1 P_1 + \cdots + \lambda_n P_n \qquad \lambda_i =$ eigenvalue, $\quad P_i =$ projection onto eigenspace.

Complex Eigenvalues of Real Matrice

Click here to expand...

For real matrices, complex λ's and x's come in "conjugate pairs."

$$\textit{If} \quad A\boldsymbol{x} = \lambda\boldsymbol{x} \quad \textit{then} \quad A\overline{\boldsymbol{x}} = \overline{\lambda}\,\overline{\boldsymbol{x}}.$$

This fact \\ = 1 holds for the eigenvalues of every orthogonal matrix. (projection matrix?)

All symmetric matrices are diagonalizable, even with repeated eigenvalues. p344

Real Eigenvalue ~~proof~~ for symmetrix matrix)

In: $Ax=\lambda x$ $A=A^T$ $\overline{\lambda x}=\bar{\lambda}\bar{x}$

Out: $\lambda=a+bi$ $b=0$.

$Ax=\lambda x$ ------ ① $\xrightarrow[\text{product}]{\text{dot } \bar{x}}$ $\cancel{\bar{x}Ax=\bar{x}\lambda x}$

$\bar{x}^TAx=\bar{x}^T\lambda x$ ---- ④

Conjugate: $\overline{Ax}=\overline{\lambda x}=\bar{\lambda}\bar{x}$

A is real: $A\bar{x}=\bar{\lambda}\bar{x}$ --- ②

Transpose: $(A\bar{x})^T=(\bar{\lambda}\bar{x})^T$

$\bar{x}^TA=\bar{x}^T\bar{\lambda}$ --- ③ $\xrightarrow[\text{product}]{\text{dot } x}$ $\bar{x}^TAx=\bar{x}^T\bar{\lambda}x$ -- ⑤.

④⑤: $\bar{x}^T\lambda x=\bar{x}^T\bar{\lambda}x$

$\bar{x}^Tx=|x_1|^2+|x_2|^2\cdots|x_n|^2$ 计 $\bar{x}^Tx\neq 0$. $x\neq 0$

$\lambda=\bar{\lambda}$ $a+bi=a-bi \Rightarrow b=0$ Q.E.D.

Orthogonal Eigenvectors for Symmetrix Matrix Proof.

In: $Ax=\lambda_1x$ $Ay=\lambda_2y$ $\lambda_1\neq\lambda_2$ $A=A^T$

Out: $x^Ty=0$.

$Ax=\lambda_1x$ $\xrightarrow{\text{Transpose}}$ $x^TA=\lambda_1x^T$ $\downarrow y.$ $x^TAy=\lambda_1x^Ty.$

$Ay=\lambda_2y$ $\downarrow x^T$ dot product $x^TAy=x^T\lambda_2y.$

$\lambda_1x^Ty=\lambda_2x^Ty$

$\lambda_1\neq\lambda_2$ $x^Ty=0$ Q.E.D.

Schur's Theorem

Every square matrix factors into $A=QTQ^{-1}$, T is upper triangular. $\bar{Q}^T=Q^{-1}$

For symmetric matrix. $A=Q\Lambda Q^{-1}$. $(T=\Lambda)$.

Proof: in: Schur's theorm $A=QTQ^{-1}$ $\bar{Q}^T=Q^{-1}$ $A^T=A$. $Q^TQ=I$ $(Q^T=Q^{-1})$

$\cancel{A=QTQ^T}$

$\cancel{A^T=(Q^{-1})^TTQ^T}$

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

$\cancel{A^T=QTQ^T}$

$A=QTQ^{-1}$

$T=QAQ^{-1}$

$T^T=(Q^{-1})^TAQ^T=QAQ^T$

$T^T=T \Rightarrow T=\Lambda$

Different Factorzation

A = LU

$$\textit{Forward from } A \textit{ to } U: \quad E_{21}A = \begin{bmatrix} 1 & 0 \\ -3 & 1 \end{bmatrix}\begin{bmatrix} 2 & 1 \\ 6 & 8 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 0 & 5 \end{bmatrix} = U$$

$$\textit{Back from } U \textit{ to } A: \quad E_{21}^{-1}U = \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix}\begin{bmatrix} 2 & 1 \\ 0 & 5 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 6 & 8 \end{bmatrix} = A.$$

The second line is our factorization $LU = A$. Instead of E_{21}^{-1} we write L. Move now to larger matrices with many E's. ***Then L will include all their inverses.***

Each step from A to U multiplies by a matrix E_{ij} to produce zero in the (i, j) position. To keep this clear, we stay with the most frequent case—***when no row exchanges are involved.*** If A is 3 by 3, we multiply by E_{21} and E_{31} and E_{32}. The multipliers ℓ_{ij} produce zeros in the $(2, 1)$ and $(3, 1)$ and $(3, 2)$ positions—all below the diagonal. Elimination ends with the upper triangular U.

Now move those E's onto the other side, *where their inverses multiply* U:

$$(E_{32}E_{31}E_{21})A = U \quad \textbf{becomes} \quad A = (E_{21}^{-1}E_{31}^{-1}E_{32}^{-1})U \quad \textbf{which is} \quad A = LU. \qquad (1)$$

The inverses go in opposite order, as they must. That product of three inverses is L. ***We have reached $A = LU$.*** Now we stop to understand it.

The triangular factorization can be written $\quad A = LU \quad$ or $\quad A = LDU$.

Whenever you see LDU, it is understood that U has 1's on the diagonal. *Each row is divided by its first nonzero entry—the pivot.* Then L and U are treated evenly in LDU:

$$\begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix}\begin{bmatrix} 2 & 8 \\ 0 & 5 \end{bmatrix} \quad \text{splits further into} \quad \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix}\begin{bmatrix} 2 & \\ & 5 \end{bmatrix}\begin{bmatrix} 1 & 4 \\ 0 & 1 \end{bmatrix}. \qquad (4)$$

A = QR

(A: independent columns, Q: orthogonal matrix R: triangular matrix)

Background + Projection

1. Orthonormal matrix: QTQ = I
2. When Q is square, QTQ = I means that QT= Q-1: transpose = inverse
3. multiplication of orthogonal matrix doesn't change length. `(Orthogonal matrices are excellent for computations—numbers can never grow too large when lengths of vectors are fixe)`

Note: cos and sin are just orthonornal bases. while i and j are the standard basis. The basis of a space no need to be orthogonal, yet orthogonal basis is good for calculation(point 3).

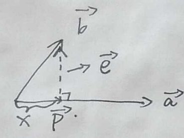
$n=1$. project of two vectors.

~~(2)~~ project $\vec{b}$ into $\vec{a}$

① what is resulting vec $\vec{P}$

② what is Matrix P

③ How it (project) related to dot · product



$\vec{P}$ is in $\vec{a}$ closest to $\vec{b}$

$\vec{P} = \vec{a} \cdot x$

use $\perp$ ($\vec{e} \cdot \vec{a} = 0$).

$\vec{e} = \vec{b} - \vec{P}$

$= \vec{b} - x\vec{a}$

$(\vec{b} - x\vec{a})\vec{a} = 0$.

$\vec{b} \cdot \vec{a} - x\vec{a} \cdot \vec{a} = 0$

$x = \frac{\vec{b} \cdot \vec{a}}{\vec{a} \cdot \vec{a}}$

① $\vec{P} = \frac{\vec{b} \cdot \vec{a}}{\vec{a} \cdot \vec{a}} \cdot \vec{a}$

② $P = \frac{\vec{a}}{\vec{a} \cdot \vec{a}} \cdot \vec{a}$.

③. $\vec{b} \cdot \vec{a} = \underbrace{x\vec{a}}_{\vec{P}} \cdot \vec{a}$

since $\vec{a}$ and $\vec{P}$ in the same ~~direction~~ space.

$\therefore \vec{b} \cdot \vec{a}$ is like $\mathrm{len}(\underset{\vec{b}\text{ project to }\vec{a}}{\vec{P}}) \cdot \mathrm{len}(\vec{a})$

$||P|| = \frac{|b||a|\cos\theta}{|a|^2} \cdot |a| = |b|\cos\theta$.

~~n project b into A subspace.~~

~~$\vec{P} = A\vec{b}$ (order matters A apply to $\vec{b}$)~~

project b into A subspace. $A: [a_1\, a_2 \cdots a_n]$

a_i vector in $\mathbb{R}^m$.

① Find $\vec{P} = x_1\vec{a_1} + \cdots x_n\vec{a_n}$ closest to $\vec{b}$

② P · matrix.

$\vec{P} = A\vec{x}$

still $\vec{e} = \vec{b} - A\vec{x}$

$A^T(\vec{b} - A\vec{x}) = 0$.

$A^Tb = A^TA\vec{x}$

$\vec{x} = (A^TA)^{-1}A^Tb$.

$\vec{P} = A(A^TA)^{-1}A^Tb$.

$P = A\underset{\sim\sim}{(A^TA)}^{-1}A^T$.

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

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

$\because A^TA$ is invertible

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

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

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

$P^2 = P$.

A^TA must revertiable. (A is rectegura)

A^TA revertiable iff cols in A are all independt.

prof: A^TA is ~~rever~~ intertible iff ① A^TA is sqaure

② Null space $= 0$.

$\because A^TA$ and A have same null space.

$\therefore$ if cols A is dependt null $\neq 0$ $N(A) \neq 0$

same as A^TA $N(A^TA) \neq 0$

prof A^TA and A have same null space.

$$\begin{cases} Ax = 0 \\ A^TAx = A^T0 \\ \quad = 0 \end{cases} \Rightarrow Ax = 0 \Rightarrow A^TAx = 0.$$

$$\left.\begin{array}{l} A^TAx = 0. \\ x^TA^TAx = x^T0 = 0 \\ \cancel{(xA)^TAx} \\ \star\ (Ax)^TAx = 0 \\ \quad ||Ax|| = 0 \end{array}\right\} \Rightarrow \begin{array}{l} A^TAx = 0 \\ Ax = 0. \end{array}$$

prof!

~~$P^2 = (A$~~

$P^2 = (A(A^TA)^{-1}A^T)^T(A(A^TA)^{-1}A^T)$

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

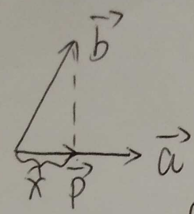
$= A(A^TA)^{-1}(A^TA)(A^TA)^{-1}A^T$

$= AI \cdot (A^TA)^{-1}A^T$

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

- Recall

Find. $\vec{p}$. P ?



$\hat{x}$ lead to scale $\vec{a}$ end with $\vec{p}$ where closest to $\vec{b}$.

def: $\vec{p} = \vec{a}_1 x_1 + \vec{a}_2 x_2 \cdots \vec{a}_n x_n$

$\vec{p} = A\underbrace{\vec{x}}_{\text{scale vector}}$

def: $\vec{p} = P\vec{b}$

$\because \perp \therefore \quad A^T(\vec{b} - A\vec{x}) = 0$

$A^T\vec{b} = A^T\underline{A\vec{x}}$

$\vec{x} = (A^TA)^{-1}A^T\vec{b}$

$\vec{p} = A\vec{x} = \underbrace{A(A^TA)^{-1}A^T}\vec{b}$

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

- When A is orthonormal

$\rightarrow Q. \quad Q^TQ = I$

$P = Q(Q^TQ)^{-1}Q^T$

$= QQ^T$

$\vec{p} = QQ^T\vec{b}$

$= q_1(q_1^Tb_1) + q_2(q_2^Tb_2)$

$\cdots q_n(q_n^Tb_n)$

- When Q is square.

$Q^T = Q^{-1}, \quad P = QQ^T = I$

QQ^T projects $\vec{b}$ onto same (whole space) of $\vec{b}$ itself $[\vec{b} = QQ^T\vec{b}]$ is the sum of its perpendicular component along $\vec{q}$'s

$b = \underset{\text{component}}{q_1}\underbrace{(q_1^Tb)}_{\text{scaler}} + q_2(q_2^Tb) \cdots q_n(q_n^Tb)$

$q_1 \perp q_2, \cdots$

$A = QR$:

- How to find Q given A?

In: 3 independent vectors $\vec{a}, \vec{b}, \vec{c}$

Out: 3 orthogonal vectors $\vec{A}\ \vec{B}\ \vec{C}$.

$\vec{A} = \vec{a}$.

$\vec{B} = \vec{b} - \dfrac{\vec{A}^Tb}{\vec{A}^T\vec{A}}\vec{A}$.

$\vec{C} = \vec{c} - \dfrac{A^T\vec{c}}{A^TA}\vec{A} - \dfrac{B^T\vec{c}}{B^TB}\vec{B}$

$\vec{Q} = [\vec{A}\ \vec{B}\ \vec{C}]$

$(q_1)(q_2)(q_3)$

- How Q Related to A

$A = QR$.

$$\begin{bmatrix} a & b & c \end{bmatrix} = \begin{bmatrix} q_1 & q_2 & q_3 \end{bmatrix} \begin{bmatrix} q_1^Ta & q_1^Tb & q_1^Tc \\ & q_2^Tb & q_2^Tc \\ & & q_3^Tc \end{bmatrix}$$

Text book -- Foundation of Fourier

There are no matrices to invert. This is the point of an orthonormal basis. The best $\widehat{x} = Q^{\mathrm{T}}b$ just has dot products of $q_1, \ldots, q_n$ with b. We have n 1-dimensional projections! The "coupling matrix" or "correlation matrix" $A^{\mathrm{T}}A$ is now $Q^{\mathrm{T}}Q = I$. There is no coupling. When A is Q, with orthonormal columns, here is $p = Q\widehat{x} = QQ^{\mathrm{T}}b$:

Projection onto q's

$$p = \begin{bmatrix} | & & | \\ q_1 & \cdots & q_n \\ | & & | \end{bmatrix} \begin{bmatrix} q_1^{\mathrm{T}}b \\ \vdots \\ q_n^{\mathrm{T}}b \end{bmatrix} = q_1(q_1^{\mathrm{T}}b) + \cdots + q_n(q_n^{\mathrm{T}}b). \quad (5)$$

Important case: When Q is square and $m = n$, the subspace is the whole space. Then $Q^{\mathrm{T}} = Q^{-1}$ and $\widehat{x} = Q^{\mathrm{T}}b$ is the same as $x = Q^{-1}b$. The solution is exact! The projection of b onto the whole space is b itself. In this case $P = QQ^{\mathrm{T}} = I$.

You may think that projection onto the whole space is not worth mentioning. But when $p = b$, our formula assembles b out of its 1-dimensional projections. If $q_1, \ldots, q_n$ is an orthonormal basis for the whole space, so Q is square, then every $b = QQ^{\mathrm{T}}b$ is the sum of its components along the q's:

$$b = q_1(q_1^{\mathrm{T}}b) + q_2(q_2^{\mathrm{T}}b) + \cdots + q_n(q_n^{\mathrm{T}}b). \quad (6)$$

That is $QQ^{\mathrm{T}} = I$. It is the foundation of Fourier series and all the great "transforms" of applied mathematics. They break vectors or functions into perpendicular pieces. Then by adding the pieces, the inverse transform puts the function back together.

6.5 Positive Definite Matrices

what is Positive Definite Matrice?

They are symmetric matrices that have positive eigenvalue.

How to test (find) Positive Definite Matrice?

Click here to expand...

The λ's are automatically real because the matrix is symmetric.

Start with 2 by 2. When does $A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$ *have* $\lambda_1 > 0$ *and* $\lambda_2 > 0$?

The eigenvalues of A are positive if and only if $a > 0$ *and* $ac - b^2 > 0$.

$A_1 = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$ is *not* positive definite because $ac - b^2 = 1 - 4 < 0$

$A_2 = \begin{bmatrix} 1 & -2 \\ -2 & 6 \end{bmatrix}$ is positive definite because $a = 1$ and $ac - b^2 = 6 - 4 > 0$

$A_3 = \begin{bmatrix} -1 & 2 \\ 2 & -6 \end{bmatrix}$ is *not* positive definite (even with $\det A = +2$) because $a = -1$

Why Positive Definite Matrice is important? / what are the applications?

First Application: The Ellipse ax2 + 2bxy + cy 2 =1

Second application: Test for a minimum

SVD

What is Singular Value Decomposition (SVD)?

SVD is a new way of factoring.

1. To find a special unit bases v1 in Row spaces that allows Av1 = u1. And u should be orthogonal to each other as well.
2. We want u is also unit vector. Thus Av1 = u1 *sigma (scaler)

Why SVD?

Why not SAS-1

1. U and V are orthogonal matrices, yet S is an eigenvector
2. S might not have enough eigenvector to factorization (might fail)
3. SAS-1 requires A is a square matrix. Yet SVD has no needs.

How to find the V and U?

1. First, reduce U and only left v. By calculating ATA, find the V and square sigma
2. U = Av/sigma

Click here to expand...

There is a neat way to remove U and see V by itself. *Multiply A^T times A.*

$$A^T A = (U \Sigma V^T)^T (U \Sigma V^T) = V \Sigma^T \Sigma V^T. \tag{7}$$

$U^T U$ disappears because it equals I. (We require $u_1^T u_1 = 1 = u_2^T u_2$ and $u_1^T u_2 = 0$.) Multiplying those diagonal Σ^T and Σ gives σ_1^2 and σ_2^2. That leaves an ordinary diagonalization of the crucial symmetric matrix $A^T A$, whose eigenvalues are σ_1^2 and σ_2^2:

Eigenvalues σ_1^2, σ_2^2
Eigenvectors v_1, v_2

$$A^T A = V \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} V^T. \tag{8}$$

Compute the eigenvectors v and eigenvalues σ^2 of $A^T A$. Then each $u = Av/\sigma$.

How u and v related to subspace of A?

v1 is the unit vector(base) in the row space of A and **u1** is the unit base in the column space

Then the square **sigma** = eig-value(ATA)

v2 is the unit vector(base) in the null space of A and **u2** is the unit base in the left null space

(**V** is an orthogonal matrix and row space is perpendicular to null space, thus if V1 is in row space, v2 will be in null space, same to **U**)
Click here to expand...

Example 4 Find the SVD of the singular matrix $A = \begin{bmatrix} 2 & 2 \\ 1 & 1 \end{bmatrix}$. The rank is $r = 1$.

Solution The row space has only one basis vector $v_1 = (1, 1)/\sqrt{2}$. The column space has only one basis vector $u_1 = (2, 1)/\sqrt{5}$. Then $Av_1 = (4, 2)/\sqrt{2}$ must equal $\sigma_1 u_1$. It does, with $\sigma_1 = \sqrt{10}$.

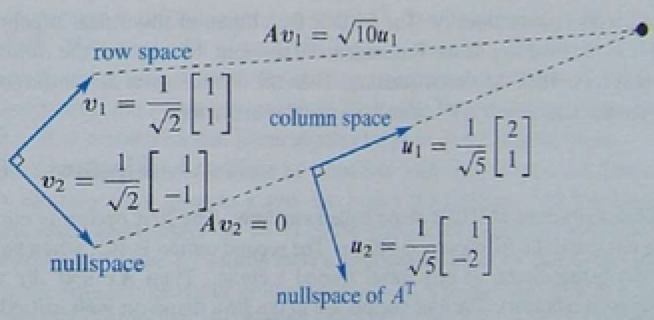


Figure 6.9: The SVD chooses orthonormal bases for 4 subspaces so that $Av_i = \sigma_i u_i$.

The matrices U and V contain orthonormal bases for all four subspaces:

first	r	columns of V:	*row space* of A
last	$n - r$	columns of V:	*nullspace* of A
first	r	columns of U:	*column space* of A
last	$m - r$	columns of U:	*nullspace* of A^T

Understand SVD from transformation: