EURECOM
Sophia Antipolis

Essential Mathematical Methods for Engineers

Lecture 7:

Linear algebra 2

Outline

- orthogonality
  - projections
  - least squares approximation
  - Gram-Scmidt
- eigenvales and eigenvectors
  - diagonalisation
  - symmetric matrices
  - singular value decomposition (SVD)
- numercial linear algebra
  - iterative methods
  - norms and condition number
  - ill conditioning

Orthogonality

- the row space is perpendicular to the nullspace
  - every row of A is perpendicular to every solution of $Ax = 0$
- the column space is perpendicular to the left nullspace
  - every column of A is perpendicular to every solution of $A^Ty = 0$
- two spaces are perpendicular when
  - $v.w = 0$ or $v^Tw = 0$ for all v in V and w in W
- about the left nullspace
  - it is never reached by Ax
  - when b is outside the column space we cannot solve $Ax = b$
  - the left nullspace contains the error in the least squares solution

2^{nd} Fundamental Theorem of Linear Algebra

- the fundamental subspaces are not only orthogonal
  - they are orthogonal complements

the nullspace is the orthogonal complement of the row space (in R^n)

the left nullspace is the orthogonal complement of the column space (in R^m)

EURECOM

2^{nd} Fundamental Theorem of Linear Algebra

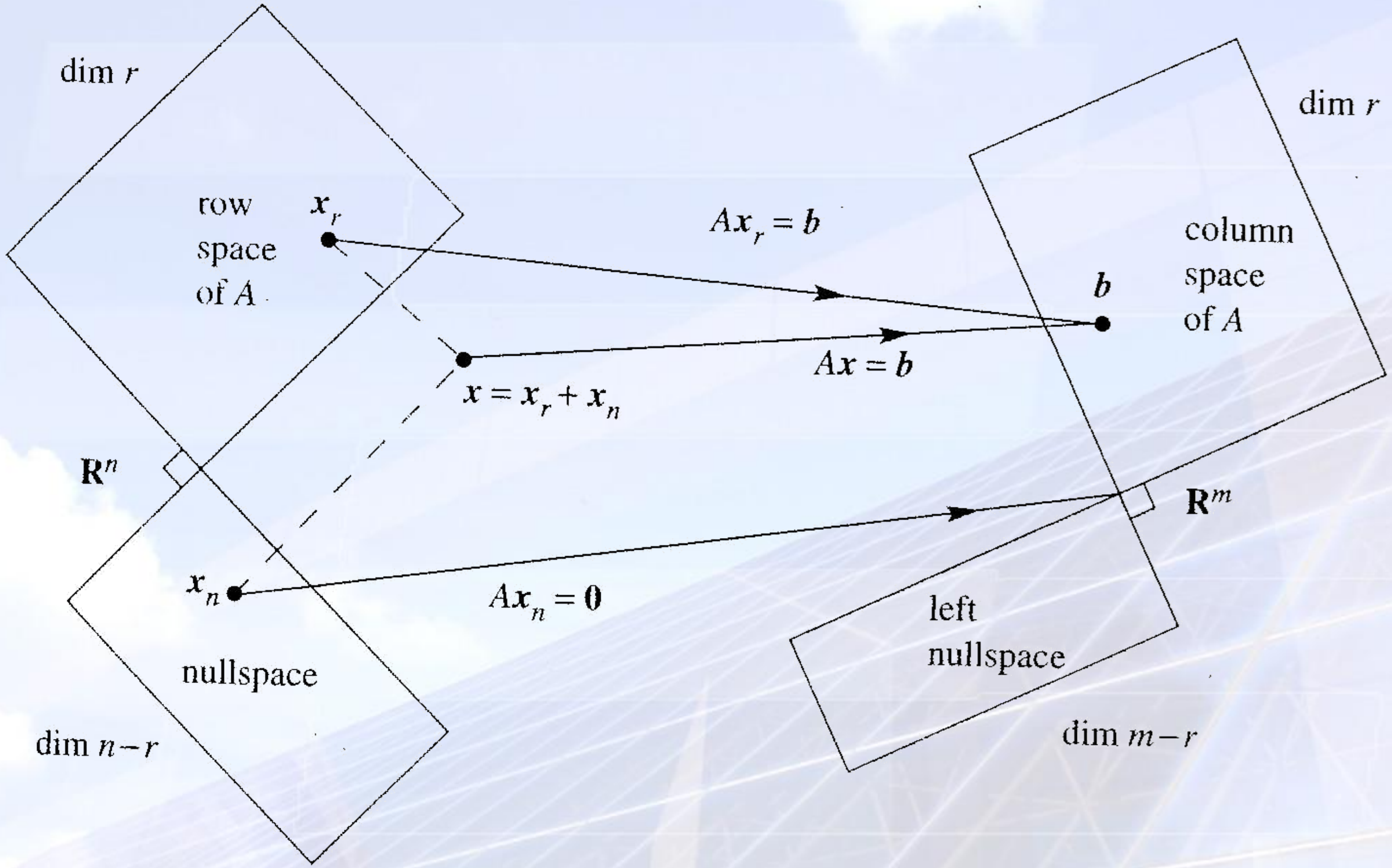
- every vector x can be split into
  - row space component x_r
  - nullspace component x_n

  so we really have $A(x_r + x_n)$
  - $Ax_r = Ax$
  - $Ax_n = 0$

- every vector goes to the column space
  - every vector in the column space comes from one and only one vector in the row space
  - there is an invertible matrix hiding in A
    - the pseudoinverse

EURECOM

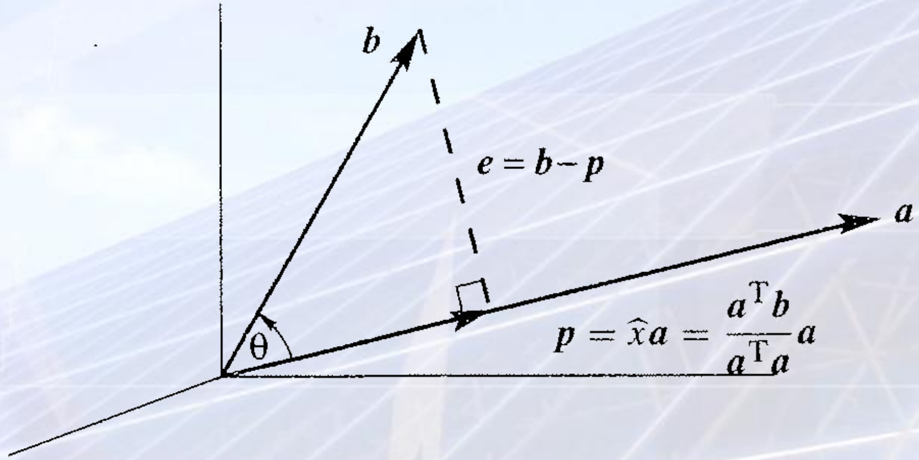
2nd Fundamental Theorem of Linear Algebra



Projections

- we are given
  - a point $b = (b_1, \ldots, b_m)$ in m-dimensional space
  - a line through the origin $a = (a_1, \ldots, a_m)$

- we want the point p on a that is closest to b
  - the line connecting p and b is perpendicular to a



subspaces &
column spaces

Projections

- p is a multiple of a : $p = \hat{x}a$
- the dotted line $b - p$ is : $b - p = b - \hat{x}a$
- noting perpendicularity

$$a \cdot \left(b - \hat{x}a\right) = 0 \quad \text{or} \quad a \cdot b - \hat{x}a \cdot a = 0 \quad \text{or} \quad \hat{x} = \frac{a \cdot b}{a \cdot a} = \frac{a^T b}{a^T a}$$

- the projection of b onto the line through a is the vector

$$p = \hat{x}a = \frac{a^T b}{a^T a} a$$

- the length of the line is $||p|| = ||b|| \cos\theta$

Projections

- the projection matrix is given by

$$p = a\hat{x} = a\frac{a^T b}{a^T a} = Pb \quad \text{thus} \quad P = \frac{aa^T}{a^T a}$$

- what is the rank of P?
- two special cases
  - if $b = a$ then $\hat{x} = 1$ – a is projected onto itself
  - if b is perpendicular to a then $a^T b = 0$

Example

Find the projection matrix onto the line through $a = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$

EURECOM

Projections

Example

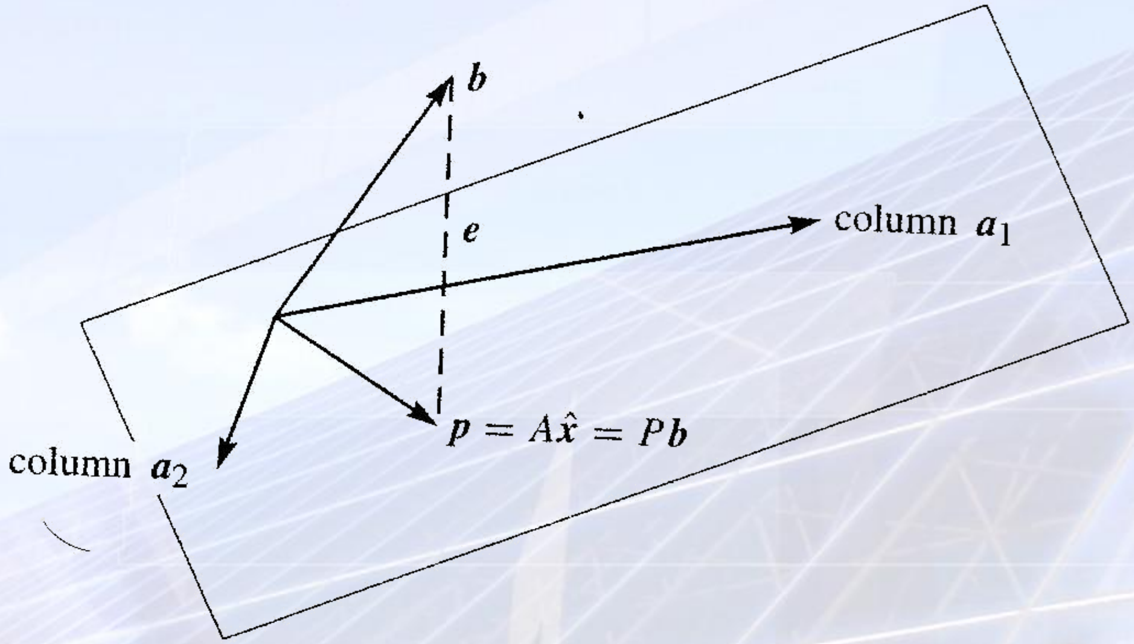
Project $b = [1\ 1\ 1]^T$ onto $a = [1\ 2\ 2]^T$ to find p

What happens if we project this new vector again?

EURECOM

Projections onto a subspace

- with n linearly independent vectors in R^m find the combination that is closest to a given vector b
  - each b in R^m is projected onto a's subspace



Projections onto a subspace

- e goes from b to the nearest point $A\hat{x}$ in the subspace
  - the error vector $e = b - A\hat{x}$ is perpendicular to the subspace
  - it makes a right angle with all the vectors $a_1, \ldots, a_n$

$$\begin{matrix} a_1^T (b - A\hat{x}) = 0 \\ \vdots \\ a_n^T (b - A\hat{x}) = 0 \end{matrix} \quad \text{or} \quad \begin{bmatrix} - a_1^T - \\ \vdots \\ - a_n^T - \end{bmatrix} \begin{bmatrix} \\ b - A\hat{x} \\ \\ \end{bmatrix} = \begin{bmatrix} \\ 0 \\ \\ \end{bmatrix}$$

$$A^T (b - A\hat{x}) = 0 \quad \text{or} \quad A^T A\hat{x} = A^T b$$

  - $A^T A$ is symmetric and it's invertible if the a's are independent

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

  - the projection of b onto the subspace is the vector

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

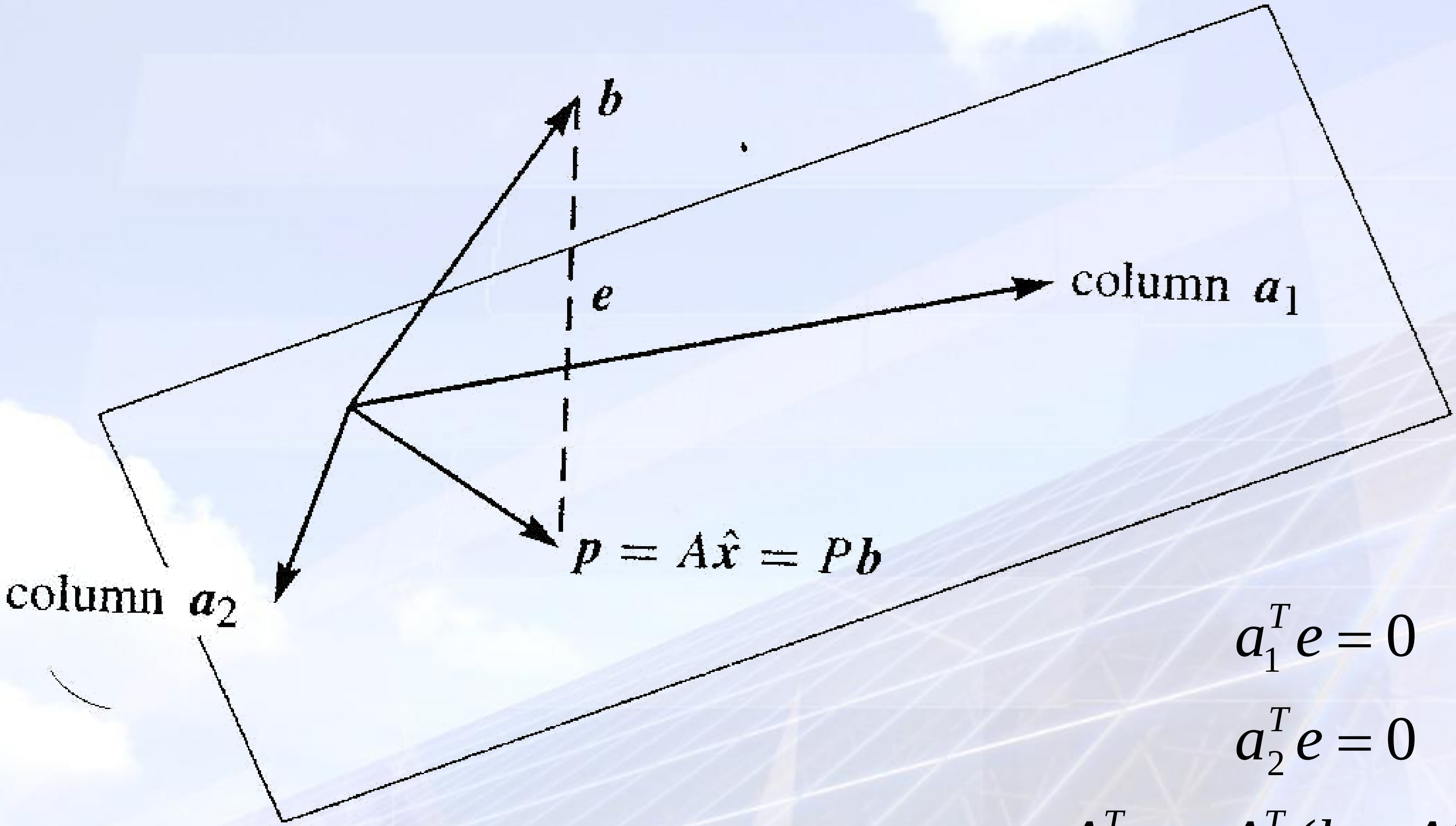
  - and the $n \times n$ projection matrix that produces $p = Pb$ is

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

Projections onto a subspace

- the key equation $A^T\left(b - A\hat{x}\right) = 0$ is given by linear algebra
  - the subspace is the column space of A
  - $e = b - A\hat{x}$ is perpendicular to the column space
  - $b - A\hat{x}$ is in the left nullspace, therefore $A^T\left(b - A\hat{x}\right) = 0$
- thus the left nullspace is important in projections and contains the error vector $e = b - A\hat{x}$
  - b is split into the projection p and the error $e = b - p$

Projections onto a subspace



$$a_1^T e = 0$$

$$a_2^T e = 0$$

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

Projections onto a subspace

Example

Find $\hat{x}$, p and P when

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \quad \text{and} \quad b = \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix}$$

Question

What is the effect of multiplying a projected vector by the same P again?

Least squares approximation

- sometimes, for example when we have too many equations, there is no solution to $Ax = b$
  - e.g. when $m > n$, the n columns span a small part of the m-dimensional space
  - unless all measurements are perfect $e = b - Ax \neq 0$
    - when $e = 0$, x is an exact solution to $Ax = b$
    - when $||e||$ is as small as possible $\hat{x}$ is a least squares solution

- when there is no solution, we instead solve $A^T A\hat{x} = A^T \mathrm{b}$
  - every vector in b has a part in the column space (p) and another perpendicular part in the left nullspace (e) – so remove e

$$Ax = b \;= p + e \quad \text{is impossible}\,; \quad Ax = p \quad \text{is solvable}$$

EURECOM

Least squares approximation

- the solution $\hat{x}$ to $Ax = p$ makes the error as small as possible because for any x

$$\| Ax - b \|^2 = \| Ax - p \|^2 + \| e \|^2$$

this is simply $a^2 + b^2 = c^2$ for a right angle triangle

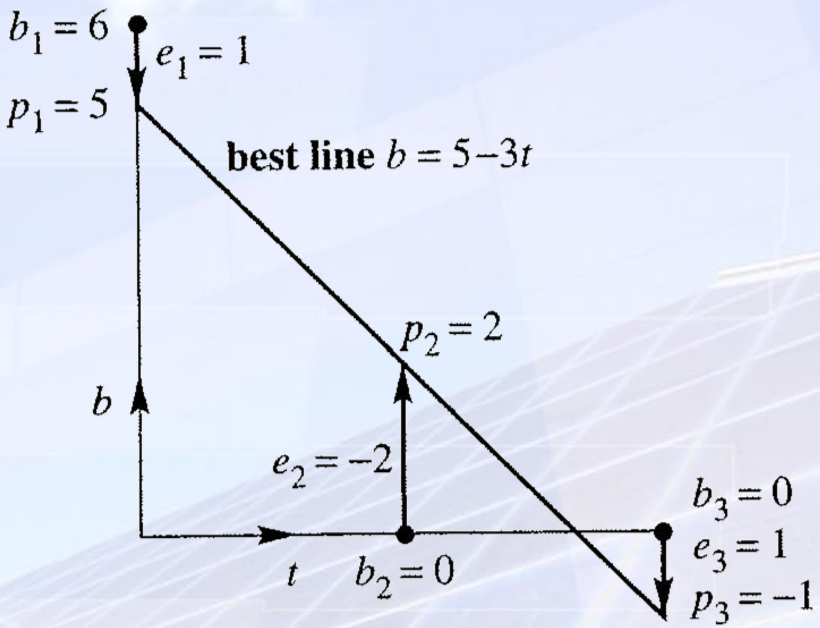
- vector $Ax - p$ in the column space is perpendicular to e in the left nullspace

- $Ax - p$ is reduced to zero by choosing x to be $\hat{x}$
  - this leaves the smallest possible error, i.e. e

Least squares approximation

Example

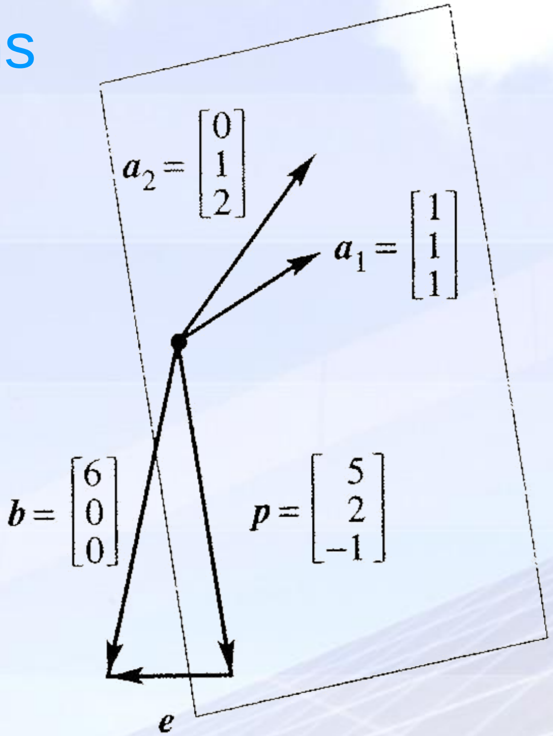
Find a line $b = C+Dt$ through the three points (0,6), (1,0) and (2,0).



- the solution is $C = 5$ and $D = -3$, i.e. $b = 5 - 3t$ is the best line
  - note the vertical error distances e_1, e_2, e_3 chosen to minimise the total error $E = e_1^2 + e_2^2 + e_3^2$

Least squares approximation

- and in 3 dimensions

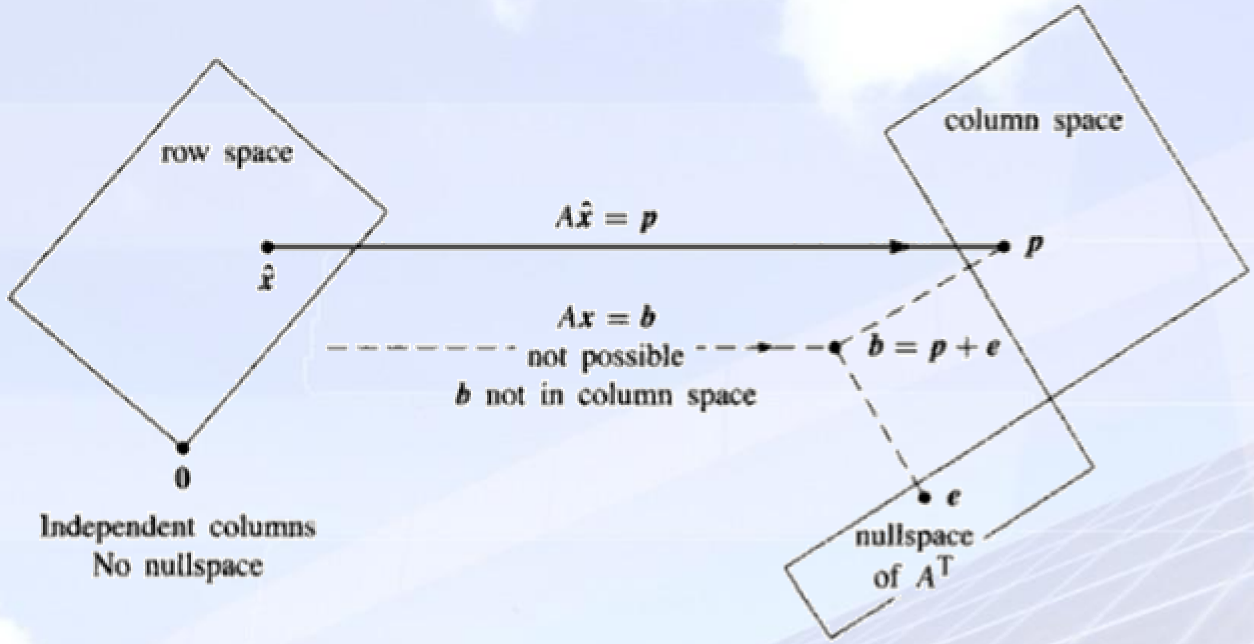


$$A^T A\hat{x} = A^T b$$

- b is not in the column space of A – hence there's no solution
  - the smallest error is $e = b - A\hat{x}$ which is perpendicular to the plane

EURECOM

Least squares approximation



- here we are splitting up b
  - instead of solving $Ax = b$ we solve $A\hat{x} = p$
  - error $e = b - p$

Fitting a straight line

- we wish to fit a line to some points (t_i,b_i)

$$Ax = b \quad \text{is} \quad \begin{array}{c} C + Dt_1 = b_1 \\ C + Dt_2 = b_2 \\ \vdots \\ C + Dt_m = b_m \end{array} \quad \text{and} \quad A = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix}$$

  the column space is so thin that b most probably lies outside

  – components of e are the vertical distances $e_1,\dots,e_m$ to the closest line which has heights $p_1,\dots,p_m$

- the equations $A^T A\hat{x} = A^T b$ give $\hat{x} = (C, D)$

  – the errors are $e_i = b_i - C - Dt_i$

Fitting a straight line

- $b = C + Dt$ exactly fits the data points if

$$\begin{matrix} C + Dt_1 = b_1 \\ \vdots \\ C + Dt_m = b_m \end{matrix} \quad \text{or} \quad \begin{bmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$$

- the columns are independent so we multiply by A^T to get

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

$$A^T A = \begin{bmatrix} 1 & \cdots & 1 \\ t_1 & \cdots & t_m \end{bmatrix} \begin{bmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix} = \begin{bmatrix} m & \Sigma t_i \\ \Sigma t_i & \Sigma t_i^2 \end{bmatrix}$$

$$A^T b = \begin{bmatrix} 1 & \cdots & 1 \\ t_1 & \cdots & t_m \end{bmatrix} \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} = \begin{bmatrix} \Sigma b_i \\ \Sigma t_i b_i \end{bmatrix}$$

Fitting a straight line

- the line $C + Dt$ which minimises $e_1^2 + \ldots + e_m^2$ is determined by

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

$$\begin{bmatrix} m & \Sigma t_i \\ \Sigma t_i & \Sigma t_i^2 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} \Sigma b_i \\ \Sigma t_i b_i \end{bmatrix}$$

- the vertical errors at the m points on the line are the components of $e = b - p = b - A\hat{x}$ which is perpendicular to the columns of A

- it is in the nullspace of A^T and the best $\hat{x} = (C, D)$ minimises the total error E, the sum of the squares

$$E(x) = \| Ax - b \|^2 = (C + Dt_1 - b_1)^2 + \cdots + (C + Dt_m - b_m)^2$$

EURECOM

Fitting a parabola

- problems with parabolas $b = C + Dt + Et^2$ are still problems in **linear** algebra

- e.g. fitting $b_1, \ldots, b_m$ at times $t_1, \ldots, t_m$ to such a parabola

$$\begin{aligned} C + Dt_1 + Et_1^2 &= b_1 \\ &\vdots \\ C + Dt_m + Et_1^2 &= b_m \end{aligned} \quad \text{has the } m \text{ by 3 matrix} \quad A = \begin{bmatrix} 1 & t_1 & t_1^2 \\ \vdots & \vdots & \vdots \\ 1 & t_m & t_m^2 \end{bmatrix}$$

the best parabola chooses $\hat{x} = (C, D, E)$ to satisfy the three normal equations

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

Orthogonal bases and Gram-Schmidt

- orthogonal vectors are good
  - can make A^TA diagonal

  we want to be able to choose orthogonal vectors

- vectors $q_1,\ldots,q_n$ are orthogonal when their dot products are zero, i.e. $q_i \cdot q_j$ or $q_i^T q_j = 0$ for $i \neq j$

- they are orthonormal if for each i, $||q_i|| = 1$

- the matrix of orthonormal vectors is written Q
  - Q is easy to work with since $Q^TQ = I$ and $Q^T = Q^{-1}$
  - if its only orthogonal then it's still diagonal

Orthogonal bases and Gram-Schmidt

- Q matrices are important for numerical computation because

$$\| Qx \|^2 = \| x \|^2$$

$$\text{because} \quad (Qx)^T (Qx) = x^T Q^T Qx = x^T Ix = x^T x$$

numbers can never grow too large when vectors of fixed length are used – computer codes which use Q matrices are numerically stable

EURECOM

Projections using orthogonal bases

- when we have Q instead of A projection formulae become

$$A^T A\hat{x} = A^T b \qquad \hat{x} = Q^T b$$

$$p = A\hat{x} \qquad p = Q\hat{x}$$

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

and there are no matrices to invert

- there are n separate 1-dimensional projections and there is no coupling (A^TA) from A matrices

$$p = Q\hat{x} = QQ^T b = \begin{bmatrix} | & & | \\ q_1 & \cdots & q_n \\ | & & | \end{bmatrix} \begin{bmatrix} q_1^T b \\ \vdots \\ q_n^T b \end{bmatrix} = q_1(q_1^T b) + \cdots + q_n(q_n^T b)$$

Projections using orthogonal bases

- when Q is square the subspace is the whole space
  - $Q^T = Q^{-1}$ and $\hat{x} = Q^T b$ is equal to $x = Q^{-1} b$
  - b's projection onto the whole space is b itself
  - $P = QQ^T = I$ and the solution is exact!

- when $p = b$ the formulae decomposes b into its constituent 1-dimensional projections

$$b = q_1(q_1^T b) + \cdots + q_n(q_n^T b)$$

  the vector is decomposed into perpendicular components
  - Fourier!

Projections using orthogonal bases

Example

The orthogonal matrix has orthonormal columns q^1, q^2 and q^3

$$Q = \frac{1}{3}\begin{bmatrix} -1 & 2 & 2 \\ 2 & -1 & 2 \\ 2 & 2 & -1 \end{bmatrix} \quad \text{has first column} \quad q_1 = \begin{bmatrix} -\frac{1}{3} \\ \frac{2}{3} \\ \frac{2}{3} \end{bmatrix}$$

the separate projections of $b = (0, 0, 1)$ onto the columns is

$$q_1(q_1^T b) = \frac{2}{3} q_1 \quad \text{and} \quad q_2(q_2^T b) = \frac{2}{3} q_2 \quad \text{and} \quad q_3(q_3^T b) = -\frac{1}{3} q_3$$

their sum is the projection of b onto the whole space, b

$$\frac{2}{3} q_1 + \frac{2}{3} q_2 - \frac{1}{3} q_3 = \frac{1}{9}\begin{bmatrix} -\ 2 + 4 - 2 \\ 4 - 2 - 2 \\ 4 + 4 + 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = b$$

The Gram-Schmidt Process

- a means of obtaining orthonormal vectors
- we start with (e.g. 3) vectors a, b, c and want to determine a number of orthonormal bases, say A, B, C
- begin by choosing $A = a$ as the first direction
- the next direction B must be perpendicular to A
  – start with b and remove its projection onto A

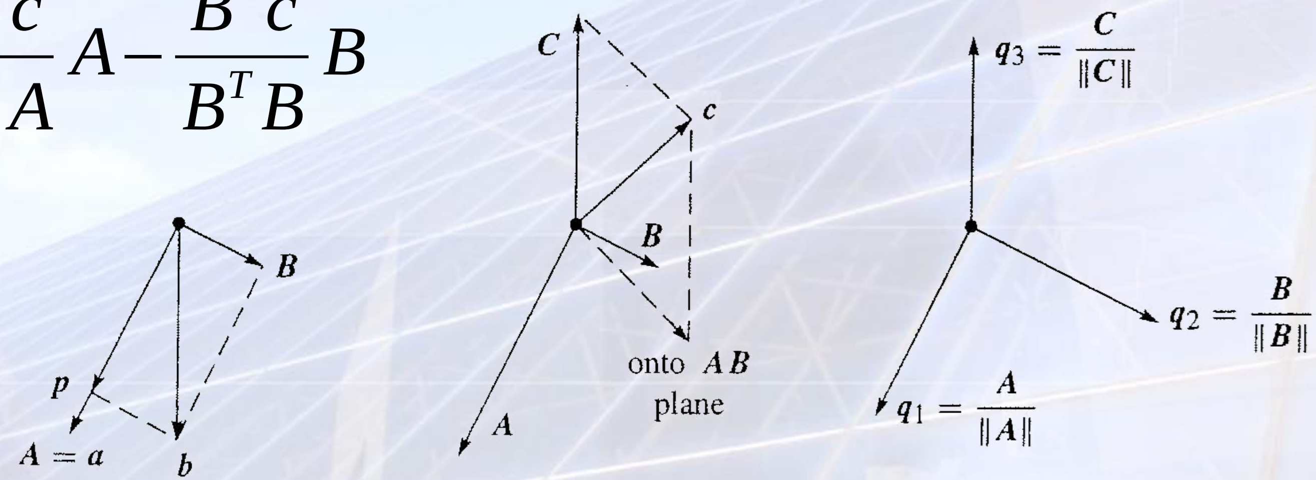
$$B = b - \frac{A^T b}{A^T A} A$$

EURECOM

The Gram-Schmidt Process

- by taking dot products we can verify that A and B are perpendicular, $A^T B = 0$
- B is the error vector e, that is perpendicular to A
- the third direction starts with c

$$C = c - \frac{A^T c}{A^T A} A - \frac{B^T c}{B^T B} B$$



EURECOM

The Gram-Schmidt Process

Example

Calculate an orthonormal basis for the three vectors

$$a = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \quad \text{and} \quad b = \begin{bmatrix} 2 \\ 0 \\ -2 \end{bmatrix} \quad \text{and} \quad c = \begin{bmatrix} 3 \\ -3 \\ 3 \end{bmatrix}$$

EURECOM

The factorisation $A = QR$

- there is a matrix connecting A and Q

- we can see it easily already

$$\begin{bmatrix} a & b & c \end{bmatrix} = \begin{bmatrix} q_1 & q_2 & q_3 \end{bmatrix} \begin{bmatrix} q_1^T a & q_1^T b & q_1^T c \\ & q_2^T b & q_2^T c \\ & & q_3^T c \end{bmatrix} \quad \text{or} \quad A = QR$$

- e.g. from the last example

$$A = \begin{bmatrix} 1 & 2 & 3 \\ -1 & 0 & -3 \\ 0 & -2 & 3 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{6} & 1/\sqrt{3} \\ -1/\sqrt{2} & 1/\sqrt{6} & 1/\sqrt{3} \\ 0 & -2/\sqrt{6} & 1/\sqrt{3} \end{bmatrix} \begin{bmatrix} \sqrt{2} & \sqrt{2} & \sqrt{18} \\ 0 & \sqrt{6} & -\sqrt{6} \\ 0 & 0 & \sqrt{3} \end{bmatrix} = QR$$

The factorisation $A = QR$

- the square matrix R is upper triangular with positive diagonal

- it's useful for least squares

$$A^T A = R^T Q^T QR = R^T R$$

- the least squares equation simplifies to

$$R^T R\hat{x} = R^T Q^T b \quad \text{or} \quad R\hat{x} = Q^T b$$

- instead of solving $Ax = b$, which is impossible, we solve $R\hat{x} = Q^T b$ by back substitution – which is very fast

EURECOM

The factorisation $A = QR$

Example

Determine the QR decomposition of

$$A = \begin{bmatrix} 12 & -51 & 4 \\ 6 & 167 & -68 \\ -4 & 24 & -41 \end{bmatrix}$$

For your own time if you wish - you will **not** have a question like this in your exam, but you **should** understand the concepts.

Eigenvalues and Eigenvectors

- take a matrix and look at its powers

$$\begin{array}{cccc} A & A^2 & A^3 & A^{100} \\ \begin{bmatrix} 0.8 & 0.3 \\ 0.2 & 0.7 \end{bmatrix} & \begin{bmatrix} 0.7 & 0.45 \\ 0.3 & 0.55 \end{bmatrix} & \begin{bmatrix} 0.650 & 0.525 \\ 0.350 & 0.475 \end{bmatrix} & \begin{bmatrix} 0.6000 & 0.6000 \\ 0.4000 & 0.4000 \end{bmatrix} \end{array}$$

- A^{100} was found by looking at eigenvalues
- certain special vectors x do not change direction when multiplied by A – they are the eigenvectors of A
- they can, however, change their length and do so according to the corresponding eigenvalue, λ

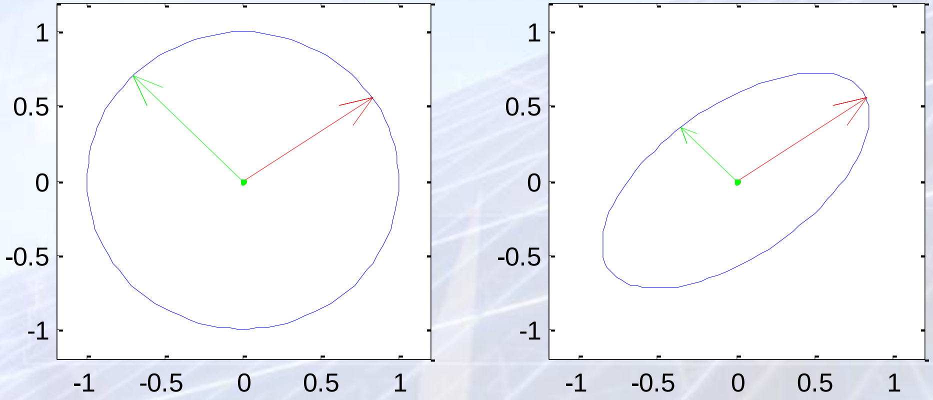
Eigenvalues and Eigenvectors

- the basic equation

$$Ax = \lambda x$$

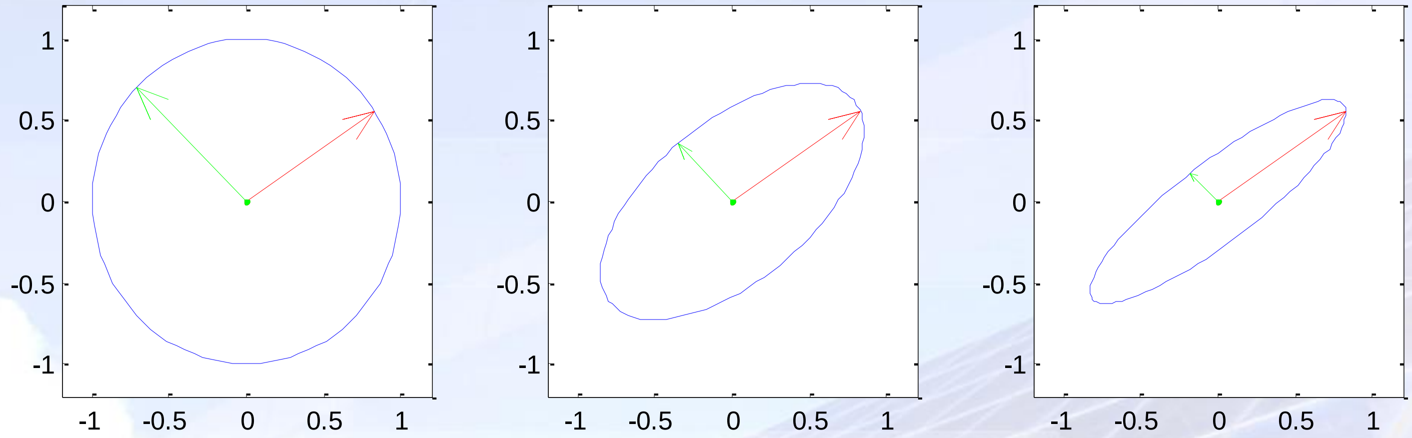
- example

$$A = \begin{bmatrix} 0.8 & 0.3 \\ 0.2 & 0.7 \end{bmatrix} \qquad S = \begin{bmatrix} 0.6 & 1 \\ 0.4 & -1 \end{bmatrix} \qquad V = \begin{bmatrix} 1 & 0 \\ 0 & 0.5 \end{bmatrix}$$



Eigenvalues and Eigenvectors

- what happens if we multiply by A again?



$$A^{99}\begin{bmatrix}0.8\\0.2\end{bmatrix} \quad \text{is really} \quad x_1 + (0.2)\left(\frac{1}{2}\right)^{99} x_2 = \begin{bmatrix}0.6\\0.4\end{bmatrix} + \begin{bmatrix}\text{very}\\\text{small}\\\text{number}\end{bmatrix}$$

Eigenvalues and Eigenvectors

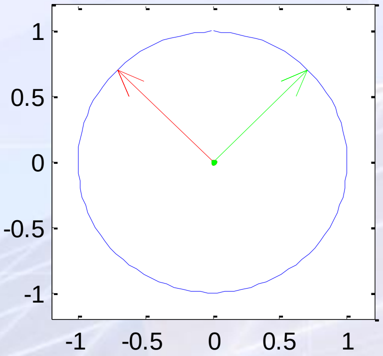
- eigenvector x_1 is a steady-state that doesn't change
  - $\lambda_1 = 1$
- eigenvector x_2 is a decaying mode
  - $\lambda_1 = 0.5$
- the higher the power of A the closer its columns approach the steady state
- A is a Markov matrix
  - largest eigenvalue of 1
  - [0.6 0.4]' is the steady state

Eigenvalues and Eigenvectors

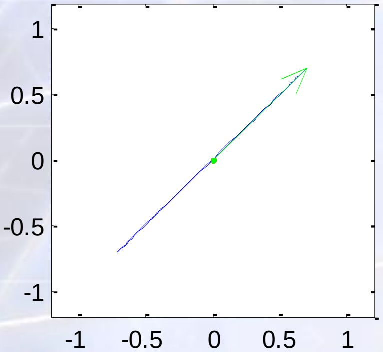
Example

The projection matrix P has eigenvalues of 0 and 1. Its eigenvectors are x_1=(1,1) and x_2=(1,-1).

$$P = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix}$$



- each column of P adds up to 1, so $\lambda = 1$ is an eigenvalue
- P is singular, so $\lambda = 0$ is an eigenvalue
- P is symmetric, so x_1 and x_2 are perpendicular



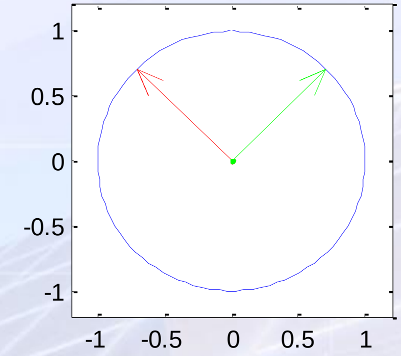
What can you say about nullspace and the column space?

Eigenvalues and Eigenvectors

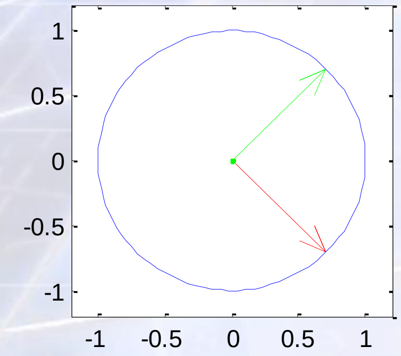
$$R = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

Example

The reflection matrix R has eigenvalues of 1 and -1.



- eigenvector x_1=(1,1) is unchanged
- eigenvector x_2=(-1,1) is reflected
- the eigenvectors are the same as for P



The eigenvalue problem

- refers to the seeking of non-trivial solutions to

$$Ax = \lambda x$$
$$(\lambda I - A)x = 0$$

  values of the scalar λ are the eigenvalues and the corresponding values of $x \neq 0$ are the eigenvectors which make up the nullspace of $A - \lambda I$

- a non-trivial solution exists if $A - \lambda I$ is not invertible, or if $\det(A - \lambda I) = 0$

- for a matrix A of size $n \times n$, there will be n eigenvalues

- each eigenvalue leads to a corresponding eigenvector

The eigenvalue problem

Example

Calculate the eigenvalues and eigenvectors of:

$$A = \begin{bmatrix} -2 & 1 \\ 1 & -2 \end{bmatrix} \qquad B = \begin{bmatrix} 1 & 1 & -2 \\ -1 & 2 & 1 \\ 0 & 1 & -1 \end{bmatrix}$$

Example

Calculate the eigenvalues and eigenvectors of:

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \text{ and } B = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

where we have repeated eigenvalues

Properties of eigenvalues for $n \times n$ matrices

- ## Property 1
  The sum of the eigenvalues:
  $$\sum_{i=1}^{n} \lambda_i = \text{trace } A = \sum_{i=1}^{n} a_{ii}$$

- ## Property 2
  The product of the eigenvalues:
  $$\prod_{i-1}^{n} \lambda_i = |A|$$

- ## Property 3
  The eigenvalues of the inverse, if it exists, are:
  $$\frac{1}{\lambda_1}, \frac{1}{\lambda_2}, \cdots, \frac{1}{\lambda_n}$$

- ## Property 4
  The eigenvalues of the transpose are:
  $$\lambda_1, \lambda_2, \cdots, \lambda_n$$

- ## Property 5
  If k is a scalar then the eigenvalues of kA are:
  $$k\lambda_1, k\lambda_2, \cdots, k\lambda_n$$

- ## Property 6
  If k is a scalar then the eigenvalues of $A + kI$ are:
  $$\lambda_1 \pm k, \lambda_2 \pm k, \cdots, \lambda_n \pm k$$

- ## Property 6
  If k is a positive integer then the eigenvalues of A^k are:
  $$\lambda_1^k, \lambda_2^k, \cdots, \lambda_n^k$$

Diagonalisation

- when x is an eigenvector multiplication by A is equivalent to multiplication by a single number: $Ax = \lambda x$

- it is like having a diagonal matrix
  - the 100^{th} power of a diagonal matrix is easy to compute

- here we show that A turns into a diagonal matrix when the eigenvalues are properly used

Diagonalisation

- suppose $n \times n$ matrix A has n linearly independent eigenvectors
- if we put them into the columns of an eigenvector matrix, S, then

$$S^{-1}AS = \Lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$$

Proof:

$$AS = A\begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} = \begin{bmatrix} \lambda_1 x_1 & \cdots & \lambda_n x_n \end{bmatrix}$$

$$\begin{bmatrix} \lambda_1 x_1 & \cdots & \lambda_n x_n \end{bmatrix} = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} = S\Lambda$$

$$AS = S\Lambda \quad \text{is} \quad S^{-1}AS = \Lambda \quad \text{or} \quad A = S\Lambda S^{-1}$$

Diagonalisation

- some points:
  - S has an inverse because we assumed the eigenvectors were linearly independent
    - without linearly independent eigenvectors we can’t diagonalise the matrix
  - matrices A and Λ have the same eigenvalues but different eigenvectors
    - the eigenvectors in S diagonalise A
    - the eigenvectors of Λ are NOT the same as those of A and are the columns of I
      - diagonalisation aligns the new eigenvectors with the coordinate axes

Diagonalisation

An example for the projection matrix P which has eigenvalues of 1 and 0 and corresponding eigenvectors (1,1) and (-1,1)

$$\begin{bmatrix} .5 & .5 \\ -.5 & .5 \end{bmatrix} \begin{bmatrix} .5 & .5 \\ .5 & .5 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

$$S^{-1} \qquad P \qquad S \qquad = \qquad \Lambda$$

Note that

- $P^2 = P$
- $\Lambda^2 = \Lambda$
- the column space has swung around from (1,1) to (1,0)
- the nullspace has swung around from (-1,1) to (0,1)
- diagonalisation lines up the **new** eigenvectors with the co-ordinate axes

Diagonalisation

- suppose all eigenvalues of A are different
  - then the eigenvectors are independent
  - then A is diagonalisable

proofs in course text

- the eigenvector matrix is not unique
  - multiplying its columns by any non-zero constant gives the same Λ
- to diagonalise A we must use an eigenvector matrix
- matrices with repeated eigenvalues (too few) are not diagonalisable

Diagonalisation

- our Markov matrix A can be diagonalised as follows

$$A=\begin{bmatrix}0.8 & 0.3\\0.2 & 0.7\end{bmatrix}=\begin{bmatrix}0.6 & 1\\0.4 & -1\end{bmatrix}\begin{bmatrix}1 & 0\\0 & 0.5\end{bmatrix}\begin{bmatrix}1 & 1\\0.4 & -0.6\end{bmatrix}=S\Lambda S^{-1}$$

- note also that, along the same lines as before

$$A^k=S\Lambda^k S^{-1}=\begin{bmatrix}0.6 & 1\\0.4 & -1\end{bmatrix}\begin{bmatrix}1^k & 0\\0 & 0.5^k\end{bmatrix}\begin{bmatrix}1 & 1\\0.4 & -0.6\end{bmatrix}$$

$$A^\infty=\begin{bmatrix}0.6 & 1\\0.4 & -1\end{bmatrix}\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix}\begin{bmatrix}1 & 1\\0.4 & -0.6\end{bmatrix}=\begin{bmatrix}0.6 & 0.6\\0.4 & 0.4\end{bmatrix}$$

Diagonalisation

- what's the point?
- we've already seen their use in calculating matrix powers
- there are more applications
  - Fibonacci sequences
  - Markov processes
  - differential equations
  - exponential of a matrix
  - quadratic forms

Symmetric matrices

- $A = S\Lambda S^{-1}$ has particular properties when A is symmetric

- the ***spectral theorem*** tells us that for a symmetric matrix
  - the eigenvalues are real
  - the eigenvectors can be chosen to be orthonormal

- we denote such a diagonalisation by $A = Q\Lambda Q^{-1} = Q\Lambda Q^{\mathrm{T}}$

- to see that $Q\Lambda Q^{\mathrm{T}}$ is symmetric take its transpose

$$\left(Q\Lambda Q^T\right)^T = \left(Q^T\right)^T \Lambda^T Q^T = Q\Lambda Q^T$$

EURECOM

Symmetric matrices

Example

Obtain the eigenvalues and corresponding orthogonal eigenvectors of the symmetric matrix

$$A = \begin{bmatrix} 2 & 2 & 0 \\ 2 & 5 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

and show that the normalised eigenvectors form an orthonormal set.

EURECOM

Singular value decomposition (SVD)

- A is an $m \times n$ matrix, square or rectangular
  - row space is r-dimensional in R^n
  - column space is r-dimensional in R^m
  - we will choose orthonormal bases for these spaces
    - row space basis $v_1,\ldots,v_r$
    - column space basis $u_1,\ldots,u_r$

- e.g. for an invertible 2×2 matrix where $m=n=r=2$
  - row space is the plane R^2
  - we want v_1 and v_2 to be perpendicular unit vectors
    - an orthogonal basis
  - also want Av_1 and Av_2 to be perpendicular
  - unit vectors $u_1=Av_1/||Av_1||$ and $u_2=Av_2/||Av_2||$ will be orthogonal

EURECOM

Singular value decomposition (SVD)

- for the matrix

$$A = \begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix}$$

  note that

  – no single orthogonal basis Q will make $Q^{-1}AQ$ diagonal

  – we cannot use the eigenvectors of A to form the basis as they aren't orthonormal

  – A is not symmetric and we need two different orthogonal matrices to diagonalise it

Singular value decomposition (SVD)

- inputs v_1 and v_2 give outputs Av_1 and Av_2
  - we want them to line up with u_1 and u_2

- the basis vectors have to give $Av_1=\sigma_1 u_1$ and $Av_2=\sigma_2 u_2$
  - σ_1 and σ_2 are just the lengths $||Av_1||$ and $||Av_2||$

- with v_1 and v_2 as columns

$$A\begin{bmatrix} \vdots & \vdots \\ v_1 & v_2 \\ \vdots & \vdots \end{bmatrix} = \begin{bmatrix} \vdots & \vdots \\ \sigma_1 u_1 & \sigma_2 u_2 \\ \vdots & \vdots \end{bmatrix} = \begin{bmatrix} \vdots & \vdots \\ u_1 & u_2 \\ \vdots & \vdots \end{bmatrix} \begin{bmatrix} \sigma_1 & \\ & \sigma_2 \end{bmatrix}$$

$$AV = U\Sigma$$

Singular value decomposition (SVD)

$$A\begin{bmatrix} \vdots & \vdots \\ v_1 & v_2 \\ \vdots & \vdots \end{bmatrix} = \begin{bmatrix} \vdots & \vdots \\ \sigma_1 u_1 & \sigma_2 u_2 \\ \vdots & \vdots \end{bmatrix} = \begin{bmatrix} \vdots & \vdots \\ u_1 & u_2 \\ \vdots & \vdots \end{bmatrix}\begin{bmatrix} \sigma_1 & \\ & \sigma_2 \end{bmatrix}$$

$$AV = U\Sigma$$

- Σ is diagonal and like the matrix Λ
  - Λ contained the eigenvalues
  - Σ contains the singular values σ_1 and σ_2

- when U and V both equal S we have $AS = S\Lambda$ which gives $S^{-1}AS = \Lambda$ - it is diagonalised
  - but the vectors in S are not generally orthogonal

Singular value decomposition (SVD)

- we require U and V to be orthogonal
  - basis vectors in their columns must be orthonormal

$$V^TV = \begin{bmatrix} \cdots & v_1^T & \cdots \\ \cdots & v_2^T & \cdots \end{bmatrix} \begin{bmatrix} \vdots & \vdots \\ v_1 & v_2 \\ \vdots & \vdots \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

  - thus $V^TV = I$ which means $V^T = V^{-1}$
    - similarly $U^TU = I$ and $U^T = U^{-1}$

- this is the SVD

$$AV = U\Sigma \quad \text{and then} \quad A = U\Sigma V^{-1} = U\Sigma V^T$$

where U and V are orthogonal

EURECOM

Singular value decomposition (SVD)

- to see V by itself multiply A^T by A

$$A^T A = \left(U\Sigma V^T\right)^T (U\Sigma V^T) = V\Sigma^T U^T U\Sigma V^T$$

$$= V\Sigma^T \Sigma V^T$$

$$= V \begin{bmatrix} \sigma_1^2 & \\ & \sigma_2^2 \end{bmatrix} V^T$$

which is an ordinary factorisation exactly like $A = Q\Lambda Q^T$

- except that A is really $A^T A$
- the columns of V are now the eigenvectors of $A^T A$!!!

EURECOM

Singular value decomposition (SVD)

Example

Show that the SVD of

$$A = \begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix}$$

is given by

$$A = U\Sigma V^T$$

$$\begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2\sqrt{2} & \\ & \sqrt{2} \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$$

Singular value decomposition (SVD)

- we can also calculate the u's first, then the v's

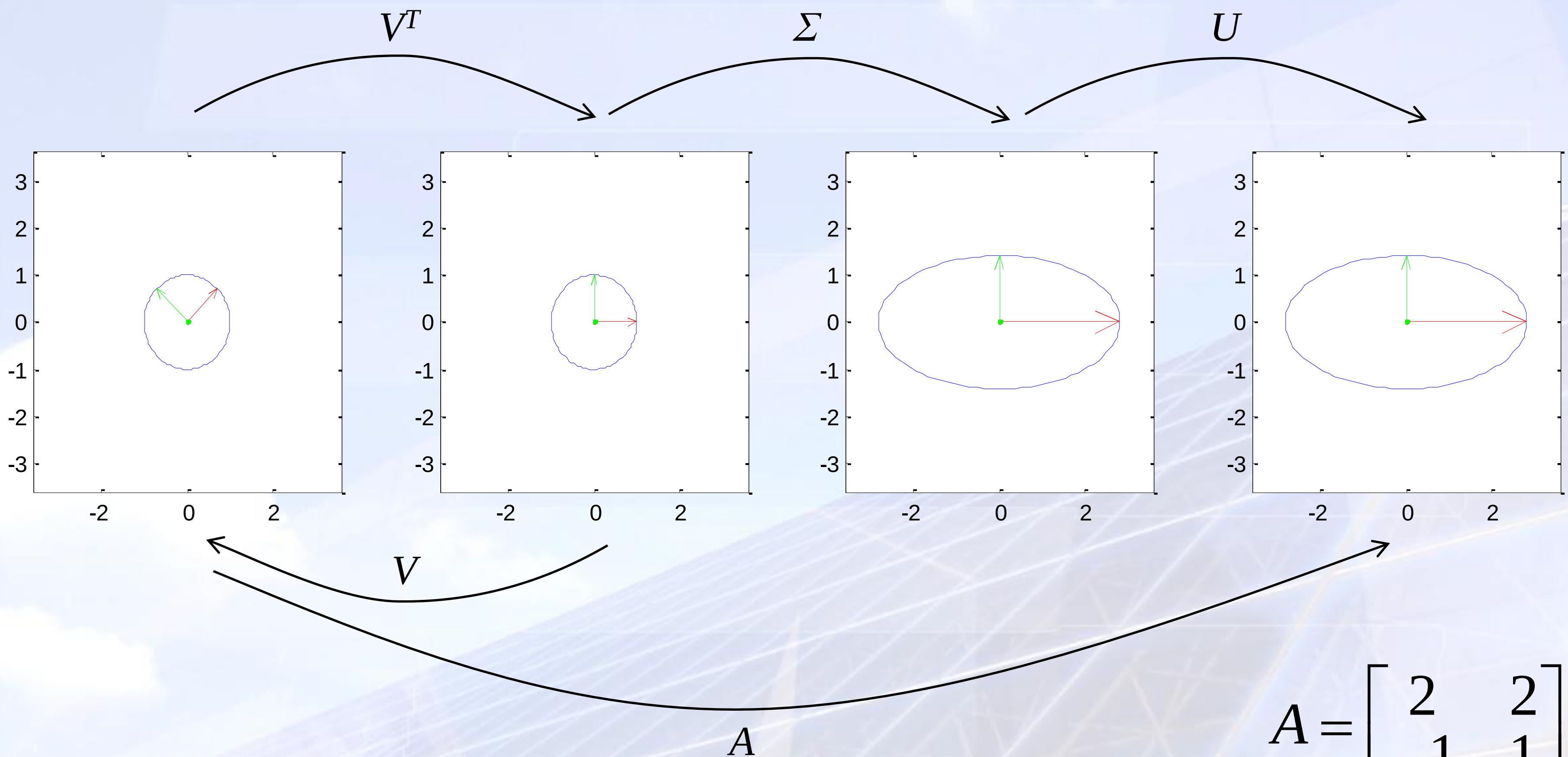
$$AA^T = (U\Sigma V^T)\left(U\Sigma V^T\right)^T = U\Sigma V^T V\Sigma^T U^T$$

$$= U\Sigma^T \Sigma U^T$$

$$= U\begin{bmatrix} \sigma_1^2 & \\ & \sigma_2^2 \end{bmatrix} U^T$$

- we have an ordinary factorisation of AA^T
  - columns of U are its eigenvectors
- we can show that it gives the same result as in the previous example

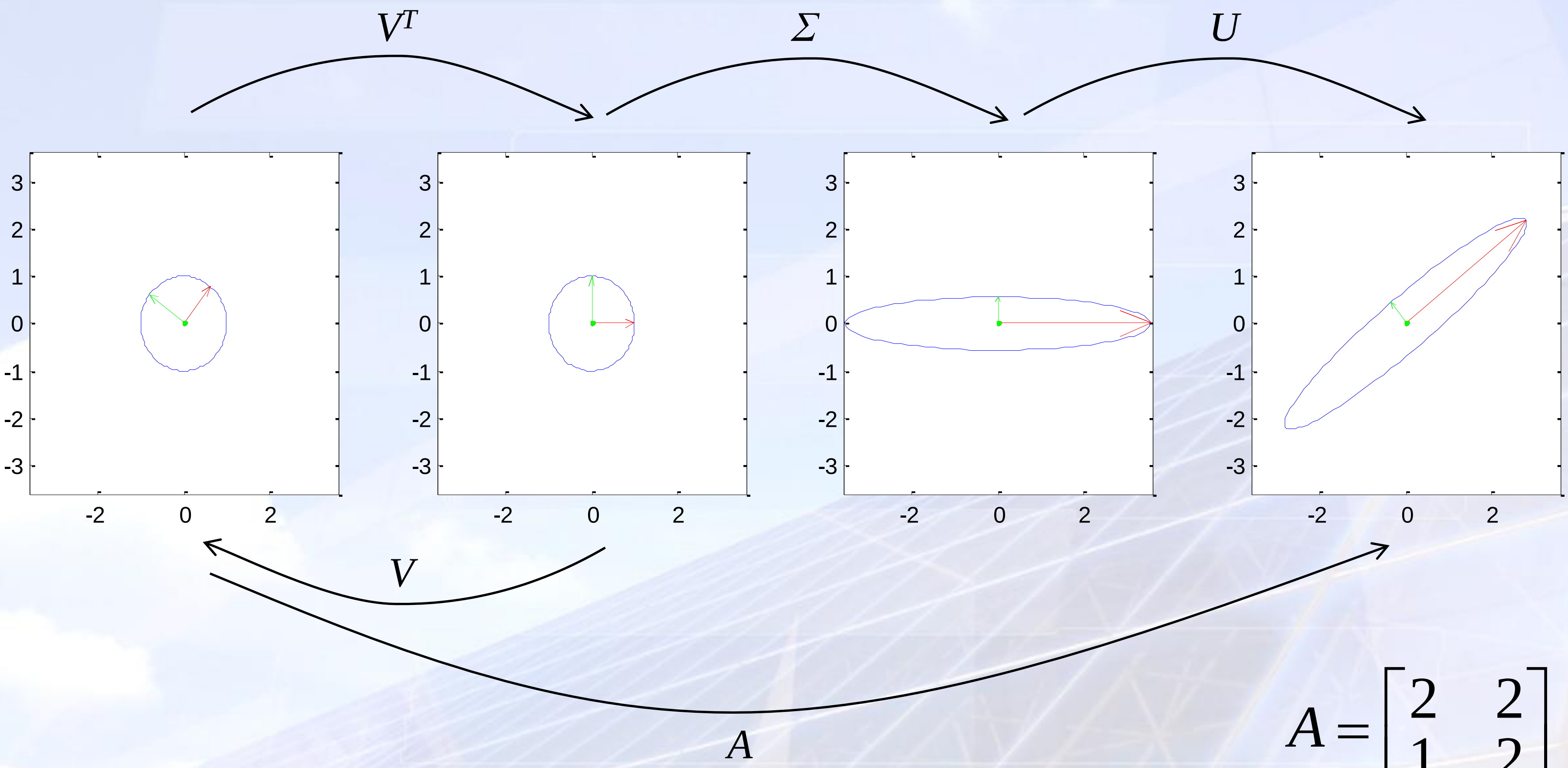
Singular value decomposition (SVD)



$$A = \begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix}$$

Singular value decomposition (SVD)

A **different** example:



$$A = \begin{bmatrix} 2 & 2 \\ 1 & 2 \end{bmatrix}$$

EURECOM

Singular value decomposition (SVD)

Example

Find the SVD of the following singular matrix:

$$A = \begin{bmatrix} 2 & 2 \\ 1 & 1 \end{bmatrix}$$

Remember that it is **singular**!

$$A = U\Sigma V^T$$

$$\begin{bmatrix} 2 & 2 \\ 1 & 1 \end{bmatrix} = \frac{1}{\sqrt{5}}\begin{bmatrix} 2 & 1 \\ 1 & -2 \end{bmatrix}\begin{bmatrix} \sqrt{10} & 0 \\ 0 & 0 \end{bmatrix}\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

Singular value decomposition (SVD)

- 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

- some points:
  - the v's are eigenvectors of A^TA
  - the u's are eigenvectors of AA^T
  - A^TA and AA^T have the same eigenvalues
  - Av_i has to fall in the direction of u_i, $Av_i=\sigma_i u_i$

Singular value decomposition (SVD)

- starting from $A^T A v_i = \sigma_i^2 v_i$ and multiplying
  - first by v_i^T

$$v_i^T A^T A v_i = \sigma_i^2 v_i^T v_i \quad \text{gives} \quad \|Av_i\|^2 = \sigma_i^2 \quad \text{so that} \quad \|Av_i\| = \sigma_i$$

  were we used $(v_i^T A^T)(Av_i) = ||Av_i||^2$

  - then by A

$$AA^T A v_i = \sigma_i^2 A v_i \quad \text{gives} \quad u_i = Av_i / \sigma_i \quad \text{as a unit eigenvector of} \quad AA^T$$

- this gives us the proof that $Av_i = \sigma_i u_i$
  - A is diagonalised by the two bases

$$A = U \Sigma V^T$$

Norms

- the norm of a matrix A is the maximum ratio

$$\|A\| = \max_{x \neq 0} \frac{\|Ax\|}{\|x\|}$$

- e.g. the norm of I is 1 since the ratio is always $||x||/||x||$
- e.g. the norm of a diagonal matrix is its largest entry

The norm of $A = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$ is $\|A\| = \max_{x \neq 0} \frac{\| Ax \|}{\| x \|} = \max_{x \neq 0} \frac{\sqrt{2^2 x_1^2 + 3^2 x_2^2}}{\sqrt{x_1^2 + x_2^2}} = 3$

its maximum is when $x_1 = 0$ and $x_2 = 1$

- $x = (0,1)^T$ is an eigenvector with $Ax = (0,3)^T$
- the eigenvalue is 3 and it equals the norm

EURECOM

Norms

- for a positive definite symmetric matrix the norm is $||A|| = \lambda_{max}$ - no other vector other than the corresponding eigenvector can make the norm larger
  - $A = Q\Lambda Q^{\mathrm{T}}$ – neither Q nor Q^T change the length
  - the ratio to maximise is really $||\Lambda x||/||x||$
- for symmetric matricies (not positive definite) $||A||=|\lambda_{max}|$
  - norm is concerned with length, $||Ax|| = ||\lambda x|| = |\lambda|.||x||$
- for unsymmetric matrices we cannot generally use eigenvalues e.g.

$$A = \begin{bmatrix} 0 & 2 \\ 0 & 0 \end{bmatrix} \quad \text{has} \quad \lambda_1 = \lambda_2 = 0 \quad \text{but} \quad \|\mathrm{A}\| = \frac{\|Ax\|}{\|x\|} = 2$$

EURECOM

Norms

- so how do we calculate norms generally?

- for the last example vector $x = (0,1)$ which gives $Ax = (2,0)$
  - $x = (0,1)$ is an eigenvector of A^TA

- the rule is

$$\|A\|^2 = \max_{x \neq 0} \frac{\|Ax\|^2}{\|x\|^2} = \max_{x \neq 0} \frac{x^T A^T Ax}{x^T x} = \lambda_{\max}(A^T A)$$

when x is the eigenvector corresponding to $\lambda_{\max}$ the ratio is $x^TA^TAx = x^T(\lambda_{\max})x$ divided by x^Tx

  - for this particular x, the ratio equals $\lambda_{\max}$

Norms

- the matrix A^TA has orthonormal eigenvectors $q_1, \ldots, q_n$
  - every x is a combination of these vectors
  - putting this combination into the ratio gives

$$\frac{x^T A^T A x}{x^T x} = \frac{(c_1 q_1 + \cdots + c_n q_n)^T (c_1 \lambda_1 q_1 + \cdots + c_n \lambda_n q_n)}{(c_1 q_1 + \cdots + c_n q_n)^T (c_1 q_1 + \cdots + c_n q_n)} = \frac{c_1^2 \lambda_1 + \cdots + c_n^2 \lambda_n}{c_1^2 + \cdots + c_n^2}$$

  - remember that $q_i^T q_j = 0$
  - this ratio cannot be larger than λ_{max} which occurs when all c's are zero except the one that multiplies λ_{max}

Norms

- the ratio on the last slide is the Rayleigh quotient for the matrix A^TA
  - the maximum is the largest eigenvalue $\lambda_{\max}(A^TA)$
  - the minimum is the smallest eigenvalue $\lambda_{\min}(A^TA)$

- the norm $||A||$ is equal to the largest singular value $\sigma_{\max}$ of A – the singular values are equal to the square roots of the positive eigenvalues of A^TA

Condition numbers

- indicates sensitivity to errors
  - when $Ax = b$ and the right side changes to $b + \Delta b$
  - the solution changes to $x + \Delta x$ – but how big is Δx?

Subtract $Ax = b$ from $A(x + \Delta x) = b + \Delta b$ to find $A(\Delta x) = \Delta b$

- the error is $\Delta x = A^{-1}\Delta b$
  - large when A^{-1} is large (it's nearly singular)
  - especially large when Δb points in the worst direction
  - the worst error has $||\Delta x|| = ||A^{-1}||.||\Delta b||$
- we really want to know the relative error

Condition numbers

- we find that the solution error is less than $c = ||A||.||A^{-1}||$ times the problem error

$$\frac{\|\Delta x\|}{\|x\|} \leq c \frac{\|\Delta b\|}{\|b\|}$$

- for proof, note that from $b = Ax$ and $\Delta x = A^{-1} \Delta b$ we note that $\|b\| \leq \|A\|.\|x\| \quad and \quad \|\Delta x\| \leq \left\|A^{-1}\right\|.\|\Delta b\|$. Multiplying left and right sides and then dividing by $||b||.||x||$
  - the left side gives relative error $||\Delta x||/ ||x||$
  - the right side is the equation above

Condition numbers

- if the error is in A then

$$\frac{\|\Delta x\|}{\|x+\Delta x\|} \le c\,\frac{\|\Delta A\|}{\|A\|}$$

- for proof we subtract $Ax = b$ from $(A + \Delta A)(x+\Delta x) = b$ to find $A(\Delta x) = -(\Delta A)(x+\Delta x)$ and multiply the last equation by A^{-1} and take norms to reach the above equation

$$\|\Delta x\| \le \|A^{-1}\|.\|\Delta A\|.\|x+\Delta x\| \quad \text{or} \quad \frac{\|\Delta x\|}{\|x+\Delta x\|} \le \|A\|.\|A^{-1}\|\frac{\|\Delta A\|}{\|A\|}$$

EURECOM

Condition numbers

- errors occur either in
  - b: depends on original measurement or computer roundoff
  - A: also depends on elimination steps
    - small pivots tend to produce large errors in L and U

- errors in A or in b are amplified into the solution error Δx
  - errors are bounded relative to x by the condition number, c
  - when the condition number is too large errors in the solution can become unacceptable

Condition numbers

Example

When A and A^{-1} are symmetric, $c = ||A||.||A^{-1}||$ comes from the eigenvalues:

$$A = \begin{bmatrix} 6 & 0 \\ 0 & 2 \end{bmatrix} \quad \text{has norm} \quad 6. \qquad A^{-1} = \begin{bmatrix} \frac{1}{6} & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \quad \text{has norm} \quad \frac{1}{2}$$

This A is symmetric positive definite. Its norm is $\lambda_{max} = 6$. The norm of A^{-1} is $1/\lambda_{min} = ½$. Multiplying these norms gives the condition number:

$$c = \| A \| . \| A^{-1} \| = \frac{\lambda_{max}}{\lambda_{min}} = 3$$

From $c = ||A||.||A^{-1}|| = 3$ the solution error is less than 3 times the problem error.

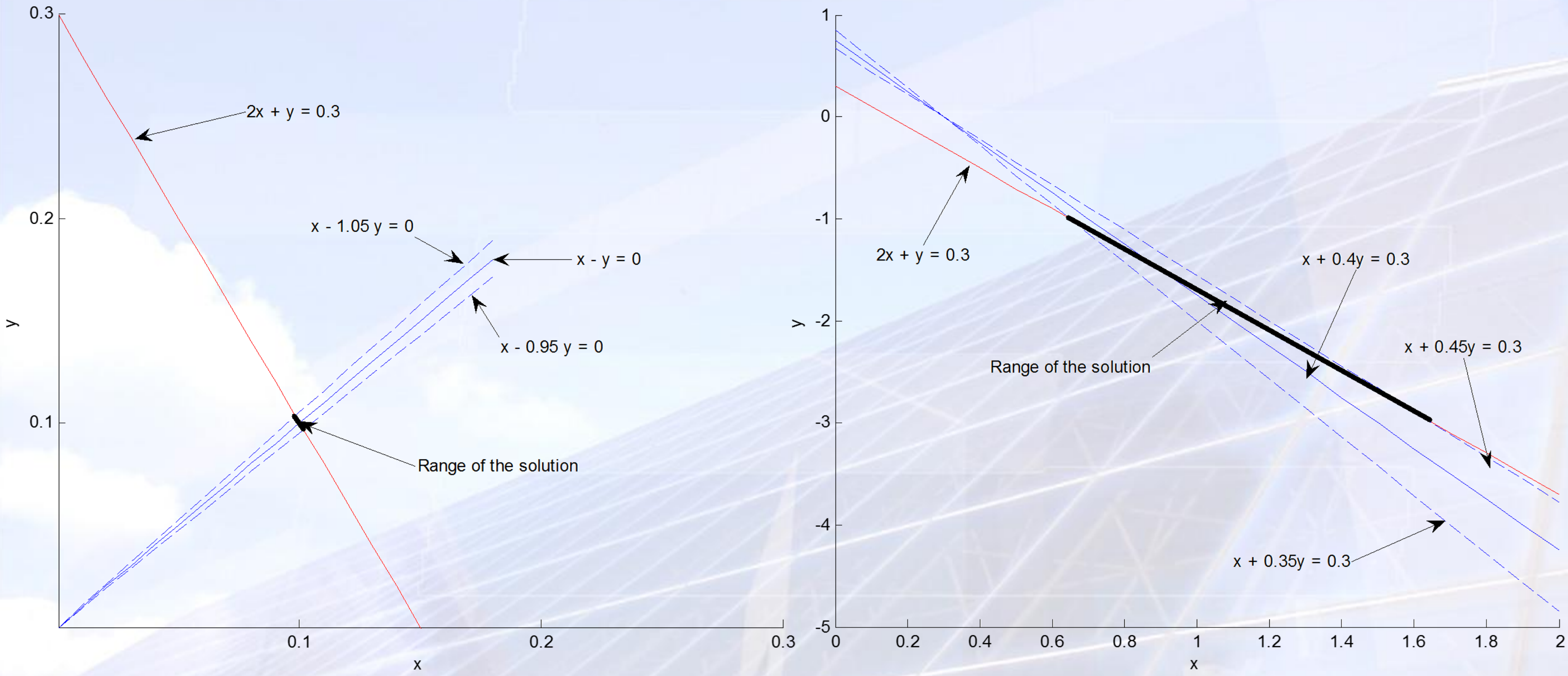
EURECOM

Ill conditioning

- suppose we have to solve $\begin{array}{l} 2x + y = 0.3 \\ x - \alpha y = 0 \end{array}$ where $\alpha = 1 \pm 0.05$ has some error
- we obtain $\begin{array}{l} x = 0.3\alpha / (1 + 2\alpha) \\ y = 0.3 / (1 + 2\alpha) \end{array}$
  - so $0.0983 \le x \le 0.1016$ and $0.0968 \le y \le 0.1034$
  - an error of $\pm$5% in the value of α produces errors in x of up to $\pm$2% and in y of up to $\pm$3%
- changing to $\begin{array}{l} 2x + y = 0.3 \\ x + \alpha y = 0.3 \end{array}$ where $\alpha = 0.4 \pm 0.05$ we have $\begin{array}{l} x = 0.3(1 - \alpha) / (1 - 2\alpha) \\ y = -0.3 / (1 - 2\alpha) \end{array}$
  - so $0.65 \le x \le 1.65$ and $-3 \le y \le -1$.
  - now an error of $\pm$12.5% in the value of α produces errors in x and y of up to 100%

Ill conditioning

- a small change in $x - \alpha y = 0$ makes only a small difference
- a large difference for $x + \alpha y = 0.3$ which is almost parallel



Ill conditioning

- one means of identifying the problem comes from looking at the determinant
  - the determinants are small thus the associated equations are nearly singular

e.g.:

$$\begin{vmatrix} 2 & 1 \\ 1 & 0.5001 \end{vmatrix} = 0.0002 \qquad \begin{vmatrix} 2 & 1 \\ 1 & 0.4999 \end{vmatrix} = -0.0002 \qquad \begin{vmatrix} 2 & 1 \\ 1 & 0.4 \pm 0.05 \end{vmatrix} = -0.2 \pm 0.1$$

Summary

- You should be able to
  - work comfortably with matrices
  - calculate matrix determinants and inverses
  - solve linear equations via matrix methods
  - understand the concept of a matrix rank
  - understand the principles of orthogonality and the 4 subspaces
  - determine eigenvalues and eigenvectors
  - determine matrix decompositions