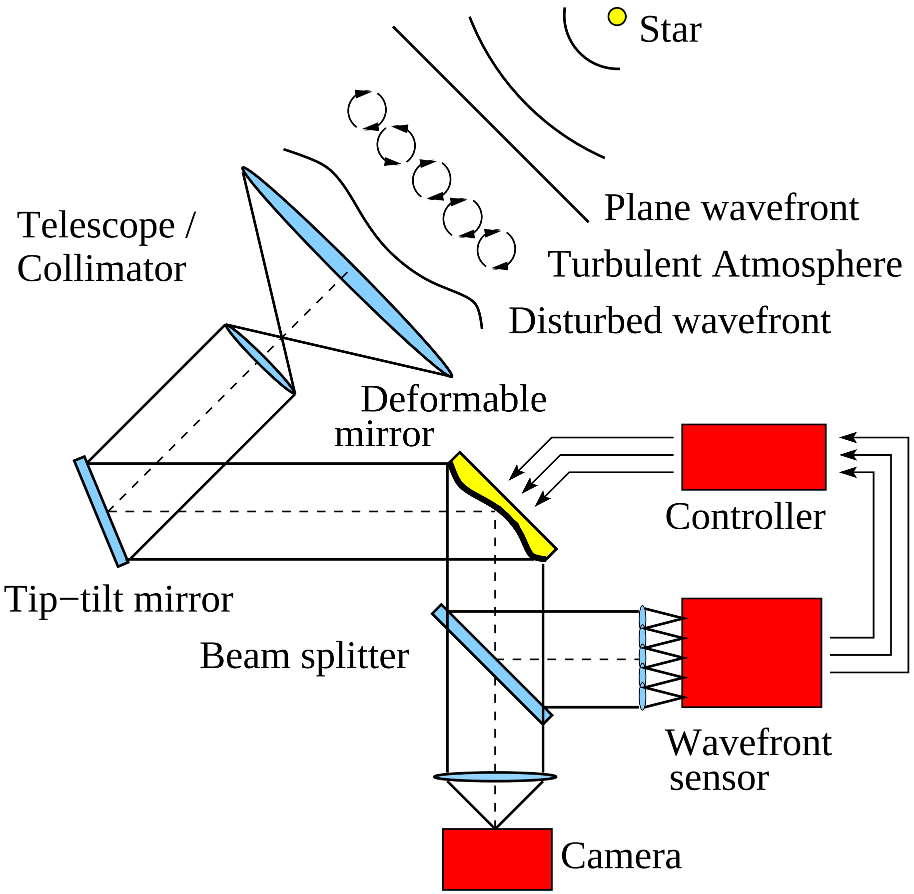
Filtering and Identification

Day 1 - Lecture 1: Introduction and refreshment LA

Michel Verhaegen

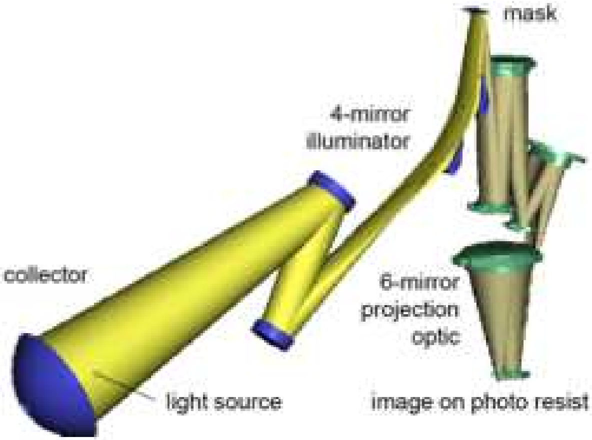
Division of Automatic Control - EE

Smart Optics Systems



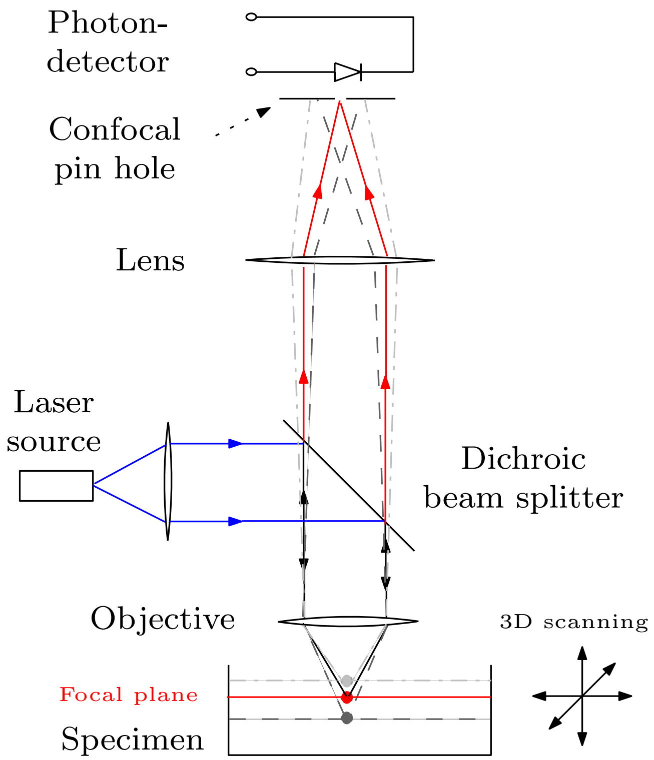
Adaptive Optics Active correction of wavefront aberrations by a deformable mirror. What is needed from a control engineer?

Lithography

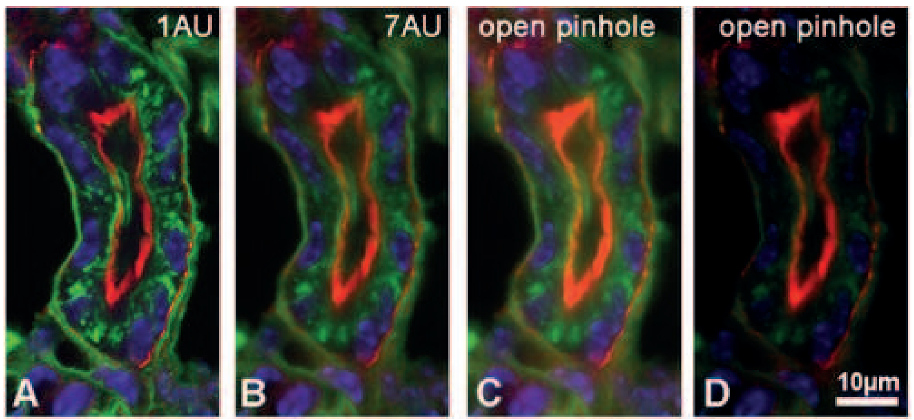


Challenge: Aberration correction due to deformations in the mirrors caused by the heating of the Light Source (pm accuracy for $32nm$ technology! Internships possible!

Microscopy



- Pin hole *con*jugated to the *focal* point (rejection of out-of-focus emission)
- 3-dimensional pointwise scanning (image formed by points)
- Confocal to widefield:



(Image courtesy: http://www.rudbeck.uu.se, AU: Airy-disk U

Teaching Staff (DCSC)

Lecturer	
Prof.dr.ir. Michel Verhaegen	m.verhaegen@tudelft.nl
Teaching Assistants (TA's):	
Jonas Calimer	calimer@isy.liu.se
Karl Granström	karl@isy.liu.se

Objectives of the course

After studying this course you should

be able to derive *estimation, filtering and identification* [a] algorithms based on the the linear least squares method

[a] And control (H_2, etc.)

Course material

- Book:
  Filtering and System Identification: An Introduction, by Michel Verhaegen and Vincent Verdult, **Cambridge University Press, 2007.**
- hand-outs or local blackboard?

Outline of the course

This intensive course will run for a week; with morning lectures and homework in the afternoon.

- Day 1: LA review and Deterministic Linear Least Squares
- Day 2: Stochastic Least Squares and Kalman filtering

Outline of the course (C'td)

- Day 3: Use of the Kalman filter and optimal predictors for input-output models
- Day 4: Deterministic Subspace Identification and a framework for consistency analysis
- Day 5: Instrumental variables in Subspace identification and probing some future developments
  No Homework!

Exam

- Four sets of homeworks:
  Hand-in sets on morning of the next day to the Lecturer.

Filering and identification
Let’s start!

System identification?

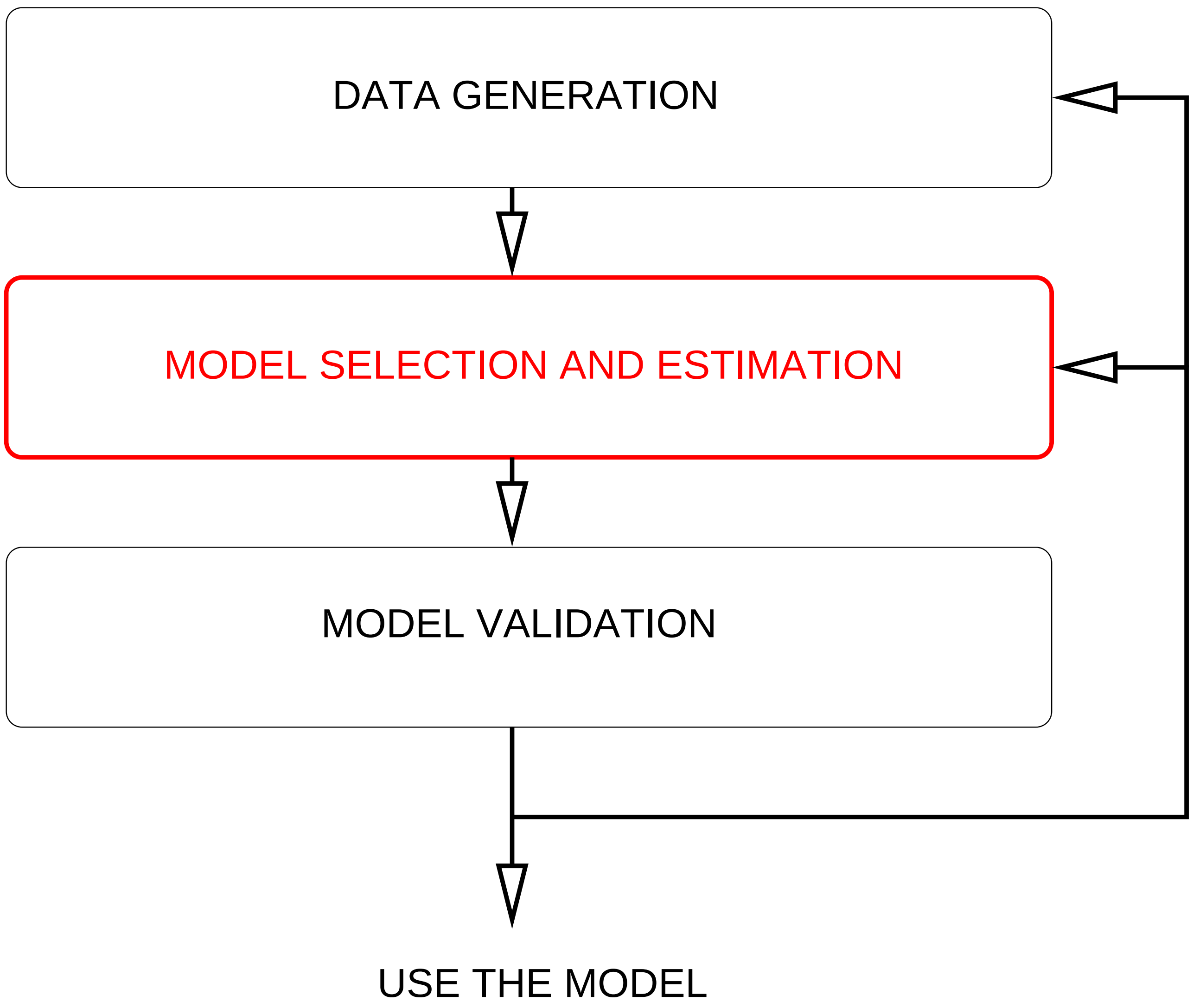
in a general context

The art to extract missing information by inspection with the goal to ...

in a scientific context

The art to extract mathematical models from measurements derived by experimentation with physical phenomenon one wants to understand/control (GOAL!)

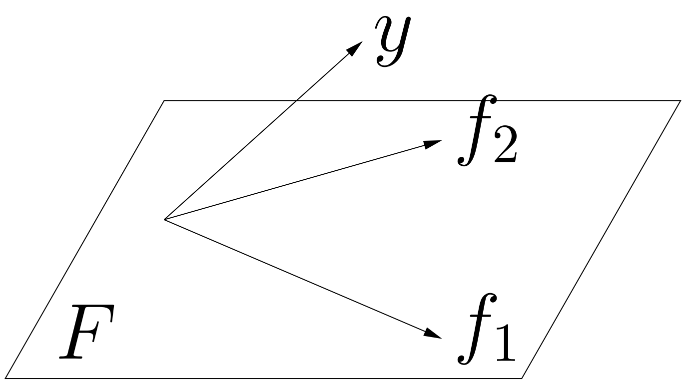
Identification cycle



Mathematical ingredients?

(Linear) Least Squares

$$\min_x \epsilon^T \epsilon \qquad y = Fx + \epsilon$$



- Matrix Theory
- Probability Theory
- Signal/System Theory
- Domain Knowledge

Overview Linear Algebra (LA)

- **T**he matrix concept!
- The Usefull matrix factorization: The SVD
- A Quick view on its potential!
- “Matrix-crimes”
- The Useful matrix Lemma

Matrix theory: Some history

Matrix is Latin for womb (matrix = “mögel”, “grogrund”,matris)

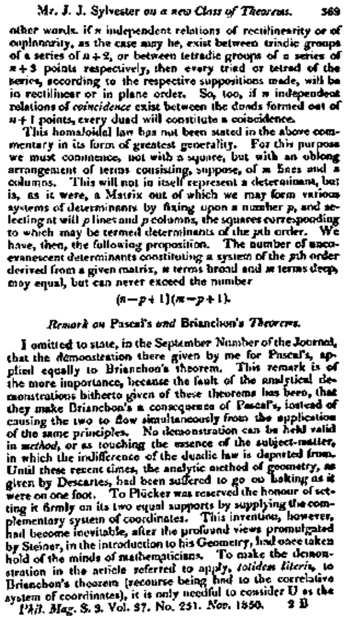
Chinese used matrix methods already in [200 BC — 300AD].

1. They used concepts like determinants of a table of numbers
2. Determinant was long known to be invented by Japanese Seki Kowa 1683.

Matrix theory: Some history

The term “Matrix” was first introduced by James Sylvester 1850

www−history.mcs.st−and.ac.uk/history/



Mr. J. J. Sylvester on a new Class of Theorems. 369

other words, if n independent relations of rectilinearity or of coplanarity, as the case may be, exist between triadic groups of a series of $n+2$, or between tetradic groups of a series of $n+3$ points respectively, then every triad or tetrad of the series, according to the respective suppositions made, will be in rectilinear or in plane order. So, too, if n independent relations of *coincidence* exist between the duads formed out of $n+1$ points, every duad will constitute a coincidence.

This homaloidal law has not been stated in the above commentary in its form of greatest generality. For this purpose we must commence, not with a square, but with an oblong arrangement of terms consisting, suppose, of m lines and n columns. This will not in itself represent a determinant, but is, as it were, a Matrix out of which we may form various systems of determinants by fixing upon a number p, and selecting at will p lines and p columns, the squares corresponding to which may be termed determinants of the pth order. We have, then, the following proposition. The number of unevanescent determinants constituting a system of the pth order derived from a given matrix, n terms broad and m terms deep, may equal, but can never exceed the number

$$(n-p+1)(m-p+1).$$

Remark on Pascal's *and* Brianchon's *Theorems.*

I omitted to state, in the September Number of the Journal, that the *demonstration* there given by me for Pascal's, applied equally to Brianchon's theorem. This remark is of the more importance, because the fault of the analytical demonstrations hitherto given of these theorems has been, that they make Brianchon's a consequence of Pascal's, instead of causing the two to flow simultaneously from the application of the same principles. No demonstration can be held valid in *method*, or as touching the essence of the subject-matter, in which the indifference of the dualic law is departed from. Until these recent times, the analytic method of geometry, as given by Descartes, had been suffered to go on halting as it were on one foot. To Plücker was reserved the honour of setting it firmly on its two equal supports by supplying the complementary system of coordinates. This invention, however, had become inevitable, after the profound views promulgated by Steiner, in the introduction to his Geometry, had once taken hold of the minds of mathematicians. To make the demonstration in the article referred to apply, *totidem literis*, to Brianchon's theorem (recourse being had to the correlative system of coordinates), it is only needful to consider U as the

Phil. Mag. S. 3. Vol. 37. No. 251. *Nov.* 1850. 2 B

Mr. J.J. Sylvester on a new Class of Theorems

Phil. Mag. S. 6, Vol 37, No. 251, Nov. 1850


MATRIX

Definition of a matrix

A matrix $A \in \mathbb{R}^{m \times n}$ is a
two-dimensional table of numbers:

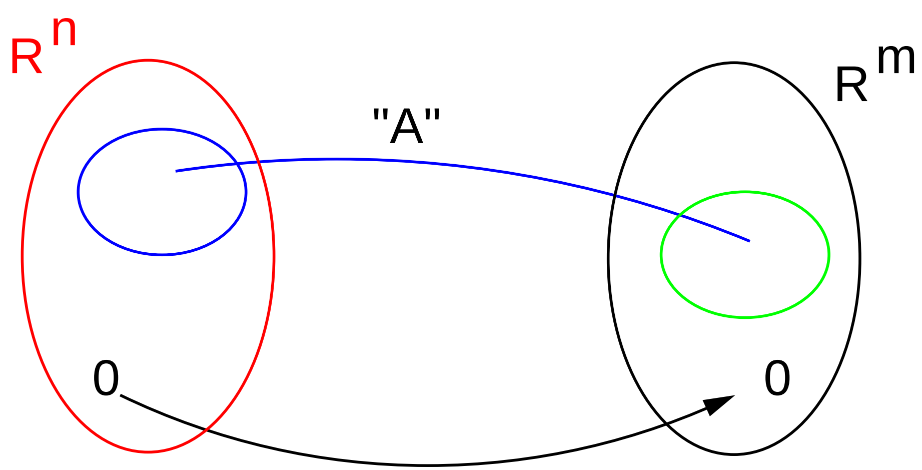
$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & & a_{2n} \\ & & \ddots & \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix}$$

with $a_{ij} \in \mathbb{R}, \;\; a_i \in \mathbb{R}^m$.

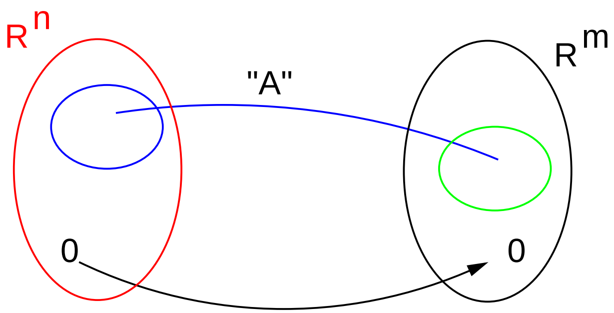
A matrix represents a (linear) mapping

A matrix is (also) a mapping between two Euclidean vector spaces:

$$A : \mathbb{R}^n \to \mathbb{R}^m : \forall x \in \mathbb{R}^n, \exists y \in \mathbb{R}^m : Ax = y$$



The “Four” key spaces of a linear mapping



The linear mapping: $A : \mathbb{R}^n$ (“domain”) $\rightarrow \mathbb{R}^m$ (“Image or Range space”) is characterized by four subspaces:

- range$(A) = \{y \in \mathbb{R}^m : y = Ax$ for some $x \in \mathbb{R}^n\}$
- range$(A^T) = \{x \in \mathbb{R}^n : x = A^T y$ for some $y \in \mathbb{R}^m\}$
- ker$(A) = \{x \in \mathbb{R}^n : Ax = 0\}$
- ker$(A^T) = \{y \in \mathbb{R}^m : A^T y = 0\}$

The rank of A equals the dimension of range(A).

Special class of matrices

Definition: An “square” matrix $Q \in \mathbb{R}^{n\times n}$ is *orthogonal* if

$$Q^T Q = Q Q^T = I_n$$

This means:

1. Each column vector of an orthogonal matrix has length $\cdots$?
2. Two different column (row) vectors of an orthogonal matrix satisfy?
3. What is the inverse of an orthogonal matrix?
4. And many more useful (numerical) advantages ...

Overview Linear Algebra (LA)

- The matrix concept!
- **T**he Usefull matrix factorization: The SVD
- A Quick view on its potential!
- “Matrix-crimes”
- The Useful matrix Lemma

The Singular value decomposition (SVD)

The SVD-Theorem: Let $A \in \mathbb{R}^{m\times n}$, then there exists a pair of orthogonal matrices:

$$
\begin{aligned}
U &= \begin{bmatrix} u_1 \cdots & u_m \end{bmatrix} \in \mathbb{R}^{m\times m} : UU^T = U^TU = I_m \\
V &= \begin{bmatrix} v_1 \cdots & v_n \end{bmatrix} \in \mathbb{R}^{n\times n} : VV^T = V^TV = I_n
\end{aligned}
$$

such that,

$$
U^TAV = \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} \in \mathbb{R}^{m\times n}, \quad \Sigma = \text{diag}(\sigma_1, \cdots, \sigma_p)
$$

with $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p \geq 0$ and $p = \min(m, n)$.

Example SVD

$$A = \begin{bmatrix} 1 & 0 & 1 \\ \frac{1}{2} & \frac{1}{2} & 1 \\ 0 & 1 & 1 \end{bmatrix} \Rightarrow$$

$$A = \underbrace{\begin{bmatrix} -\frac{\sqrt{3}}{3} & \frac{\sqrt{2}}{2} & \frac{\sqrt{6}}{6} \\ -\frac{\sqrt{3}}{3} & 0 & \frac{\sqrt{6}}{3} \\ -\frac{\sqrt{3}}{3} & -\frac{\sqrt{2}}{2} & \frac{\sqrt{6}}{6} \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} \frac{3\sqrt{2}}{2} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \boxed{0} \end{bmatrix}}_{\Sigma} \underbrace{\begin{bmatrix} -\frac{\sqrt{6}}{6} & \frac{\sqrt{2}}{2} & -\frac{\sqrt{3}}{3} \\ -\frac{\sqrt{6}}{6} & -\frac{\sqrt{2}}{2} & -\frac{\sqrt{3}}{3} \\ -\frac{\sqrt{6}}{3} & 0 & \frac{\sqrt{3}}{3} \end{bmatrix}^T}_{V^T}$$

[U,Sigma,V]=svd(A);

- Column vectors of the matrix U: left singular vectors
- Column vectors of the matrix V: right singular vectors
- Diagonal elements of Σ: the singular values

RangeDemo.m

Observations from RangeDemo.m

- columns of A lie in a plane $\subset \mathbb{R}^3 \Leftrightarrow$ dim$\Big(\mathrm{span}_{\mathrm{col}}(A)\Big) = 2 \Leftrightarrow$ # non-zero singular values (sv's) $= 2$
- the left singular vectors u_1, u_2 corresponding to the non-zero singular values:

$$A = \sum_{i=1}^{2} \sigma_i u_i v_i^T$$

  form an orthogonal basis for $\mathrm{span}_{\mathrm{col}}(A)$.
- the left singular vector u_3 corresponding to the zero singular value ($i = 3$) is a basis for $\ker(A^T)$.
- the left (and right) singular vectors are orthogonal and are of unit length.

The four key subspaces

Let the SVD of the matrix A be given as,

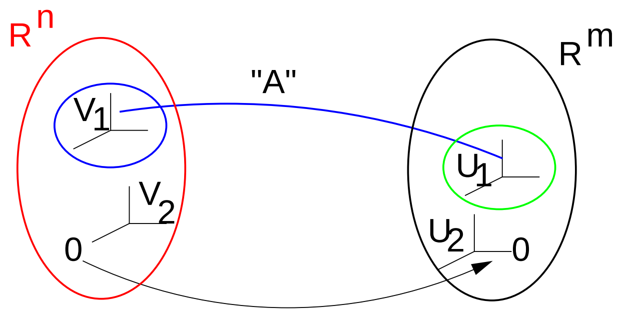
$$A = \left[\begin{array}{c|c} U_1 & U_2 \end{array}\right] \left[\begin{array}{c|c} \Sigma_1 & 0 \\ \hline 0 & 0 \end{array}\right] \left[\begin{array}{c} V_1^T \\ \hline V_2^T \end{array}\right] \quad \text{with } \Sigma_1 > 0$$

then, since $Ax = \left(U_1 \left(\Sigma_1 (V_1^T x)\right)\right)$,

$$\text{range}(A) = \{y \in \mathbb{R}^m : y = Ax \text{ for some } x \in \mathbb{R}^n\} = \text{span}(U_1)$$

Further, since for $x = V_2 \alpha$:
$Ax = U_1 \Sigma_1 V_1^T V_2 \alpha = 0$,

$$\ker(A) = \{x \in \mathbb{R}^n : Ax = 0\} = \text{span}(V_2)$$



The SVD: the “workhorse” for reliable calculations

Contrary to the eigenvalue decomposition, the determinant, etc. the SVD allows for a numerically reliable “calculus”. **Example:**

Checking the singularity of a matrix A: The notion $det(A)$ is “often” used to signal the singularity of a matrix. This is only true in the case it is “exactly” zero!

Checking Singularity (Ct'd)

Example: Consider the "square" matrix:

$$A = \begin{bmatrix} 1 & -1 & \cdots & -1 \\ 0 & 1 & \cdots & -1 \\ \vdots & & \ddots & \\ 0 & 0 & \cdots & 1 \end{bmatrix} \in \mathbb{R}^{n\times n}$$

Then $det(A)$ equals 1. But the condition number of the matrix A defined as:

$$\kappa_\alpha(A) = \|A\|_\alpha \|A^{-1}\|_\alpha$$

for $\alpha = 1, 2, \infty$ and $\|A\|_\alpha = \sup_{x\neq 0} \frac{\|Ax\|_\alpha}{\|x\|_\alpha}$ equals:

$$\kappa_\infty(A) = n2^{n-1}$$

Condition number of a matrix

Definition: For a general matrix $A \in \mathbb{R}^{m\times n}$ $(m \geq n)$, its condition number $\kappa_2(A)$ (in short $\kappa(A)$) is given as:

$$\kappa(A) = \|A\|_2 \|A^\dagger\|_2$$

where $A^\dagger$ denotes the pseudo-inverse of a matrix, i.e. satisfying,

$$AA^\dagger A = A \quad A^\dagger AA^\dagger = A^\dagger \quad (AA^\dagger)^T = AA^\dagger \quad (A^\dagger A)^T = A^\dagger A$$

If A is full rank, then $A^\dagger = (A^T A)^{-1} A^T$.

Exercise: Check that $\kappa(A) = \frac{\sigma_1}{\sigma_n}$!

Overview Linear Algebra (LA)

- The matrix concept!
- The Usefull matrix factorization: The SVD
- **A** Quick view on its potential
- “Matrix-crimes”
- The Useful matrix Lemma

"Optimal" low rank approximation

Theorem: Let the SVD in the SVD-theorem be given and let $k < \text{rank}(A)$ and let the following approximation A_k of A be given:

$$A_k = \sum_{i=1}^{k} \sigma_i u_i v_i^T$$

then,

$$\min_{\text{rank}(B)=k} \|A - B\|_2 = \|A - A_k\|_2 = \sigma_{k+1}$$

Spiegelman.m

Overview Linear Algebra (LA)

- The matrix concept!
- The Usefull matrix factorization: The SVD
- A Quick view on its potential
- ‘‘Matrix-crimes”
- The Useful matrix Lemma

[a] Matrix-crimes: Syntax crimes

1. Non-compatibility of dimensions: $A + B$ when $A \in \mathbb{R}^{2\times 3}$ and $B \in \mathbb{R}^{3\times 3}$ and the same for $A^T B$.
2. Matrix products do (in general) not commute: $AB \neq BA$.
3. Matrix inverse of the product of matrices: $(AB)^{-1} \neq A^{-1}B^{-1}$ in stead of $(AB)^{-1} = B^{-1}A^{1}$ - provided inverses exist!
4. $(A + B)^2 \neq A^2 + 2AB + B^2$!

---

[a] Typical violations of Stanford students [S. Boyd - EE 263], our TUD students “too often” join the club ...

Matrix-crimes: Semantic crimes

Matrix expressions that simply do not make sense. Examples:

1. Let $x \in \mathbb{R}^n$, then xx^T exists but $(xx^T)^{-1}$ not, why?
2. If the matrix $Q \in \mathbb{R}^{m \times n}$ for $m > n$, then

$$QQ^T$$

can never be the identity matrix.

Overview Linear Algebra (LA)

- The matrix concept!
- The Usefull matrix factorization: The SVD
- A Quick view on its potential
- “Matrix-crimes”
- **T**he Useful matrix Lemma

Lemma 2.3 p. 19

Schur Complements: Block Triangular Factorizations

Let the block matrix $A \in \mathbb{R}^{n \times n}$ (symmetric) be invertible, then a very useful matrix factorization of matrix consisting of different blocks is the following ($C \in \mathbb{R}^{m \times m}$):

$$\begin{bmatrix} A & B \\ B^T & C \end{bmatrix} = \begin{bmatrix} I & 0 \\ B^T A^{-1} & I \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & C - B^T A^{-1} B \end{bmatrix} \begin{bmatrix} I & A^{-1}B \\ 0 & I \end{bmatrix}$$

Therefore the following holds,

$$\begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \geq 0 \Leftrightarrow A > 0 \text{ and } C - B^T A^{-1} B \geq 0$$

Exercise

Given $\begin{bmatrix} A & B \\ B^T & C \end{bmatrix} (A \in \mathbb{R}^{n\times n} and C \in \mathbb{R}^{m\times m})$

(symmetric) with $A > 0$ and $C - B^T A^{-1} B > 0$, then show:

$$\operatorname{rank}\left(\begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \right) = n + m$$

Lemma 2.3 p. 19

Schur Complements: Block Triangular Factorizations

When C is invertible, then we have:

$$\begin{bmatrix} A & B \\ B^T & C \end{bmatrix} = \begin{bmatrix} I & BC^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} A - BC^{-1}B^T & 0 \\ 0 & C \end{bmatrix} \begin{bmatrix} I & 0 \\ C^{-1}B^T & I \end{bmatrix}$$

Therefore the following holds,

$$\begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \geq 0 \Leftrightarrow C > 0 \text{ and } A - BC^{-1}B^T \geq 0$$

The condition $Matrix \geq 0$ among others means that a square root of the matrix exists: $Matrix = Matrix^{1/2}Matrix^{T/2}$

Summary of Lecture 1

To start the discovery tour for retrieving system information from measured data records:

What we just have done is a brief review of linear algebra. Next we briefly review probability theory and filtering of stochastic processes! We will also start with analysing the derterministic least squares problem !

Reading of the course book of first Day Lecture:

Study Chapters 1, 2(2.1-2.5), 3, 4(4.1-4.3)