MATH280

Applied Linear Algebra

1st lecture 2018

We will use Lays book: *"Linear algebra and its applications, 5ed"*. (The 3ed and 4ed of Lays book are also consistent with the notes, but there will be some differences in the exercises of these editions.)

We will also use parts of: Stephen Boyd and Lieven Vandenberghe: "Introduction to Applied Linear Algebra - Vectors, Matrices, and Least Squares"

Motivation

- Linear algebra play an important computational part in many useful applications:
  - Solving systems of equations (many applications simply come down to this)
  - Encoding and Encryption (how information is reshaped to be unintelligible and then brought back.)
  - Signal processing / compression (Noise reduction or compaction while keeping important information)
  - Multivariate Statistics/ Machine Learning/
    (Prediction modelling, Clustering and Explorative data analysis)
  - Image Analysis, Computer graphics and animations

Key prerequisites (Math113/131, Lay):

- Linear equations with geometric interpretation of solutions, Lay: 1.1-1.3.
- Matrix - vector representation ($A\mathbf{x} = \mathbf{b}$), Lay: 1.4.
- Elementary row operations and description of solutions, Lay: 1.1 to 1.2.
- Solutions for linear equations, Lay: 1.5.
- Applications of linear systems, Lay: 1.6.
- Linear dependence and independence, Lay: 1.7.
- Linear transformations with matrix representation, Lay: 1.8 to 1.9.
- Matrix Operations and inverse matrices, Lay: 2.1 to 2.2.
- Characterization of invertible matrices, Lay: 2.3.

- Vector and subspace of $\mathbf{R}^n$, Lay: 2.8.
- Size, rank and coordinate systems, Lay: 2.9.
- Determinants - definition and calcuation, Lay: 3.1-3-2.
- Eigenvectors and eigenvalues, Lay: 5.1 to 5.2.
- Diagonalization and similarity, Lay: 5.3.
- Inner products, Length and Orthogonality Lay: 6.1.
- Orthogonal sets, Lay: 6.2.
- Orthogonal projections, Lay: 6.3.
- Least squares problems and applications, Lay: 6.5-6.6.

A summary of this material (minus chapter 5) follows:

Systems of linear equations (1.1)

A **system of linear equations** (also called a **linear system**) is a collection of one or more linear equations that use a common set of variables

Example of a system linear of equations:

- $2x_1 - x_2 + 1.5x_3 = 8$
- $x_1 \qquad - 4x_3 = -7$

is a system with $n = 3$, unknown variables x_1, x_2, x_3 and 2 equations.

Generally, a system of equations with m equations and n unknowns (often called an $m \times n$ - system) is given by

$$\begin{array}{c} a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + ... + a_{2n}x_n = b_2 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + ... + a_{mn}x_n = b_m \end{array}$$

A solution of a linear system is a list of numbers $\mathbf{s} = \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{bmatrix}$ satisfying the equations in the system if they are inserted as the corresponding variables $x_1, x_2, ..., x_n$.

The corresponding coefficients for the system above have the form

$$A = \begin{bmatrix} a_{11} & .. & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & .. & a_{mn} \end{bmatrix}$$

The right hand side of the system is given by the column vector

$\mathbf{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$ and the unknown variables as $\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$.

We can describe the entire system in the compact form

$$A\mathbf{x} = \mathbf{b}.$$

The associated augmented matrix for the system is a $m \times (n+1)$ - matrix with the form:

$$[A \ \mathbf{b}] = \begin{bmatrix} a_{11} & .. & a_{1n} & b_1 \\ \vdots & & & \vdots \\ a_{m1} & .. & a_{mn} & b_m \end{bmatrix}.$$

Solving Linear Systems

A simple example of a 2×2 - system from section 1.1 of the textbook :

- $x_1 - 2x_2 = -1$
- $-x_1 + 3x_2 = 3$

Here the representation of each equation is by a line in the (x_1, x_2)-plane

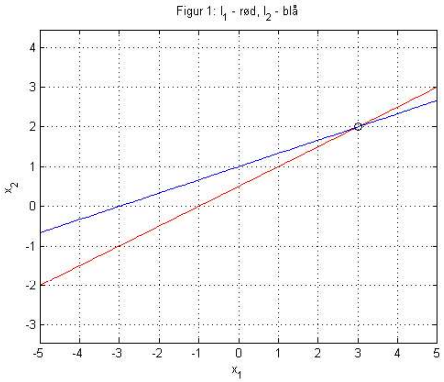


Figure 1: Graphical solution

Each point on line l_1 gives a pair of numbers that fulfills the first equation. The same for every point on on line l_2 for the other equation. The two lines intersect at $(3, 2)$, the only set of values that satisfies both equations.

Thus we can say that this system has exactly one solution (called a *unique solution*).

From the geometric interpretation of this example, one can for 2×2 - systems easily realize the two other possibilities that may occur:

- The lines for each of the two equations can be parallel versions of each other, ie they never crosses and the system is no solution.
- lines for each of the two equations can have no divergence, i.e. the two equations describing one and the same line. Thus, the system has infinitely many solutions.

This holds generally for linear systems, a system of linear equations has either:

1. No solutions, referred to as *inconsistent*.
2. Exactly one solution.
3. Infinite number of solutions.

In cases 2 and 3, we say that the system is *consistent*.

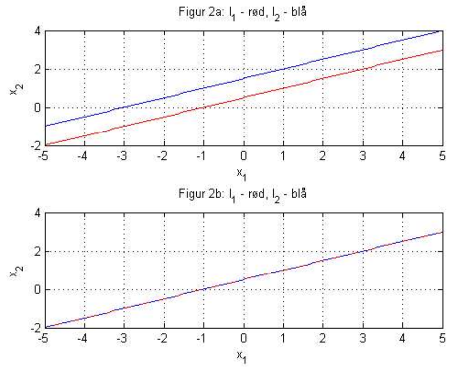
The system (see figure 2 on page 3 of the textbook)

- $x_1 - 2x_2 = -1$
- $-x_1 + 2x_2 = 3$

is an example of a system with no solution, but the system

- $x_1 - 2x_2 = -1$
- $-x_1 + 2x_2 = 1$

is an example of a system with an infinite number of solutions.



Solution of linear systems by means of elementary row operations

Definition: Two linear systems are called <u>*equivalent*</u> if the have the same solution.

Example (see also example in the text book)
Let us solve the 2×2 - system below by replacing it with a sequence of equivalent systems. (to the right is given the equivalent augmented matrix):

1) $\begin{array}{rcrcr} 2x_1 & + & 4x_2 & = & 8 \\ 4x_1 & + & 16x_2 & = & 40 \end{array} \qquad \begin{bmatrix} 2 & 4 & 8 \\ 4 & 16 & 40 \end{bmatrix}$

Multiply the first row in the system with 1/2 to produce the equivalent system:

2) $\begin{array}{rcrcr} x_1 & + & 2x_2 & = & 4 \\ 4x_1 & + & 16x_2 & = & 40 \end{array} \qquad \begin{bmatrix} 1 & 2 & 4 \\ 4 & 16 & 40 \end{bmatrix}$

Add (-4) times the first row to the second row. The equivalent system becomes:

3) $\begin{array}{rcrcr} x_1 & + & 2x_2 & = & 4 \\ & & 8x_2 & = & 24 \end{array} \qquad \begin{bmatrix} 1 & 2 & 4 \\ 0 & 8 & 24 \end{bmatrix}$

Multiply the second row by 1/8. The equivalent system becomes:

4) $\begin{array}{rcrcr} x_1 & + & 2x_2 & = & 4 \\ & & x_2 & = & 3 \end{array} \qquad \begin{bmatrix} 1 & 2 & 4 \\ 0 & 1 & 3 \end{bmatrix}$

Add (-2) times the second row to the first row, the equivalent system in now:

5) $\begin{array}{rcr} x_1 & = & -2 \\ x_2 & = & 3 \end{array} \qquad \begin{bmatrix} 1 & 0 & -2 \\ 0 & 1 & 3 \end{bmatrix}$

The system 5) is trivial in the sense that we can directly read out the solution, and the example above shows step by step the normal solution strategy of a system of equations that take place by replacing a system with an equivalent system that is easier to solve.

Observation
If a system of equations is subjected to the following three types of operations, an equivalent system is created:

I) a multiple of one equation is added to another.
II) arranging of order of equations.
III) An equation is multiplied by a constant $\neq 0$.

These operations are called **elementary row operations**, and lead to the same conclusion if we apply them directly on the equations or the rows of the augmemnted matrix of the system. In the example above we used operations of types I and III.

An important property of the three row operations is that they are *reversible*. Transposition of two rows leads back to the starting point if they are switched again. A row which is scaled by multiplication with a constant $c \neg 0$, can be recreated by a subsequent multiplication by the inverse constant $1/c$. A row which is modified through addition of a multiple k of another row , can be recreated by a subsequent subtraction of a multiple k (addition of a multiple $-k$) of the same row. We say that the systems and the extended matrix <u>*row-equivalent*</u>.

Consequence:
If the extended matrices of two different linear systems are be row-equivalent, the systems are equivalent, and therefore has the same solution quantity. Let us conclude with investigating a system which proves to be inconsistent:

Example 3 (1.1)
Let us try to solve the following 3×3 system):

1) $\begin{array}{rcrcrcr} & & x_2 & - & 4x_3 & = & 8 \\ 2x_1 & - & 3x_2 & + & 2x_3 & = & 1 \\ 5x_1 & - & 8x_2 & + & 7x_3 & = & 1 \end{array} \qquad \begin{bmatrix} 0 & 1 & -4 & 8 \\ 2 & -3 & 2 & 1 \\ 5 & -8 & 7 & 1 \end{bmatrix}$

Swith the first two rows:

2) $\begin{array}{rcrcrcr} 2x_1 & - & 3x_2 & + & 2x_3 & = & 1 \\ & & x_2 & - & 4x_3 & = & 8 \\ 5x_1 & - & 8x_2 & + & 7x_3 & = & 1 \end{array} \qquad \begin{bmatrix} 2 & -3 & 2 & 1 \\ 0 & 1 & -4 & 8 \\ 5 & -8 & 7 & 1 \end{bmatrix}$

Add $(-5/2)$ time the first row to the third row:

3) $\begin{array}{rcrcrcr} 2x_1 & - & 3x_2 & + & 2x_3 & = & 1 \\ & & x_2 & - & 4x_3 & = & 8 \\ & - & \frac{x_2}{2} & + & 2x_3 & = & \frac{-3}{2} \end{array} \qquad \begin{bmatrix} 2 & -3 & 2 & 1 \\ 0 & 1 & -4 & 8 \\ 0 & \frac{-1}{2} & 2 & \frac{-3}{2} \end{bmatrix}$

Add (1/2) times the second row to the third row:

4) $\begin{array}{rcrcrcr} 2x_1 & - & 3x_2 & + & 2x_3 & = & 1 \\ & & x_2 & - & 4x_3 & = & 8 \\ & & & & 0 & = & \frac{5}{2} \end{array} \qquad \begin{bmatrix} 2 & -3 & 2 & 1 \\ 0 & 1 & -4 & 8 \\ 0 & 0 & 0 & \frac{5}{2} \end{bmatrix}$

The bottom equation $0 = 0x_1 + 0x_2 + 0x_3 = 5/2$ can not be satisfied with any solution candidate, $\mathrm{s} = \begin{bmatrix} s_1 \\ s_2 \\ s_3 \end{bmatrix}$ for x_1, x_2, x_3, and since systems 1) and 4) are equivalent, this means that the original system is inconsistent.

Essentially, solving a system of equations involves transforming the problem into an equivalent system of so-called *row-echelon form*, and further to the reduced row-echelon form. This theme in handled in the text book, section 1.2.

Row reduction and row-echelon form, Gaussian Elimination (1.2)

The rows in an $m \times n$ - matrix are called *row vectors* of length n, and the columns of the same matrix are called the *column vectors* of length m. The *leading* element of a row vector corresponds to the first position (reading from left) containing an integer other than 0. A *pivot position* in a matrix A is a location in A that corresponds to a leading 1 in the reduced echelon form of A. A *pivot column* is a column of A that contains a pivot position.

Definition (Row-echelon form)
A $m \times n$ - matrix is in *row-echelon form* if

- All rows consisting entirely of zeros are at the bottom.
- In each row, the first non-zero entry form the left is a 1, called the *leading* 1.
- The leading 1 in each row is to the right of all leading 1's in the rows above it.

A matrix is in the *reduced row-echelon form* if in addition:

- each leading 1 is the only non-zero entry in its column.

Practical solution of linear systems

Theorem 1 - uniqueness of reduced row-echelon form
Each matrix is row - equivalent to exactly one matrix in reduced row-echelon form.

Assume the extended matrix in reduced row-echelon from below add is row - equivalent with the extended matrix of a system you want to solve.

$$\begin{bmatrix} 1 & 0 & -5 & 1 \\ 0 & 1 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

The corresponding equations are:

$$\begin{array}{ccccccc} x_1 & & & - & 5x_3 & = & 1 \\ & & x_2 & + & x_3 & = & 4 \\ & & & & 0 & = & 0 \end{array},$$

and from this we can read system solution amount. The third row $0 = 0$ tells us nothing of value and can be ignored. Since the third column in the matrix in reduced row-echelon form contains no leading element, x_3 is not dependent on other variables or constant values. This means that the values of x_3 can be selected freely irrespective of x_1 and x_2.

$$\begin{array}{lcl} x_1 & = & 1 + 5x_3 \\ x_2 & = & 4 - x_3 \\ x_3 & \text{er} & \text{fri} \end{array},$$

The fact that we can freely assign values to x_3 has implications for x_1 and x_2. This can be illustrated by using t to specify values for x_3 (formally called introducing t as *parameter* x_3). The *parametric solution* of the system can be described as

$$\begin{array}{lcl} x_1(t) & = & 1 + 5t \\ x_2(t) & = & 4 - t \\ x_3(t) & = & t \end{array}.$$

with $t = 2 = s_3$ for example,the solution vector $\mathbf{s} = \begin{bmatrix} 11 \\ 2 \\ 2 \end{bmatrix}$. With a choice of $t = 0 = s_3$ or $t = 1 = s_3$, solutions are $\mathbf{s} = \begin{bmatrix} 1 \\ 4 \\ 0 \end{bmatrix}$ og $\mathbf{s} = \begin{bmatrix} 6 \\ 3 \\ 1 \end{bmatrix}$.

We introduce the notation $\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{bmatrix}$ to indicate that the general solution of the system can be defined in the vector form as

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{bmatrix} = \begin{bmatrix} 1+5t \\ 4-t \\ t \end{bmatrix} = \begin{bmatrix} 1 \\ 4 \\ 0 \end{bmatrix} + t \begin{bmatrix} 5 \\ -1 \\ 1 \end{bmatrix}.$$

Existence and uniqueness of solutions

Theorem 2 - Existence and uniqueness
A linear system has the solution (the system is consistent) if and only if the rearmost column in a row echelon form derived from the systems extended matrix contains no leading element. If the system is consistent, it has either

1. exactly one solution (the case where no free variables occur, ie all columns except the last in a reduced echelon form has a leading element), or
2. infinitely many solutions (case where one or more free variables occur, ie there are additional columns missing leading element)

This gives us the following formula for calculating the complete solution set of a linear system:

General method for solving linear systems

1. Set up the systems extended matrix.
2. Use elementary row operations to calculate an equivalent echelon form matrix. If the rearmost column contains a leading element, the system is inconsistent, and the solution set is empty. In other cases we continue with Step 3.
3. Continue with appropriate elementary row operations to change the system to reduced row echelon form. Columns that contain a leading element (pivot columns) corresponds to the dependent variables, while those containing no leading element correspond to the free variables.
4. Write the equations corresponding to the matrix you found in step 3, and add if necessary any free variables in parameter notation (t, s osv.)
5. Rewrite each nonzero equation from the previous step so that its one basic variable is expressed in terms of any parameters appearing in the equation.

Vectors in $\mathbf{R}^n$ (1.3)

n - tuples ($n \geq 1$) can be defined in vector form as $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$, and can be interpreted as a point in $\mathbf{R}^n$.

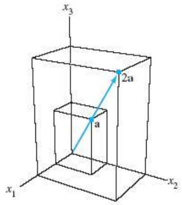


Figure 1.3.6: Vector scaling in $\mathbf{R}^3$

Figure 1.3.6 shows an example of a vector $\mathbf{a} = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} \in \mathbf{R}^3$ together with

the scaled $2\mathbf{a} = \begin{bmatrix} 4 \\ 6 \\ 8 \end{bmatrix}$. Scaling of a general vector $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \in \mathbf{R}^n$ with a number c is defined as

$$c\mathbf{u} = \begin{bmatrix} cu_1 \\ cu_2 \\ \vdots \\ cu_n \end{bmatrix}.$$

The nullvector $\mathbf{0}$ in $\mathbf{R}^n$ is the vector where all components are zero:

$$\mathbf{0} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Properties of vectors in $\mathbf{R}^n$
Assume that $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbf{R}^n$, and that $c, d \in \mathbf{R}$ are normal real numbers (scalars). The following algebraic properties are valid in i $\mathbf{R}^n$:

1. $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$
2. $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$
3. $\mathbf{u} + \mathbf{0} = \mathbf{u}$
4. $\mathbf{u} + (-\mathbf{u}) = -\mathbf{u} + \mathbf{u} = \mathbf{0}$, where $-\mathbf{u}$ is interpreted as $(-1)\mathbf{u}$
5. $c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v}$
6. $(c + d)\mathbf{u} = c\mathbf{u} + d\mathbf{u}$
7. $c(d\mathbf{u}) = (cd)\mathbf{u}$
8. $1\mathbf{u} = \mathbf{u}$

It should be mentioned that we also often operate with vector subtraction. The notation $\mathbf{u} - \mathbf{v}$ means the sum of $\mathbf{u}$ and $-\mathbf{v}$, therefore $\mathbf{u} - \mathbf{v} \equiv \mathbf{u} + (-\mathbf{v})$.

Linear Combinations

For the vectors $\mathbf{v}_1, ..., \mathbf{v}_p \in \mathbf{R}^n$ and scalars $c_1, ..., c_p$ we call the vector

$$\mathbf{y} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + ... + c_p\mathbf{v}_p$$

a linear combination of vectors $\mathbf{v}_1, ..., \mathbf{v}_p$ with *weightings* $c_1, ..., c_p$. Rule number 2 for vectors i $\mathbf{R}^n$ guarantee that we don't need to be concerned about parentheses here.

The notation with linear combinations is also useful in consideration of linear equations. The vector equation

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + ... + x_p\mathbf{a}_p = \mathbf{b}$$

is equivalent to the matrix - vector equation.

$$A\mathbf{x} = \mathbf{b},$$

and can be solved via elementary row operations on the extended matrix

$$[\mathbf{a}_1 \ \mathbf{a}_2 \ ... \ \mathbf{a}_p \ \mathbf{b}] = [A \ \mathbf{b}].$$

Specifically the vector $\mathbf{b}$ can be written as a linear combination of the columns $\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_n$ in A, if and only if the linear system has a solution.

Matrix Equation $A\mathbf{x} = \mathbf{b}$ (1.4)

Definition
Span(A) $= \{\mathbf{b} : \mathbf{b} = c_1\mathbf{a}_1 + ... + c_n\mathbf{a}_n,\ c_1, ..., c_n \in \mathbf{R}\}$ describes the group of vectors $\mathbf{b}$ that can be expressed as a linear combination of the columns in the $m \times n$ matrix $A = [\mathbf{a}_1\ \mathbf{a}_2\ ...\ \mathbf{a}_n]$. This group is called the *column space* to A. Alternatively use the notation Col(A) to specify the column space.

Observation:
If all vectors $\mathbf{b} \in \mathbf{R}^m$ can be written as linear combinations of the column vectors $\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_n$, then $\mathbf{R}^m =$ Span(A).

Theorem 4
Let A be an $m \times n$ - matrix. The following statements are equivalent:

1. For every $\mathbf{b} \in \mathbf{R}^m$ så $A\mathbf{x} = \mathbf{b}$ has a solution.

2. Every $\mathbf{b} \in \mathbf{R}^m$ is a linear combination of the columns of A.

3. The columns in A span $\mathbf{R}^m$.

4. An echelon form row equivalent matrix to A has pivots in each row.

Description of Solution Method

Theorem 6

Assume that the system $A\mathbf{x} = \mathbf{b}$ is consistent for a given $\mathbf{b}$, and let $\mathbf{p}$ be a solution. Then consists solution set of the system $A\mathbf{x} = \mathbf{b}$ of all vectors of the form $\mathbf{w} = \mathbf{p} + \mathbf{v}_h$, where $\mathbf{v}_h$ is a vector that is in the solution set for the homogeneous system $A\mathbf{x} = \mathbf{0}$.

Let's look at the example 4, p. 18. This system has expanded matrix

$$\begin{bmatrix} 1 & 6 & 2 & -5 & -2 & -4 \\ 0 & 0 & 2 & -8 & -1 & 3 \\ 0 & 0 & 0 & 0 & 1 & 7 \end{bmatrix} \sim \begin{bmatrix} 1 & 6 & 0 & 3 & 0 & 0 \\ 0 & 0 & 1 & -4 & 0 & 5 \\ 0 & 0 & 0 & 0 & 1 & 7 \end{bmatrix}$$

The equivalent matrix at reduced echelon form shows leading elements for columns 1, 3 and 5, correspondent dependent variables x_1, x_3 and x_5. Columns 2 and 4 do not have leading elements, then the corresponding x_2 and x_4 are

therefore free variable. The solution can be written

$$\begin{array}{rcrcrcl} x_1 & + & 6x_2 & + & 3x_4 & = & 0 \\ & & & & x_2 & = & s \\ & & x_3 & - & 4x_4 & = & 5 \\ & & & & x_4 & = & t \\ & & & & x_5 & = & 7 \end{array} \Rightarrow \begin{array}{rcl} x_1(s,t) & = & -6s - 3t \\ x_2(s,t) & = & s \\ x_3(s,t) & = & 5 + 4t \\ x_4(s,t) & = & t \\ x_5(s,t) & = & 7 \end{array}$$

The corresponding homogeneous system gives

$$\begin{bmatrix} 1 & 6 & 2 & -5 & -2 & 0 \\ 0 & 0 & 2 & -8 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 6 & 0 & 3 & 0 & 0 \\ 0 & 0 & 1 & -4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

with solution

$$\begin{array}{ccccccccccc} x_1 & + & 6x_2 & + & 3x_4 & = & 0 & & x_1(s,t) & = & -6s-3t \\ & & & & x_2 & = & s & & x_2(s,t) & = & s \\ & & x_3 & - & 4x_4 & = & 0 & \Rightarrow & x_3(s,t) & = & 4t \\ & & & & x_4 & = & t & & x_4(s,t) & = & t \\ & & & & x_5 & = & 0 & & x_5(s,t) & = & 0 \end{array}$$

We see that expressed in vector form, the solution of the homogeneous system written as

$$\mathbf{x}(s,t) = s\begin{bmatrix} -6 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + t\begin{bmatrix} -3 \\ 0 \\ 4 \\ 1 \\ 0 \end{bmatrix} = s\mathbf{v}_1 + t\mathbf{v}_2,$$

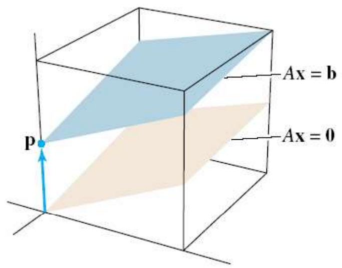
while the solution to the original system becomes

$$\mathbf{x}(s,t) = \begin{bmatrix} 0 \\ 0 \\ 5 \\ 0 \\ 7 \end{bmatrix} + s\begin{bmatrix} -6 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + t\begin{bmatrix} -3 \\ 0 \\ 4 \\ 1 \\ 0 \end{bmatrix} = \mathbf{p} + s\mathbf{v}_1 + t\mathbf{v}_2.$$

Summary (1.5)

The solution to a homogeneous $m \times n$ - system $A\mathbf{x} = \mathbf{0}$ is described by Span$\{\mathbf{v}_1, ..., \mathbf{v}_q\}$ $(q < n)$, where $\mathbf{v}_1, ..., \mathbf{v}_q$ are the vectors for the free variables of the system.
The solution for the corresponding inhomogeneous system (normal linear system) $A\mathbf{x} = \mathbf{b}(\neq \mathbf{0})$ where we know a specific solution $\mathbf{p}$, is given by

$$\{\mathbf{w} = \mathbf{p} + \mathbf{v}_h | \mathbf{v}_h \in \text{Span}\{\mathbf{v}_1, ..., \mathbf{v}_q\}\}.$$



Graphical representation of solution quantities for a homogeneous system with two - dimensional solution quantity, and solution amounts for the corresponding non - homogeneous system.

Linear Independence(1.7)

A homogeneous system such as

$$\begin{bmatrix} 1 & 2 & -3 \\ 3 & 5 & 9 \\ 5 & 9 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

can be expressed as the vector sum

$$x_1 \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix} + x_2 \begin{bmatrix} 2 \\ 5 \\ 9 \end{bmatrix} + x_3 \begin{bmatrix} -3 \\ 9 \\ 3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

This has the trivial solution of $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$, but there may be other solutions.

Definition:
A collection of vectors $\{\mathbf{v}_1, ..., \mathbf{v}_p \in \mathbf{R}^n\}$ is called *linear independent* if the vector equation

$$x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + ... + x_p\mathbf{v}_p = \mathbf{0}$$

only has the trivial solution $\begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$. If there is a choice of weights $c_1, c_2, ..., c_p$ such that all of the choices are not 0 and $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + ... + c_p\mathbf{v}_p = \mathbf{0}$, we say that the collection $\{\mathbf{v}_1, ..., \mathbf{v}_p\}$ is *linearly dependent*.

Observastion 1:
The question of whether a collection of vectors $\{\mathbf{v}_1, ..., \mathbf{v}_p \in \mathbf{R}^n\}$ are linearly independent is equivalent to determining if the corresponding homogeneous system $A\mathbf{x} = \mathbf{0}$ has a non-trivial solution. (Implied that the matrix $A = [\mathbf{v}_1 \ \dots \ \mathbf{v}_p]$).

Observation 2:
The vectors $\{\mathbf{v}_1, ..., \mathbf{v}_p \in \mathbf{R}^n\}$ are linearly independent if and only if the corresponding homogeneous system has only the trivial solution.

Theorem 7
A vector collection $S = \{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_p \in \mathbf{R}^n\}$, der $p \geq 2$ is linearly dependent if and only if at least one of the vectors in S can be expressed as a linear combination of the other vectors in S.

Theorem 8
If the vector collection $S = \{\mathbf{v}_1, ..., \mathbf{v}_p \in \mathbf{R}^n\}$ contains more vectors than n, so that $p > n$, then the vectors S are linearly dependent.

Theorem 9
If the vector collection $S = \{\mathbf{v}_1, ..., \mathbf{v}_p \in \mathbf{R}^n\}$ contains the null vector, so that $\mathbf{0} \in S$, then S is a linearly dependent collection.

Linear Transformations (1.8)

Functions that act between the spaces $\mathbf{R}^m$ and $\mathbf{R}^n$ can be described with the help of a matrix-vector product. Assume A is $n \times m$ - matrix, and let $\mathbf{x} \in \mathbf{R}^m$. Then the matrix product $A\mathbf{x} \in \mathbf{R}^n$, and we can describe the matrix vector multiplications

$$T(\mathbf{x}) = A\mathbf{x}$$

as a transformation of a m-dimensional vector over to a n-dimensional vector.

Example
Let $A = \begin{bmatrix} 2 & -4 \\ 3 & -6 \\ 1 & -2 \end{bmatrix}$, and $\mathbf{x} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$. Then $A\mathbf{x} = \begin{bmatrix} -8 \\ -12 \\ -4 \end{bmatrix}$. With $\mathbf{x} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ then $A\mathbf{x} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$.

Note that when we find the solution to the system $A\mathbf{x} = \mathbf{b}$ we find the set of all $\mathbf{x} \in \mathbf{R}^m$ that produce element $\mathbf{b} \in \mathbf{R}^n$ through the transformation $T(\mathbf{x}) = A\mathbf{x}$.

Matrix Transformations

A *transformation* $T : \mathbf{R}^m \to \mathbf{R}^n$ is nothing more than a mathematical rule, that for every vector $\mathbf{x} \in \mathbf{R}^m$ a vector $T(\mathbf{x}) \in \mathbf{R}^n$ is produced.

Terminology
For a *transformation* $T : \mathbf{R}^m \to \mathbf{R}^n$ we call $\mathbf{R}^m$ the *Domain* to T, while $\mathbf{R}^n$ we call the for *codomain* to T.

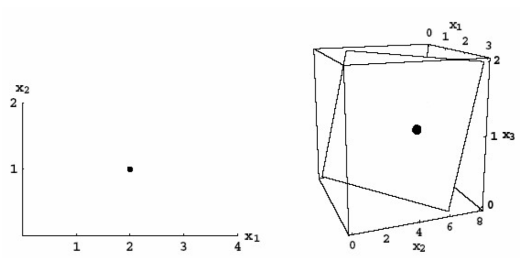
$T(\mathbf{x}) \in \mathbf{R}^n$ we call *the image* of $\mathbf{x}$ through the transformation T.

The set $T(\mathbf{R}^m) = \{\mathbf{b} = T(\mathbf{x}) \in \mathbf{R}^n | \mathbf{x} \in \mathbf{R}^m\}$ of all possible images $T(\mathbf{x})$ we call the*range* of $\mathbf{R}^m$ though the transformation T.

Example

Let $A = \begin{bmatrix} 1 & 0 \\ 2 & 1 \\ 0 & 1 \end{bmatrix}$ and define the transformation $T : \mathbf{R}^2 \to \mathbf{R}^3$ through $T(\mathbf{x}) = A\mathbf{x}$. With $\mathbf{x} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ we have

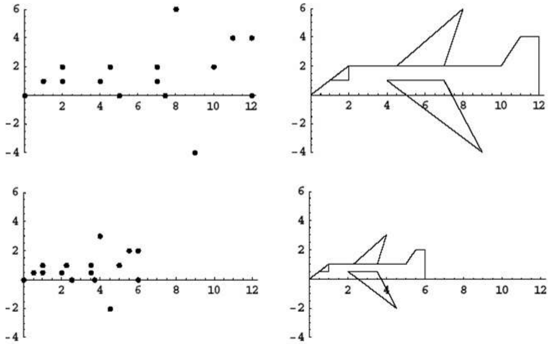
$$T(\mathbf{x}) = A\mathbf{x} = \begin{bmatrix} 1 & 0 \\ 2 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 5 \\ 1 \end{bmatrix}$$



Simple Application

Matrix transformations have many applications. Perhaps the best known examples are in computer graphics.

Example
Let $A = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}$, and define the transformation $T : \mathbf{R}^2 \to \mathbf{R}^2$ through $T(\mathbf{x}) = A\mathbf{x}$. This is an example of a *contraction*. Here we can describe the transformation $T(\mathbf{x}) = A\mathbf{x}$ as acting on all points in $\mathbf{R}^2$. With $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$
Then $T(\mathbf{x}) = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \frac{x_1}{2} \\ \frac{x_2}{2} \end{bmatrix}$.



Rules for linear transformations

With the $n \times m$-matrix A we know from established properties of matrix - vector product that for $T(\mathbf{x}) = A\mathbf{x}$, the following rules hold:

- $T(\mathbf{u}+\mathbf{v}) = A(\mathbf{u}+\mathbf{v}) = A\mathbf{u} + A\mathbf{v} = T(\mathbf{u}) + T(\mathbf{v})$
- $T(c\mathbf{u}) = A(c\mathbf{u}) = cA\mathbf{u} = cT(\mathbf{u})$

for all $\mathbf{u}, \mathbf{v} \in \mathbf{R}^m$, and scalars (numbers) c.

These rules allow us to define a linear transformation .

Definition
We refer to a transformation T as linear if:

1. $T(\mathbf{u}+\mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$ for all $\mathbf{u}, \mathbf{v}$ in the domain to T.
2. $T(c\mathbf{u}) = cT(\mathbf{u})$ for all $\mathbf{u}$ in the domain to T, and scalars c.

Therefore, for a linear transformation T the following holds

$$T(\mathbf{0}) = \mathbf{0} \; og \; T(c\mathbf{u} + d\mathbf{v}) = cT(\mathbf{u}) + dT(\mathbf{v}).$$

Example

Let $\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $\mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, $\mathbf{y}_1 = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}$ and $\mathbf{y}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$. assume that $T : \mathbf{R}^2 \to \mathbf{R}^3$ is a linear transformation that transforms $\mathbf{e}_1$ to $\mathbf{y}_1$ og $\mathbf{e}_2$ to $\mathbf{y}_2$. Use this information to find the image of the vectors $\mathbf{v} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$ og $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$.

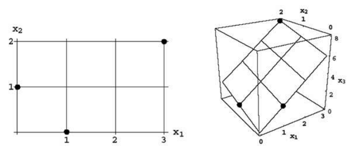
Solution:
In the example above we said $T(\mathbf{e}_1) = \mathbf{y}_1$ og $T(\mathbf{e}_2) = \mathbf{y}_2$. Therefore we can write

$$\mathbf{v} = \begin{bmatrix} 3 \\ 2 \end{bmatrix} = 3\begin{bmatrix} 1 \\ 0 \end{bmatrix} + 2\begin{bmatrix} 0 \\ 1 \end{bmatrix} = 3\mathbf{e}_1 + 2\mathbf{e}_2$$

and it holds that

$$T(\mathbf{v}) = T(3\mathbf{e}_1 + 2\mathbf{e}_2) = 3T(\mathbf{e}_1) + 2T(\mathbf{e}_2)$$

$$= 3\mathbf{y}_1 + 2\mathbf{y}_2 = 3\begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} + 2\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \\ 8 \end{bmatrix}$$



What about $T(\mathbf{x})$??

Matrix as a Linear Transformation (1.9)

Theorem 10
Let $T : \mathbf{R}^m \to \mathbf{R}^n$ be a linear transformation. We want to find the matrix A such that $T(\mathbf{x}) = A\mathbf{x}$ for all $\mathbf{x} \in \mathbf{R}^m$. We can simply construct the matrix A through the help of the images of the the columns in the i $m \times m$ - identity matrix $I_m = [\mathbf{e}_1\ \mathbf{e}_2\ ...\ \mathbf{e}_m]$:

$$A = [T(\mathbf{e}_1)\ T(\mathbf{e}_2)\ ....\ T(\mathbf{e}_n)]$$

<u>Proof:</u>
Idea: We know that $\mathbf{x} = I_m\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + ... + x_m\mathbf{e}_m$. Since T is linear it holds that

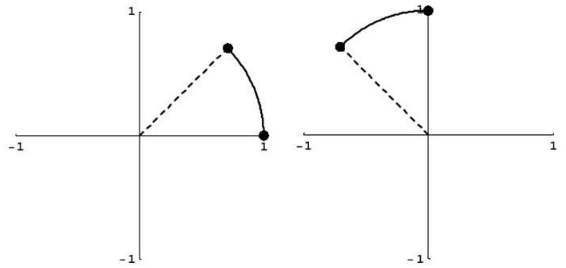
$$\begin{aligned} T(\mathbf{x}) &= T(x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + ... + x_m\mathbf{e}_m) \\ &= x_1T(\mathbf{e}_1) + x_2T(\mathbf{e}_2) + ... + x_mT(\mathbf{e}_m) \\ &= [T(\mathbf{e}_1)\ T(\mathbf{e}_2)\ ...\ T(\mathbf{e}_m)] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} = A\mathbf{x}. \end{aligned}$$

See example 1.9.33

Example 3

Choose a matrix representation A as a linear transformation $T : \mathbf{R}^2 \to \mathbf{R}^2$ that rotates a real vector $\mathbf{x}$ clockwise through the angle $\pi/4$.

solution:



Again use the rule that $A = [T(\mathbf{e}_1)\ T(\mathbf{e}_2)]$. With support from the figure

above we can choose the coordinates for $T(\mathbf{e}_1) = \begin{bmatrix} \cos(\pi/4) \\ \sin(\pi/4) \end{bmatrix}$ og $T(\mathbf{e}_2) = \begin{bmatrix} -\sin(\pi/4) \\ \cos(\pi/4) \end{bmatrix}$, and we can describe the matrix representation

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

Matrix Operations and Inverses

Theorem 1
Let A, B, C be $m \times n$ - matrices, and r, s a scalar (number). Then the following rules apply:

a.	$A + B = B + A$	d.	$r(A + B) = rA + rB$
b.	$(A + B) + C = A + (B + C)$	e.	$(r + s)A = rA + sA$
c.	$A + 0 = A$	f.	$r(sA) = (rs)A$

Row/Column rules for the matrix product AB:

The following (and equivalent) definition is often useful for manual calculations:

A is an $n \times m$ - matrix, and B is an $m \times p$ - matrix, and let $[c_{ij}] = C = AB$, where $1 \le i \le n, 1 \le j \le p$. Then we can calculate c_{ij} as

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + ... + a_{im}b_{mj}$$

$$C = AB = \begin{bmatrix} \\ a_{i1} & a_{i2} & .. & a_{im} \\ \\ \end{bmatrix} \begin{bmatrix} & b_{1j} & \\ & b_{2j} & \\ & : & \\ & b_{mj} & \end{bmatrix} = [c_{ij}]$$

therefore the element c_{ij} (number in the i-th row and j-th column in C) is calculated from the sum of the products get by the elementwise multiplication of the number in the i - th row in A with the numbers from the j - th column in B.

Theorem 2
Let A be an $m \times n$ - matrix, and assume that B og C such that their sizes makes sums and products that are meaningful. The following rules apply:

1. $A(BC) = (AB)C = ABC$ (associative law of multiplication)
2. $A(B + C) = AB + AC$ (left distributive law)
3. $(B + C)A = BA + CA$ (right distributive law)
4. $r(AB) = (rA)B = A(rB)$ for arbitrary values (scalars) r.

5. $I_m A = A = AI_n$ (identity for matrix multiplication)

Proof: See Lay (1.9) and exercises 29 and 30.

WARNING: All characteristics stated above holds for ordinary real numbers, BUT not all properties that hold for real numbers will be valid for matrices.

The following rules do not generally apply:

1. It can be that $AB \neq BA$ see example 7.
2. Even if $AB = AC$ it does not require that $B = C$, see excercise 10.
3. There can exist $AB = 0$ such that $A \neq 0$ and $B \neq 0$, see excercise 12.

Powers and Transposes

Suppose A is a $n \times n$ - matrix (square matrix) then the k - th *power* of A is defined as

$$A^k = AA \cdots A, \ (k \text{ times})$$

Example

$$\begin{bmatrix} 1 & 0 \\ 3 & 2 \end{bmatrix}^3 = \begin{bmatrix} 1 & 0 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 3 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 9 & 4 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 3 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 21 & 8 \end{bmatrix}$$

Given A is an $m \times n$ - matrix, then the *transposed* A^t $n \times m$ - matrix vi make by switching the rows and columns of A.

Example

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 6 & 7 & 8 & 9 & 8 \\ 7 & 6 & 5 & 4 & 3 \end{bmatrix} \Rightarrow A^t = \begin{bmatrix} 1 & 6 & 7 \\ 2 & 7 & 6 \\ 3 & 8 & 5 \\ 4 & 9 & 4 \\ 5 & 8 & 3 \end{bmatrix}.$$

Example

Let $A = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 0 & 1 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 2 \\ 0 & 1 \\ -2 & 4 \end{bmatrix}$. Let us calculate AB, $(AB)^t$, A^tB^t og B^tA^t.

$$AB = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 0 & 1 \\ -2 & 4 \end{bmatrix} = \begin{bmatrix} -5 & 16 \\ 1 & 10 \end{bmatrix}$$

$$(AB)^t = \begin{bmatrix} -5 & 1 \\ 16 & 10 \end{bmatrix}$$

$$A^tB^t = \begin{bmatrix} 1 & 3 \\ 2 & 0 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & -2 \\ 2 & 1 & 4 \end{bmatrix} = \begin{bmatrix} 7 & 3 & 10 \\ 2 & 0 & -4 \\ 5 & 1 & -2 \end{bmatrix}$$

$$B^tA^t = \begin{bmatrix} 1 & 0 & -2 \\ 2 & 1 & 4 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 2 & 0 \\ 3 & 1 \end{bmatrix} = \begin{bmatrix} -5 & 1 \\ 16 & 10 \end{bmatrix}$$

You should recognize the last result!

Rules for transposes

Theorem 3
Let A and B be matrices that add up to totals and multiplied according to the above definitions :

1. $(A^t)^t = A$
2. $(A+B)^t = A^t + B^t$
3. For arbitrary number r, så it holds that $(rA)^t = rA^t$
4. $(AB)^t = B^tA^t$

Proof: See Lay (1.9) excercise 33.

Example
Show that $(ABC)^t = C^tB^tA^t$.

Via theorem 2a) and 3d):

$$(ABC)^t = ((AB)C)^t = C^t(AB)^t = C^t(B^tA^t) = C^tB^tA^t.$$

Matrix Inversion (2.2)

Remember that the multiplicative inverse to a real number $r \neq 0$ is written as r^{-1} or $\frac{1}{r}$. For inverse numbers: $r \cdot r^{-1} = r^{-1} \cdot r = 1$. A square $(n \times n)$matrix A is called *invertible* if there exists a square matrix C such that

$$AC = CA = I_n,$$

where I_n is $n \times n$ identity matrix. We call the matrix C the *inverse* to A.

Propostion
If A is invertible then the inverse to A is unique.
Proof: Assume that B og C are inverses to A. Then:

$$B = BI_n = B(AC) = (BA)C = I_nC = C,$$

so the assumption that both matrices are inverse to A leads that they must be identical. So the inverse to an invertible matrix uniquely determined.

This justifies the use of the notation A^{-1} for the inverse to A, and we have

$$AA^{-1} = A^{-1}A = I_n$$

NOTE: Not all $n \times n$ - matrices are invertible. *Non invertible* matrices can

be called *singular* matrices. Conversely, invertible matrices are also called *nonsingular* matrices

Matrix Inversion cont.

Theorem 4
Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. If $ad - bc \neq 0$, then A is invertible, and

$$A^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \begin{bmatrix} \frac{d}{ad-bc} & \frac{-b}{ad-bc} \\ \frac{-c}{ad-bc} & \frac{a}{ad-bc} \end{bmatrix}.$$

If $ad-bc = 0$, så er A non invertible (singular). Proof: See excercise 25 og 26.

Assume that the coefficient matrix A for the linear $n \times n$ $A\mathbf{x} = \mathbf{b}$ is invertible. We can solve the system as follows:

$$A\mathbf{x} = \mathbf{b} \Leftrightarrow A^{-1}A\mathbf{x} = A^{-1}\mathbf{b}$$

$$\Leftrightarrow I_n\mathbf{x} = A^{-1}\mathbf{b} \Leftrightarrow \mathbf{x} = A^{-1}\mathbf{b}$$

This reasoning gives the following results:

Theorem 5
Let A be an invertible $n \times n$ - matrix. For every $\mathbf{b} \in \mathbf{R}^n$ for which there is a unique solution to $A\mathbf{x} = \mathbf{b}$, this solution is given by $\mathbf{x} = A^{-1}\mathbf{b}$.

Example:
Find the inversion to the matrix

$$A = \begin{bmatrix} -7 & 3 \\ 5 & -2 \end{bmatrix}$$

and use it to solve the system

$$\begin{array}{rcrcl} -7x_1 & + & 3x_2 & = & 2 \\ 5x_1 & - & 2x_2 & = & 1 \end{array}$$

The matrix form of the system is

$$\begin{bmatrix} -7 & 3 \\ 5 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

Through the formula in Theorem 4 for the inverse of a 2×2 - matrix we have

$$A^{-1} = \frac{1}{14-15} \begin{bmatrix} -2 & -3 \\ -5 & -7 \end{bmatrix} = (-1) \begin{bmatrix} -2 & -3 \\ -5 & -7 \end{bmatrix} = \begin{bmatrix} 2 & 3 \\ 5 & 7 \end{bmatrix}$$

We can now solve the system

$$\mathbf{x} = A^{-1}\mathbf{b} = \begin{bmatrix} 2 & 3 \\ 5 & 7 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 7 \\ 17 \end{bmatrix}$$

Matrix Inversion cont.

Theorem 6
Assume that A og B are invertible $n \times n$ - matrices. The following applies:

1. A^{-1} is invertible and $(A^{-1})^{-1} = A$ (therefore that A is the inverse of A^{-1}).
2. AB is invertible, and $(AB)^{-1} = B^{-1}A^{-1}$.
3. A^t is invertible and $(A^t)^{-1} = (A^{-1})^t$

Proof: For point 2 (See 2.2 in Lay for the other two points.)

$$(AB)B^{-1}A^{-1} = A(BB^{-1})A^{-1} = AI_nA^{-1} = AA^{-1} = I_n$$
$$B^{-1}A^{-1}(AB) = B^{-1}(A^{-1}A)B = B^{-1}I_nB = B^{-1}B = I_n.$$

Point 2. of Theorem 6 can be generalized to the products of more than two invertible matrices:

$$(ABC)^{-1} = ((AB)C)^{-1} = C^{-1}(AB)^{-1}$$

$$= C^{-1}(B^{-1}A^{-1}) = C^{-1}B^{-1}A^{-1}$$

We have seen the formula for calculating the inverese of a 2×2 - matrix. It is of considerable interest to calculate the inverse of a general invertible $n \times n$ - matrix. To handle this we will go through the so-called *elementary* matrices

Elementary Matrices

Definition
A *elementary* matrix is a matrix that performs a single elementary row operation on a an identity matrix.

Example: Let $E_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $E_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$, $E_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 3 & 0 & 1 \end{bmatrix}$ og $A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$.

Explain how E_1, E_2 og E_3 are elementary matrices!

Let us study the following matrix products with a focus on interpretation in terms of elementary row operations

$$E_1A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = \begin{bmatrix} a & b & c \\ 2d & 2e & 2f \\ g & h & i \end{bmatrix}$$

$$E_2A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = \begin{bmatrix} a & b & c \\ g & h & i \\ d & e & f \end{bmatrix}$$

$$E_3A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 3 & 0 & 1 \end{bmatrix} \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = \begin{bmatrix} a & b & c \\ d & e & f \\ 3a+g & 3b+h & 3c+i \end{bmatrix}$$

Conclusion:
If elementary row operations are performed on an $m \times n$ - matrix A, then the resulting matrix is expressed as EA where E is an $m \times m$ - matrix which is obtained by performing the same operations on the identity matrix I_m.

Note that the elementary matrices will be *invertible*, since the elementary row operations are reversible.

Check that $E_3^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -3 & 0 & 1 \end{bmatrix}$, $E_2^{-1} = E_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$ og $E_1^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

Example:
La $A = \begin{bmatrix} 1 & 0 & 0 \\ -\frac{3}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{bmatrix}$. Then

$$E_3(E_2E_1A) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 3 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -3 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

We have with this shown that

$$E_3E_2E_1A = I_3 \Leftrightarrow E_3E_2E_1AA^{-1} = I_3A^{-1}$$

$$\Updownarrow$$

$$E_3E_2E_1 = A^{-1}$$

General method for calculating an inverse matrix

Fact:
The elementary row operations that reduce A to I_n are exactly the same row operations that transform I_n to A^{-1}.

Theorem 7
A $n \times n$ - matrix A is invertible if and only if A is row equivalent to I_n, therefore that A can be reduced by a series of elementary row reductions to the identity matrix I_n. Moreover, precisely the same sequence of elementary row operations transform I_n to A^{-1}.
Proof:
See Lay (2.2)

A method for finding A^{-1} for a $n \times n$ - matrix A
. Create an extended matrix by setting A and I_n in sequence, $U = [A \;\; I_n]$. Therefore A is invertable by Theorem 7 if we can find the elementary row operations such that

$$U \sim [I_n \;\; A^{-1}] \text{ (U is row equivalent with } [I_n \;\; A^{-1}])$$

Example:
Let us try to find the inverse to

$$A = \begin{bmatrix} 2 & 0 & 0 \\ -3 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

$$U = [A\ I_3] = \begin{bmatrix} 2 & 0 & 0 & 1 & 0 & 0 \\ -3 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix} \sim$$

$$... \sim \begin{bmatrix} 1 & 0 & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & \frac{3}{2} & 1 & 0 \end{bmatrix} = L$$

Thus we have performed the row reduction of A to I_3, and from the three last columns of L we find

$$A^{-1} = \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & 0 & 1 \\ \frac{3}{2} & 1 & 0 \end{bmatrix}$$

Check the calculation of A^{-1} through matrix multiplication AA^{-1} og $A^{-1}A$!

Theorem 8 (extended)
Assume that A is a $n \times n$ - matrix. Then the following are equivalent:

1. A is invertible.
2. A er row equivalent with I_n.
3. En echelon form row equivalent matrix to A has n pivot elements.
4. The system $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.
5. columns in A er linearly independent.
6. (linear transformations $T(\mathbf{x}) = A\mathbf{x}$ are unique.)
7. The system $A\mathbf{x} = \mathbf{b}$ has at least one solution for each $\mathbf{b} \in \mathbf{R}^n$.
8. The columns in A span out $\mathbf{R}^n$.
9. One can find an $n \times n$ - matrix C such taht $CA = I_n$

10. One can find an $n \times n$ - matrix D such that $AD = I_n$.
11. A^t is invertible.
12. The columns in A form a basis for $\mathbf{R}^n$.
13. $Col(A) = \mathbf{R}^n$.
14. $\dim(Col(A)) = n$.
15. $rank(A) = n$.
16. $Nul(A) = \{\mathbf{0}\}$.
17. $\dim(Nul(A)) = 0$.

Proof: See Lay (2.2).

Example 1: Let us use theorem 8 to go through whether the following 3×3 - matrix A is invertible:

$$A = \begin{bmatrix} 1 & -3 & 0 \\ -4 & 11 & 1 \\ 2 & 7 & 3 \end{bmatrix} \sim \begin{bmatrix} 1 & -3 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & 16 \end{bmatrix}.$$

Conclusion?

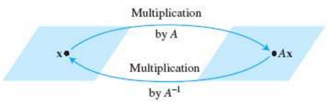
Example 2: Let H be an 5×5 - matrix, and assume that there is a vector $\mathbf{v} \in \mathbf{R}^5$ that can not be formed from some linear combination of the columns in H. Does the system $H\mathbf{x} = \mathbf{0}$ only have the a trivial solution?

Invertible linear transformations

For an invertible $n \times n$ - matrix A it holds that

$$A^{-1}A\mathbf{x} = AA^{-1}\mathbf{x} = \mathbf{x} \text{ for all } \mathbf{x} \in \mathbf{R}^n$$

Illustrastion:



A^{-1} inverse transformation $\mathbf{y} = A\mathbf{x}$ to $\mathbf{x}$.

A linear transformation $T : \mathbf{R}^n \to \mathbf{R}^n$ is called *invertible* if one can find another linear transformation $T : \mathbf{R}^n \to \mathbf{R}^n$ such that

$$S(T(\mathbf{x})) = T(S(\mathbf{x})) = \mathbf{x} \text{ for all } \mathbf{x} \in \mathbf{R}^n.$$

From theorem 10 in chapter 1.9 we know in particular that a linear transformation $T : \mathbf{R}^n \to \mathbf{R}^n$ can be uniquely represented by an $n \times n$ - matrix A. This can be summarized as:

Theorem 9
La $T : \mathbf{R}^n \to \mathbf{R}^n$ be a linear transformation with matrix representation A. T is invertible if and only if matrix A is invertible, and if T is invertible then thee inverse transformation is $S(\mathbf{x}) = A^{-1}\mathbf{x}$ and

$$S(T(\mathbf{x})) = A^{-1}A\mathbf{x} = \mathbf{x} = AA^{-1}\mathbf{x} = T(S(\mathbf{x})) \text{ for all } \mathbf{x} \in \mathbf{R}^n.$$

Subspaces in $\mathbf{R}^n$ (2.8)

The term subspace closely links up the description of the solution set of two systems of equations $A\mathbf{x} = \mathbf{b}$.

Definition
A *subspace* of $\mathbf{R}^n$ is a partial set $H \subseteq \mathbf{R}^n$ that meets the following requirments:

1. $\mathbf{0} \in H$.
2. If $\mathbf{u}$, $\mathbf{v} \in H$, then the sum $\mathbf{u} + \mathbf{v} \in H$.
3. Hvis $\mathbf{u} \in H$, og c is a scalar (number), then $c\mathbf{u} \in H$.

Definition
The Column Space to a matrix A is referred to $Col(A)$, and depicts all linear combinations of the columns in A.

To determine whether $\mathbf{b} \in Col(A)$ is the same as to whether the system $A\mathbf{x} = \mathbf{b}$ has solutions.

The Null space to a matrix

Definition
The Null space to a matrix A is designated $Nul(A)$, and describes all solution vectors $\mathbf{x}$ for the homogeneous system $A\mathbf{x} = \mathbf{0}$.

<u>Observation:</u> If A is $m \times n$ then the Null space of A a subspace of $\mathbf{R}^n$.

Theorem 12
If A is a $m \times n$ - matrix then the null space to A is a subspace of $\mathbf{R}^n$. Equivalently, the collection of all solutions to the homogeneous system $A\mathbf{x} = \mathbf{0}$ with m equations and n uknowns is a subspace in $\mathbf{R}^n$.

<u>Proof:</u> Since $A\mathbf{0} = \mathbf{0}$, then $\mathbf{0} \in Nul(A)$. Assume further that $\mathbf{u}, \mathbf{v} \in Nul(A)$, dvs $A\mathbf{u} = \mathbf{0}$ and $A\mathbf{v} = \mathbf{0}$. This means

$$A(\mathbf{u} + \mathbf{v}) = A\mathbf{u} + A\mathbf{v} = \mathbf{0} + \mathbf{0} = \mathbf{0},$$

therefore $\mathbf{u} + \mathbf{v} \in Nul(A)$. Further for an arbitrary scalar c and $\mathbf{u} \in Nul(A)$, it holds that

$$A(c\mathbf{u}) = cA\mathbf{u} = c\mathbf{0} = \mathbf{0},$$

so we also have that $c\mathbf{0} \in Nul(A)$. This shows that $Nul(A)$ fills all the definitions of a subspace.

Basis for a subspace

Remember the definition of linear independence og dependence:

Definition:
A collection of vectors $\mathbf{v}_1, ..., \mathbf{v}_p \in \mathbf{R}^n$ is called *linearly independent* if the vector equation

$$x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + ... + x_p\mathbf{v}_p = \mathbf{0}$$

only has the trivial solution $\begin{bmatrix} x_1 \\ : \\ x_p \end{bmatrix} = \begin{bmatrix} 0 \\ : \\ 0 \end{bmatrix}$. If one can find a solution with weights $c_1, c_2, ..., c_p$ such that these are all not 0 and $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + ... + c_p\mathbf{v}_p = \mathbf{0}$, then we say the set is $\{\mathbf{v}_1, ..., \mathbf{v}_p\}$ er *linearly dependent*.

Definition:
A *basis* for a subspace H in $\mathbf{R}^n$ is a set of linearly independent vectors $\mathbf{v}_1, ..., \mathbf{v}_p \in H$ such that $H = Span(\{\mathbf{v}_1, ..., \mathbf{v}_p\})$.

Basis for a subspace - Nullspace

Example
Find a basis for the null space to the matrix

$$A = \begin{bmatrix} 5 & 2 & 0 & 1 & -6 \\ 1 & 5 & 1 & 5 & -2 \\ -2 & \frac{3}{2} & \frac{1}{2} & 2 & 2 \end{bmatrix}.$$

The extended matrix to the homogenous system $A\mathbf{x} = \mathbf{0}$ is

$$[A\ \mathbf{0}] \sim \begin{bmatrix} 1 & 5 & 1 & 5 & -2 & 0 \\ 0 & -23 & -5 & -24 & 4 & 0 \\ 0 & \frac{23}{2} & \frac{5}{2} & 12 & -2 & 0 \end{bmatrix}$$

$$\sim \begin{bmatrix} 1 & 5 & 1 & 5 & -2 & 0 \\ 0 & -23 & -5 & -24 & 4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\sim \begin{bmatrix} 1 & 5 & 1 & 5 & -2 & 0 \\ 0 & 1 & \frac{5}{23} & \frac{24}{23} & -\frac{4}{23} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\sim \begin{bmatrix} 1 & 0 & -\frac{2}{23} & -\frac{5}{23} & -\frac{26}{23} & 0 \\ 0 & 1 & \frac{5}{23} & \frac{24}{23} & -\frac{4}{23} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

From the reduced echelon form we can write the system as ut får vi the system:

$$\begin{array}{ccccccccc} x_1 & & & - & \frac{2x_3}{23} & - & \frac{5x_4}{23} & - & \frac{26x_5}{23} & = & 0 \\ & & x_2 & + & \frac{5x_3}{23} & + & \frac{24x_4}{23} & - & \frac{4x_5}{23} & = & 0 \\ & & & & & & & & 0 & = & 0 \end{array}$$

where we have x_3, x_4, x_5 as free variable. In vector form the solution becomes

$$\mathbf{x}(r,s,t) = \begin{bmatrix} x_1(r,s,t) \\ x_2(r,s,t) \\ x_3(r,s,t) \\ x_4(r,s,t) \\ x_5(r,s,t) \end{bmatrix} = \begin{bmatrix} \frac{2r}{23} + \frac{5s}{23} + \frac{26t}{23} \\ -\frac{5r}{23} - \frac{24s}{23} + \frac{4t}{23} \\ r \\ s \\ t \end{bmatrix}$$

$$= r\begin{bmatrix} \frac{2}{23} \\ -\frac{5}{23} \\ 1 \\ 0 \\ 0 \end{bmatrix} + s\begin{bmatrix} \frac{5}{23} \\ -\frac{24}{23} \\ 0 \\ 1 \\ 0 \end{bmatrix} + t\begin{bmatrix} \frac{26}{23} \\ \frac{4}{23} \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

This means that $Nul(A) = Span(\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\})$, where

$$\mathbf{v}_1 = \begin{bmatrix} \frac{2}{23} \\ -\frac{5}{23} \\ 1 \\ 0 \\ 0 \end{bmatrix},\ \mathbf{v}_2 = \begin{bmatrix} \frac{5}{23} \\ -\frac{24}{23} \\ 0 \\ 1 \\ 0 \end{bmatrix},\ \mathbf{v}_3 = \begin{bmatrix} \frac{26}{23} \\ \frac{4}{23} \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

It is easy to show that $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ are linearly independent (how?), and form a basis for $Nul(A)$.

Basis for a subspace - Column space

To find a basis for the Column Space to a matrix A in easier than to find he Null space for the same matrix, but the method requires a careful explanation. Let us look at an examplel.

Example
We shall find the Column Space to the matrix

$$B = \begin{bmatrix} 1 & 0 & -2 & 3 & 0 \\ 0 & 1 & 5 & 4 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

that is in reduced echelon form. Let us call the columns in B, $\mathbf{b}_1, ..., \mathbf{b}_5$.

We have $\mathbf{b}_3 = -2\mathbf{b}_1 + 5\mathbf{b}_2 + 0\mathbf{b}_5$ and $\mathbf{b}_4 = 3\mathbf{b}_1 + 4\mathbf{b}_2 + 0\mathbf{b}_5$.

Notice that the matrix B is in reduced echelon form, and that $\mathbf{b}_3$ and $\mathbf{b}_4$ are linear combinations of the columns with pivot elements ($\mathbf{b}_1$, $\mathbf{b}_2$ og $\mathbf{b}_5$).

In general any linear combinationof the columns of B can be written as a

linear combination of the columns with pivots (when B is in reduced echelon form). For example we could write out the linear combination

$$\mathbf{v} = c_1\mathbf{b}_1 + c_2\mathbf{b}_2 + c_3\mathbf{b}_3 + c_4\mathbf{b}_4 + c_5\mathbf{b}_5$$

as

$$\begin{aligned} \mathbf{v} &= c_1\mathbf{b}_1 + c_2\mathbf{b}_2 + c_3(-2\mathbf{b}_1 + 5\mathbf{b}_2) + c_4(3\mathbf{b}_1 + 4\mathbf{b}_2) + c_5\mathbf{b}_5 \\ &= (c_1 - 2c_3 + 3c_4)\mathbf{b}_1 + (c_2 + 5c_3 + 4c_4)\mathbf{b}_2 + c_5\mathbf{b}_5. \end{aligned}$$

This means that the vectors $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_5$ span out the Column space $Col(B)$, and since the are linearly independent (how?), these vectors form a basis for $Col(B)$

The matrix B in the example was in reduced echelon form, and this makes finding the basis for the column space simple. How should this be done for the general matrix A? Let us first note the following:

We have learned that the solution to the homogeneous system $A\mathbf{x} = \mathbf{0}$ can be found through a row equivalent matrix B ($A \sim B$), such that $A\mathbf{x} = \mathbf{0}$ if and only if $B\mathbf{x} = \mathbf{0}$, therefore the two systems have the same common solution.

But it means that the columns of matrix A and B have exactly the same linear independence.

Example
Check that the matrix

$$A = \begin{bmatrix} 2 & 0 & -4 & 6 & 0 \\ 1 & 1 & 3 & 7 & 0 \\ 3 & 2 & 4 & 17 & 1 \\ 5 & 2 & 0 & 23 & 1 \end{bmatrix}$$

is row equivalent with B from the example by bringing it to reduced echelon form.

A basis for the column space of A can be made by identifying the columns in A with pivot elements in the row equivalent reduced echelon form matrix B.

Let us call the columns in A as $\mathbf{a}_1, ..., \mathbf{a}_5$. From analysis of matrix B we know that $\mathbf{b}_3 = -2\mathbf{b}_1 + 5\mathbf{b}_2$ and $\mathbf{b}_4 = 3\mathbf{b}_1 + 4\mathbf{b}_2$. Let us check that the same linear dependence holds for the corresponding columns in A:

$$\mathbf{a}_3 = -2\mathbf{a}_1 + 5\mathbf{a}_2 \text{ and } \mathbf{a}_4 = 3\mathbf{a}_1 + 4\mathbf{a}_2.$$

This gives us the basis for the following theorem:

Theorem 13
We can find a basis for the Column Space to a matrix A by choosing the columns that correspond to pivot elements for the row equivalent matrix B, ($A \sim B$), where B is in reduced echelon form.

Coordinate Systems (2.9)

Definition
Assume that $B = \{\mathbf{b}_1, ..., \mathbf{b}_p\}$ forms the basis for the subspace $H \subseteq \mathbf{R}^n$. For every $\mathbf{y} \in H$ we can find weights $c_1, c_2, ..., c_p$ such that $\mathbf{y} = c_1\mathbf{b}_1 + ... + c_p\mathbf{b}_p$, and the vector

$$[\mathbf{y}]_B = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_p \end{bmatrix}$$

is called *the coordinate vector* to $\mathbf{y}$ relative to the base B.

Example

Let $\mathbf{b}_1 = \begin{bmatrix} 6 \\ 2 \\ 0 \\ 3 \end{bmatrix}, \mathbf{b}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ -1 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 12 \\ 7 \\ 0 \\ 3 \end{bmatrix}$,

and $B = \{\mathbf{b}_1, \mathbf{b}_2\}$. Since $\mathbf{b}_1, \mathbf{b}_2$ is linearly independent, B forms a basis for $H = Span(B)$. Let us find whether $\mathbf{y} \in H$, and what is the coordinate vector to $\mathbf{y}$ relative to the basis B.

If $\mathbf{y} \in H$ we will find weights, c_1, c_2 such that

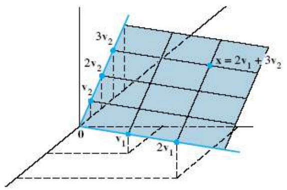
$$c_1 \begin{bmatrix} 6 \\ 2 \\ 0 \\ 3 \end{bmatrix} + c_2 \begin{bmatrix} 0 \\ 1 \\ 0 \\ -1 \end{bmatrix} = \begin{bmatrix} 12 \\ 7 \\ 0 \\ 3 \end{bmatrix}.$$

the extended matrix for the solution of the system is

$$\begin{bmatrix} 6 & 0 & 12 \\ 2 & 1 & 7 \\ 0 & 0 & 0 \\ 3 & -1 & 3 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

This give $c_1 = 2, c_2 = 3$ og $[\mathbf{y}]_B = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$.

note: Mark that the vectors $\mathbf{b}_1,\ \mathbf{b}_2 \in H \subseteq \mathbf{R}^4$, and the coordinates for these vectors and linear combinations of them identify with points in $\mathbf{R}^2$! The figure below shows and example of a 2-dimenstional coordinate system on a plane H i $\mathbf{R}^3$:



Grid structure on the plane in the figure indicates that we can think about the plane H as a "copy" of $\mathbf{R}^2$.

Dimension (2.9)

Definition
The dimension of a subspace $H \subseteq \mathbf{R}^n$ is given by $\dim(H)$, and is an integer corresponding to the number of elements in a any basis for H. For $H = \{\mathbf{0}\}$ (contains only the nullvector) we define $\dim(H) = 0$.

We have mastered the technique to be used to find the basis for the null space for a matrix by finding the solution set to the homogeneous system $A\mathbf{x} = \mathbf{0}$.

Previously we found the null space of the matrix

$$A = \begin{bmatrix} 5 & 2 & 0 & 1 & -6 \\ 1 & 5 & 1 & 5 & -2 \\ -2 & \frac{3}{2} & \frac{1}{2} & 2 & 2 \end{bmatrix}$$

and determined the linearly independent vectors $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$, where

$$\mathbf{v}_1 = \begin{bmatrix} \frac{2}{23} \\ -\frac{5}{23} \\ 1 \\ 0 \\ 0 \end{bmatrix}, \ \mathbf{v}_2 = \begin{bmatrix} \frac{5}{23} \\ -\frac{24}{23} \\ 0 \\ 1 \\ 0 \end{bmatrix}, \ \mathbf{v}_3 = \begin{bmatrix} \frac{26}{23} \\ \frac{4}{23} \\ 0 \\ 0 \\ 1 \end{bmatrix},$$

such that $Nul(A) = Span(\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\})$ and the dimension to $Nul(A)$, $\dim(Nul(A)) = 3$.
Definition
Rank to a matrix A is called $rank(A)$ and is defined as the dimension to the space spanned out in the columns of A.

Dimension cont.

Since the columns that correspond to the columns with pivot elements in row equivalent echelon form B to A ($A \sim B$) form a basis for $Col(A)$, so will the rank of A always be given by the total number of columns with pivot elements.

Example
Matrix Rank

$$A = \begin{bmatrix} 5 & 2 & 0 & 1 & -6 \\ 1 & 5 & 1 & 5 & -2 \\ -2 & \frac{3}{2} & \frac{1}{2} & 2 & 2 \end{bmatrix} \sim ... \sim \begin{bmatrix} 1 & 0 & -\frac{2}{23} & -\frac{5}{23} & -\frac{26}{23} \\ 0 & 1 & \frac{5}{23} & \frac{24}{23} & -\frac{4}{23} \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

we find the columns with pivot elements, therefore $rank(A) = 2$ (note that A has $n = 5$ columns). In the last example we found $\dim(Nul(A)) = 3$ and $rank(A) = 2$, so $\dim(Nul(A)) + rank(A) = 5(= n)$. This relationship holds for any matrix:

Theorem 14 - Rank Theorem
IF A is a matrix with n columns, it always holds that

$$rank(A) + \dim(Nul(A)) = n.$$

The Basis theorem

Theorem 15 - The Basis Theorem
assume that $H \subseteq \mathbf{R}^n$ is an p - dimensional subspace ($p \leq n$). Then every linearly independent set of p vectors i H automatically form a basis for H. Therefore we have many sets consisting of p vectors in H that span H, all of which form a basis for H.

Introduction to determinants (3.1/3.2)

The determinant to a 2×2 - matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is designated by "$\det(A)$" and is defined as

$$\det(A) = ad - cb.$$

In the trivial case of a 1×1 - matrix $[a]$ (nothing more than a simple number) we define $\det([a]) = a$.

Assume now more generally a $n \times n$ - matrix A, and let A_{ij} be $(n-1) \times (n-1)$ - matrices that arise when we eliminate the i - th row and j - th column A.

Example
With $A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & 16 \end{bmatrix}$ then $A_{23} = \begin{bmatrix} 1 & 2 & 4 \\ 9 & 10 & 12 \\ 13 & 14 & 16 \end{bmatrix}$.

Definition

For $n \geq 2$ the *determinant* to a $n \times n$ - matrix $A = [a_{ij}]$ is defined as

$$\det(A) = a_{11}\det(A_{11}) - a_{12}\det(A_{12}) + ... + (-1)^{n+1}a_{1n}\det(A_{1n})$$
$$= \sum_{j=1}^{n}(-1)^{j+1}a_{1j}\det(A_{1j}).$$

Example

Let us calculate the determinant to $A = \begin{bmatrix} 1 & 2 & 0 \\ 3 & -1 & 2 \\ 2 & 0 & 1 \end{bmatrix}$:

$$\det(A) = a_{11}\det(A_{11}) - a_{12}\det(A_{12}) + a_{13}\det(A_{13})$$
$$= 1\det(\begin{bmatrix} -1 & 2 \\ 0 & 1 \end{bmatrix}) - 2\det(\begin{bmatrix} 3 & 2 \\ 2 & 1 \end{bmatrix}) + 0\det(\begin{bmatrix} 3 & -1 \\ 2 & 0 \end{bmatrix})$$
$$= 1(-1-0) - 2(3-4) + 0(0-(-2)) = 1.$$

A common method to designate the the determinant is to use vertical lines in place of the matrix parentheses,

$$\det(\begin{bmatrix} 1 & 2 & 0 \\ 3 & -1 & 2 \\ 2 & 0 & 1 \end{bmatrix}) = \begin{vmatrix} 1 & 2 & 0 \\ 3 & -1 & 2 \\ 2 & 0 & 1 \end{vmatrix}$$

An example in this notation is -

$$\begin{vmatrix} 1 & 2 & 0 \\ 3 & -1 & 2 \\ 2 & 0 & 1 \end{vmatrix} = 1\begin{vmatrix} -1 & 2 \\ 0 & 1 \end{vmatrix} - 2\begin{vmatrix} 3 & 2 \\ 2 & 1 \end{vmatrix} + 0\begin{vmatrix} 3 & -1 \\ 2 & 0 \end{vmatrix}.$$

Cofactor expansion

Definition
The (i,j) - th *cofactors* to A is designated C_{ij} and is given as

$$C_{ij} = (-1)^{i+j}\det(A_{ij}).$$

With this definition we have

$$\det(A) = a_{11}C_{11} + a_{12}C_{12} + ... + a_{1n}C_{1n},$$

and this for is referred to as *cofactor expansion* to $\det(A)$ along the first row in A.

Theorem 1
The determinant to a $n \times n$ - matrix A can be calculated through a cofactor expansion along any chosen row or column in matrix A. Along the i-th row the cofactor expansion becomes

$$\det(A) = a_{i1}C_{i1} + a_{i2}C_{i2} + ... + a_{in}C_{in}.$$

Along the j-th column the cofactor expansion becomes

$$\det(A) = a_{1j}C_{1j} + a_{2j}C_{2j} + ... + a_{nj}C_{nj}.$$

cofactor expansion cont.

To keep the signs clean in the cofactor expansion one can make use of the sign matrix -

$$\begin{bmatrix} + & - & + & \dots \\ - & + & - & \dots \\ + & - & + & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

Example

Let us calculate the determinant to $A = \begin{bmatrix} 1 & 2 & 0 \\ 3 & -1 & 2 \\ 2 & 0 & 1 \end{bmatrix}$ Through a cofactor expansion along the 3rd column:

$$\begin{vmatrix} 1 & 2 & 0 \\ 3 & -1 & 2 \\ 2 & 0 & 1 \end{vmatrix}$$

$$= 0(-1)^{1+3}\begin{vmatrix} 3 & -1 \\ 2 & 0 \end{vmatrix} + 2(-1)^{2+3}\begin{vmatrix} 1 & 2 \\ 2 & 0 \end{vmatrix} + 1(-1)^{3+3}\begin{vmatrix} 1 & 2 \\ 3 & -1 \end{vmatrix}$$

$$= 0(2) - 2(-4) + 1(-7) = 1.$$

cofactor expansion cont.

Example

Let us calculate the determinant to $A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 2 & 1 & 5 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 3 & 5 \end{bmatrix}$.

Here it is wise to chose a cofactor expansion that gives the fewest calculations. We should therefore examine the rows and columns to find the entries of 0.

$$\begin{vmatrix} 1 & 2 & 3 & 4 \\ 0 & 2 & 1 & 5 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 3 & 5 \end{vmatrix}$$

$$= 1\begin{vmatrix} 2 & 1 & 5 \\ 0 & 2 & 1 \\ 0 & 3 & 5 \end{vmatrix} - 0\begin{vmatrix} 2 & 3 & 4 \\ 0 & 2 & 1 \\ 0 & 3 & 5 \end{vmatrix} + 0\begin{vmatrix} 2 & 3 & 4 \\ 2 & 1 & 5 \\ 0 & 3 & 5 \end{vmatrix} - 0\begin{vmatrix} 2 & 3 & 4 \\ 2 & 1 & 5 \\ 0 & 2 & 1 \end{vmatrix}$$

$$= 1 \cdot 2\begin{vmatrix} 2 & 1 \\ 3 & 5 \end{vmatrix} = 1 \cdot 2(10 - 3) = 14.$$

Be aware that the technique of determinant calculation is not efficient with large matrices. (see note in textbook)

The determinant to triangular matrices

We have to types of *triangular* matrices:

$$\begin{bmatrix} a_{11} & * & \dots & * & * \\ 0 & a_{22} & \dots & * & * \\ 0 & 0 & \ddots & * & * \\ 0 & 0 & 0 & a_{n-1,n-1} & * \\ 0 & 0 & 0 & 0 & a_{nn} \end{bmatrix} \quad \begin{bmatrix} a_{11} & 0 & 0 & 0 & 0 \\ \star & a_{22} & 0 & 0 & 0 \\ \star & \star & \ddots & 0 & 0 \\ \star & \star & \dots & a_{n-1,n-1} & 0 \\ \star & \star & \dots & \star & a_{nn} \end{bmatrix}$$

Upper triangular Lower triangular

Theorem 2
Given A is a triangular $n \times n$ - matrix with the diagonal elements $a_{11}, a_{22}, ..., a_{nn}$, then

$$det(A) = a_{11} \cdot a_{22} \cdot ... \cdot a_{nn}.$$

For the identity matrix I_n then

$$\det(I_n) = 1$$

Example

$$\begin{vmatrix} 2 & 3 & 4 & 5 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & -3 & 5 \\ 0 & 0 & 0 & 4 \end{vmatrix} = 2 \cdot 1 \cdot (-3) \cdot 4 = -24.$$

Rules for determinants

Assume that A, B are two $n \times n$ - matrices. Given matrix B can be created through elementary row operations on the matrix A gives rise to the following theorem:

Theorem 3

1. Row operations that add a multiple of one row in A to another row in a A to generate B: Then $\det(B) = \det(A)$.

2. Row operation of the type that switch two rows in A to produce B: Then $\det(B) = -\det(A)$.

3. Row operations that multiple a row in A by a constant k to produce B: Then $\det(B) = k \det(A)$.

   Example

Finn

$$\begin{vmatrix} 1 & 2 & 3 & 4 \\ 0 & 5 & 0 & 0 \\ 2 & 7 & 6 & 10 \\ 2 & 9 & 7 & 11 \end{vmatrix} = 5 \begin{vmatrix} 1 & 3 & 4 \\ 2 & 6 & 10 \\ 2 & 7 & 11 \end{vmatrix} = 5 \begin{vmatrix} 1 & 3 & 4 \\ 0 & 0 & 2 \\ 2 & 7 & 11 \end{vmatrix}$$

$$= 5 \begin{vmatrix} 1 & 3 & 4 \\ 0 & 0 & 2 \\ 0 & 1 & 3 \end{vmatrix} = -5 \begin{vmatrix} 1 & 3 & 4 \\ 0 & 1 & 3 \\ 0 & 0 & 2 \end{vmatrix} = (-5) \cdot 1 \cdot 1 \cdot 2 = -10$$

More rules for for determinants

Theorem 3, third point indicates thatt

$$\begin{vmatrix} \star & \star & \star \\ -2k & 5k & 4k \\ \star & \star & \star \end{vmatrix} = k \begin{vmatrix} \star & \star & \star \\ -2 & 5 & 4 \\ \star & \star & \star \end{vmatrix}.$$

Example
Let us calculate

$$\begin{vmatrix} 2 & 4 & 6 \\ 5 & 6 & 7 \\ 7 & 6 & 10 \end{vmatrix} = 2 \begin{vmatrix} 1 & 2 & 3 \\ 5 & 6 & 7 \\ 7 & 6 & 10 \end{vmatrix} = 2 \begin{vmatrix} 1 & 2 & 3 \\ 0 & -4 & -8 \\ 0 & -8 & -11 \end{vmatrix}$$

$$= 2(-4) \begin{vmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & -8 & -11 \end{vmatrix} = 2(-4) \begin{vmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 5 \end{vmatrix}$$

$$= 2 \cdot (-4) \cdot 1 \cdot 1 \cdot 5 = -40.$$

Example
Find $\begin{vmatrix} 2 & 3 & 0 & 1 \\ 4 & 7 & 0 & 3 \\ 7 & 9 & -2 & 4 \\ 1 & 2 & 0 & 4 \end{vmatrix}$ through use of elementary row operations and cofactor expansion.

$$\begin{vmatrix} 2 & 3 & 0 & 1 \\ 4 & 7 & 0 & 3 \\ 7 & 9 & -2 & 4 \\ 1 & 2 & 0 & 4 \end{vmatrix} = -2 \begin{vmatrix} 2 & 3 & 1 \\ 4 & 7 & 3 \\ 1 & 2 & 4 \end{vmatrix} = -2 \begin{vmatrix} 2 & 3 & 1 \\ 0 & 1 & 1 \\ 1 & 2 & 4 \end{vmatrix}$$

$$2 \begin{vmatrix} 2 & 3 & 1 \\ 1 & 2 & 4 \\ 0 & 1 & 1 \end{vmatrix} = -2 \begin{vmatrix} 1 & 2 & 4 \\ 2 & 3 & 1 \\ 0 & 1 & 1 \end{vmatrix} = -2 \begin{vmatrix} 1 & 2 & 4 \\ 0 & -1 & -7 \\ 0 & 1 & 1 \end{vmatrix}$$

$$-2 \begin{vmatrix} 1 & 2 & 4 \\ 0 & -1 & -7 \\ 0 & 0 & -6 \end{vmatrix} = (-2) \cdot 1 \cdot (-1) \cdot (-6) = -12.$$

assume that matrices A og U are row equivalent ($A \sim U$), that U is in echelon form, through use of only the first two types of elementary row operations described in theorem 3.

$$U = \begin{bmatrix} u_1 & * & \dots & * & * \\ 0 & u_2 & \dots & * & * \\ 0 & 0 & \ddots & * & * \\ 0 & 0 & 0 & u_{n-1} & * \\ 0 & 0 & 0 & 0 & u_n \end{bmatrix}$$

and assume that row operation type 2 is used r times. Then

$$\det(A) = (-1)^r \det(U) = (-1)^r \cdot u_1 \cdot u_2 \cdot \ldots \cdot u_n.$$

Note that if U does not have pivot elements in all columns, we will end up having a 0 in the diagonal to U. This means that a can not be inverted and that $\det(A) = 0$. With this the following theorem is proven:

Theorem 4
A $n \times n$-matrix A is invertible if and only if

$$\det(A) \neq 0$$

.

Theorem 5
For a $n \times n$-matrix A,

$$\det(A) = \det(A^t)$$

Proof:
The general proof is based on mathematical induction with cofactor expansion along the first row in A and the first column in A^t, see Lay (3.2).

In the special case of a 2×2 - matrix we can show the following: Med $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ er $A^t = \begin{bmatrix} a & c \\ b & d \end{bmatrix}$ and $\det(A^t) = ad - bc = ad - cb = \det(A)$.

NOTE: An implication of theorem 5 is that theorem 3 continues to hold if we replace all statements about operations on rows with operations on the columns.

Theorem 6
For $n \times n$ - matrices A and B the following formula applies

$$\det(AB) = \det(A)\det(B)$$

Proof: See Lay (3.2).

Example:
Assume A is invertible. What happens with $\det(A^{-1})$? We know that $I_n = A^{-1}A$, and with help from theorem 6 we have -

$$1 = \det(I_n) = \det(A^{-1}A) = \det(A^{-1})\det(A)$$

$$\Downarrow$$

$$\det(A^{-1}) = \frac{1}{\det(A)}$$

Example:
If $\det(A) = 5$ and we need to find $\det(A^3)$ we can use $A^3 = A^2A$ together with theorem 6:

$$\det(A^3) = \det(A^2)\det(A) = \det(AA)\det(A)$$

$$= (\det(A)\det(A))\det(A) = \det(A)^3 = 5^3 = 125.$$

Example:
Assume A and B are two $n \times n$ - matrices and that $\det(B) \neq 0$. If $\det(AB) = 0$ it follows that

$$\det(A) = \frac{\det(A)\det(B)}{\det(B)} = \frac{\det(AB)}{\det(B)} = \frac{0}{\det(B)} = 0.$$

Through theorem 4 we can conclude that A is not invertible.

The next theorems summarize the more important results from sections 3.1 and 3.2 :

Theorem - rules for determinants
assume that A, B er to $n \times n$ - matrixr. Da holder følgende:

1. A is invertible if and only if $\det(A) \neq 0$.
2. $\det(AB) = \det(A)\det(B)$.
3. $\det(A^t) = \det(A)$.
4. If A is triangular, then $\det(A)$ is the product of the elements along the diagonal of A.
5. Elementary row operations which add a multiple of one row to another do not change the determinant.
6. Elementary row operations of the type that switch two rows only switch the sign of the determinant
7. Elementary row operations which multiply one row by a scalar scale the determinant by the same factor.

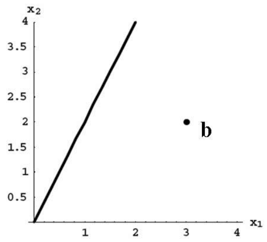
Inner product, norm, length and orthogonality (6.1)

What if $A\mathbf{x} = \mathbf{b}$ has no solution?

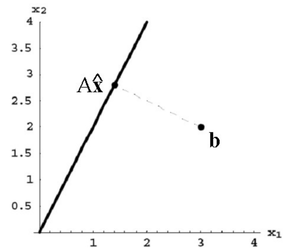
Example

$$\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$$

Note that $\mathbf{v} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ spans the column space of A and that $\mathbf{b} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$ is not contained in $Col(A)$.



No exact solution $\mathbf{x}$ can be found. However, it makes sense to seek a $\hat{\mathbf{x}}$ so that $A\hat{\mathbf{x}}$ is as close to $\mathbf{b}$ as possible:



Proper formalization of concepts like *lenght*, *orthogonality* and *orthogonal projections* are required.

The Inner Product

The *Inne Product* between two vectors in $\mathbf{R}^n$ $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$ og $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$

is given by

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^t \mathbf{v} = u_1 v_1 + u_2 v_2 + \ldots + u_n v_n.$$

Note that

$$\mathbf{v} \cdot \mathbf{u} = \mathbf{v}^t \mathbf{u} = v_1 u_1 + v_2 u_2 + \ldots + v_n u_n$$
$$= u_1 v_1 + u_2 v_2 + \ldots + u_n v_n = \mathbf{u}^t \mathbf{v} = \mathbf{u} \cdot \mathbf{v}.$$

Theorem 1
Let $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbf{R}^n$, and let c be a scalar. Then

1. $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$
2. $(\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = \mathbf{u} \cdot \mathbf{w} + \mathbf{v} \cdot \mathbf{w}$
3. $(c\mathbf{u}) \cdot \mathbf{v} = c(\mathbf{u} \cdot \mathbf{v}) = \mathbf{u} \cdot (c\mathbf{v})$

4. $\mathbf{u} \cdot \mathbf{u} \geq 0$, og $\mathbf{u} \cdot \mathbf{u} = 0$ if and only if $\mathbf{u} = \mathbf{0}$

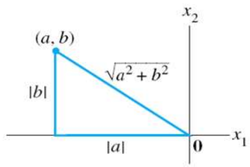
From 2 and 3 it follows that

$$(c_1\mathbf{u}_1 + \ldots + c_p\mathbf{u}_p) \cdot \mathbf{w} = c_1(\mathbf{u}_1 \cdot \mathbf{w}) + \ldots + c_p(\mathbf{u}_p \cdot \mathbf{w})$$

Definition
The *length* (or *norm*) of $\mathbf{v} \in \mathbf{R}^n$ is the non-negative number $\|\mathbf{v}\|$ defined by

$$\|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}} = \sqrt{v_1^2 + \ldots v_n^2}, \text{ og } \|\mathbf{v}\|^2 = \mathbf{v} \cdot \mathbf{v}$$

See example 1 (6.1) in Lay.



For any scalar c, $c\mathbf{v} = \begin{bmatrix} cv_1 \\ cv_2 \\ \vdots \\ cv_n \end{bmatrix}$ og

$$\|c\mathbf{v}\| = \sqrt{(cv_1)^2 + \ldots + (cv_n)^2} = \sqrt{c^2}\sqrt{v_1^2 + \ldots v_n^2} = |c|\|\mathbf{v}\|$$

A vector of length 1 is called a *unit vector*. By appropriate scaling of $\mathbf{v} \neq \mathbf{0}$, we obtain a proportional unit vector $\mathbf{u} = \frac{1}{\|\mathbf{v}\|}\mathbf{v}$ (see ex. 2 & 3 (6.1) in Lay.)

Distance between vectors in $\mathbf{R}^n$

Definition
The *distance* between the vectors $\mathbf{u}, \mathbf{v} \in \mathbf{R}^n$ is defined as

$$dist(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|$$

Recall form $\mathbf{R}^2$ (and $\mathbf{R}^3$):

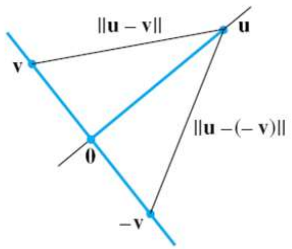
For $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$, we have $\mathbf{u} - \mathbf{v} = \begin{bmatrix} u_1 - v_1 \\ u_2 - v_2 \end{bmatrix}$ and

$$dist(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\| = \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2}.$$

See ex. 4 og 5 (6.1) in Lay.

Orthogonal vectors

In the figure $\mathbf{u}$ and $\mathbf{v}$ are orthogonal by design:



Note that

$$dist(\mathbf{u}, \mathbf{v})^2 = \|\mathbf{u} - \mathbf{v}\|^2 = (\mathbf{u} - \mathbf{v}) \cdot (\mathbf{u} - \mathbf{v})$$
$$\mathbf{u} \cdot (\mathbf{u} - \mathbf{v}) + (-\mathbf{v}) \cdot (\mathbf{u} - \mathbf{v})$$
$$\mathbf{u} \cdot \mathbf{u} - \mathbf{u} \cdot \mathbf{v} - \mathbf{v} \cdot \mathbf{u} + \mathbf{v} \cdot \mathbf{v}$$

$$\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - 2\mathbf{u} \cdot \mathbf{v}$$

By a similar argument we also have

$$dist(\mathbf{u}, -\mathbf{v})^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 + 2\mathbf{u} \cdot \mathbf{v}.$$

By symmetry we must have

$$dist(\mathbf{u}, -\mathbf{v}) = dist(\mathbf{u}, \mathbf{v}).$$

This implies that $2\mathbf{u} \cdot \mathbf{v} = -2\mathbf{u} \cdot \mathbf{v} \Rightarrow \mathbf{u} \cdot \mathbf{v} = 0$.

Definition
Two vectors $\mathbf{u}, \mathbf{v}$ are called (mutually) *orthogonal* if

$$\mathbf{u} \cdot \mathbf{v} = 0.$$

Note that $\mathbf{0} \in \mathbf{R}^n$ is orthogonal to any vector $\mathbf{v} \in \mathbf{R}^n$.

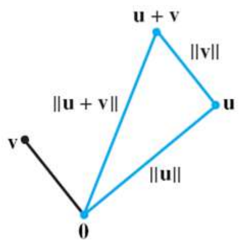
Also note that if $\mathbf{u}$ and $\mathbf{v}$ are orthogonal than

$$\|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u} - (-\mathbf{v})\|^2 = dist(\mathbf{u}, -\mathbf{v})^2$$
$$= \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 + 2\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2.$$

This result is known as

Teorem 2 - Pythagoras theorem
Two vektors $\mathbf{u}, \mathbf{v} \in \mathbf{R}^n$ are orthogonal if and only if

$$\|\mathbf{u}+\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2.$$



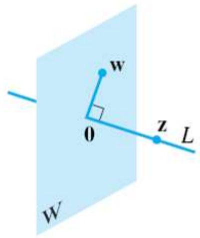
Orthogonal complements

If $\mathbf{z} \in \mathbf{R}^n$ is orthogonal to any vector in the subspace $W \subseteq \mathbf{R}^n$ we say that $\mathbf{z}$ is *orthogonal to* W.

The set of W-orthogonal vectors is denoted

$$W^\perp = \{\mathbf{z} \in \mathbf{R}^n | \ \mathbf{z} \cdot \mathbf{w} = 0 \text{ for all } \mathbf{w} \in W\}$$

and referred to as *the orthogonal complement* of W.



Two important properties regarding orthogonal complements in $\mathbf{R}^n$:

1. $\mathbf{x} \in W^\perp$ if and only if $\mathbf{x}$ is orthogonal to every vector in a set $\{\mathbf{w}_1, ..., \mathbf{w}_k\}$ spanning $W = Span(\{\mathbf{w}_1, ..., \mathbf{w}_k\})$.

2. $W^\perp$ is a subspace of $\mathbf{R}^n$.

(See exercise 6.1.29.)

Theorem 3
Assume that the matrix A is $m \times n$. Then the orthogonal complement to the subspace spanned by the rows in A is identical to the null space of A, and the orthogonal complement of the subspace spanned by the columns of A is identical to the null space of A^t:

$$Row(A)^\perp = Nul(A) \text{ og } Col(A)^\perp = Nul(A^t).$$

Here,
$Col(A) = Span(A)$, $Row(A) = Span(A^t) = Col(A^t)$, and $Nul(A) = \{\mathbf{x} \in \mathbf{R}^n | \ A\mathbf{x} = \mathbf{0}\}$.

Proof:

Anta at $A\mathbf{x} = \mathbf{0}$. Note that

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} = \begin{bmatrix} \mathbf{r}_1^t \\ \mathbf{r}_2^t \\ \vdots \\ \mathbf{r}_m^t \end{bmatrix}$$

where $\mathbf{r}_k^t = [a_{k1} \ a_{k2} \ \dots \ a_{kn}]$ is the k-th row of A. Using this notation we have

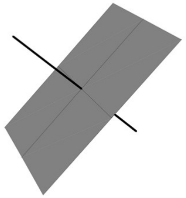
$$A\mathbf{x} = \begin{bmatrix} \mathbf{r}_1^t \\ \mathbf{r}_2^t \\ \vdots \\ \mathbf{r}_m^t \end{bmatrix} \mathbf{x} = \begin{bmatrix} \mathbf{r}_1^t\mathbf{x} \\ \mathbf{r}_2^t\mathbf{x} \\ \vdots \\ \mathbf{r}_m^t\mathbf{x} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

meaning that $\mathbf{x}$ is contained in $Nul(A)$ and is orthogonal to the rows of A, i.e. $\mathbf{x} \in Nul(A)$ if and only if $\mathbf{x}$ is orthogonal to $Row(A) = Span(\{\mathbf{r}_1, ..., \mathbf{r}_m\})$. In other words $\mathbf{x} \in Row(A)^\perp$. By setting $B = A^t$ we also have $Row(B)^\perp = Nul(B)$, and because $Row(B) = Col(A)$ $Nul(B) = Nul(A^t)$, $Col(A)^\perp = Nul(A^t)$ also holds.

Example
Let $A = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \end{bmatrix}$.

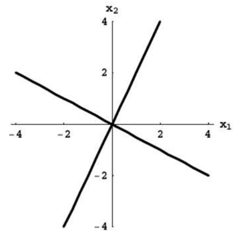
A basis for $Nul(A)$ is given by $\mathbf{v}_1 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$. The $Nul(A) = Span(\{\mathbf{v}_1, \mathbf{v}_2\})$ is a plane in $\mathbf{R}^3$.

The vector $\mathbf{u} = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$ represents a basis for $Row(A) = Span(\mathbf{u})$ that is a line in $\mathbf{R}^3$.



$Nul(A)$ (the plane) and $Row(A)$ (the line).

Note that $\mathbf{w}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ represents a basis for $Col(A) = Span(\mathbf{w}_1)$ corresponding to a line in $\mathbf{R}^2$, and $\mathbf{w}_2 = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$ represents a basis for $Nul(A^t) = Span(\mathbf{w}_2)$ - another line $\mathbf{R}^2$.



$Nul(A^t)$ (line) and $Col(A)$ (also a line).

Orthogonal sets (6.2)

A set of vectos $\{\mathbf{u}_1, ..., \mathbf{u}_p\}$ i $\mathbf{R}^n$ is called an *orthogonal set* if $\mathbf{u}_i \cdot \mathbf{u}_j = 0$ for $i \neq j$.

Example

Decide if $\left\{ \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$ is an orthogonal set:

By referring to these vectors as $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ we must check if the inder product between every pair of vektors $\mathbf{u}_i \cdot \mathbf{u}_j = 0$ for $1 \leq i < j \leq 3$. There are three possibilities:

$$\mathbf{u}_1 \cdot \mathbf{u}_2 = 1 \cdot 1 + (-1) \cdot 1 + 0 \cdot 0 = 0$$

$$\mathbf{u}_1 \cdot \mathbf{u}_3 = 1 \cdot 0 + (-1) \cdot 0 + 0 \cdot 1 = 0$$

$$\mathbf{u}_2 \cdot \mathbf{u}_3 = 1 \cdot 0 + 1 \cdot 0 + 0 \cdot 1 = 0$$

Consequently, $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is an orthogonal set.

Theorem 4
Assume that $S = \{\mathbf{u}_1, ..., \mathbf{u}_p\}$ is an orthogonal set in $\mathbf{R}^n$ noth containing $\mathbf{0}$, and that $W = span(S)$. Then S is a linearly independent set, and consequently a basis for W.

Proof: Assume that the coefficients $c_1, ..., c_p$ have been chosen so that

$$c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + ... + c_p\mathbf{u}_p = \mathbf{0}.$$

For any $\mathbf{u}_k$ $(1 \leq k \leq p)$ we must have

$$(c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + ... + c_p\mathbf{u}_p) \cdot \mathbf{u}_k = \mathbf{0} \cdot \mathbf{u}_k = 0$$

$$\Downarrow$$

$$c_1(\mathbf{u}_1 \cdot \mathbf{u}_k) + c_2(\mathbf{u}_2 \cdot \mathbf{u}_k) + ... + c_k(\mathbf{u}_k \cdot \mathbf{u}_k) + ... + c_p(\mathbf{u}_p \cdot \mathbf{u}_k) = 0$$

$$\Downarrow \text{ because } \mathbf{u}_i \cdot \mathbf{u}_k = 0 \text{ for } i \neq k$$

$$c_k(\mathbf{u}_k \cdot \mathbf{u}_k) = 0$$

$$\Downarrow$$

$$c_k = 0$$

Thus $c_1 = c_2 = ... = c_p = 0$, meaning that the vectors in S are linearly independent and therefore a basis for W.

Definition
An *orthogonal basis* for a subspace W in $\mathbf{R}^n$ is a basis for W that also is an orthogonal set.

Example
Assume that $S = \{\mathbf{u}_1, ..., \mathbf{u}_p\}$ is an orthogonal basis for $W = span(S)$, and that $\mathbf{y} \in W$. Find weights $c_1, c_2, ..., c_p$ (the coordinates of $\mathbf{y}$ w.r.t. the basis S) so that

$$\mathbf{y} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + ... + c_p\mathbf{u}_p.$$

Solution: Note that for $1 \le k \le p$ we have

$$\mathbf{y} \cdot \mathbf{u}_k = (c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + ... + c_p\mathbf{u}_p) \cdot \mathbf{u}_k$$

$$\Updownarrow$$

$$\mathbf{y} \cdot \mathbf{u}_k = c_1(\mathbf{u}_1 \cdot \mathbf{u}_k) + c_2(\mathbf{u}_2 \cdot \mathbf{u}_k) + ... + c_k(\mathbf{u}_k \cdot \mathbf{u}_k) + ... + c_p(\mathbf{u}_p \cdot \mathbf{u}_k)$$

$$\Downarrow \text{ because } \mathbf{u}_i \cdot \mathbf{u}_k = 0 \text{ for } i \neq k$$

$$\mathbf{y} \cdot \mathbf{u}_k = c_k(\mathbf{u}_k \cdot \mathbf{u}_k)$$

$$\Downarrow$$

$$c_k = \frac{\mathbf{y} \cdot \mathbf{u}_k}{\mathbf{u}_k \cdot \mathbf{u}_k} \text{ for } k = 1, ..., p.$$

This proves

Teorem 5
La $\{\mathbf{u}_1, ..., \mathbf{u}_p\}$ være en ortogonal basis for et underrom W av $\mathbf{R}^n$. Da har enhver $\mathbf{y} \in W$ en entydig representasjon som en lineær kombinasjon av $\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_p$ med

$$\mathbf{y} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + ... + c_p\mathbf{u}_p \text{ der}$$

$$c_k = \frac{\mathbf{y} \cdot \mathbf{u}_k}{\mathbf{u}_k \cdot \mathbf{u}_k} \text{ for } k = 1, ..., p.$$

Example

Express $\mathbf{y} = \begin{bmatrix} 3 \\ 7 \\ 4 \end{bmatrix}$ as a linear combination of vectors in the orthogonal basis

$$\left\{ \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

SOlution: By theorem 5 we have $c_1 = \frac{\mathbf{y}\cdot\mathbf{u}_1}{\mathbf{u}_1\cdot\mathbf{u}_1} = \frac{-4}{2} = -2$, $c_2 = \frac{\mathbf{y}\cdot\mathbf{u}_2}{\mathbf{u}_2\cdot\mathbf{u}_2} = \frac{10}{2} = 5$ and $c_3 = \frac{\mathbf{y}\cdot\mathbf{u}_3}{\mathbf{u}_3\cdot\mathbf{u}_3} = \frac{4}{1} = 4$. Thus

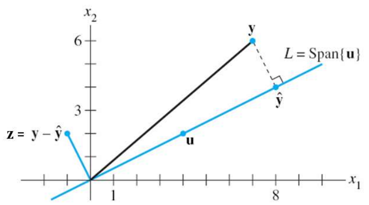
$$\mathbf{y} = -2\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} + 5\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + 4\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

Orthogonal projections - onto a vector

Let $\mathbf{u} \in \mathbf{R}^n$. We would like to express $\mathbf{y} \in \mathbf{R}^n$ as a linear combination

$$\mathbf{y} = c\mathbf{u} + \mathbf{z} \Leftrightarrow \mathbf{z} = \mathbf{y} - c\mathbf{u}$$

where $\mathbf{z} \cdot \mathbf{u} = 0$ (i.e. $\mathbf{z}$ is chosen to be orthogonal to $\mathbf{u}$).



The orthogonal projection of $\mathbf{y}$ onto the line spanned by $\mathbf{u}$. Requiring orthogonality between $\mathbf{u}$ and $\mathbf{z}$ implies

$$(\mathbf{y} - c\mathbf{u}) \cdot \mathbf{u} = 0$$

$$\Downarrow$$

$$\mathbf{y}\cdot\mathbf{u} - c(\mathbf{u}\cdot\mathbf{u}) = 0 \Rightarrow c = \frac{\mathbf{y}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}$$

Therefore

$\hat{\mathbf{y}} = \frac{\mathbf{y}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}\mathbf{u} = c\mathbf{u}$ (the orthogonal projection of $\mathbf{y}$ onto $\mathbf{u}$)

$\mathbf{z} = \mathbf{y} - c\mathbf{u} = \mathbf{y} - \hat{\mathbf{y}}$ (the component in $\mathbf{y}$ orthogonal to $\mathbf{u}$)

Orthonormal vectors

A set of vectors $S = \{\mathbf{u}_1, ..., \mathbf{u}_p\}$ in $\mathbf{R}^n$ are called an *orthonormal set* if S consists of orthogonal unit vectors.

If $W = Span(S)$ then $S = \{\mathbf{u}_1, ..., \mathbf{u}_p\}$ is an orthonormal basis for the subspace W.

Recall that if $\mathbf{v}$ is a unit vector, then $\|\mathbf{v}\| = 1$.

Assume that $U = [\mathbf{u}_1\ \mathbf{u}_2\ \mathbf{u}_3]$, where $S = \{\mathbf{u}_1,\ \mathbf{u}_2,\ \mathbf{u}_3\}$ is an orthonormal set in $\mathbf{R}^3$. Then

$$U^tU = \begin{bmatrix} \mathbf{u}_1^t \\ \mathbf{u}_2^t \\ \mathbf{u}_3^t \end{bmatrix} [\mathbf{u}_1\ \mathbf{u}_2\ \mathbf{u}_3]$$

$$= \begin{bmatrix} \mathbf{u}_1^t\mathbf{u}_1 & \mathbf{u}_1^t\mathbf{u}_2 & \mathbf{u}_1^t\mathbf{u}_3 \\ \mathbf{u}_2^t\mathbf{u}_1 & \mathbf{u}_2^t\mathbf{u}_2 & \mathbf{u}_2^t\mathbf{u}_3 \\ \mathbf{u}_3^t\mathbf{u}_1 & \mathbf{u}_3^t\mathbf{u}_2 & \mathbf{u}_3^t\mathbf{u}_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

See also ex. 5 (6.2) in Lay.

If the matrix U is $n \times n$ and $U^tU = I$, we also have $UU^t = I$. Consequently $U^{-1} = U^t$. Such matrices are referred to as *orthogonal matrices*.

Two theorems

Matrices with orthonormal columns are an important part of many numerical algorithms involving matrices. Some important properties are formulated in the following theorems:

Theorem 6
A $m \times n$-matrix U has orthonormal columns if and only if $U^tU = I_n$.

Teorem 7
Assume that the $m \times n$-matrix U has orthonormal columns, and let $\mathbf{x}, \mathbf{y} \in \mathbf{R}^n$. Then

1. $\|U\mathbf{x}\| = \|\mathbf{x}\|$
2. $(U\mathbf{x}) \cdot (U\mathbf{y}) = \mathbf{x} \cdot \mathbf{y}$
3. $(U\mathbf{x}) \cdot (U\mathbf{y}) = 0$ if and only if $\mathbf{x} \cdot \mathbf{y} = 0$

Proof:

1. Holds because $\|U\mathbf{x}\| = \sqrt{U\mathbf{x}\cdot U\mathbf{x}} = \sqrt{\mathbf{x}^tU^tU\mathbf{x}} = \sqrt{\mathbf{x}^tI\mathbf{x}} = \sqrt{\mathbf{x}^t\mathbf{x}} = \|\mathbf{x}\|$.
2. Holds because $(U\mathbf{x}) \cdot (U\mathbf{y}) = \mathbf{x}^tU^tU\mathbf{y} = \mathbf{x}^tI\mathbf{y} = \mathbf{x}^t\mathbf{y} = \mathbf{x} \cdot \mathbf{y}$.
3. This is a special case of the previous point showing that $(U\mathbf{x})\cdot(U\mathbf{y}) = \mathbf{x}\cdot\mathbf{y}$ always holds.

Se examples 6 and 7 (6.2) in Lay.

Orthogonal projections (6.3)

Teorem 8 - orthogonal decompositions
Let W be a subspace of $\mathbf{R}^n$. Then any vector $\mathbf{y} \in \mathbf{R}^n$ can be expressed uniquely as a sum

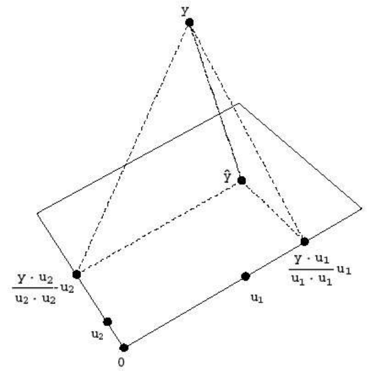
$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z},$$

where $\hat{\mathbf{y}} \in W$ and $\mathbf{z} \in W^{\perp}$. If $\{\mathbf{u}_1, ..., \mathbf{u}_p\}$ is an orthogonal basis for W,

$$\hat{\mathbf{y}} = \frac{\mathbf{y}\cdot\mathbf{u}_1}{\mathbf{u}_1\cdot\mathbf{u}_1}\mathbf{u}_1 + ... + \frac{\mathbf{y}\cdot\mathbf{u}_p}{\mathbf{u}_p\cdot\mathbf{u}_p}\mathbf{u}_p$$

$$\text{and } \mathbf{z} = \mathbf{y} - \hat{\mathbf{y}}$$

The vector $\hat{\mathbf{y}}$ is called the *orthogonal projection of* $\mathbf{y}$ *onto* W, see the figure:

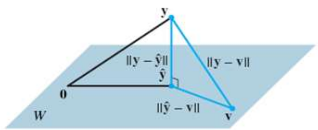


Best approximation

Theorem 9 - Best approximation theorem
Let W be a subspace of $\mathbf{R}^n$, and $\mathbf{y} \in \mathbf{R}^n$. If $\hat{\mathbf{y}}$ is the orthogonal projection of $\mathbf{y}$ onto W, then $\hat{\mathbf{y}}$ is the vector in W closest to $\mathbf{y}$ in the sense that

$$\|\mathbf{y} - \hat{\mathbf{y}}\| < \|\mathbf{y} - \mathbf{v}\|$$

for all $\mathbf{v} \in W$, where $\mathbf{v} \neq \hat{\mathbf{y}}$.



<u>Proof:</u> Choose any $\mathbf{v} \in W$, where $\mathbf{v} \neq \hat{\mathbf{y}}$. Because W is a vector space, then the difference $\mathbf{v} - \hat{\mathbf{y}} \in W$, and $\mathbf{z} = \mathbf{y} - \hat{\mathbf{y}}$ which is er orthogonal to W is in particular orthogonal to $\mathbf{v} - \hat{\mathbf{y}}$. From

$$\mathbf{y} - \mathbf{v} = (\mathbf{y} - \hat{\mathbf{y}}) + (\hat{\mathbf{y}} - \mathbf{v})$$

and Theorem 2 (Pythagoras theorem) it follows that

$$\|\mathbf{y} - \mathbf{v}\|^2 = \|\mathbf{y} - \hat{\mathbf{y}}\|^2 + \|\hat{\mathbf{y}} - \mathbf{v}\|^2$$

$$\Downarrow$$

$$\|\mathbf{y} - \mathbf{v}\|^2 > \|\mathbf{y} - \hat{\mathbf{y}}\|^2 \Rightarrow \|\mathbf{y} - \hat{\mathbf{y}}\| < \|\mathbf{y} - \mathbf{v}\|$$

Best approximation (cont.)

Example

Let $\mathbf{y} = \begin{bmatrix} 2 \\ 4 \\ 0 \\ -2 \end{bmatrix}, \mathbf{u}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}$ and $\mathbf{u}_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$. Let's compute the orthogonal projection $\hat{\mathbf{y}} \in W = span(\{\mathbf{u}_1, \mathbf{u}_2\})$.

Solution: Because

$$\hat{\mathbf{y}} = \frac{\mathbf{y} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1}\mathbf{u}_1 + \frac{\mathbf{y} \cdot \mathbf{u}_2}{\mathbf{u}_2 \cdot \mathbf{u}_2}\mathbf{u}_2$$

and $\mathbf{y} \cdot \mathbf{u}_1 = 6$, $\mathbf{y} \cdot \mathbf{u}_2 = -2$, $\mathbf{u}_1 \cdot \mathbf{u}_1 = 2$ og $\mathbf{u}_2 \cdot \mathbf{u}_2 = 2$, we find

$$\hat{\mathbf{y}} = (3) \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} + (-1) \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \\ -1 \\ -1 \end{bmatrix}$$

For the special case where $\{\mathbf{u}_1, ..., \mathbf{u}_p\}$ is an orthonormal basis, the expression

$$\hat{\mathbf{y}} = \frac{\mathbf{y} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1}\mathbf{u}_1 + ... + \frac{\mathbf{y} \cdot \mathbf{u}_p}{\mathbf{u}_p \cdot \mathbf{u}_p}\mathbf{u}_p$$

in theorem 8 simplifies to

$$\hat{\mathbf{y}} = (\mathbf{y} \cdot \mathbf{u}_1)\mathbf{u}_1 + ... + (\mathbf{y} \cdot \mathbf{u}_p)\mathbf{u}_p$$

because $\mathbf{u}_k \cdot \mathbf{u}_k = 1$ for $k = 1, ..., p$. The first part of the following theorem follows directly from this observation:

Theorem 10
If $\{\mathbf{u}_1, ..., \mathbf{u}_p\}$ is an orthonormal basis for a subspace $W \subseteq \mathbf{R}^n$ and $\mathbf{y} \in \mathbf{R}^n$, then

$$\hat{\mathbf{y}} = proj_W(\mathbf{y}) = (\mathbf{y} \cdot \mathbf{u}_1)\mathbf{u}_1 + ... + (\mathbf{y} \cdot \mathbf{u}_p)\mathbf{u}_p,$$

and for $U = [\mathbf{u}_1\ \mathbf{u}_2\ ...\ \mathbf{u}_p]$ we obtain the formula

$$\hat{\mathbf{y}} = proj_W(\mathbf{y}) = UU^t\mathbf{y}.$$

Proof: Note that $U^t\mathbf{y}$ is a $p \times 1$-vector with entries

$$U^t\mathbf{y} = \begin{bmatrix} \mathbf{u}_1^t\mathbf{y} \\ \mathbf{u}_2^t\mathbf{y} \\ \vdots \\ \mathbf{u}_p^t\mathbf{y} \end{bmatrix} = \begin{bmatrix} \mathbf{u}_1 \cdot \mathbf{y} \\ \mathbf{u}_2 \cdot \mathbf{y} \\ \vdots \\ \mathbf{u}_p \cdot \mathbf{y} \end{bmatrix} = \begin{bmatrix} \mathbf{y} \cdot \mathbf{u}_1 \\ \mathbf{y} \cdot \mathbf{u}_2 \\ \vdots \\ \mathbf{y} \cdot \mathbf{u}_p \end{bmatrix}.$$

By the rules of matrix-vector multiplication

$$UU^t\mathbf{y} = U(U^t\mathbf{y}) = U \begin{bmatrix} \mathbf{y} \cdot \mathbf{u}_1 \\ \mathbf{y} \cdot \mathbf{u}_2 \\ \vdots \\ \mathbf{y} \cdot \mathbf{u}_p \end{bmatrix}$$

$$= (\mathbf{y} \cdot \mathbf{u}_1)\mathbf{u}_1 + (\mathbf{y} \cdot \mathbf{u}_2)\mathbf{u}_2 + \ldots + (\mathbf{y} \cdot \mathbf{u}_p)\mathbf{u}_p = proj_W(\mathbf{y}) = \hat{\mathbf{y}}.$$

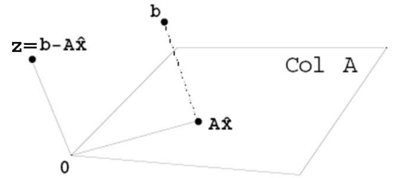
Least squares problems (6.5)

Problem: What can we do when the system

$$A\mathbf{x} = \mathbf{b}$$

has no solution $\mathbf{x}$ (i.e. is inconsistent)?

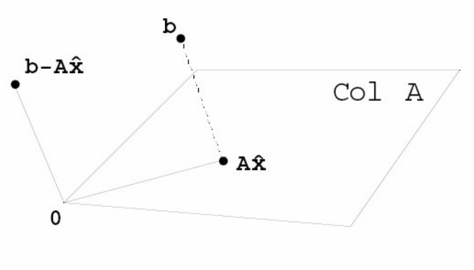
Answer: We seek the "best possible" approximate (least squares) solution $\hat{\mathbf{x}}$ in the sense that $\hat{\mathbf{b}} = A\hat{\mathbf{x}} \in Col(A)$ is as close to $\mathbf{b}$ as possible: The solution $\hat{\mathbf{x}} \in \mathbf{R}^n$ satisfies $\|\mathbf{b} - A\hat{\mathbf{x}}\| \leq \|\mathbf{b} - A\mathbf{x}\|$ holder for every choice of $\mathbf{x} \in \mathbf{R}^n$.



Least squares problems cont.

Because $\widehat{\mathbf{b}} \in W = Col(A)$ we know from Theorem 9 (best approximation) that $\widehat{\mathbf{b}} = proj_W(\mathbf{b})$ is the best achievable approximation to $\mathbf{b}$, and that the difference $\mathbf{z} = \mathbf{b} - \widehat{\mathbf{b}} \in W^\perp$.

STILL WE HAVE NOT YET SEEN HOW TO CALCULATE the least squares solution $\widehat{\mathbf{x}}$.



By the uniqueness in Theorem 8 (orthogonal decomposition theorem) it is sufficient to find a $\widehat{\mathbf{x}}$ so that $\mathbf{z} = \mathbf{b} - A\widehat{\mathbf{x}}$ becomes orthogonal to $W = Col(A)$.

In particular $\mathbf{z} = \mathbf{b} - A\widehat{\mathbf{x}}$ must be orthogonal to the columns of $A = [\mathbf{a}_1\ \mathbf{a}_2\ \ldots\ \mathbf{a}_n]$:

$$\mathbf{a}_1^t\mathbf{z} = \mathbf{a}_1^t(\mathbf{b} - A\widehat{\mathbf{x}}) = 0, \quad \mathbf{a}_2^t\mathbf{z} = \mathbf{a}_2^t(\mathbf{b} - A\widehat{\mathbf{x}}) = 0, \ \ldots, \ \mathbf{a}_n^t\mathbf{z} = \mathbf{a}_n^t(\mathbf{b} - A\widehat{\mathbf{x}}) = 0.$$

Because

$$A^t = \begin{bmatrix} \mathbf{a}_1^t \\ \mathbf{a}_2^t \\ \vdots \\ \mathbf{a}_n^t \end{bmatrix}$$

these equations can be summarized as

$$\begin{bmatrix} \mathbf{a}_1^t \\ \mathbf{a}_2^t \\ \vdots \\ \mathbf{a}_n^t \end{bmatrix} (\mathbf{b} - A\widehat{\mathbf{x}}) = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\Updownarrow$$

$$A^t(\mathbf{b} - A\widehat{\mathbf{x}}) = \mathbf{0}$$

$$\Updownarrow$$

$$A^t\mathbf{b} - A^tA\widehat{\mathbf{x}} = \mathbf{0}$$

$$\Updownarrow$$

$$A^tA\widehat{\mathbf{x}} = A^t\mathbf{b}$$

This derivation shows that $\widehat{\mathbf{x}}$ must be an exact solution of the system

$$A^tA\mathbf{x} = A^t\mathbf{b}$$

that is often referred to as the *normal equations* of the least squares problem $A\mathbf{x} = \mathbf{b}$. We summarize this as:

Theorem 13
A least squares solution of the system

$$A\mathbf{x} = \mathbf{b}$$

can always be found, and the set of solutions (one or infinetely many) is found by solving the normal equations

$$A^tA\widehat{\mathbf{x}} = A^t\mathbf{b}.$$

Note that A^tA is always $n \times n$ (i.e. a square matrix) when A is $m \times n$. Uniqueness of the least squares solution $\widehat{\mathbf{x}}$ (one solution only) therfore follows if A^tA is invertible.

Teorem 14
The matrix A^tA is invertible if and only if the columns of A are linearly independent, and the unique least squares solution is then

$$\widehat{\mathbf{x}} = (A^tA)^{-1}A^t\mathbf{b}.$$

Proof: The result follows from the exercises 6.5.19-21.

Definition (projection matrix)
When A^tA is invertible (A has linearly independent columns), the matrix

$$P = A(A^tA)^{-1}A^t$$

is called the *projection matrix* onto the column space $Col(A)$.

Note that the projection $\hat{\mathbf{b}}$ of $\mathbf{b}$ onto $Col(A)$ is $\hat{\mathbf{b}} = A\hat{\mathbf{x}} = A(A^tA)^{-1}A^t\mathbf{b} = P\mathbf{b}$, and that $P^2 = P$ (verify!). The last property means that after projecting $\mathbf{b}$ onto $Col(A)$ there is (as we should expect) no additional effect of projecting $\hat{\mathbf{b}}$ onto $Col(A)$, i.e.

$$\hat{\mathbf{b}} = P\mathbf{b} = P^2\mathbf{b} = PP\mathbf{b} = P\hat{\mathbf{b}}.$$

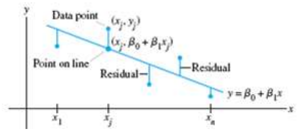
Linear regression (6.6)

In Statistics, Machine Learning and Data analysis least squares problems frequently occur, and the method for solving them is often referred to as *Linear Regression*. In applications the notation is somewhat different from the formulations of (6.5). Rather than $A\mathbf{x} = \mathbf{b}$, the notation

$$\mathbf{y} = X\beta,$$

is preferred. Here X (replacing the coefficient matrix A) is often referred to as the *data matrix*, $\mathbf{y}$ (replacing the right hand side $\mathbf{b}$) is called the *responses* and β (replacing the unknown solution $\mathbf{x}$) is referred to as the *regression coefficients.*

The simplest case of linear regression is the fitting of a line $y = \beta_0 + \beta_1 x$. to a set of points in the (x, y)-plane:



Here the challenge is finding the regression coefficients β_0 and β_1. We want the line to be as close as possible to a set of observed points $\{(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)\}$ in the sense that the sum of squared errors $SSE = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ where $\hat{y}_k = \beta_0 + \beta_1 x_k$.
This problem can be represented as a linear system by the matrix-vector equation

$$X\beta = \mathbf{y}, \text{ where } X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \ \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \text{ and } \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix},$$

and solved by the least squares method by approaching the. normal equations

$$X^tX\beta = X^t\mathbf{y}.$$

If X^tX is invertible the *estimated* least squares regression coefficients (the least squares solution)

$$\hat{\beta} = (X^tX)^{-1}X^t\mathbf{y} \text{ and the fitted values } \hat{\mathbf{y}} = X\hat{\beta}.$$

Example (exercise 6.6.2 in Lay)
We want to fit a straight line to the points

$$(1, 0),\ (2, 1),\ (4, 2),\ (5, 3)$$

by using the least squares method. The corresponding system is

$$X = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 4 \\ 1 & 5 \end{bmatrix} \text{ and } \mathbf{y} = \begin{bmatrix} 0 \\ 1 \\ 2 \\ 3 \end{bmatrix}.$$

From

$$X^tX = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & 5 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 4 \\ 1 & 5 \end{bmatrix} = \begin{bmatrix} 4 & 12 \\ 12 & 46 \end{bmatrix} \text{ and } X^t\mathbf{y} = \begin{bmatrix} 6 \\ 25 \end{bmatrix},$$

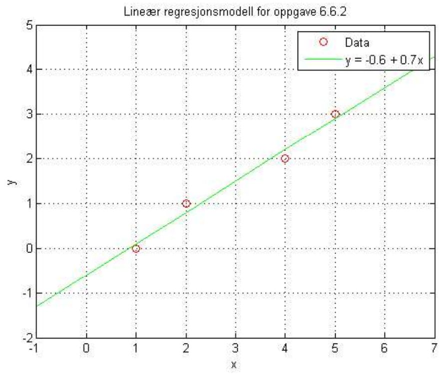
the resulting normal equations are

$$\begin{bmatrix} 4 & 12 \\ 12 & 46 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 6 \\ 25 \end{bmatrix}.$$

The solution in this case can be easily be found by Gaussian elimination (verify the indicated elementary row operations) with the extended matrix

$$[X^tX \ X^t\mathbf{y}] = \begin{bmatrix} 4 & 12 & 6 \\ 12 & 46 & 25 \end{bmatrix} \sim \begin{bmatrix} 4 & 12 & 6 \\ 0 & 10 & 7 \end{bmatrix}$$

$$\sim \begin{bmatrix} 1 & 3 & 1.5 \\ 0 & 1 & 0.7 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -0.6 \\ 0 & 1 & 0.7 \end{bmatrix}, \text{ implying the solution } \widehat{\beta} = \begin{bmatrix} -0.6 \\ 0.7 \end{bmatrix}.$$



An extension to the general case based on p measurements for each response value is straightforward. Estimating the least squares regression coefficients $\beta_0, \beta_1, ..., \beta_p$ of the model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p$$

for the data points

$$(x_{11}, x_{12}, ..., x_{1p}, y_1)$$
$$(x_{21}, x_{22}, ..., x_{2p}, y_2)$$
$$\vdots$$
$$(x_{n1}, x_{n2}, ..., x_{np}, y_n)$$

goes as follows:
The associated linear system is

$$X\beta = \mathbf{y}$$

$$\text{where } X = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix}, \ \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} \text{ and } \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}.$$

The associated Normal equations are

$$X^t X\beta = X^t \mathbf{y},$$

and if $X^t X$ is invertible, the unique least squares solution results in the regression coefficients

$$\widehat{\beta} = (X^t X)^{-1} X^t \mathbf{y}.$$