What can go wrong in (Linear) Least Squares Problems

$$A_{ij} = \frac{X_j(x_i)}{\sigma_i}$$

$$b_i = \frac{y_i}{\sigma_i}$$

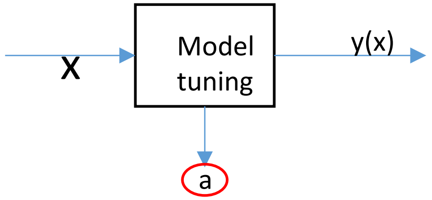
$$\begin{array}{cc}
 & \begin{array}{cccc} X_0(\;) & X_1(\;) & \cdots & X_{M-1}(\;) \end{array} \\
\text{data points}\;\begin{array}{c} x_0 \\ x_1 \\ \vdots \\ x_{N-1} \end{array} &
\begin{pmatrix}
\frac{X_0(x_0)}{\sigma_0} & \frac{X_1(x_0)}{\sigma_0} & \cdots & \frac{X_{M-1}(x_0)}{\sigma_0} \\
\frac{X_0(x_1)}{\sigma_1} & \frac{X_1(x_1)}{\sigma_1} & \cdots & \frac{X_{M-1}(x_1)}{\sigma_1} \\
\vdots & & & \vdots \\
\frac{X_0(x_{N-1})}{\sigma_{N-1}} & \frac{X_1(x_{N-1})}{\sigma_{N-1}} & \cdots & \frac{X_{M-1}(x_{N-1})}{\sigma_{N-1}}
\end{pmatrix}
\end{array}$$

Residuals: $\|\mathbf{A}\mathbf{x} - \mathbf{b}\|$ (the only test we have until now)

Estimated model parameter: $\mathbf{x}$

Input-output relation

Model calibration



The design matrix cause most of the problems due to

- *Bad models*.
  - Will lead to large residuals due to bad data fitting.
  - Easy to observe.

- *Overparametrized models (too many basis functions)*.
  - If we are only interested in data fitting, overparametrized models are not so problematic as residuals will mostly be small as long as the data to be fitted is densely sampled over the region of relevance for the fitting.
  - However, overparametrized models should be avoided when using models to fit to data from other domains and for obtaining knowledge as estimated model parameters will be bad.
  - On this, read about *Occam's razor* and related statements by Isaac Newton, Bertrand Russell and others.

- *Inadequate sampling.*
  - If the experimental points are not sampled good enough, they may not excite all parameters although we have created a good model.
  - Will lead to good residuals, but estimated model parameters can be bad. Can also again lead to bad results outside the domain used to fit the parameters

We need to be able to test all this by numerical methods !!!

Definitions from Linear algebra

Slides 3-12 defines some basic concepts from linear algebra. If you do not already understand these concepts, please search on google and in textbooks for additional material. It is important for you to check that you can solve the exercises. We will introduce

- **Linearly independent** vectors
- **Rank** of a matrix and **singular** matrices
- **Orthogonal/orthonormal** vectors
- **Orthonormal matrices** and **column orthonormal matrices**
- **Vector spaces** and **subspaces**
- **Dimension** of a vector space
- **Basis** and **orthonormal basis** of a vector space. Subspaces **spanned by** the basis.
- **Coordinates** of a vector with respect to an (orthonormal) basis

1. Basic Matrix and vector computations

Definition 1: An $m \times n$ **matrix A** consists of m rows and n columns of numbers. The number in the i'th row and j'te column is denoted $\mathbf{A}_{ij}$ or a_{ij}. An n-dimensional **vector x** consists of n numbers, $\mathbf{x} = (x_1, \ldots, x_n)$. We can multiply an $m \times n$ matrix $\mathbf{A}$ with an $n \times k$ matrix $\mathbf{B}$ to get an $m \times k$ matrix $\mathbf{AB}$ or with an n-dimensional vector $\mathbf{x}$ to obtain an m-dimensional vector $\mathbf{Ax}$. The multiplication formulas are shown to the right.

$$(\mathbf{AB})_{ij} = \sum_{s=1}^{n} a_{is} b_{sj}$$

$$(\mathbf{Ax})_i = \sum_{j=1}^{n} a_{ij} x_j$$

In todays presentation, indices runs from 1 to m,n unlike in NR where they run from 0 to m-1,n-1.

Theorem 1: The following results hold

i) Consider arbitrary $m \times j$, $j \times k$ and $k \times n$ matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$. Then $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$.

ii) Consider arbitrary $m \times n$ matrix $\mathbf{A}$, n-dimensional vectors $\mathbf{x}$ and $\mathbf{y}$ and scalars a, b. We then have $\mathbf{A}(a\mathbf{x} + b\mathbf{y}) = a\mathbf{Ax} + b\mathbf{Ay}$

iii) Consider an arbitrary $m \times n$ matrix $\mathbf{A}$ and $n \times k$ matrix $\mathbf{B}$. Then $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$. As a special case, we have for and an an arbitrary $m \times n$ matrix $\mathbf{A}$ and an arbitrary n-dimensional vector $\mathbf{x}$ that $(\mathbf{Ax})^T = \mathbf{x}^T\mathbf{A}^T$.

Definition 2: Consider an n-dimensional vector $\mathbf{x} = (x_1, \ldots, x_n)$. The **norm** of $\mathbf{x}$, written as $\|\mathbf{x}\|$ is defined as

$$\|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2 + \ldots + x_n^2}$$

Vector Norm

2. Linearly independent vectors

Definition 3: The vectors $\mathbf{x}_1, \ldots, \mathbf{x}_K$ in $\mathbb{R}^n$ are **linearly independent** if the expression

Linearly Independent

$$c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \ldots + c_K\mathbf{x}_K = 0$$

implies that

$$c_1 = c_2 = \ldots = c_K = 0$$

Definition 4: Let $\mathbf{A}$ be an arbitrary $m \times n$ matrix where $m \geq n$. Now perceive the columns of $\mathbf{A}$ as vectors in $\mathbb{R}^m$. The maximum number of linearly independent columns in $\mathbf{A}$ is denoted the **rank** of $\mathbf{A}$. Furthermore, if the n columns of $\mathbf{A}$ all are linearly independent $\mathbf{A}$ is said to have full rank, i.e. the rank is n.

Rank of an mxn matrix

If $\mathbf{A}$ does not have have full rank it is said to be **singular**.

Singular Matrix

Theorem 2: For an arbitrary $m \times n$ matrix $\mathbf{A}$, where $m \geq n$, the matrix $\mathbf{A}^* = \mathbf{A}^T\mathbf{A}$ is singular if and only if $\mathbf{A}$ is singular.

Ex. 2.1 Check if the following vectors are linearly independent

i) $\mathbf{x}_1 = (1, 1, 0)$; $\mathbf{x}_2 = (0, 1, 1)$; $\mathbf{x}_3 = (1, 2, 1)$

ii) $\mathbf{x}_1 = (1, 1, 0, 0)$; $\mathbf{x}_2 = (0, 1, 1, 0)$; $\mathbf{x}_3 = (0, 0, 1, 1)$

iii) $\mathbf{x}_1 = (1, 1, 8)$; $\mathbf{x}_2 = (8, 1, -5)$; $\mathbf{x}_3 = (0, 0, 0)$

iv) $\mathbf{x}_1 = (1, 1, 8, 2, 4)$; $\mathbf{x}_2 = (8, 1, -5, 3, 2)$; $\mathbf{x}_3 = (4, 5, 1, -2, 3)$; $\mathbf{x}_4 = (2, 7, -4, 3, 8)$; $\mathbf{x}_5 = (-4, 9, 2, -21, -8)$

Ex. 2.2 Assume the columns of a 4×3 matrix $\mathbf{A}$ is given by the vectors in Exercise 2.1 ii). Compute $\mathbf{A}^T\mathbf{A}$.

3. Orthogonality and Orthonormality

Definition 5: The vectors $\mathbf{x}_1, \mathbf{x}_2$ in a vectorspace $\mathbb{R}^n$ are said to be **orthogonal** if $\mathbf{x}_1 \cdot \mathbf{x}_2 = 0$. Furthermore, if $\mathbf{x}_1 \cdot \mathbf{x}_1 = \mathbf{x}_2 \cdot \mathbf{x}_2 = 1$ then $\mathbf{x}_1, \mathbf{x}_2$ are said to be **orthonormal**.

Orthogonal vectors

Theorem 3: Orthonormal vectors are always linearly independent.

Definition 6: An $m \times n$ matrix $\mathbf{U}$ $(m \geq n)$ is said to be **column orthonormal** if it consists of n pairwise orthonormal columns (the columns are perceived as vectors in $\mathbb{R}^m$).

Column Orthonormal

If $\mathbf{U}$ is an $m \times n$ column orthonormal matrix then

$$(\mathbf{U}^T\mathbf{U})_{ij} = \sum_{k=1}^{m} \mathbf{U}_{ki}\mathbf{U}_{kj} = \begin{cases} 1 \text{ if } i = j \\ 0 \text{ if } i \neq j \end{cases}$$

$$\mathbf{U}^T\mathbf{U} = \mathbf{I}_n$$

$\mathbf{I}_n$ is the $n \times n$ identitymatrix

Definition 7: An $n \times n$ matrix $\mathbf{V}$ is said to be **orthonormal** if it consists of n pairwise orthonormal columns (the columns are perceived as vectors in $\mathbb{R}^n$).

Orthonormal Matrix

Ex. 3.1 Check if the following vectors are orthogonal, respectively orthonormal

i) $\mathbf{x}_1 = (1, 1, 0, 0)$; $\quad \mathbf{x}_2 = (0, 0, 1, 1)$;

ii) $\mathbf{x}_1 = (1, 1, 0)$; $\quad \mathbf{x}_2 = (0, 1, 1)$;

iii) $\mathbf{x}_1 = (\frac{\sqrt{3}}{2}, \frac{1}{2})$; $\quad \mathbf{x}_2 = (-\frac{1}{2}, \frac{\sqrt{3}}{2})$;

iv) $\mathbf{x}_1 = (1, 3, 8)$; $\quad \mathbf{x}_2 = (0, 0, 0)$;

Ex. 3.2 Consider a 3×3 rotation matrix $\mathbf{R}$. Is this matrix always orthonormal ?

4. Subspace, Dimension and basis

Definition 8: Consider $S \subseteq \mathbb{R}^n$. If for arbitrary elements $\mathbf{x}_1, \mathbf{x}_2 \in S$ and arbitrary scalars $c_1, c_2 \in \mathbb{R}$ applies that $c_1\mathbf{x}_1 + c_2\mathbf{x}_2 \in S$ then S is said to be a **vector space** (also sometimes called *linear space*). The vector space S is also said to be a **subspace** of $\mathbb{R}^n$.

Vector space/subspace

Notice: $\mathbb{R}^n$ is thus a vector space.

Definition 9: Let $S \subseteq \mathbb{R}^n$ be a vector space. The **dimension** of the vector space is the maximum number of linearly independent vectors, that can be found in S. Or stated differently: if $\mathbf{u}_1, \ldots, \mathbf{u}_K \in S$ are linearly independent and meet the condition that every $y \in S$ can be written as $y = c_1\mathbf{u}_1 + \ldots + c_K\mathbf{u}_K$ then S has dimension K. The vectors $\mathbf{u}_1, \ldots, \mathbf{u}_K$ is then said to form a **basis** for S. If $\mathbf{u}_1, \ldots, \mathbf{u}_K$ are orthonormal, $\mathbf{u}_1, \ldots, \mathbf{u}_K$ is said to form an orthonormal **basis** for S. Conversely, we call S the subspace **spanned by** the basis $\mathbf{u}_1, \ldots, \mathbf{u}_K$.

Dimension, Basis, Span

Examples…

Notice: Hence $\mathbb{R}^n$ has dimension n.

5. Coordinates and Norm computation

Assume that the vectors $\mathbf{u}_1, \ldots, \mathbf{u}_n$ form an orthonormal basis for $\mathbb{R}^n$. We know then that an arbitrary $\mathbf{x}$ can be written as

Coordinates

$$\mathbf{x} = \alpha_1 \mathbf{u}_1 + \alpha_2 \mathbf{u}_2 + \ldots \alpha_n \mathbf{u}_n$$

We call $(\alpha_1, \ldots, \alpha_n)$ the **coordinates** of $\mathbf{x}$ with respect to the base $\mathbf{u}_1, \ldots, \mathbf{u}_n$
We now immediately get for all i that

$$\mathbf{x} \cdot \mathbf{u}_i = (\alpha_1 \mathbf{u}_1 + \alpha_2 \mathbf{u}_2 + \ldots \alpha_n \mathbf{u}_n) \cdot \mathbf{u}_i = \alpha_i$$

Theorem 4: Let $\mathbf{u}_1 \ldots, \mathbf{u}_K$ be orthonormal vectors and let

Norm Computation

$$\mathbf{x} = \sum_{k=1}^{K} \alpha_k \mathbf{u}_k$$

Then $\|\mathbf{x}\|^2 \equiv \mathbf{x} \cdot \mathbf{x} = \sum_{k=1}^{K} \alpha_k^2$.

6. Gram-Schmidt method

Assume that the vectors $\mathbf{x}_1, \ldots, \mathbf{x}_k$ are linearly independent and let S be the subspace for which these vectors form a basis. Now we can construct a set of orthonormal vectors $\mathbf{e}_1, \ldots, \mathbf{e}_k$, which form an orthonormal basis for S in this way:

The Gram-Schmidt method:

$\mathbf{e}_1 := \mathbf{x}_1 / \|\mathbf{x}_1\|$

For $i := 2, \ldots, k$ do {

$\quad \mathbf{e}_i := \mathbf{x}_i - \sum_{j=1}^{i-1} (\mathbf{x}_i \cdot \mathbf{e}_j)\mathbf{e}_j$

$\quad \mathbf{e}_i := \mathbf{e}_i / \|\mathbf{e}_i\|$

}

Range, Null Space and the Least Squares solution

Range, Null Space

Definition 10: Let $\mathbf{A}$ be an arbitrary $m \times n$ $(m \geq n)$ matrix. The function $f(\mathbf{x}) = \mathbf{Ax}$ is then said to be a **linear mapping** from $\mathbb{R}^n \rightarrow \mathbb{R}^m$. The **range** of the linear mapping is the set $B(\mathbf{A}) \subseteq \mathbb{R}^m$, that meets the condition that for any $\mathbf{y} \in B(\mathbf{A})$ exists an $\mathbf{x} \in \mathbb{R}^n$ such that $\mathbf{Ax} = \mathbf{y}$. **The null space** of the linear mapping is the set $N(\mathbf{A}) \subseteq \mathbb{R}^n$, that meets the condition that for any $\mathbf{x} \in N(\mathbf{A})$ it applies that $\mathbf{Ax} = 0$.

From Theorem 1 ii), we immediately see that $B(\mathbf{A})$ is a subspace of $\mathbb{R}^m$ and $N(\mathbf{A})$ is a subspace of $\mathbb{R}^n$.

Theorem 1: The following results hold

i) Consider arbitrary $m \times j$, $j \times k$ and $k \times n$ matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$. Then $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$.

ii) Consider arbitrary $m \times n$ matrix $\mathbf{A}$, n-dimensional vectors $\mathbf{x}$ and $\mathbf{y}$ and scalars a, b. We then have $\mathbf{A}(a\mathbf{x} + b\mathbf{y}) = a\mathbf{Ax} + b\mathbf{Ay}$

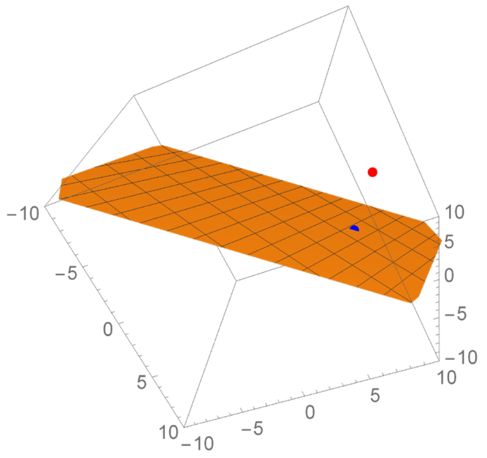
iii) Consider an arbitrary $m \times n$ matrix $\mathbf{A}$ and $n \times k$ matrix $\mathbf{B}$. Then $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$. As a special case, we have for and an an arbitrary $m \times n$ matrix $\mathbf{A}$ and an arbitrary n-dimensional vector $\mathbf{x}$ that $(\mathbf{Ax})^T = \mathbf{x}^T\mathbf{A}^T$.

Range, Null Space and the Least Squares solution

Range, Null Space

Definition 10: Let $\mathbf{A}$ be an arbitrary $m \times n$ $(m \geq n)$ matrix. The function $f(\mathbf{x}) = \mathbf{Ax}$ is then said to be a **linear mapping** from $\mathbb{R}^n \to \mathbb{R}^m$. The **range** of the linear mapping is the set $B(\mathbf{A}) \subseteq \mathbb{R}^m$, that meets the condition that for any $\mathbf{y} \in B(\mathbf{A})$ exists an $\mathbf{x} \in \mathbb{R}^n$ such that $\mathbf{Ax} = \mathbf{y}$. **The null space** of the linear mapping is the set $N(\mathbf{A}) \subseteq \mathbb{R}^n$, that meets the condition that for any $\mathbf{x} \in N(\mathbf{A})$ it applies that $\mathbf{Ax} = 0$.

From Theorem 1 ii), we immediately see that $B(\mathbf{A})$ is a subspace of $\mathbb{R}^m$ and $N(\mathbf{A})$ is a subspace of $\mathbb{R}^n$.



Theorem 5 Let $\mathbf{u}_1 \ldots, \mathbf{u}_K$ be an arbitrary orthonormal basis for $B(\mathbf{A})$. Then the least squares solution $\mathbf{x}$ that minimizes $\|\mathbf{Ax} - \mathbf{b}\|$ satisfies

$$\mathbf{Ax} = \sum_{k=1}^{K} (\mathbf{b} \cdot \mathbf{u}_k)\mathbf{u}_k \equiv \mathbf{b}_{LS}$$

(projection onto B(A)…)

Ex. 6.1 Write a computer program (in any computer language you like) that implements the Gram Schmidt method.

Ex. 6.2 Without proof, we can assume that the vectors $\mathbf{x}_1 = (2, 8, 4, 2, 1)$; $\mathbf{x}_2 = (1, 1, 5, 7, 8)$; $\mathbf{x}_3 = (4, -5, 1, -4, 3)$ are linearly independent vectors in $\mathbb{R}^5$. Use the program from Ex. 6.1 to compute an orthonormal basis for the subspace of $\mathbb{R}^5$ spanned by $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$.

Ex. 6.3 Compute the point in the subspace spanned by $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$ from Ex. 6.2 that is nearest to $(5, 6, 1, 2, 3)$ (HINT: Consider the statement in Theorem 5 in the next section and use your result from Ex. 6.2).

Ex. 6.4 For each of the sets below, determine whether the set is a subspace of $\mathbb{R}^3$. If it is a subspace, determine the dimension and find an orthonormal basis for the subspace.

i) The sphere $x_1^2 + x_2^2 + x_3^2 = 1$

ii) The origin $(x_1, x_2, x_3) = (0, 0, 0)$

iii) The set $(x_1, x_2, x_3) = (0, s, 0); \quad s \in \mathbb{R}$

iv) The set $(x_1, x_2, x_3) = (3s, s, -2s); \quad s \in \mathbb{R}$

v) The set $(x_1, x_2, x_3) = (s_1, s_2, 0); \quad s_1, s_2 \in \mathbb{R}$

vi) The set given by all (x_1, x_2, x_3) satisfying $x_1 + 2x_2 - 2x_3 = 0$

vii) The set given by all (x_1, x_2, x_3) satisfying $x_1 + 2x_2 - 2x_3 = 1$

Ex. 6.5 Consider the matrix $\mathbf{A}$ given as

$$\mathbf{A} = \begin{pmatrix} 2 & 0 & 2 \\ 0 & 4 & 4 \\ 2 & 4 & 4 \\ -1 & -4 & -3 \end{pmatrix}$$

i) Determine the Null Space $N(\mathbf{A})$ and an orthonormal basis for $N(\mathbf{A})$

ii) Determine the range $B(\mathbf{A})$ and an orthonormal basis for $B(\mathbf{A})$

Singular Value Decomposition

Theorem 6 Consider an arbitrary $m \times n$ matrix $\mathbf{A}$. Then we can write $\mathbf{A}$ as $\mathbf{A} = \mathbf{U}\mathbf{W}\mathbf{V}^T$, where $\mathbf{U}$ is an $m \times n$ column othonormal matrix, $\mathbf{V}$ is an $n \times n$ orthonormal matrix and $\mathbf{W}$ is an $n \times n$ diagonal matrix having non-negative diagonal elements $w_1, \ldots w_n$ ordered such that $w_1 \geq w_2 \geq \ldots \geq w_n$. This is said to be a **Singular Value Decomposition** (SVD) of $\mathbf{A}$.

$$\left(\quad \mathbf{A} \quad \right) = \left(\quad \mathbf{U} \quad \right) \cdot \begin{pmatrix} w_1 & & & 0 \\ & w_2 & & \\ & & \ddots & \\ 0 & & & w_n \end{pmatrix} \cdot \left(\quad \mathbf{V}^T \quad \right)$$

$\mathbf{A} = \mathbf{U}\mathbf{W}\mathbf{V}^T$, where $\mathbf{U}$ is an $m \times n$ column othonormal matrix, $\mathbf{V}$ is an $n \times n$ orthonormal matrix and $\mathbf{W}$ is an $n \times n$ diagonal matrix having non-negative diagonal elements $w_1, \ldots w_n$ ordered such that $w_1 \geq w_2 \geq ... \geq w_n$.

Theorem 7 Consider an arbitrary $m \times n$ matrix $\mathbf{A}$ and assume that for $\mathbf{W}$ it applies that $w_1, \ldots, w_K$ are positive and $w_{K+1}, \ldots, w_n$ are equal to zero. Then it applies that

i) $N(\mathbf{A})$ has dimension $n - K$ and the last $n - K$ columns of $\mathbf{V}$ form an orthonormal basis for $N(\mathbf{A})$.

ii) $B(\mathbf{A})$ has dimension K and the first K columns of $\mathbf{U}$ form an orthonormal basis for $B(\mathbf{A})$.

iii) The SVD solution $\mathbf{x} = \mathbf{V}\tilde{\mathbf{W}}^{-1}\mathbf{U}^T\mathbf{b}$, where $[\tilde{\mathbf{W}}^{-1}]_{jj} = 0$ if $\mathbf{W}_{jj} = 0$, otherwise $[\tilde{\mathbf{W}}^{-1}]_{jj} = 1/\mathbf{W}_{jj}$, is the least squares solution to $\mathbf{A}\mathbf{x} = \mathbf{b}$. (Notice that it then follows that if all of the $\mathbf{W}_{jj}$'s are positive, i.e. $\mathbf{A}$ has full rank it applies that $\mathbf{x} = \mathbf{V}\mathbf{W}^{-1}\mathbf{U}^T\mathbf{b}$ is the least squares solution to $\mathbf{A}\mathbf{x} = \mathbf{b}$).

Exercises for this week

- Perform the exercises from this presentation to make sure that you understand and can work with the various concepts
- Perform SVD on the problems ”Pontius” and ”Filip” that you worked on last week. Compute the SVD solution