Slides 2-7 contain the exercises that will be handled with Jens on Wednesday !
Remember to REALLY consider the exercises before lecture 4.
The lecture 4 material is from slide 8 onwards

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

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

implies that

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

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 3×3 matrix $\mathbf{A}$ is given by the vectors in Exercise 2.1 ii). Compute $\mathbf{A}^T\mathbf{A}$.

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**.

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 $\mathbb{R}^m$).

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 $\mathbb{R}^n$).

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 ?

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 $\mathbb{R}^n$.

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 $\mathbf{u}_1, \ldots, \mathbf{u}_K$.

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

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 $x_1 + 2x_2 - 2x_3 = 1$

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 {

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

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

}

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 $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$.

Definition 10: Let $\mathbf{A}$ be an arbitrary $m \times n$ $(m \geq n)$ matrix. The function $f(\mathbf{x}) = \mathbf{A}\mathbf{x}$ 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 $\mathbf{A}\mathbf{x} = \mathbf{y}$. **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 $\mathbf{A}\mathbf{x} = 0$.

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 $B(\mathbf{A})$

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

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

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).

Singular Value Decomposition

Theorem 6 Consider an arbitrary $m \times n$ matrix $\mathbf{A}$. Then we can write $\mathbf{A}$ as $\mathbf{A} = \mathbf{UWV}^T$, 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 $\mathbf{A}$.

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

$\mathbf{A} = \mathbf{U}\mathbf{W}\mathbf{V}^T$, 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 $\mathbf{A}\mathbf{x} = \mathbf{b}$).

$\mathbf{A} = \mathbf{U}\mathbf{W}\mathbf{V}^T$, 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 $w_1 \geq w_2 \geq \ldots \geq w_n$.

IF !!! $w_1, \ldots, w_K$ are positive and $w_{K+1}, \ldots, w_n$ are equal to zero

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 $B(\mathbf{A})$.

Proof: As $\mathbf{V}$ is orthonormal the columns $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are orthonormal and thus form an orthonormal basis for $\mathbb{R}^n$. Then we can write an arbitrary $\mathbf{x} \in \mathbb{R}^n$ as $\mathbf{x} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \ldots + c_n\mathbf{v}_n$. Now we can calculate

$$\begin{aligned} \mathbf{Ax} &= \mathbf{U}\mathbf{W}\mathbf{V}^T(c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \ldots + c_n\mathbf{v}_n) \\ &= c_1\mathbf{U}\mathbf{W}\mathbf{V}^T\mathbf{v}_1 + c_2\mathbf{U}\mathbf{W}\mathbf{V}^T\mathbf{v}_2 + \ldots c_n\mathbf{U}\mathbf{W}\mathbf{V}^T\mathbf{v}_n \end{aligned}$$

$$\mathbf{A}(a\mathbf{x} + b\mathbf{y}) = a\mathbf{Ax} + b\mathbf{Ay}$$

We notice that $\mathbf{V}^T\mathbf{v}_i = (0, \ldots, 0, 1, 0, \ldots, 0)$ where the 1 is at the i'th position. Insertion gives us

$$\begin{aligned} \mathbf{Ax} &= c_1 w_1 \mathbf{u}_1 + c_2 w_2 \mathbf{u}_2 + \ldots c_n w_n \mathbf{u}_n \\ &= c_1 w_1 \mathbf{u}_1 + c_2 w_2 \mathbf{u}_2 + \ldots c_K w_K \mathbf{u}_K \end{aligned}$$

where $\mathbf{u}_1, \ldots, \mathbf{u}_n$ are the columns of $\mathbf{U}$. Since the terms with $c_{K+1}, \ldots, c_n$ disappeared in $\mathbf{Ax}$, the vector

$$c_{K+1}\mathbf{v}_{K+1} + \ldots + c_n\mathbf{v}_n$$

belongs to N($\mathbf{A}$), which proves i). Furthermore, since $\mathbf{Ax}$ becomes a linear combination of the K first vectors in $\mathbf{U}$, we have also shown ii).

$\mathbf{A} = \mathbf{UWV}^T$, 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 $w_1 \geq w_2 \geq \ldots \geq w_n$.

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 $\mathbf{Ax} = \mathbf{b}$.

In order to show iii), i.e. that the SVD solution is the same as the least squares solution we exploit that we know that the first K columns of $\mathbf{U}$ form an orthonormal basis for the range of $\mathbf{A}$. I.e. that the nearest point in $B(\mathbf{A})$ (the least squares mapping) according to Theorem 5 is given by $\mathbf{b}_{LS} = \sum_{j=1}^{K}(\mathbf{u}_j \cdot \mathbf{b})\mathbf{u}_j$. Now let $\mathbf{x}$ be the SVD solution. I.e.

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

$$\begin{aligned}
\mathbf{Ax} &= (\mathbf{UWV}^T)(\mathbf{V}[\tilde{\mathbf{W}}^{-1}]\mathbf{U}^T\mathbf{b}) = \mathbf{U}(\mathbf{W}[\tilde{\mathbf{W}}^{-1}]\mathbf{U}^T\mathbf{b}) \\
&= [\mathbf{u}_1 \ldots \mathbf{u}_K | \mathbf{u}_{K+1} \ldots \mathbf{u}_n] \begin{bmatrix} \mathbf{u}_1 \cdot \mathbf{b} \\ \mathbf{u}_2 \cdot \mathbf{b} \\ \ldots \\ \mathbf{u}_K \cdot \mathbf{b} \\ - - - \\ 0 \\ \ldots \\ 0 \end{bmatrix} = \sum_{j=1}^{K}(\mathbf{u}_j \cdot \mathbf{b})\mathbf{u}_j = \mathbf{b}_{LS}
\end{aligned}$$

3 linear equations in 2 unknowns

In[4]:=
```
A // MatrixForm
```

Out[4]//MatrixForm=

$$\begin{pmatrix} 0.3 & 0.1 \\ 0.2 & 0.4 \\ 0.8 & 0.1 \end{pmatrix}$$

In[6]:=
```
{U // MatrixForm, W // MatrixForm, V // MatrixForm}
```

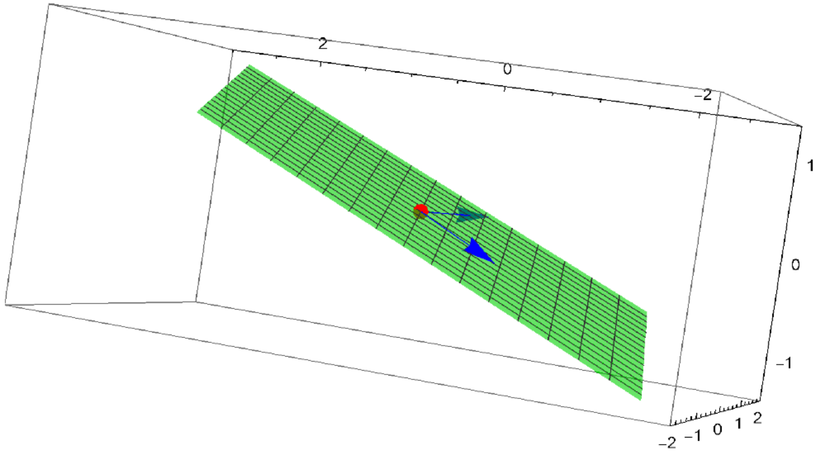
Out[6]=

$$\left\{\begin{pmatrix} -0.347746 & 0.0319972 \\ -0.335342 & 0.929061 \\ -0.875567 & -0.368539 \end{pmatrix}, \begin{pmatrix} 0.908786 & 0. \\ 0. & 0.35229 \end{pmatrix}, \begin{pmatrix} -0.959353 & -0.28221 \\ -0.28221 & 0.959353 \end{pmatrix}\right\}$$

```
b = {1, 0, -2}; xLS = V.Wm1.Transpose[U].b; bLS = A.xLS;
```

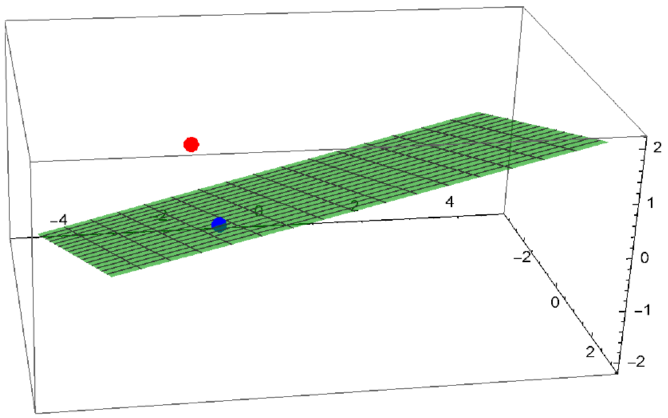
In[7]:=
```
Show[ParametricPlot3D[A.x, {x1, -3, 3}, {x2, -3, 3}, PlotStyle → {Green, Opacity[0.6]}],
 ListPointPlot3D[{{0, 0, 0}}, PlotStyle → {Red, PointSize[0.02]}],
 vektorplot[Table[{0, 0, 0}, n], Transpose[U], Blue]]
```

Out[7]=



In[10]:=
```
Show[ParametricPlot3D[A.x, {x1, -5, 5}, {x2, -5, 5}, PlotStyle → {Green, Opacity[0.6]}],
 ListPointPlot3D[{b}, PlotStyle → {Red, PointSize[0.025]}],
 ListPointPlot3D[{bLS}, PlotStyle → {Blue, PointSize[0.025]}]]
```

Out[10]=



3 singular linear equations in 2 unknowns

In[12]:= `A // MatrixForm`

Out[12]//MatrixForm=

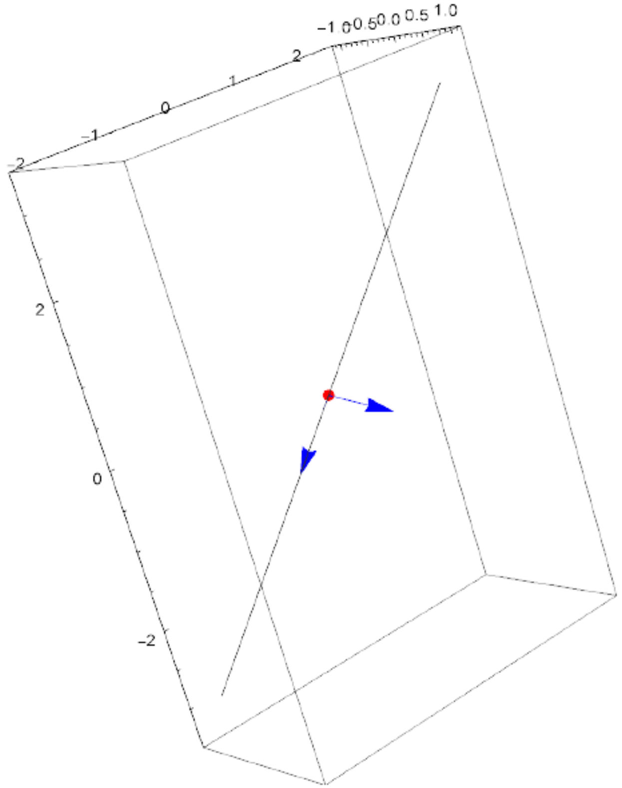
$$\begin{pmatrix} 0.15 & 0.2 \\ 0.3 & 0.4 \\ 0.45 & 0.6 \end{pmatrix}$$

In[14]:= `{U // MatrixForm, W // MatrixForm, V // MatrixForm}`

Out[14]= $$\left\{ \begin{pmatrix} -0.267261 & -0.00509709 \\ -0.534522 & 0.832823 \\ -0.801784 & -0.553516 \end{pmatrix}, \begin{pmatrix} 0.935414 & 0. \\ 0. & 0. \end{pmatrix}, \begin{pmatrix} -0.6 & -0.8 \\ -0.8 & 0.6 \end{pmatrix} \right\}$$

In[15]:=
```
Show[ParametricPlot3D[A.x, {x1, -3, 3}, {x2, -3, 3}],
 ListPointPlot3D[{{0, 0, 0}}, PlotStyle → {Red, PointSize[0.02]}],
 vektorplot[Table[{0, 0, 0}, n], Transpose[U], Blue]]
```

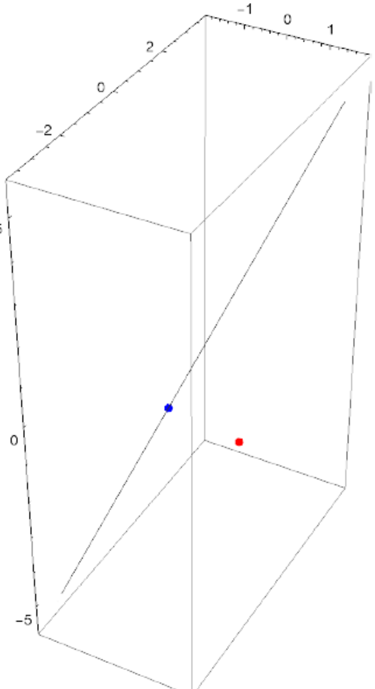
Out[15]=



In[17]:=
```
b = {1, 0, -2}; xLS = V.Wm1.Transpose[U].b; bLS = A.xLS;
```

In[18]:=
```
Show[ParametricPlot3D[A.x, {x1, -5, 5}, {x2, -5, 5}],
 ListPointPlot3D[{b}, PlotStyle → {Red, PointSize[0.025]}],
 ListPointPlot3D[{bLS}, PlotStyle → {Blue, PointSize[0.025]}]]
```

Out[18]=



3 near singular linear equations in 2 unknowns

In[20]:= `A // MatrixForm`

Out[20]//MatrixForm=

$$\begin{pmatrix} 0.15 & 0.2 \\ 0.3 & 0.399999 \\ 0.45 & 0.600001 \end{pmatrix}$$

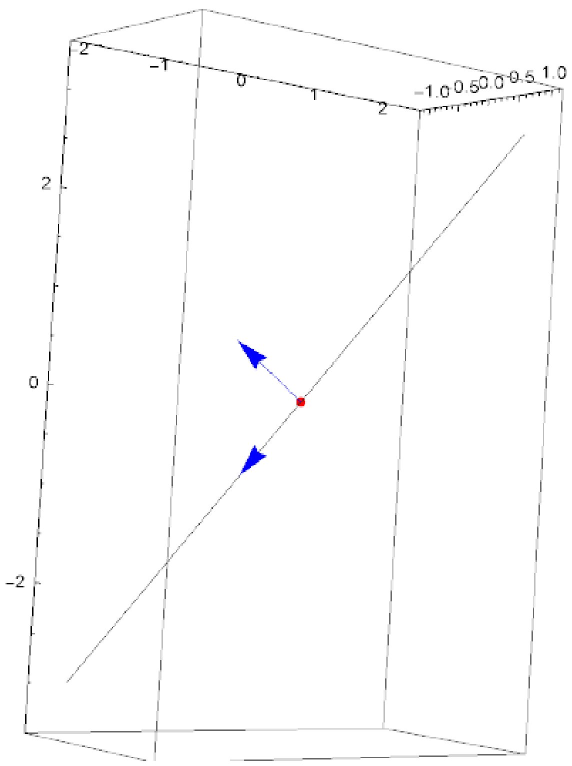
In[22]:= `{U // MatrixForm, W // MatrixForm, V // MatrixForm}`

Out[22]=

$$\left\{\begin{pmatrix} -0.267261 & -0.0514348 \\ -0.534522 & -0.822952 \\ -0.801784 & 0.565778 \end{pmatrix}, \begin{pmatrix} 0.935415 & 0. \\ 0. & 8.33238\times 10^{-7} \end{pmatrix}, \begin{pmatrix} -0.6 & -0.8 \\ -0.8 & 0.6 \end{pmatrix}\right\}$$

In[23]:=
```
Show[ParametricPlot3D[A.x, {x1, -3, 3}, {x2, -3, 3}],
 ListPointPlot3D[{{0, 0, 0}}, PlotStyle → {Red, PointSize[0.02]}],
 vektorplot[Table[{0, 0, 0}, n], Transpose[U], Blue]]
```

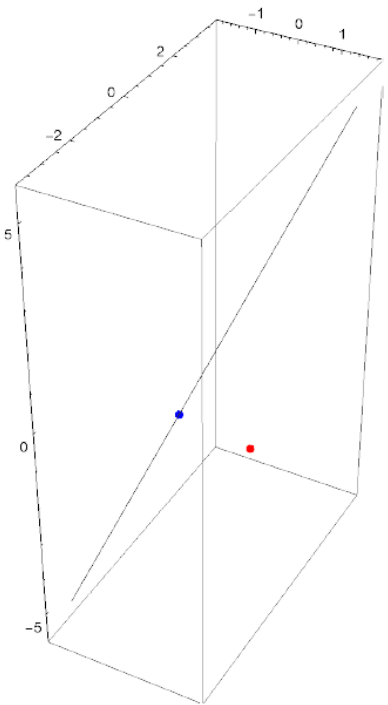
Out[23]=



In[25]:=
```
b = {1, 0, -2}; xLS = V.Wm1.Transpose[U].b; bLS = A.xLS;
```

In[26]:=
```
Show[ParametricPlot3D[A.x, {x1, -5, 5}, {x2, -5, 5}],
 ListPointPlot3D[{b}, PlotStyle → {Red, PointSize[0.025]}],
 ListPointPlot3D[{bLS}, PlotStyle → {Blue, PointSize[0.025]}]]
```

Out[26]=



Error analysis for systems of linear equations. Residual errors

The residual error should be computed as a relative error, namely

$$\epsilon_{residual} = \frac{\|\mathbf{Ax} - \mathbf{b}\|}{\|\mathbf{b}\|}$$

If $m = n$, the residual error should be very close to zero unless the matrix is near singular. For $m > n$, the linear equations are typically from some sort of fitting problem such as a Least Squares Problem. The value $\epsilon_{residual}$ indicates how good the fitting model is. It is easy to see that a random fitting model would produce $\epsilon_{residual} \simeq \sqrt{\frac{m-n}{m}}$. If your result is not much better than that, you should consider the quality of your model.

Error analysis for systems of linear equations. Errors on solution

Even though solving a set of linear equations seem very deterministic, it is relevant to consider the error $\delta\mathbf{x}$ on the result $\mathbf{x}$. In typical applications, there are two very different sources to this error.

The first source is the error on the right hand side $\delta\mathbf{b}$. The error $\delta\mathbf{b}$ is typically is some kind of measurement error and therefore may be quite large.

The second source is the error on the matrix $\delta\mathbf{A}$ which is typically due to the real number precision. Hence, $\|\delta\mathbf{A}\|$ is mostly of the order $\|\delta\mathbf{A}\| \simeq 10^{-18}$.

We consider only the contribution from the error on the right hand side.

Error on solution. Impact from error on right hand side

Consider the design matrix and right hand side:

$$
\begin{aligned}
\mathbf{A}_{ij} &:= \mathbf{A}_{ij}/\sigma_i \quad i = 1, \ldots m, \quad j = 1, \ldots n \\
\mathbf{b}_i &:= \mathbf{b}_i/\sigma_i \quad i = 1, \ldots m
\end{aligned}
$$

where σ_i is the inaccuracy on $\mathbf{b}_i$.

The error estimate $\delta\mathbf{x}$ is then purely given by the SVD matrices using Eq.15.4.19

$$
[\delta\mathbf{x}]_j \simeq \sqrt{\sum_{i=1}^{n} \left(\frac{V_{ji}}{w_i}\right)^2} \quad j = 1, \ldots, n \tag{4}
$$

If we compute $\mathbf{A}^T\mathbf{A}$ using SVD, we get

$$
\mathbf{A}^T\mathbf{A} = (\mathbf{U}\mathbf{W}\mathbf{V}^T)^T(\mathbf{U}\mathbf{W}\mathbf{V}^T) = (\mathbf{V}\mathbf{W}\mathbf{U}^T)(\mathbf{U}\mathbf{W}\mathbf{V}^T) = \mathbf{V}\mathbf{W}^2\mathbf{V}^T
$$

which is itself the SVD of $\mathbf{A}^T\mathbf{A}$.

If w_n is e.g. 10^-8, we get with the Normal Equations w_n^2=10^-16 !!!!

Updating the model based on the SVD analysis

Residual error: $\epsilon_{residual} = \frac{||\mathbf{Ax} - \mathbf{b}||}{||\mathbf{b}||}$

Should be SIGNIFICANTLY smaller than the random fitting $\sqrt{\frac{m-n}{m}}$

If it is not, *reconsider your model completely*.

If the residual error is significantly smaller than random fitting, but the residual error is too big for your application, search for a *model modification/extension*.

Small last singular value w_n (compared to w_1):

Consider the last column in V which give you the linear combination of your model functions that causes the problem. Try to resolve the reason and if that is impossible, try to remove the model function corresponding to the numerically largest value in the column (*model reduction*).

Setting the threshold in SVD (finalized model)

Residual error:

$$\epsilon_{residual} = \frac{||\mathbf{A}\mathbf{x} - \mathbf{b}||}{||\mathbf{b}||}$$

Will become bigger for bigger thresholds as near-singular linear combinations of the basis functions are taken out of the optimization

Error on solution:

$$[\delta \mathbf{x}]_j \simeq \sqrt{\sum_{i=1}^{n} \left(\frac{V_{ji}}{w_i} \right)^2} \quad j = 1, \ldots, n$$

Will become smaller for bigger thresholds. Notice that the contributions for w_i's that are "removed" by the threshold should be omitted from the sum and the model is "numerically reduced". But the error on the reduced model solution due to right hand side uncertainties is completely controlled by this formula

Data fitting: Residual error is important, hence go with a small threshold (e.g. default from NR).
Model estimation: Solution error is important, hence go with a bigger threshold (of course still small) and so that both residuals and the error on the solution is acceptable.
Remember that if the threshold is active, the error is on the reduced model parameters and not on the original model.
IMPORTANT: There may be two reasons for a near-singularity, overparametrized models such as in Filip or inadequate data sampling where not all parameters are excited. If the near-singularity is due to bad data sampling, it is VERY wrong to do numerical reduction by increasing the threshold. Always search for the reason for the near-singularity first.

Exercises

- Slide 2-7
- Estimate the errors on the solution parameters for Pontius and Filip

- This finalizes systems of linear equations in the sense that we will essentially not present any additional methods. But we will discuss efficient versions of LU/Cholesky for so-called band-matrices when they become relevant.
- First mandatory exercise is out. Deadline 10/3