Lecture 15

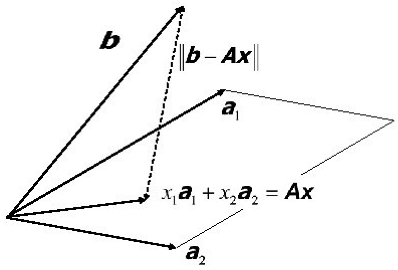
Least Squares Problems

Learning Outcomes

- **Comprehend Least Squares Problems:**
  - You will be able to understand and formulate least squares problems, particularly focusing on the minimization of the function $g(x) = \frac{1}{2}\|Ax - b\|_2^2$. You will be able to apply this knowledge to real-world scenarios such as linear regression to find the best-fitting line through a set of points.
- **Apply Singular Value Decomposition (SVD):**
  - Gain proficiency in applying Singular Value Decomposition (SVD) to matrices, acquiring the ability to extract both algebraic and geometrical information from matrices. You will also be able to leverage SVD for practical applications like image compression.
- **Solve Least Squares Problems Efficiently:**
  - Develop the skills to solve least squares problems using various methods, including the normal equation, QR factorization, and SVD. You will understand the implications of condition numbers and will be able to choose the most efficient and practical method for solving least squares problems in different contexts.

Least Squares Problems:

A SLE $Ax = b$ can be solved if and only if $b \in \text{range}(A)$ (which is also know as the column space of A). Although there is no exact solution when $b \notin \text{range}(A)$, an estimate x that makes Ax as close as possible to b can be found. This is called the least square solution.



Column and row spaces A column space (or range) of matrix A is the space that is spanned by As columns. Similarly, a row space is the space spanned by As rows. In particular, if $A \in \mathbb{C}^{m\times n}$, with column vectors $v_1, v_2, \dots, v_n$, then a linear combination of these vectors is any vector of the form,

$$c_1v_1 + c_2v_2 + \cdots + c_nv_n$$

where $c_1, c_2, \dots, c_n$ are scalars. The set of all possible linear combinations of $v_1, v_2, \dots, v_n$ is called the range(A). This means that the range(A) is the span of the vectors $v_1, v_2, \dots, v_n$.
Any linear combination of the column vectors of a matrix A can be written as the product of A with a column vector,

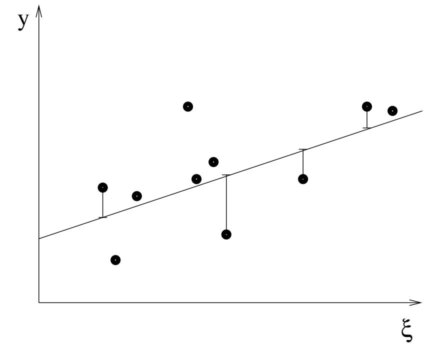
$$A\begin{pmatrix} c_1 \\ c_2 \\ \dots \\ c_n \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}\begin{pmatrix} c_1 \\ c_2 \\ \dots \\ c_n \end{pmatrix} = c_1\underbrace{\begin{pmatrix} a_{11} \\ a_{21} \\ \dots \\ a_{m1} \end{pmatrix}}_{v_1} + c_2\underbrace{\begin{pmatrix} a_{12} \\ a_{22} \\ \dots \\ a_{m2} \end{pmatrix}}_{v_2} + \cdots + c_n\underbrace{\begin{pmatrix} a_{1n} \\ a_{2n} \\ \dots \\ a_{mn} \end{pmatrix}}_{v_n}$$

Definition 15.1. *Given a matrix $A \in \mathbb{R}^{m\times n}$ and a vector $b \in \mathbb{R}^m$, the least squares problem* LSQ *consists of minimising the function*

$$g : \mathbb{R}^n \to \mathbb{R}, \quad g(x) = \frac{1}{2}\|Ax - b\|_2^2.$$

Example: Recall the linear regression problem. Given points $(\xi_i, y_i)_{i=1}^m$ find a linear function $\xi \to x_1 + x_2\xi$ such that

$$g(x) = \frac{1}{2}\sum_{i=1}^{m}\big(x_1 + x_2\xi_i - y_i\big)^2 \quad \text{is minimal.}$$



In this case,

$$A = \begin{pmatrix} 1 & \xi_1 \\ \vdots & \vdots \\ 1 & \xi_m \end{pmatrix}, \quad b = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix}.$$

Example: Solve,

$$\begin{aligned} x_1 + 2x_2 &= 3 \\ x_1 + 3x_2 &= 5 \\ x_1 + 4x_2 &= 5 \\ x_1 + 5x_2 &= 6 \end{aligned} \quad \Longrightarrow \underbrace{\begin{pmatrix} 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \end{pmatrix}}_{A} \underbrace{\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}}_{x} = \underbrace{\begin{pmatrix} 3 \\ 2 \\ 6 \\ 5 \end{pmatrix}}_{b}$$

which can be written as,

$$x_1 \underbrace{\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}}_{a_1} + x_2 \underbrace{\begin{pmatrix} 2 \\ 3 \\ 4 \\ 5 \end{pmatrix}}_{a_2} = \underbrace{\begin{pmatrix} 3 \\ 2 \\ 6 \\ 5 \end{pmatrix}}_{y}$$

Theorem 15.1. ([7.1], *$x \in \mathbb{R}^n$ solves the least squares problem if and only if $Ax - b \perp \mathrm{range}(A)$, which is the case if and only if the normal equation*

$$A^T A x = A^T b \tag{7.1}$$

is satisfied.

Proof. If x minimises g then for all $y \in \mathbb{R}^n$

$$\begin{aligned} 0 &= \frac{d}{d\varepsilon} g(x + \varepsilon y)\Big|_{\varepsilon=0} \\ &= \frac{d}{d\varepsilon}\Big(\frac{1}{2}\langle Ax + \varepsilon Ay - b, Ax + \varepsilon Ay - b\rangle\Big)\Big|_{\varepsilon=0} \\ &= \frac{d}{d\varepsilon}\Big(\frac{1}{2}\langle Ax - b, Ax - b\rangle + \frac{1}{2}\varepsilon\big(\langle Ay, Ax - b\rangle + \langle Ax - b, Ay\rangle\big) + \frac{1}{2}\varepsilon^2\langle Ay, Ay\rangle\Big)\Big|_{\varepsilon=0} \\ &= \frac{d}{d\varepsilon}\Big(\frac{1}{2}\|Ax - b\|_2 + \varepsilon\langle Ay, Ax - b\rangle + \frac{1}{2}\varepsilon^2\|Ay\|_2\Big)\Big|_{\varepsilon=0} \\ &= \langle Ay, Ax - b\rangle \end{aligned}$$

which means that $Ax - b \perp Ay$, for all $y \in \mathbb{R}^n$. The other way round, if $Ax - b \perp \operatorname{range}(A)$ then $\langle Ax - b, Ay - Ax \rangle = 0$ for all $y \in \mathbb{R}^n$, hence with Pythagoras

$$2g(y) = \|Ay - b\|_2^2 = \|Ay - Ax\|_2^2 + \|Ax - b\|_2^2 \geq \|Ax - b\|_2^2 = 2g(x).$$

For the second assertion we use that $Ax - b \perp \operatorname{range}(A) \Leftrightarrow Ax - b \perp a_i$ where the a_i, $i = 1, \ldots, m$, are the column vectors of A. But this is equivalent to

$$\big(\langle a_i, Ax \rangle\big)_{i=1}^m = \big(\langle a_i, b \rangle\big)_{i=1}^m \quad \Leftrightarrow \quad (\boxed{7.1})$$

□

Example: In the linear regression problem the normal equation is a 2×2 system where

$$A^T A = \begin{pmatrix} m & \sum_{i=1}^m \xi_i \\ \sum_{i=1}^m \xi_i & \sum_{i=1}^m \xi_i^2 \end{pmatrix}, \quad A^T b = \begin{pmatrix} \sum_{i=1}^m y_i \\ \sum_{i=1}^m \xi_i y_i \end{pmatrix}.$$

Example:

$$A = \begin{pmatrix} 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \end{pmatrix}, \qquad b = \begin{pmatrix} 3 \\ 2 \\ 6 \\ 5 \end{pmatrix}$$

$$A^T = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 2 & 3 & 4 & 5 \end{pmatrix} \implies A^T A = \begin{pmatrix} 4 & 14 \\ 14 & 54 \end{pmatrix}, \quad A^T b = \begin{pmatrix} 16 \\ 61 \end{pmatrix}$$

Thus,

$$A^T A x = A^T b$$
$$\begin{pmatrix} 4 & 14 \\ 14 & 54 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 16 \\ 61 \end{pmatrix}$$

Which can be solved using LU factorisation as follows,

$$A^T A = LU$$
$$\begin{pmatrix} 4 & 14 \\ 14 & 54 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 3.5 & 1 \end{pmatrix} \begin{pmatrix} 4 & 14 \\ 0 & 5 \end{pmatrix}$$

Then solve $Ly = A^T b$ using forward substitution,

$$\begin{pmatrix} 1 & 0 \\ 3.5 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 16 \\ 61 \end{pmatrix} \implies y_1 = 16, \; y_2 = 5$$

Finally solve $Ux = y$,

$$\begin{pmatrix} 4 & 14 \\ 0 & 5 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 16 \\ 5 \end{pmatrix} \implies x_1 = 0.5, \; x_2 = 1$$

which yields,

$$Ax = \begin{pmatrix} 2.5 \\ 3.5 \\ 4.5 \\ 5.5 \end{pmatrix}, \quad b - Ax = b_\perp = \begin{pmatrix} -0.5 \\ 1.5 \\ -1.5 \\ 0.5 \end{pmatrix}$$

It is certainly possible to use (7.1) to solve LSQ, for example one could use Cholesky since A^TA is positive definite provided that A has full rank. But, as we will see later on, we have

$$\kappa_2(A^TA) = (\kappa_2(A))^2$$

for the condition number, and even the condition number $\kappa_2(A)$ (which is not yet defined for $m \neq n$) can be big in practical applications. There are better approaches, based on the QR factorisation (later) or on the singular value decomposition that we consider next.

Singular Value Decomposition (SVD) [2.3]

SVD gives insight into the least square problems. The least square is,

$$\underbrace{\min}_{x} \frac{1}{2}\|Ax - b\|_2^2$$

We have seen that the solution in this case is given by,

$$\begin{aligned} A^TAx &= A^Tb \\ \implies x &= (A^TA)^{-1}A^Tb \end{aligned} \tag{15.1}$$

Which means that the estimated measured values are,

$$\begin{aligned} \hat{b} &= Ax \\ &= A(A^TA)^{-1}A^Tb \\ &= P_{A_\parallel}b \end{aligned}$$

which is the projection of b into the column space of A. The error is then,

$$\begin{aligned} \text{error} &= b - \hat{b} \\ &= b - A(A^TA)^{-1}A^Tb \\ &= P_{A_\perp}b \end{aligned}$$

where $P_{A_\perp}$ is the orthogonal projection of b onto the column space of A.

SVD can be used to factorise the matrix A. Any arbitrary matrix $A \in \mathbb{R}^{m\times n}$ factors into,

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

where $U \in \mathbb{R}^{m\times m}$ and $V \in \mathbb{R}^{n\times n}$ are orthogonal matrices and $\Sigma \in \mathbb{R}^{m\times n}$ is a diagonal matrix with entries $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p \geq 0$ where $p = \min(m,n)$. Those values σ_i are called singular values. Physically this means a rotation, a stretch and a second rotation. Note that in this decomposition we have two different matrices. But these matrices are in fact orthogonal matrices. Also this decomposition can be done for rectangular matrices, however eigenvalues really worked for square matrices. However, now we have input matrix and an output matrix. In those spaces m and n can have different dimension. So by allowing two separate basis we can factorise rectangular matrices using orthogonal factors and a diagonal matrix consists of singular values instead of eigenvalues. U is called the left singular vector while V is called the right singular vector.

Denoting by u_i, $i = 1 = 1, \dots, m$, the column vectors of U and by v_i, $i = 1, \dots, n$, the column vectors of V the SVD says that $Av_i = \sigma_i u_i$, $i = 1, \dots, p$.

SVD: Here we outline the procedure for obtaining the U and V matrices, for an arbitrary matrix $\mathbf{X} \in \mathbb{R}^{m \times n}$. Let us look at $\mathbf{X}^T\mathbf{X}$,

$$\begin{aligned} \mathbf{X}^T\mathbf{X} &= V\Sigma^T(U^TU)\Sigma V^T && (15.2) \\ &= V(\Sigma^T\Sigma)V^T. && (15.3) \end{aligned}$$

This is the usual eigenvector decomposition. But here the matrix is $\mathbf{X}^T\mathbf{X}$. Note that although $\mathbf{X}$ is rectangular and completely general, its diagonalisation is not possible. However the transition to $\mathbf{X}^T\mathbf{X}$ provided a positive semidefinite and symmetric matrix $(\mathbf{X}^T\mathbf{X})$ which has orthogonal eigenvectors and positive eigenvalues. Also note that $(\Sigma^T\Sigma)$ are the square of the singular values, i.e λ for $\mathbf{X}^T\mathbf{X}$ are the σ^2 for $\mathbf{X}$. This gives us V and Σ. However, U disappears here because $U^TU = I$ where I is the identity. How can we get hold of U?

Let us look at $\mathbf{X}\mathbf{X}^T$,

$$\begin{aligned} \mathbf{X}\mathbf{X}^T &= U\Sigma(V^TV)\Sigma U^T && (15.4) \\ &= U(\Sigma^T\Sigma)U^T. && (15.5) \end{aligned}$$

This clearly shows that U is the eigenvector matrix for $\mathbf{X}\mathbf{X}^T$. So $\mathbf{X}^T\mathbf{X}$ and $\mathbf{X}\mathbf{X}^T$ have the same eigenvalues ((AB) has the same eigenvalues as (BA)). However $\mathbf{X}^T\mathbf{X}$ has eigenvectors V while $\mathbf{X}\mathbf{X}^T$ has eigenvectors U and those are the U and the V in the singular value decomposition.

Example:

$$X = \begin{bmatrix} 3 & 1 & 1 \\ 1 & 3 & -1 \end{bmatrix}$$

$$XX^T = \frac{1}{\sqrt{5}}\begin{bmatrix} 11 & 5 \\ 5 & 11 \end{bmatrix} \qquad X^TX = \begin{bmatrix} 10 & 6 & 2 \\ 6 & 10 & -2 \\ 2 & -2 & 2 \end{bmatrix}$$

Eigenvalues and eigenvectors of XX^T,

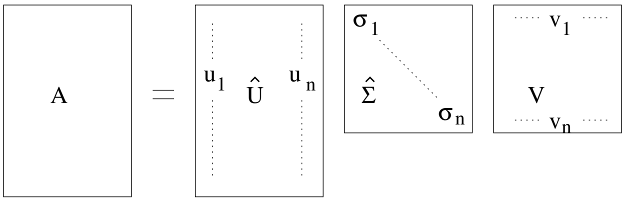
$$u_1 = \begin{bmatrix} -0.7071 \\ 0.7071 \end{bmatrix} \quad u_2 = \begin{bmatrix} 0.7071 \\ 0.7071 \end{bmatrix} \quad \lambda_1 = 6,\ \lambda_2 = 16$$

Eigenvalues and eigenvectors of X^TX,

$$v_1 = \begin{bmatrix} 0.4082 \\ -0.4082 \\ -0.8165 \end{bmatrix} \quad v_2 = \begin{bmatrix} -0.5774 \\ 0.5774 \\ -0.5774 \end{bmatrix} \quad v_3 = \begin{bmatrix} 0.7071 \\ 0.7071 \\ 0 \end{bmatrix} \quad \lambda_1 = 0,\ \lambda_2 = 6,\ \lambda_3 = 16.$$

$$\begin{aligned} X &= U\Sigma V \\ &= \begin{bmatrix} 0.7071 & -0.7071 \\ 0.7071 & 0.7071 \end{bmatrix}\begin{bmatrix} 4 & 0 & 0 \\ 0 & 2.45 & 0 \end{bmatrix}\begin{bmatrix} 0.7071 & 0.7071 & 0 \\ -0.5774 & 0.5774 & -0.5774 \\ 0.4082 & -0.4082 & -0.8165 \end{bmatrix} \end{aligned}$$

In the case $m \geq n(=p)$ which is of most interest to us the reduced singular value decomposition is defined by dropping the last $m-n$ columns of U and the last $m-n$ rows of Σ. Defining $\hat{U} := (u_1, \ldots, u_n) \in \mathbb{R}^{m\times n}$ and $\hat{\Sigma} := \mathrm{diag}(\sigma_1, \ldots, \sigma_n) \in \mathbb{R}^{n\times n}$



the matrix A can be expressed as

$$A = \hat{U}\hat{\Sigma}V^T. \tag{15.6}$$

Least square solution can then be obtained by substituting (15.6) into (15.1) which yields,

$$\begin{aligned}
x &= (A^TA)^{-1}A^Tb \\
&= (V\hat{\Sigma}^T\hat{U}^T\hat{U}\hat{\Sigma}V^T)^{-1}V\hat{\Sigma}^T\hat{U}^Tb \\
&= (V\hat{\Sigma}^2V^T)^{-1}V\hat{\Sigma}^T\hat{U}^Tb \\
&= V\hat{\Sigma}^{-2}V^TV\hat{\Sigma}^T\hat{U}^Tb \\
&= V\hat{\Sigma}^{-1}\hat{U}^Tb \\
&= \sum_{i=1}^{p}\frac{1}{\sigma_i}v_i(\hat{u}_i^T)b
\end{aligned}$$

Note that $(A^TA)^{-1}A^T = A^\dagger$ is referred to as the pseudo inverse of A.

Note that multiplying $A^\dagger$ by A on the right yields ,

$$\{(A^TA)^{-1}A^T\}A = (A^TA)^{-1}(A^TA) = I$$

Similarly, For the SVD, we have,

$$V\hat{\Sigma}^{-1}\hat{U}^T\hat{U}\hat{\Sigma}V^T = I$$

Exercise: Using the SVD for the matrix A, evaluate the estimated measured values, $\hat{b}$.

Exercise: Using the SVD for the matrix A, evaluate the error in the estimated measured values.

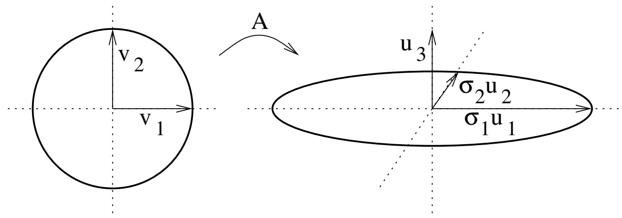
The SVD is a very powerful decomposition for analytical purposes as it provides much insight into the properties of the linear map associated with the matrix. The rank is the number of non-vanishing singular values, i.e., the biggest integer $r \leq p$ such that $\sigma_r > 0$ but $\sigma_{r+1} = 0$ (if $r+1 \leq p$). Moreover, the image of A is spanned by the first r left singular vectors,

$$\mathrm{range}(A) = \mathrm{span}\{u_1, \ldots, u_r\}$$

and the kernel is

$$\mathrm{kernel}(A) = \mathrm{span}\{v_{r+1}, \ldots, v_p\}.$$

In addition to this algebraic information the SVD also reveals some geometrical information. Recalling the identities $Av_i = u_i\sigma_i$, $i = 1, \ldots, p$, we see that the unit sphere in $\mathbb{R}^p$ containing the vectors v_i is mapped to an ellipsoid which has semi-axes in the directions of the u_i that have the extensions σ_i.



Example: Images can be compressed using the SVD. This will be demonstrated through an example on Python.

Lecture 16

More on the Singular Value Decomposition

Learning Outcomes

- **Understand Singular Value Decomposition (SVD) Uniqueness and Existence:**
  - Gain insight into the existence and uniqueness of the singular values in the Singular Value Decomposition of any matrix $A \in \mathbb{R}^{m\times n}$, and understand the implications of these properties in practical applications.
- **Apply SVD to Different Matrix Types:**
  - Acquire the ability to apply SVD to various types of matrices, including symmetric matrices and regular matrices, and understand the relationships between the eigenvalues and eigenvectors of different matrix forms and their SVD components.
- **Compute SVD Efficiently:**
  - Learn the computational aspects of SVD, including preprocessing steps and computational costs, to efficiently compute the SVD of a given matrix, especially focusing on the relation between the eigenspaces of a specific matrix form H and the SVD of A.

Singular Value Decomposition

Theorem 16.1. [2.12] *Every matrix $A \in \mathbb{R}^{m\times n}$ has a SVD, and the singular values are uniquely determined.*

Proof. We first show the existence by induction on $p = \min(m, n)$.
For $p = 1$ we may choose $u = 1$, and if $a_{11} \neq 0$ we may set $v = a_{11}/|a_{11}|$ and $\sigma_1 = |a_{11}|$ whilst in the case $a_{11} = 0$ we choose $v = 1$ and $\sigma_1 = 0$.
Let now $p > 1$ and assume without loss of generality that $A \neq 0$ (since the SVD is trivial otherwise with U and V being the identity and $\Sigma = 0$). The map $x \mapsto \|Ax\|_2$ on $\mathbb{R}^n$ is continuous. When restricted to the compact unit sphere $S^{n-1} = \{\|x\|_2 = 1 \,|\, x \in \mathbb{R}^n\}$ it attains a maximum which we denote by v_1. Further, we define

$$\sigma_1 := \|Av_1\|_2 = \max_{x\in S^{n-1}} \|Ax\|_2 = \|A\|_2.$$

Since $A \neq 0$ we have that $\sigma_1 > 0$ and we can define

$$u_1 := \frac{1}{\sigma_1} A v_1.$$

Let us now extend v_1 to an orthonormal basis (ONB) $\{v_1, \ldots, v_n\}$ of $\mathbb{R}^n$, and u_1 to an ONB $\{u_1, \ldots, u_m\}$ of $\mathbb{R}^m$. The matrices $U_1 := (u_1, \ldots, u_m) \in \mathbb{R}^{m \times m}$ and $V_1 := (v_1, \ldots, v_n) \in \mathbb{R}^{n \times n}$ then are orthogonal. Let

$$C := U_1^T A V_1 =: \begin{pmatrix} \sigma_1 & w^T \\ 0 & B \end{pmatrix}$$

with $w \in \mathbb{R}^{n-1}$ and $B \in \mathbb{R}^{m-1,n-1}$ where the zeros in the first column below the diagonal arise from the fact that $Av_1 = \sigma_1 u_1$ is orthogonal to the columns $u_2, \ldots, u_m$ of U_1. Then

$$\|C\|_2 = \max_{\|x\|_2=1} \|U_1^T A V_1 x\|_2 = \max_{\|V_1 x\|_2=1} \|A V_1 x\|_2 = \|A\|_2 = \sigma_1.$$

Since

$$\|C\|_2 \Big\| \begin{pmatrix} \sigma_1 \\ w \end{pmatrix} \Big\|_2 \geq \Big\| C \begin{pmatrix} \sigma_1 \\ w \end{pmatrix} \Big\|_2 = \Big\| \begin{pmatrix} \sigma_1^2 + \|w\|_2^2 \\ Bw \end{pmatrix} \Big\|_2 \geq \sigma_1^2 + \|w\|_2^2 \geq \sqrt{\sigma_1^2 + \|w\|_2^2} \Big\| \begin{pmatrix} \sigma_1 \\ w \end{pmatrix} \Big\|_2$$

we conclude that

$$\sigma_1 = \|C\|_2 \geq \sqrt{\sigma_1^2 + \|w\|_2^2} \quad \Rightarrow \quad w = 0.$$

By the induction hypothesis there is a singular value decomposition $B = U_2 \Sigma_2 V_2^T$ of the $(m-1) \times (n-1)$ matrix B. Writing $\Sigma_2 = \mathrm{diag}(\sigma_2, \ldots, \sigma_p)$ we observe that

$$\sigma_1 = \max_{\|x\|_2=1} \|Cx\|_2 \geq \max_{\|y\|_2=1,\, y \in \mathbb{R}^{n-1}} \Big\| C \begin{pmatrix} 0 \\ y \end{pmatrix} \Big\|_2 = \|By\|_2 = \sigma_2.$$

Therefore

$$A = U_1 C V_1^T = \underbrace{U_1 \begin{pmatrix} 1 & 0 \\ 0 & U_2 \end{pmatrix}}_{=:U} \underbrace{\begin{pmatrix} \sigma_1 & 0 \\ 0 & \Sigma_2 \end{pmatrix}}_{=:\Sigma} \underbrace{\begin{pmatrix} 1 & 0 \\ 0 & V_2^T \end{pmatrix} V_1^T}_{=:V^T}$$

is a SVD of A.
Coming to the second claim, we remark that for any SVD $A = U \Sigma V^T$

$$\|A\|_2 = \max_{\|x\|_2=1} \|U \Sigma V^T x\|_2 = \max_{\|V^T x\|_2=1} \|\Sigma V^T x\|_2 = \max_{\|y\|_2=1} = \|\Sigma y\|_2 = \sigma_1.$$

Using an induction argument again one can show the uniqueness of the singular values. □

Corollary 16.1. $\|A\|_2 = \sigma_1$.

Examples:

1. A symmetric matrix $A \in \mathbb{R}^{n \times n}$ can be diagonalised in the form $A = Q \Lambda Q^T$ with $Q \in \mathbb{R}^{n \times n}$ orthogonal and $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n) \in \mathbb{R}^{n \times n}$. Without loss of generality we may assume that $|\lambda_1| \geq \cdots \geq |\lambda_n|$. Otherwise perform a similarity transformation of Λ with an appropriate permutation matrix which then is absorbed into Q. Denote the columns of Q by $q_1, \ldots, q_n$. A SVD of A is obtained by setting $U := (u_1, \ldots, u_n)$ where $u_i = \mathrm{sign}(\lambda_i) q_i$, $\Sigma = \mathrm{diag}(|\lambda_1|, \ldots, |\lambda_n|)$, and $V := Q$.

2. Let $A \in \mathbb{R}^{m\times n}$ with $m \geq n$ (so that $p = \min(m, n) = n$). The matrix $A^T A \in \mathbb{R}^{n\times n}$ has eigenvalues σ_i^2 with corresponding eigenvectors v_i, $i = 1, \dots, p$. The matrix $AA^T \in \mathbb{R}^{m\times m}$ has eigenvectors $\{u_1, \dots, u_m\}$ with corresponding eigenvalues $\{\sigma_1^2, \dots, \sigma_p^2, 0, \dots, 0\}$ (with $m - n$ zeros).
   To see this, let us exemplary consider the latter case. We have that
   $$AA^T = U\Sigma V^T V\Sigma^T U^T \quad \Rightarrow \quad AA^T U = U\Sigma\Sigma^T,$$
   but $\Sigma\Sigma^T = \mathrm{diag}(\sigma_1^2, \dots, \sigma_p^2, 0, \dots, 0)$ with exactly $m - n$ zeros.

3. Assume now that $A \in \mathbb{R}^{n\times n}$ is regular. Then the matrix
   $$H := \begin{pmatrix} 0 & A^T \\ A & 0 \end{pmatrix}$$
   has $2n$ eigenvalues $\{\sigma_1, -\sigma_1, \sigma_2, -\sigma_2, \dots, \sigma_n, -\sigma_n\}$
   with corresponding eigenvectors $\{\binom{v_1}{u_1}, \binom{v_1}{-u_1}, \binom{v_2}{u_2}, \binom{v_2}{-u_2}, \dots, \binom{v_n}{u_n}, \binom{v_n}{-u_n}\}$.
   To show this, assume that $Hx = \lambda x$ for some $\lambda \in \mathbb{R}$ and some $x \in \mathbb{R}^{2n}\backslash\{0\}$. Writing $x = \binom{y}{z}$ with $y, z \in \mathbb{R}^n$ this means that
   $$\begin{matrix} A^T z = \lambda y, \\ Ay = \lambda z, \end{matrix} \quad \Rightarrow \quad \begin{matrix} AA^T z = \lambda A y = \lambda^2 z, \\ A^T A y = \lambda A^T z = \lambda^2 y. \end{matrix}$$
   From the previous example we know that the eigenvalues of $A^T A$ and AA^T are $\{\sigma_1^2, \dots, \sigma_n^2\}$ where $\sigma_n > 0$ by the regularity of A. Hence $\lambda = \pm\sigma_i$ for some $i \in \{1, \dots, n\}$. That $\binom{v_i}{\pm u_i}$ is a corresponding eigenvector is easy to show.

Remark 16.2. *The above results hold true analogously for complex matrices if the transposed matrices are replaced by the adjoint matrices.*