Advanced Machine Learning

Singular Value Decomposition (SVD)

$$\begin{pmatrix} 0 & X \\ X^{\mathsf{T}} & 0 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = s \begin{pmatrix} u \\ v \end{pmatrix}$$

Singular Valued Decomposition, SVD, general linear maps

Outline

1. **Singular Value Decomposition**

2. General Linear Mappings

3. Linear Regression Revisited

$$\begin{pmatrix} 0 & X \\ X^\mathsf{T} & 0 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = s \begin{pmatrix} u \\ v \end{pmatrix}$$

Singular Valued Decomposition

- Consider an arbitrary $n \times m$ matrix $\mathbf{X}$, and construct the $(n+m) \times (n+m)$ symmetric matrix, $\mathbf{B}$,

$$\begin{pmatrix} \mathbf{0} & \mathbf{X} \\ \mathbf{X}^{\mathsf{T}} & \mathbf{0} \end{pmatrix}$$

Singular Valued Decomposition

- Consider an arbitrary $n \times m$ matrix $\mathbf{X}$, and construct the $(n+m) \times (n+m)$ symmetric matrix, $\mathbf{B}$,

$$
\begin{pmatrix} \mathbf{0} & \mathbf{X} \\ \mathbf{X}^{\mathsf{T}} & \mathbf{0} \end{pmatrix}
\begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{pmatrix}
= s
\begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{pmatrix}
$$

Singular Valued Decomposition

- Consider an arbitrary $n \times m$ matrix $\mathbf{X}$, and construct the $(n+m) \times (n+m)$ symmetric matrix, $\mathbf{B}$,

$$\begin{pmatrix} 0 & \mathbf{X} \\ \mathbf{X}^\mathsf{T} & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{pmatrix} = s \begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{pmatrix}$$

$\begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{pmatrix}$ is an eigenvector of $\mathbf{B}$ with eigenvalue s

Singular Valued Decomposition

- Consider an arbitrary $n \times m$ matrix $\mathbf{X}$, and construct the $(n+m) \times (n+m)$ symmetric matrix, $\mathbf{B}$,

$$\begin{pmatrix} \mathbf{0} & \mathbf{X} \\ \mathbf{X}^\mathsf{T} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{pmatrix} = s \begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{pmatrix}$$

$\begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{pmatrix}$ is an eigenvector of $\mathbf{B}$ with eigenvalue s

- We observe that

$$\mathbf{X}\boldsymbol{v} = s\boldsymbol{u} \qquad\qquad \mathbf{X}^\mathsf{T}\boldsymbol{u} = s\boldsymbol{v}$$

Singular Valued Decomposition

- Consider an arbitrary $n \times m$ matrix $\mathbf{X}$, and construct the $(n+m) \times (n+m)$ symmetric matrix, $\mathbf{B}$,

$$\begin{pmatrix} \mathbf{0} & \mathbf{X} \\ \mathbf{X}^\mathsf{T} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{pmatrix} = s \begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{pmatrix}$$

$\begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{pmatrix}$ is an eigenvector of $\mathbf{B}$ with eigenvalue s

- We observe that

$$\mathbf{X}\boldsymbol{v} = s\boldsymbol{u} \qquad\qquad \mathbf{X}^\mathsf{T}\boldsymbol{u} = s\boldsymbol{v}$$

$$\mathbf{X}^\mathsf{T}\mathbf{X}\boldsymbol{v} = s\mathbf{X}^\mathsf{T}\boldsymbol{u}$$

Singular Valued Decomposition

- Consider an arbitrary $n \times m$ matrix $\mathbf{X}$, and construct the $(n+m) \times (n+m)$ symmetric matrix, $\mathbf{B}$,

$$\begin{pmatrix} 0 & \mathbf{X} \\ \mathbf{X}^\mathsf{T} & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{pmatrix} = s \begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{pmatrix}$$

$\begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{pmatrix}$ is an eigenvector of $\mathbf{B}$ with eigenvalue s

- We observe that

$$\mathbf{X}\boldsymbol{v} = s\boldsymbol{u} \qquad\qquad \mathbf{X}^\mathsf{T}\boldsymbol{u} = s\boldsymbol{v}$$

$$\mathbf{X}^\mathsf{T}\mathbf{X}\boldsymbol{v} = s\mathbf{X}^\mathsf{T}\boldsymbol{u} = s^2\boldsymbol{v}$$

Singular Valued Decomposition

- Consider an arbitrary $n \times m$ matrix $\mathbf{X}$, and construct the $(n+m) \times (n+m)$ symmetric matrix, $\mathbf{B}$,

$$\begin{pmatrix} 0 & \mathbf{X} \\ \mathbf{X}^{\mathsf{T}} & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{pmatrix} = s \begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{pmatrix}$$

$\begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{pmatrix}$ is an eigenvector of $\mathbf{B}$ with eigenvalue s

- We observe that

$$\mathbf{X}\boldsymbol{v} = s\boldsymbol{u} \qquad\qquad \mathbf{X}^{\mathsf{T}}\boldsymbol{u} = s\boldsymbol{v}$$

$$\mathbf{X}^{\mathsf{T}}\mathbf{X}\boldsymbol{v} = s\mathbf{X}^{\mathsf{T}}\boldsymbol{u} = s^2\boldsymbol{v} \qquad\qquad \mathbf{X}\mathbf{X}^{\mathsf{T}}\boldsymbol{u} = s\mathbf{X}\boldsymbol{v}$$

Singular Valued Decomposition

- Consider an arbitrary $n \times m$ matrix $\mathbf{X}$, and construct the $(n+m) \times (n+m)$ symmetric matrix, $\mathbf{B}$,

$$\begin{pmatrix} \mathbf{0} & \mathbf{X} \\ \mathbf{X}^\mathsf{T} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{pmatrix} = s \begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{pmatrix}$$

$\begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{pmatrix}$ is an eigenvector of $\mathbf{B}$ with eigenvalue s

- We observe that

$$\mathbf{X}\boldsymbol{v} = s\boldsymbol{u} \qquad\qquad \mathbf{X}^\mathsf{T}\boldsymbol{u} = s\boldsymbol{v}$$

$$\mathbf{X}^\mathsf{T}\mathbf{X}\boldsymbol{v} = s\mathbf{X}^\mathsf{T}\boldsymbol{u} = s^2\boldsymbol{v} \qquad\qquad \mathbf{X}\mathbf{X}^\mathsf{T}\boldsymbol{u} = s\mathbf{X}\boldsymbol{v} = s^2\boldsymbol{u}$$

Eigenvectors

- Note that as $\mathbf{X}\boldsymbol{v} = s\boldsymbol{u}$ and $\mathbf{X}^{\mathsf{T}}\boldsymbol{u} = s\boldsymbol{v}$ then

$$\mathbf{X}(-\boldsymbol{v}) = (-s)\boldsymbol{u} \qquad \mathbf{X}^{\mathsf{T}}\boldsymbol{u} = (-s)(-\boldsymbol{v})$$

  if $\begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{pmatrix}$ is an eigenvector of $\mathbf{B}$ with eigenvalue s then so is $\begin{pmatrix} \boldsymbol{u} \\ -\boldsymbol{v} \end{pmatrix}$ with eigenvalue $-s$

- If $n < m$ then $\mathbf{X}^{\mathsf{T}}\mathbf{X}$ is not full rank so some eigenvalues are zero

- As a consequence $m - n$ vectors exist such that $\boldsymbol{X}\boldsymbol{v} = 0$

- The eigenvalues and eigenvectors are

$$n \times \left(s_i, \begin{pmatrix} \boldsymbol{u}_i \\ \boldsymbol{v}_i \end{pmatrix} \right) \quad n \times \left(-s_i, \begin{pmatrix} \boldsymbol{u}_i \\ -\boldsymbol{v}_i \end{pmatrix} \right) \quad m - n \times \left(0, \begin{pmatrix} 0 \\ \boldsymbol{v}_k \end{pmatrix} \right)$$

Eigenvectors

- Note that as $\mathbf{X}\boldsymbol{v} = s\boldsymbol{u}$ and $\mathbf{X}^\mathsf{T}\boldsymbol{u} = s\boldsymbol{v}$ then

$$\mathbf{X}(-\boldsymbol{v}) = (-s)\boldsymbol{u} \qquad\qquad \mathbf{X}^\mathsf{T}\boldsymbol{u} = (-s)(-\boldsymbol{v})$$

if $\begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{pmatrix}$ is an eigenvector of $\mathbf{B}$ with eigenvalue s then so is $\begin{pmatrix} \boldsymbol{u} \\ -\boldsymbol{v} \end{pmatrix}$ with eigenvalue $-s$

- If $n < m$ then $\mathbf{X}^\mathsf{T}\mathbf{X}$ is not full rank so some eigenvalues are zero

- As a consequence $m - n$ vectors exist such that $\boldsymbol{X}\boldsymbol{v} = 0$

- The eigenvalues and eigenvectors are

$$n \times \left(s_i, \begin{pmatrix} \boldsymbol{u}_i \\ \boldsymbol{v}_i \end{pmatrix} \right) \quad n \times \left(-s_i, \begin{pmatrix} \boldsymbol{u}_i \\ -\boldsymbol{v}_i \end{pmatrix} \right) \quad m - n \times \left(0, \begin{pmatrix} 0 \\ \boldsymbol{v}_k \end{pmatrix} \right)$$

Eigenvectors

- Note that as $\mathbf{X}\boldsymbol{v} = s\boldsymbol{u}$ and $\mathbf{X}^{\mathsf{T}}\boldsymbol{u} = s\boldsymbol{v}$ then

$$\mathbf{X}(-\boldsymbol{v}) = (-s)\boldsymbol{u} \qquad \mathbf{X}^{\mathsf{T}}\boldsymbol{u} = (-s)(-\boldsymbol{v})$$

if $\begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{pmatrix}$ is an eigenvector of $\mathbf{B}$ with eigenvalue s then so is $\begin{pmatrix} \boldsymbol{u} \\ -\boldsymbol{v} \end{pmatrix}$ with eigenvalue $-s$

- If $n < m$ then $\mathbf{X}^{\mathsf{T}}\mathbf{X}$ is not full rank so some eigenvalues are zero

- As a consequence $m - n$ vectors exist such that $\boldsymbol{X}\boldsymbol{v} = 0$

- The eigenvalues and eigenvectors are

$$n \times \left(s_i, \begin{pmatrix} \boldsymbol{u}_i \\ \boldsymbol{v}_i \end{pmatrix} \right) \quad n \times \left(-s_i, \begin{pmatrix} \boldsymbol{u}_i \\ -\boldsymbol{v}_i \end{pmatrix} \right) \quad m - n \times \left(0, \begin{pmatrix} 0 \\ \boldsymbol{v}_k \end{pmatrix} \right)$$

Eigenvectors

- Note that as $\mathbf{X}\boldsymbol{v} = s\boldsymbol{u}$ and $\mathbf{X}^\mathsf{T}\boldsymbol{u} = s\boldsymbol{v}$ then

$$\mathbf{X}(-\boldsymbol{v}) = (-s)\boldsymbol{u} \qquad\qquad \mathbf{X}^\mathsf{T}\boldsymbol{u} = (-s)(-\boldsymbol{v})$$

  if $\begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{pmatrix}$ is an eigenvector of $\mathbf{B}$ with eigenvalue s then so is $\begin{pmatrix} \boldsymbol{u} \\ -\boldsymbol{v} \end{pmatrix}$ with eigenvalue $-s$

- If $n < m$ then $\mathbf{X}^\mathsf{T}\mathbf{X}$ is not full rank so some eigenvalues are zero

- As a consequence $m - n$ vectors exist such that $\boldsymbol{X}\boldsymbol{v} = 0$

- The eigenvalues and eigenvectors are

$$n \times \left(s_i, \begin{pmatrix} \boldsymbol{u}_i \\ \boldsymbol{v}_i \end{pmatrix} \right) \quad n \times \left(-s_i, \begin{pmatrix} \boldsymbol{u}_i \\ -\boldsymbol{v}_i \end{pmatrix} \right) \quad m - n \times \left(0, \begin{pmatrix} 0 \\ \boldsymbol{v}_k \end{pmatrix} \right)$$

Matrix Decomposition

- **Stacking the eigenvectors into a matrix**

$$\begin{pmatrix} 0 & X \\ X^\mathsf{T} & 0 \end{pmatrix} \begin{pmatrix} u & u & 0 \\ v & -v & V_0 \end{pmatrix} = \begin{pmatrix} u & u & 0 \\ v & -v & V_0 \end{pmatrix} \begin{pmatrix} S & 0 & 0 \\ 0 & -S & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

- Since the vectors $\begin{pmatrix} \boldsymbol{u}_i \\ \boldsymbol{v}_i \end{pmatrix}$ are eigenvectors of a symmetric matrix they from an orthogonal matrix if they are normalised.

- Multiply on the right by the transpose of the orthogonal matrix

$$\begin{pmatrix} 0 & X \\ X^\mathsf{T} & 0 \end{pmatrix} = \begin{pmatrix} u & u & 0 \\ v & -v & V_0 \end{pmatrix} \begin{pmatrix} S & 0 & 0 \\ 0 & -S & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} u^\mathsf{T} & v^\mathsf{T} \\ u^\mathsf{T} & -v^\mathsf{T} \\ 0 & V_0^\mathsf{T} \end{pmatrix}$$

Matrix Decomposition

- Stacking the eigenvectors into a matrix

$$\begin{pmatrix} 0 & X \\ X^\mathsf{T} & 0 \end{pmatrix} \begin{pmatrix} U & U & 0 \\ V & -V & V_0 \end{pmatrix} = \begin{pmatrix} U & U & 0 \\ V & -V & V_0 \end{pmatrix} \begin{pmatrix} S & 0 & 0 \\ 0 & -S & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

- Since the vectors $\begin{pmatrix} u_i \\ v_i \end{pmatrix}$ are eigenvectors of a symmetric matrix they from an orthogonal matrix if they are normalised.

- Multiply on the right by the transpose of the orthogonal matrix

$$\begin{pmatrix} 0 & X \\ X^\mathsf{T} & 0 \end{pmatrix} = \begin{pmatrix} U & U & 0 \\ V & -V & V_0 \end{pmatrix} \begin{pmatrix} S & 0 & 0 \\ 0 & -S & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} U^\mathsf{T} & V^\mathsf{T} \\ U^\mathsf{T} & -V^\mathsf{T} \\ 0 & V_0^\mathsf{T} \end{pmatrix}$$

Matrix Decomposition

- Stacking the eigenvectors into a matrix

$$\begin{pmatrix} 0 & X \\ X^\mathsf{T} & 0 \end{pmatrix} \begin{pmatrix} u & u & 0 \\ v & -v & v_0 \end{pmatrix} = \begin{pmatrix} u & u & 0 \\ v & -v & v_0 \end{pmatrix} \begin{pmatrix} S & 0 & 0 \\ 0 & -S & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

- Since the vectors $\begin{pmatrix} u_i \\ v_i \end{pmatrix}$ are eigenvectors of a symmetric matrix they from an orthogonal matrix if they are normalised.

- Multiply on the right by the transpose of the orthogonal matrix

$$\begin{pmatrix} 0 & X \\ X^\mathsf{T} & 0 \end{pmatrix} = \begin{pmatrix} u & u & 0 \\ v & -v & v_0 \end{pmatrix} \begin{pmatrix} S & 0 & 0 \\ 0 & -S & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} u^\mathsf{T} & v^\mathsf{T} \\ u^\mathsf{T} & -v^\mathsf{T} \\ 0 & v_0^\mathsf{T} \end{pmatrix}$$

Normalisation Subtlety

$$\begin{pmatrix} 0 & X \\ X^\mathsf{T} & 0 \end{pmatrix} = \begin{pmatrix} u & u & 0 \\ v & -v & V_0 \end{pmatrix} \begin{pmatrix} S & 0 & 0 \\ 0 & -S & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} u^\mathsf{T} & v^\mathsf{T} \\ u^\mathsf{T} & -v^\mathsf{T} \\ 0 & v_0^\mathsf{T} \end{pmatrix}$$

- Multiplying out we have
$$X = 2uSv^\mathsf{T} \qquad\qquad X^\mathsf{T} = 2vSu^\mathsf{T}$$

- Now the vectors $\boldsymbol{u}_i$ and $\boldsymbol{v}_i$ form an orthogonal set as it satisfy
$$X^\mathsf{T}X\boldsymbol{v} = s^2\boldsymbol{v} \qquad\qquad XX^\mathsf{T}\boldsymbol{u} = s^2\boldsymbol{u}$$

- But they are not normalised (since $\begin{pmatrix}\boldsymbol{u}_i \\ \boldsymbol{v}_i\end{pmatrix}$ is normalised). If we define $\tilde{U} = \sqrt{2}U$ and $\tilde{V} = \sqrt{2}V$ we find
$$X = \tilde{U}S\tilde{V}^\mathsf{T} \qquad\qquad X^\mathsf{T} = \tilde{V}S\tilde{U}^\mathsf{T}$$

Normalisation Subtlety

$$\begin{pmatrix} 0 & X \\ X^\mathsf{T} & 0 \end{pmatrix} = \begin{pmatrix} U & U & 0 \\ V & -V & V_0 \end{pmatrix} \begin{pmatrix} S & 0 & 0 \\ 0 & -S & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} U^\mathsf{T} & V^\mathsf{T} \\ U^\mathsf{T} & -V^\mathsf{T} \\ 0 & V_0^\mathsf{T} \end{pmatrix}$$

- Multiplying out we have
$$X = 2USV^\mathsf{T} \qquad\qquad X^\mathsf{T} = 2VSU^\mathsf{T}$$

- Now the vectors u_i and v_i form an orthogonal set as it satisfy
$$X^\mathsf{T}Xv = s^2 v \qquad\qquad XX^\mathsf{T}u = s^2 u$$

- But they are not normalised (since $\begin{pmatrix} u_i \\ v_i \end{pmatrix}$ is normalised). If we define $\tilde{U} = \sqrt{2}U$ and $\tilde{V} = \sqrt{2}V$ we find
$$X = \tilde{U}S\tilde{V}^\mathsf{T} \qquad\qquad X^\mathsf{T} = \tilde{V}S\tilde{U}^\mathsf{T}$$

Normalisation Subtlety

- Multiplying out we have

$$\mathsf{X} = 2\mathsf{u}\mathsf{S}\mathsf{v}^\mathsf{T} \qquad\qquad \mathsf{X}^\mathsf{T} = 2\mathsf{v}\mathsf{S}\mathsf{u}^\mathsf{T}$$

- Now the vectors $\boldsymbol{u}_i$ and $\boldsymbol{v}_i$ form an orthogonal set as it satisfy

$$\mathsf{X}^\mathsf{T}\mathsf{X}\boldsymbol{v} = s^2\boldsymbol{v} \qquad\qquad \mathsf{X}\mathsf{X}^\mathsf{T}\boldsymbol{u} = s^2\boldsymbol{u}$$

- But they are not normalised (since $\begin{pmatrix}\boldsymbol{u}_i \\ \boldsymbol{v}_i\end{pmatrix}$ is normalised). If we define $\tilde{\mathsf{U}} = \sqrt{2}\mathsf{U}$ and $\tilde{\mathsf{V}} = \sqrt{2}\mathsf{V}$ we find

$$\mathsf{X} = \tilde{\mathsf{U}}\mathsf{S}\tilde{\mathsf{V}}^\mathsf{T} \qquad\qquad \mathsf{X}^\mathsf{T} = \tilde{\mathsf{V}}\mathsf{S}\tilde{\mathsf{U}}^\mathsf{T}$$

SVD

- Any matrix, $\mathbf{X}$, can be written as $\mathbf{X} = \mathbf{U} \mathbf{S} \mathbf{V}^\mathsf{T}$

    - $\star$ $\mathbf{U}$, $\mathbf{V}$ are orthogonal matrices
    - $\star$ $\mathbf{S} = \mathrm{diag}(s_1, s_2, \ldots, s_n)$

- s_i can always be chosen to be positive and are known as **singular values**

- Singular value decomposition applies to both square and non-square matrices—they describe general linear mappings

SVD

- Any matrix, $\mathbf{X}$, can be written as $\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^{\mathsf{T}}$

  - $\star$ $\mathbf{U}$, $\mathbf{V}$ are orthogonal matrices
  - $\star$ $\mathbf{S} = \operatorname{diag}(s_1, s_2, \ldots, s_n)$

- s_i can always be chosen to be positive and are known as **singular values**

- Singular value decomposition applies to both square and non-square matrices—they describe general linear mappings

SVD

- Any matrix, $\mathbf{X}$, can be written as $\mathbf{X} = \mathbf{U} \mathbf{S} \mathbf{V}^{\mathsf{T}}$

  ⋆ $\mathbf{U}$, $\mathbf{V}$ are orthogonal matrices
  ⋆ $\mathbf{S} = \mathrm{diag}(s_1, s_2, \ldots, s_n)$

- s_i can always be chosen to be positive and are known as **singular values**

- Singular value decomposition applies to both square and non-square matrices—they describe general linear mappings

Finding SVD

- Most libraries will compute the SVD for you

- They can do this by choosing the smaller of two matrices $\mathbf{X}\mathbf{X}^\mathsf{T}$ and $\mathbf{X}^\mathsf{T}\mathbf{X}$ and then compute the eigenvalues

- The singular values are the square root of the eigenvalues (notice that $\mathbf{X}\mathbf{X}^\mathsf{T}$ and $\mathbf{X}^\mathsf{T}\mathbf{X}$ are both positive semi-definite so the eigenvalues will be non-negative)

- It can compute the $\mathbf{U}$ matrix or $\mathbf{V}$ matrix by multiplying through by $\mathbf{X}$ or $\mathbf{X}^\mathsf{T}$ ($\mathbf{U} = \mathbf{X}\mathbf{V}\mathbf{S}^{-1}$ and $\mathbf{V} = \mathbf{X}^\mathsf{T}\mathbf{U}\mathbf{S}^{-1}$)

- In practice to perform PCA most people subtract the mean from their data and then perform SVD

Finding SVD

- Most libraries will compute the SVD for you

- They can do this by choosing the smaller of two matrices $\mathbf{X}\mathbf{X}^\mathsf{T}$ and $\mathbf{X}^\mathsf{T}\mathbf{X}$ and then compute the eigenvalues

- The singular values are the square root of the eigenvalues (notice that $\mathbf{X}\mathbf{X}^\mathsf{T}$ and $\mathbf{X}^\mathsf{T}\mathbf{X}$ are both positive semi-definite so the eigenvalues will be non-negative)

- It can compute the $\mathbf{U}$ matrix or $\mathbf{V}$ matrix by multiplying through by $\mathbf{X}$ or $\mathbf{X}^\mathsf{T}$ ($\mathbf{U} = \mathbf{X}\mathbf{V}\mathbf{S}^{-1}$ and $\mathbf{V} = \mathbf{X}^\mathsf{T}\mathbf{U}\mathbf{S}^{-1}$)

- In practice to perform PCA most people subtract the mean from their data and then perform SVD
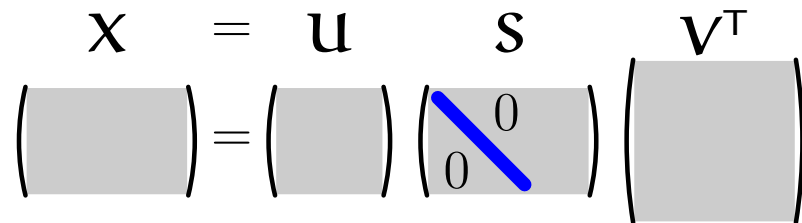
Finding SVD

- Most libraries will compute the SVD for you

- They can do this by choosing the smaller of two matrices $\mathbf{X}\mathbf{X}^\mathsf{T}$ and $\mathbf{X}^\mathsf{T}\mathbf{X}$ and then compute the eigenvalues

- The singular values are the square root of the eigenvalues (notice that $\mathbf{X}\mathbf{X}^\mathsf{T}$ and $\mathbf{X}^\mathsf{T}\mathbf{X}$ are both positive semi-definite so the eigenvalues will be non-negative)

- It can compute the $\mathbf{U}$ matrix or $\mathbf{V}$ matrix by multiplying through by $\mathbf{X}$ or $\mathbf{X}^\mathsf{T}$ ($\mathbf{U} = \mathbf{X}\mathbf{V}\mathbf{S}^{-1}$ and $\mathbf{V} = \mathbf{X}^\mathsf{T}\mathbf{U}\mathbf{S}^{-1}$)

- In practice to perform PCA most people subtract the mean from their data and then perform SVD

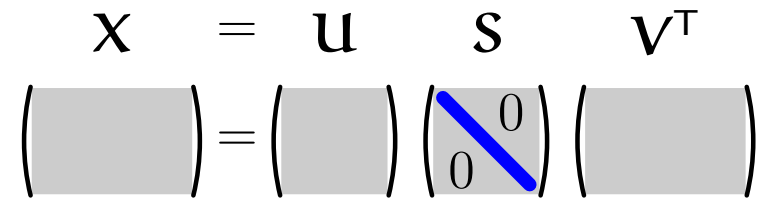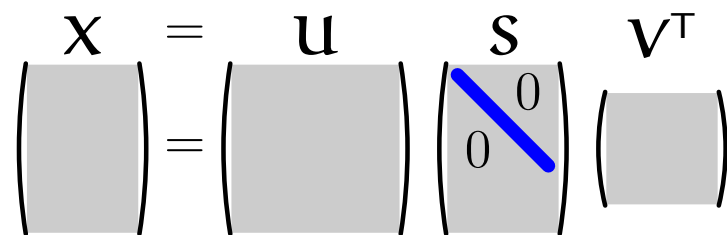# Finding SVD

- Most libraries will compute the SVD for you

- They can do this by choosing the smaller of two matrices $\mathbf{X}\mathbf{X}^\mathsf{T}$ and $\mathbf{X}^\mathsf{T}\mathbf{X}$ and then compute the eigenvalues

- The singular values are the square root of the eigenvalues (notice that $\mathbf{X}\mathbf{X}^\mathsf{T}$ and $\mathbf{X}^\mathsf{T}\mathbf{X}$ are both positive semi-definite so the eigenvalues will be non-negative)

- It can compute the $\mathbf{U}$ matrix or $\mathbf{V}$ matrix by multiplying through by $\mathbf{X}$ or $\mathbf{X}^\mathsf{T}$ ($\mathbf{U} = \mathbf{X}\mathbf{V}\mathbf{S}^{-1}$ and $\mathbf{V} = \mathbf{X}^\mathsf{T}\mathbf{U}\mathbf{S}^{-1}$)

- In practice to perform PCA most people subtract the mean from their data and then perform SVD

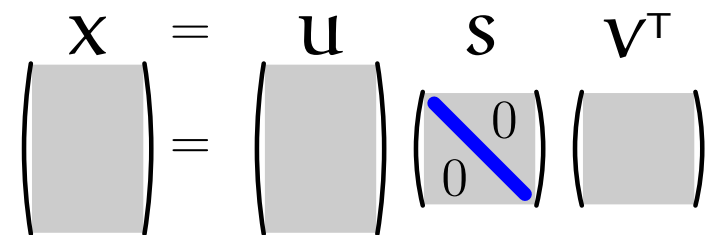# Finding SVD

- Most libraries will compute the SVD for you

- They can do this by choosing the smaller of two matrices XX^T and $X^\mathsf{T}X$ and then compute the eigenvalues

- The singular values are the square root of the eigenvalues (notice that XX^T and $X^\mathsf{T}X$ are both positive semi-definite so the eigenvalues will be non-negative)

- It can compute the U matrix or V matrix by multiplying through by X or X^T ($U = XVS^{-1}$ and $V = X^\mathsf{T}US^{-1}$)

- In practice to perform PCA most people subtract the mean from their data and then perform SVD

Economical Forms of SVD

- Often the rows or columns of the orthogonal matrices **U** and **V** that are not associated with a singular value are ignored
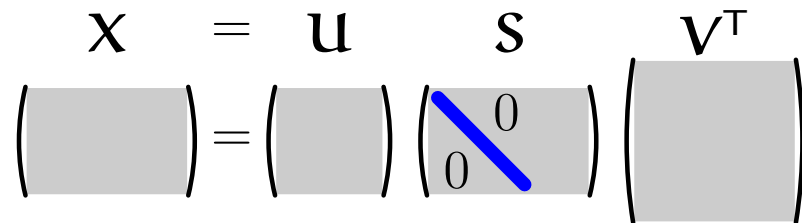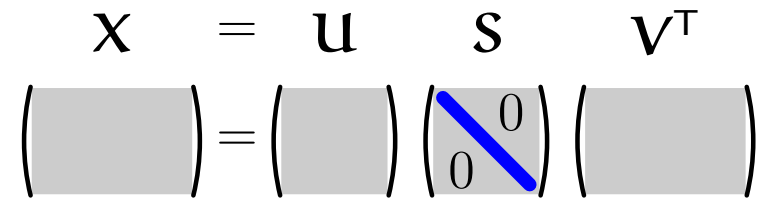

- In Matlab these are obtained using
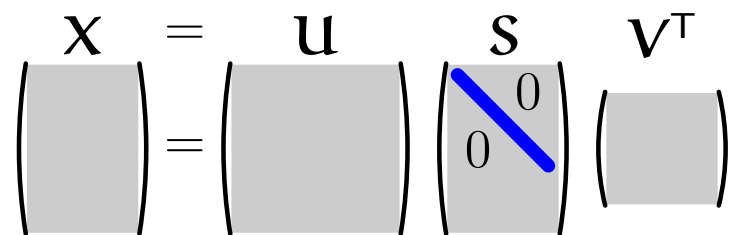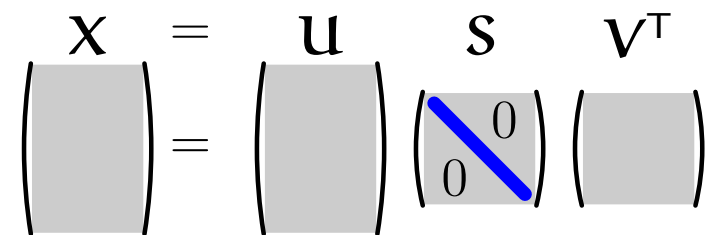
```
>> [U, S, V] = svd(X)
>> [U, S, V] = svd(X,'econ'))
```

Economical Forms of SVD

- Often the rows or columns of the orthogonal matrices U and V that are not associated with a singular value are ignored

- In Matlab these are obtained using

```
>> [U, S, V] = svd(X)
>> [U, S, V] = svd(X,'econ'))
```

Outline

1. Singular Value Decomposition

2. **General Linear Mappings**

3. Linear Regression Revisited

$$\begin{pmatrix} 0 & X \\ X^{\mathsf{T}} & 0 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = s \begin{pmatrix} u \\ v \end{pmatrix}$$

General Matrix

- Recall that we can compute the SVD for any matrix, $\mathbf{X}$

- As matrices describe the most general linear mapping

$$v \rightarrow \mathcal{T}[v] = \mathbf{X}v$$

- We can use SVD to understand any linear mapping

- Thus any linear mapping can be seen as a rotation followed by a squashing or expansion independently in each coordinate followed by another rotation

---

General Matrix

- Recall that we can compute the SVD for any matrix, $\mathbf{X}$

- As matrices describe the most general linear mapping

$$v \to \mathcal{T}[v] = \mathbf{X}v$$

- We can use SVD to understand any linear mapping

- Thus any linear mapping can be seen as a rotation followed by a squashing or expansion independently in each coordinate followed by another rotation

General Matrix

- Recall that we can compute the SVD for any matrix, $\mathbf{X}$

- As matrices describe the most general linear mapping

$$v \to \mathcal{T}[v] = \mathbf{X}v$$

- We can use SVD to understand any linear mapping

- Thus any linear mapping can be seen as a rotation followed by a squashing or expansion independently in each coordinate followed by another rotation

General Matrix

- Recall that we can compute the SVD for any matrix, $\mathbf{X}$

- As matrices describe the most general linear mapping

$$v \to \mathcal{T}[v] = \mathbf{X}v$$

- We can use SVD to understand any linear mapping

- Thus any linear mapping can be seen as a rotation followed by a squashing or expansion independently in each coordinate followed by another rotation
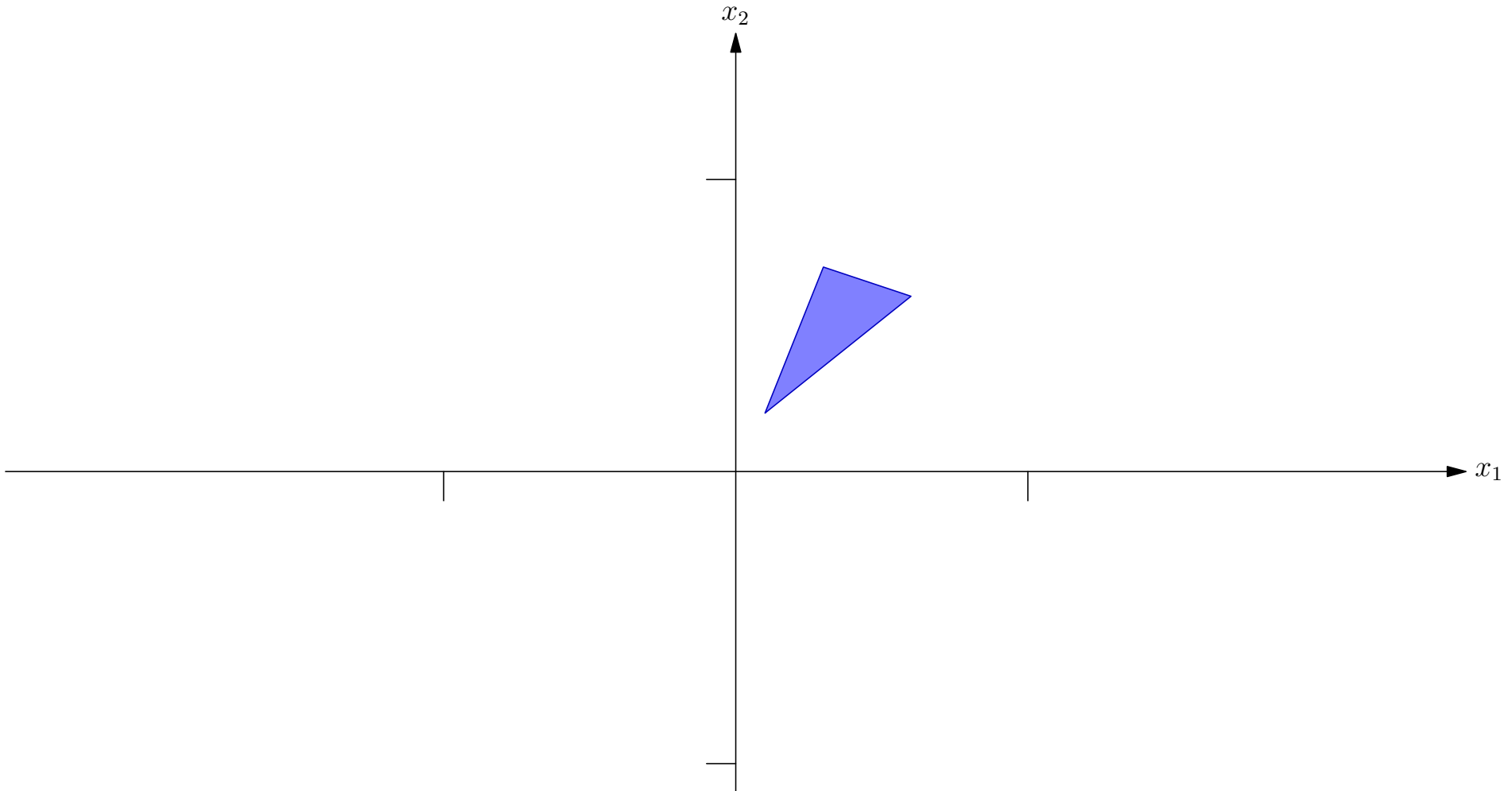
Matrices

$$\mathsf{M} = \begin{pmatrix} -0.45 & 1.9 \\ -0.77 & -0.025 \end{pmatrix} = \mathsf{U}\,\mathsf{S}\,\mathsf{V}^\mathsf{T} = \begin{pmatrix} \cos(-175) & \sin(-175) \\ -\sin(-175) & \cos(-175) \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & 0.75 \end{pmatrix} \begin{pmatrix} \cos(75) & \sin(75) \\ -\sin(75) & \cos(75) \end{pmatrix}$$

Matrices

$$\mathsf{M} = \begin{pmatrix} -0.45 & 1.9 \\ -0.77 & -0.025 \end{pmatrix} = \mathsf{U}\,\mathsf{S}\,\mathsf{V}^{\mathsf{T}} = \begin{pmatrix} \cos(-175) & \sin(-175) \\ -\sin(-175) & \cos(-175) \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & 0.75 \end{pmatrix} \begin{pmatrix} \cos(75) & \sin(75) \\ -\sin(75) & \cos(75) \end{pmatrix}$$
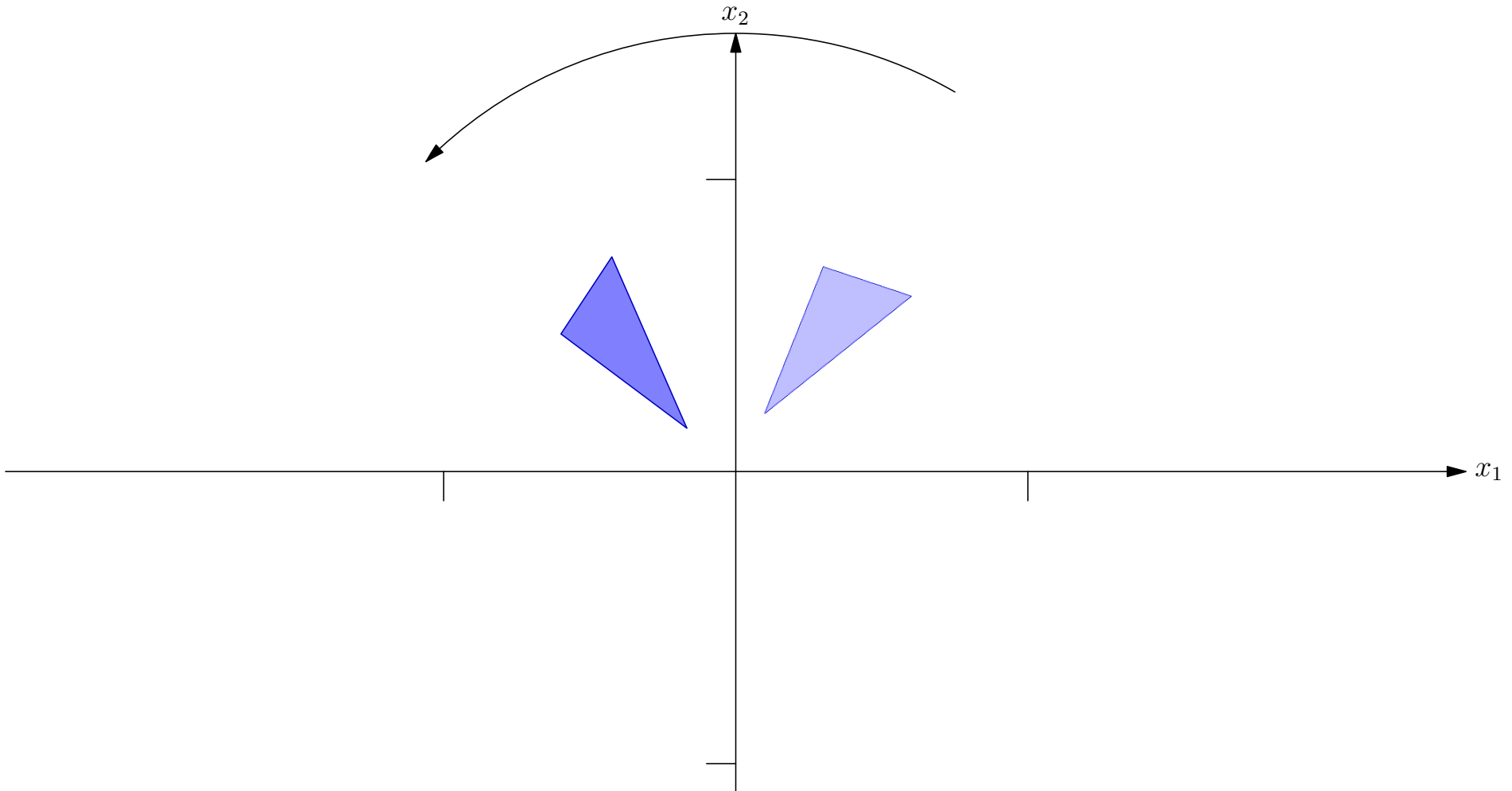
Matrices

$$\mathbf{M} = \begin{pmatrix} -0.45 & 1.9 \\ -0.77 & -0.025 \end{pmatrix} = \mathbf{U}\,\mathbf{S}\,\mathbf{V}^\mathsf{T} = \begin{pmatrix} \cos(-175) & \sin(-175) \\ -\sin(-175) & \cos(-175) \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & 0.75 \end{pmatrix} \begin{pmatrix} \cos(75) & \sin(75) \\ -\sin(75) & \cos(75) \end{pmatrix}$$

Matrices

$$\mathbf{M} = \begin{pmatrix} -0.45 & 1.9 \\ -0.77 & -0.025 \end{pmatrix} = \mathbf{U}\,\mathbf{S}\,\mathbf{V}^{\mathsf{T}} = \begin{pmatrix} \cos(-175) & \sin(-175) \\ -\sin(-175) & \cos(-175) \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & 0.75 \end{pmatrix} \begin{pmatrix} \cos(75) & \sin(75) \\ -\sin(75) & \cos(75) \end{pmatrix}$$
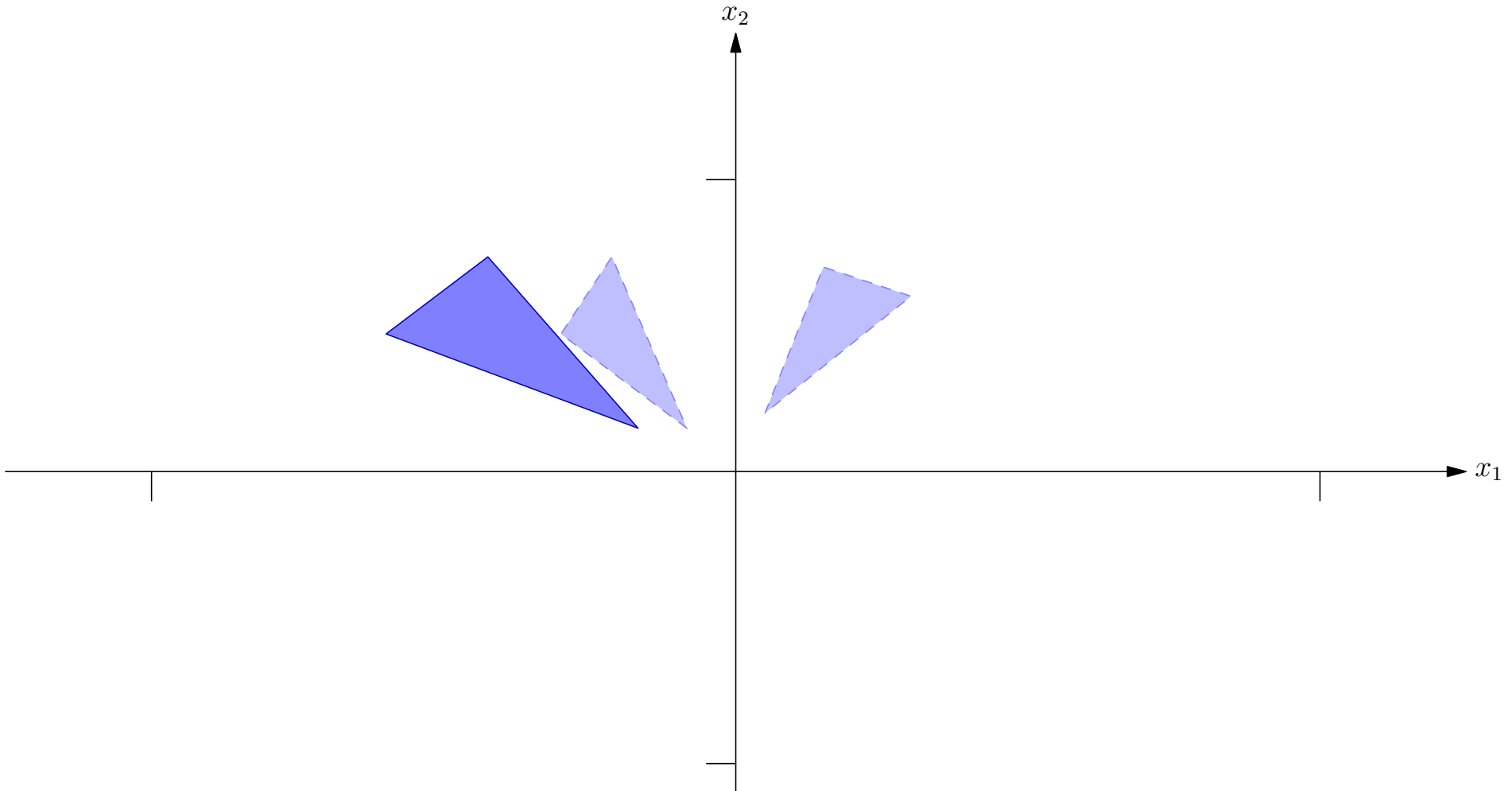
Matrices

$$\mathbf{M} = \begin{pmatrix} -0.45 & 1.9 \\ -0.77 & -0.025 \end{pmatrix} = \mathbf{U}\,\mathbf{S}\,\mathbf{V}^\mathsf{T} = \begin{pmatrix} \cos(-175) & \sin(-175) \\ -\sin(-175) & \cos(-175) \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & 0.75 \end{pmatrix} \begin{pmatrix} \cos(75) & \sin(75) \\ -\sin(75) & \cos(75) \end{pmatrix}$$
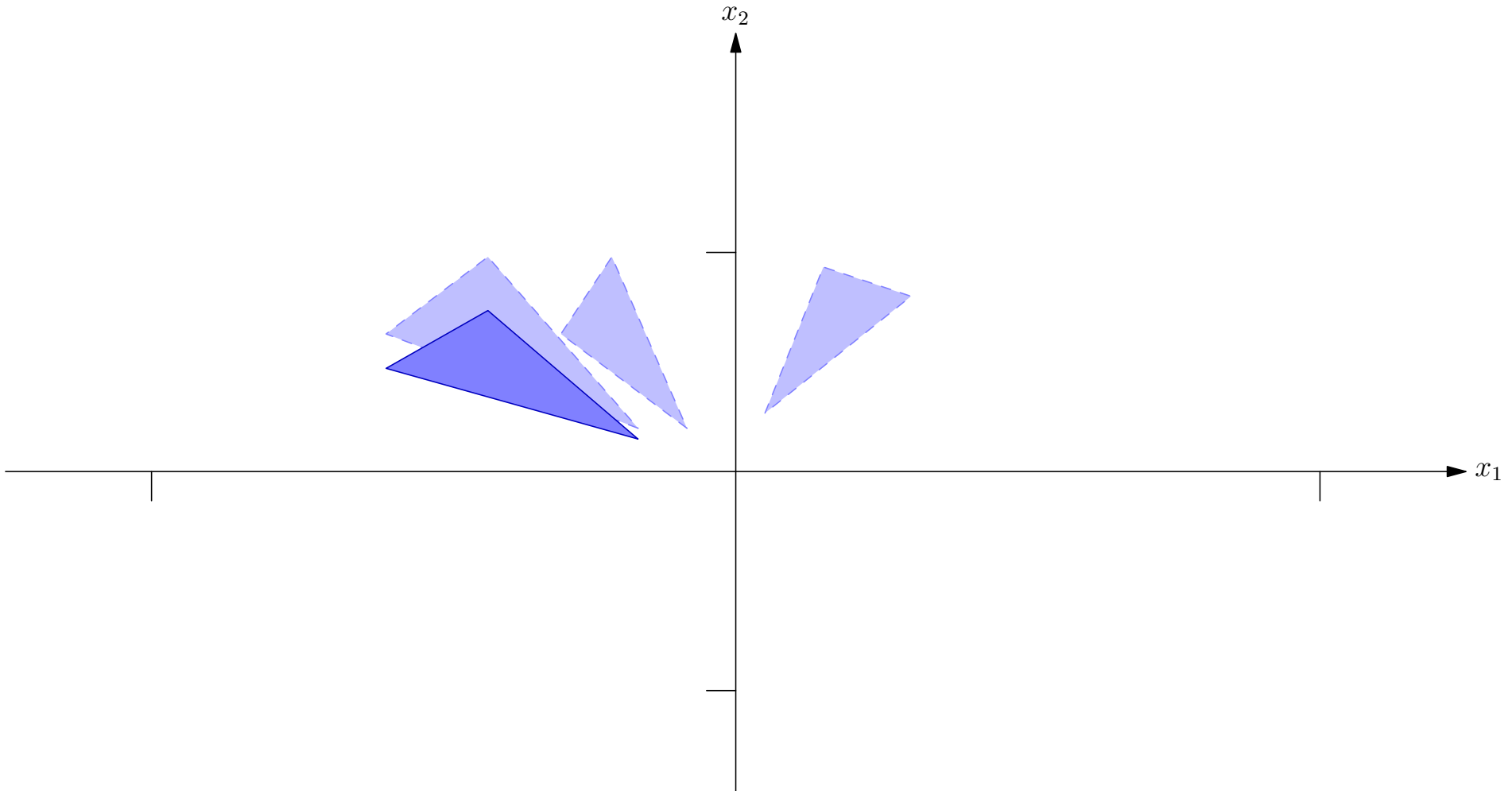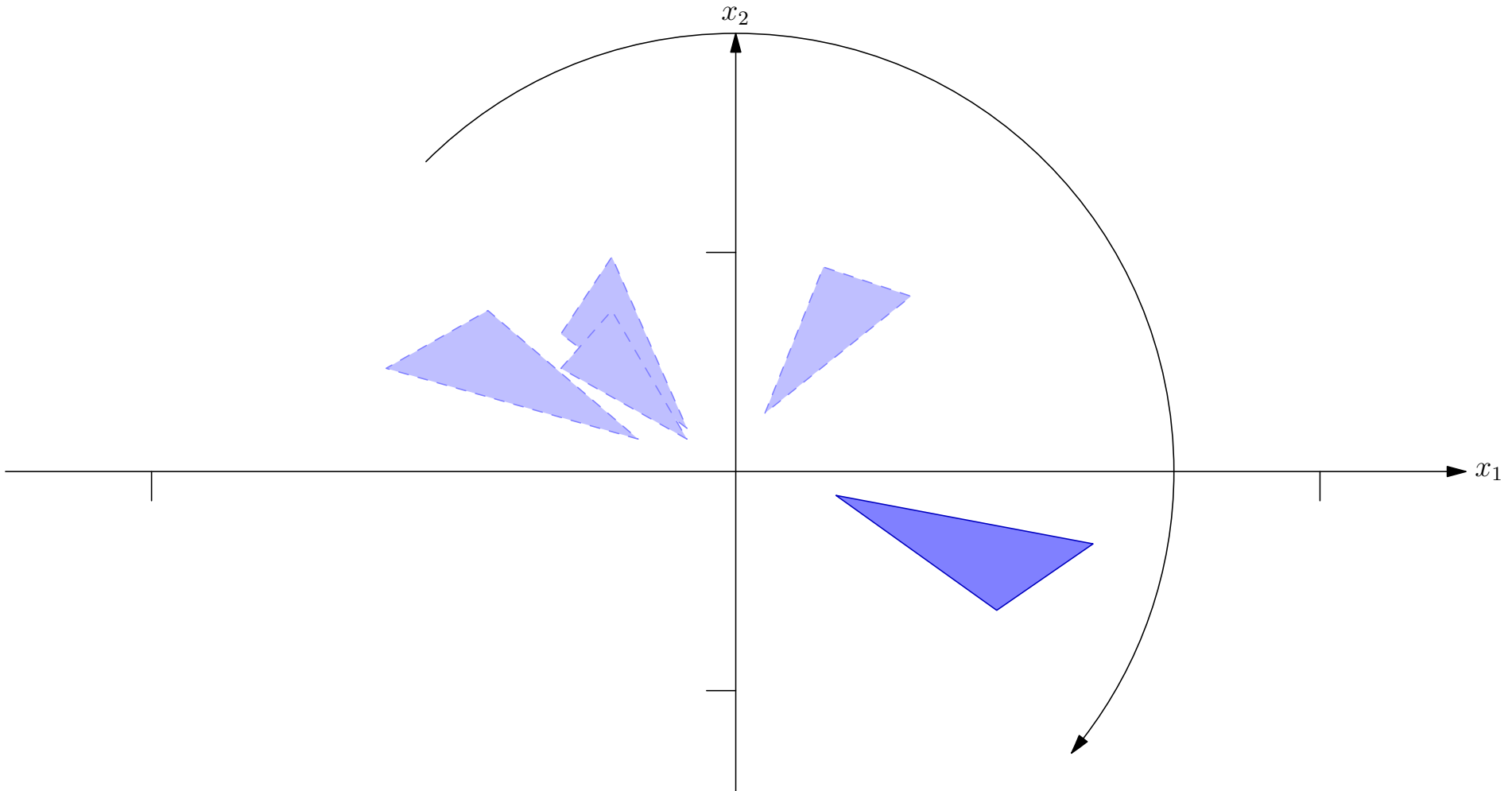
Determinants

- The determinant, $|\mathbf{M}|$ of a matrix $\mathbf{M}$ is defined for square matrices

- It describes the change in volume under the mapping

- Now for any two matrices $|\mathbf{A}\mathbf{B}| = |\mathbf{A}||\mathbf{B}|$

- Thus $|\mathbf{M}| = |\mathbf{U}||\mathbf{S}||\mathbf{V}^{\mathsf{T}}|$

- For and orthogonal matrix $|\mathbf{U}| = \pm 1$

- Thus

$$|\mathbf{M}| = \pm|\mathbf{S}|$$

Determinants

- The determinant, $|\mathbf{M}|$ of a matrix $\mathbf{M}$ is defined for square matrices

- It describes the change in volume under the mapping

- Now for any two matrices $|\mathbf{A}\mathbf{B}| = |\mathbf{A}||\mathbf{B}|$

- Thus $|\mathbf{M}| = |\mathbf{U}||\mathbf{S}||\mathbf{V}^{\mathsf{T}}|$

- For and orthogonal matrix $|\mathbf{U}| = \pm 1$

- Thus

$$|\mathbf{M}| = \pm|\mathbf{S}|$$

Determinants

- The determinant, $|\mathbf{M}|$ of a matrix $\mathbf{M}$ is defined for square matrices

- It describes the change in volume under the mapping

- Now for any two matrices $|\mathbf{A}\mathbf{B}| = |\mathbf{A}||\mathbf{B}|$

- Thus $|\mathbf{M}| = |\mathbf{U}||\mathbf{S}||\mathbf{V}^{\mathsf{T}}|$

- For and orthogonal matrix $|\mathbf{U}| = \pm 1$

- Thus

$$|\mathbf{M}| = \pm|\mathbf{S}|$$

Determinants

- The determinant, $|\mathbf{M}|$ of a matrix $\mathbf{M}$ is defined for square matrices

- It describes the change in volume under the mapping

- Now for any two matrices $|\mathbf{A}\mathbf{B}| = |\mathbf{A}||\mathbf{B}|$

- Thus $|\mathbf{M}| = |\mathbf{U}||\mathbf{S}||\mathbf{V}^{\mathsf{T}}|$

- For and orthogonal matrix $|\mathbf{U}| = \pm 1$

- Thus

$$|\mathbf{M}| = \pm|\mathbf{S}|$$

Determinants

- The determinant, $|\mathbf{M}|$ of a matrix $\mathbf{M}$ is defined for square matrices

- It describes the change in volume under the mapping

- Now for any two matrices $|\mathbf{A}\mathbf{B}| = |\mathbf{A}||\mathbf{B}|$

- Thus $|\mathbf{M}| = |\mathbf{U}||\mathbf{S}||\mathbf{V}^{\mathsf{T}}|$

- For and orthogonal matrix $|\mathbf{U}| = \pm 1$

- Thus

$$|\mathbf{M}| = \pm|\mathbf{S}|$$

Determinants

- The determinant, $|\mathbf{M}|$ of a matrix $\mathbf{M}$ is defined for square matrices

- It describes the change in volume under the mapping

- Now for any two matrices $|\mathbf{A}\mathbf{B}| = |\mathbf{A}||\mathbf{B}|$

- Thus $|\mathbf{M}| = |\mathbf{U}||\mathbf{S}||\mathbf{V}^{\mathsf{T}}|$

- For and orthogonal matrix $|\mathbf{U}| = \pm 1$ since
  $$\mathbf{U}\mathbf{U}^{\mathsf{T}} = \mathbf{I}$$

- Thus

$$|\mathbf{M}| = \pm|\mathbf{S}|$$

Determinants

- The determinant, $|\mathbf{M}|$ of a matrix $\mathbf{M}$ is defined for square matrices

- It describes the change in volume under the mapping

- Now for any two matrices $|\mathbf{A}\mathbf{B}| = |\mathbf{A}||\mathbf{B}|$

- Thus $|\mathbf{M}| = |\mathbf{U}||\mathbf{S}||\mathbf{V}^{\mathsf{T}}|$

- For and orthogonal matrix $|\mathbf{U}| = \pm 1$ since
$$\mathbf{U}\mathbf{U}^{\mathsf{T}} = \mathbf{I} \Rightarrow |\mathbf{U}\mathbf{U}^{\mathsf{T}}| = |\mathbf{I}|$$

- Thus

$$|\mathbf{M}| = \pm|\mathbf{S}|$$

Determinants

- The determinant, $|\mathbf{M}|$ of a matrix $\mathbf{M}$ is defined for square matrices

- It describes the change in volume under the mapping

- Now for any two matrices $|\mathbf{A}\mathbf{B}| = |\mathbf{A}||\mathbf{B}|$

- Thus $|\mathbf{M}| = |\mathbf{U}||\mathbf{S}||\mathbf{V}^{\mathsf{T}}|$

- For and orthogonal matrix $|\mathbf{U}| = \pm 1$ since
$$\mathbf{U}\mathbf{U}^{\mathsf{T}} = \mathbf{I} \Rightarrow |\mathbf{U}\mathbf{U}^{\mathsf{T}}| = |\mathbf{I}| \Rightarrow |\mathbf{U}||\mathbf{U}^{\mathsf{T}}| = 1$$

- Thus

$$|\mathbf{M}| = \pm|\mathbf{S}|$$

Determinants

- The determinant, $|\mathbf{M}|$ of a matrix $\mathbf{M}$ is defined for square matrices

- It describes the change in volume under the mapping

- Now for any two matrices $|\mathbf{A}\mathbf{B}| = |\mathbf{A}||\mathbf{B}|$

- Thus $|\mathbf{M}| = |\mathbf{U}||\mathbf{S}||\mathbf{V}^\mathsf{T}|$

- For and orthogonal matrix $|\mathbf{U}| = \pm 1$ since
$$\mathbf{U}\mathbf{U}^\mathsf{T} = \mathbf{I} \Rightarrow |\mathbf{U}\mathbf{U}^\mathsf{T}| = |\mathbf{I}| \Rightarrow |\mathbf{U}||\mathbf{U}^\mathsf{T}| = 1 \text{ or } |\mathbf{U}|^2 = 1$$

- Thus

$$|\mathbf{M}| = \pm|\mathbf{S}|$$

Determinants

- The determinant, $|\mathbf{M}|$ of a matrix $\mathbf{M}$ is defined for square matrices

- It describes the change in volume under the mapping

- Now for any two matrices $|\mathbf{A}\mathbf{B}| = |\mathbf{A}||\mathbf{B}|$

- Thus $|\mathbf{M}| = |\mathbf{U}||\mathbf{S}||\mathbf{V}^\mathsf{T}|$

- For and orthogonal matrix $|\mathbf{U}| = \pm 1$ since
  $\mathbf{U}\mathbf{U}^\mathsf{T} = \mathbf{I} \Rightarrow |\mathbf{U}\mathbf{U}^\mathsf{T}| = |\mathbf{I}| \Rightarrow |\mathbf{U}||\mathbf{U}^\mathsf{T}| = 1$ or $|\mathbf{U}|^2 = 1$

- Thus
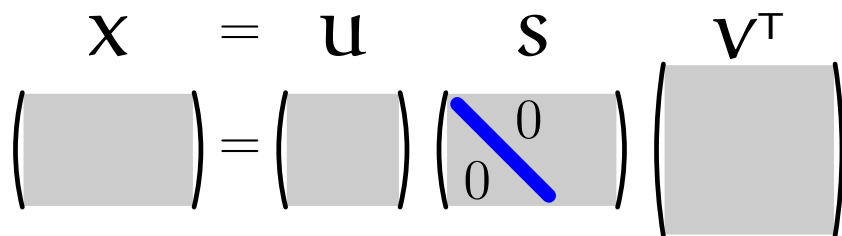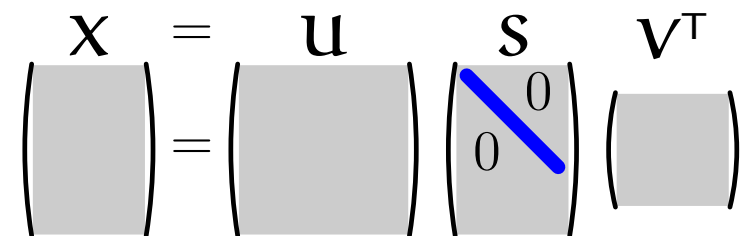
$$|\mathbf{M}| = \pm|\mathbf{S}|$$

Determinants

- The determinant, $|\mathbf{M}|$ of a matrix $\mathbf{M}$ is defined for square matrices

- It describes the change in volume under the mapping

- Now for any two matrices $|\mathbf{A}\mathbf{B}| = |\mathbf{A}||\mathbf{B}|$

- Thus $|\mathbf{M}| = |\mathbf{U}||\mathbf{S}||\mathbf{V}^{\mathsf{T}}|$

- For and orthogonal matrix $|\mathbf{U}| = \pm 1$ since
  $\mathbf{U}\mathbf{U}^{\mathsf{T}} = \mathbf{I} \Rightarrow |\mathbf{U}\mathbf{U}^{\mathsf{T}}| = |\mathbf{I}| \Rightarrow |\mathbf{U}||\mathbf{U}^{\mathsf{T}}| = 1$ or $|\mathbf{U}|^2 = 1$

- Thus

$$|\mathbf{M}| = \pm|\mathbf{S}| = \pm \prod_i s_i$$

Non-Square Matrices

- When the matrices are non-square then the matrix of singular value matrix will either

  ⋆ Squash some directions to zero
  ⋆ Introduce new dimensions orthogonal to the vector

$$\mathbf{X} = \mathbf{U} \quad \mathbf{S} \quad \mathbf{V}^\mathsf{T} \qquad \qquad \mathbf{X} = \mathbf{U} \quad \mathbf{S} \quad \mathbf{V}^\mathsf{T}$$

- The rank of an arbitrary matrix is the number of non-zero singular values (also number of linearly independent rows or columns)

Non-Square Matrices

- When the matrices are non-square then the matrix of singular value matrix will either

  ⋆ Squash some directions to zero
  ⋆ Introduce new dimensions orthogonal to the vector

$$X = U \quad S \quad V^{\mathsf{T}}$$

$$X = U \quad S \quad V^{\mathsf{T}}$$

- The rank of an arbitrary matrix is the number of non-zero singular values (also number of linearly independent rows or columns)

Non-Square Matrices

- When the matrices are non-square then the matrix of singular value matrix will either

  - ⋆ Squash some directions to zero
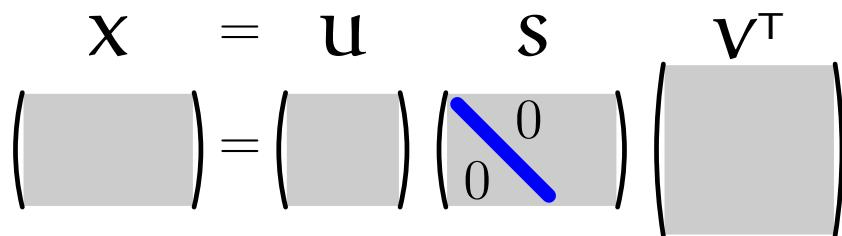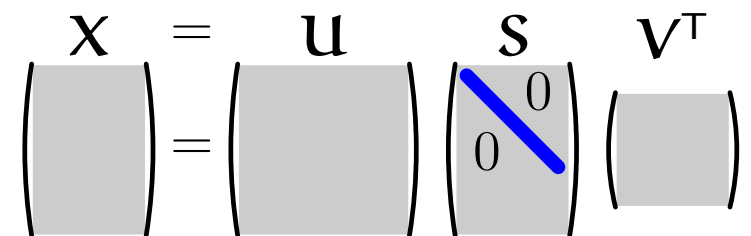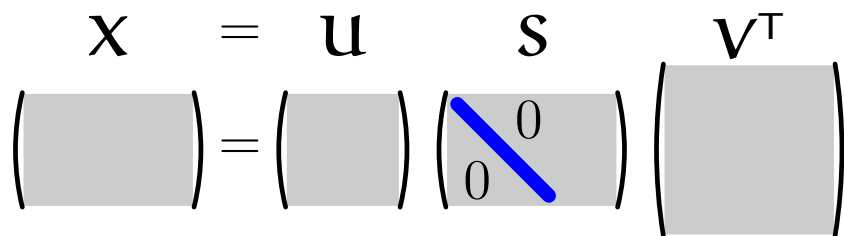 - ⋆ Introduce new dimensions orthogonal to the vector

- The rank of an arbitrary matrix is the number of non-zero singular values (also number of linearly independent rows or columns)

Duality Revisited

- If $\mathbf{X} = \mathbf{USV}^\mathsf{T}$ then

$$\mathbf{C} = \mathbf{XX}^\mathsf{T} \qquad\qquad \mathbf{D} = \mathbf{X}^\mathsf{T}\mathbf{X}$$

$$= \mathbf{USV}^\mathsf{T}\mathbf{VS}^\mathsf{T}\mathbf{U}^\mathsf{T} \qquad\qquad = \mathbf{VS}^\mathsf{T}\mathbf{U}^\mathsf{T}\mathbf{USV}^\mathsf{T}$$

$$= \mathbf{U}(\mathbf{SS}^\mathsf{T})\mathbf{U}^\mathsf{T} \qquad\qquad = \mathbf{V}(\mathbf{S}^\mathsf{T}\mathbf{S})\mathbf{V}^\mathsf{T}$$

- If $\mathbf{X}$ is an $p \times m$ matrix then $\mathbf{SS}^\mathsf{T}$ is a $p \times p$ diagonal matrix with elements $S_{ii}^2 = s_i^2$

- $\mathbf{S}^\mathsf{T}\mathbf{S}$ is an $m \times m$ matrix with elements $S_{ii}^2 = s_i^2$

- $\mathbf{U}$ and $\mathbf{V}$ are matrices of eigenvectors for $\mathbf{C}$ and $\mathbf{D}$

- The eigenvalues are $\lambda_i = S_{ii}^2 = s_i^2$

Duality Revisited

- If $X = USV^\mathsf{T}$ then

$$C = XX^\mathsf{T} \qquad\qquad D = X^\mathsf{T}X$$

$$ = USV^\mathsf{T}VS^\mathsf{T}U^\mathsf{T} \qquad\qquad = VS^\mathsf{T}U^\mathsf{T}USV^\mathsf{T}$$

$$ = U(SS^\mathsf{T})U^\mathsf{T} \qquad\qquad = V(S^\mathsf{T}S)V^\mathsf{T}$$

- If X is an $p \times m$ matrix then SS^T is a $p \times p$ diagonal matrix with elements $S_{ii}^2 = s_i^2$

- $S^\mathsf{T}S$ is an $m \times m$ matrix with elements $S_{ii}^2 = s_i^2$

- U and V are matrices of eigenvectors for C and D

- The eigenvalues are $\lambda_i = S_{ii}^2 = s_i^2$

Duality Revisited

- If $\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^\mathsf{T}$ then

$$\mathbf{C} = \mathbf{X}\mathbf{X}^\mathsf{T} \qquad\qquad \mathbf{D} = \mathbf{X}^\mathsf{T}\mathbf{X}$$
$$= \mathbf{U}\mathbf{S}\mathbf{V}^\mathsf{T}\mathbf{V}\mathbf{S}^\mathsf{T}\mathbf{U}^\mathsf{T} \qquad\qquad = \mathbf{V}\mathbf{S}^\mathsf{T}\mathbf{U}^\mathsf{T}\mathbf{U}\mathbf{S}\mathbf{V}^\mathsf{T}$$
$$\textcolor{red}{= \mathbf{U}(\mathbf{S}\mathbf{S}^\mathsf{T})\mathbf{U}^\mathsf{T}} \qquad\qquad \textcolor{red}{= \mathbf{V}(\mathbf{S}^\mathsf{T}\mathbf{S})\mathbf{V}^\mathsf{T}}$$

- If $\mathbf{X}$ is an $p \times m$ matrix then $\mathbf{S}\mathbf{S}^\mathsf{T}$ is a $p \times p$ diagonal matrix with elements $S_{ii}^2 = s_i^2$

- $\mathbf{S}^\mathsf{T}\mathbf{S}$ is an $m \times m$ matrix with elements $S_{ii}^2 = s_i^2$

- $\mathbf{U}$ and $\mathbf{V}$ are matrices of eigenvectors for $\mathbf{C}$ and $\mathbf{D}$

- The eigenvalues are $\lambda_i = S_{ii}^2 = s_i^2$

Duality Revisited

- If $\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^\mathsf{T}$ then

$$\mathbf{C} = \mathbf{X}\mathbf{X}^\mathsf{T} \qquad\qquad \mathbf{D} = \mathbf{X}^\mathsf{T}\mathbf{X}$$

$$= \mathbf{U}\mathbf{S}\mathbf{V}^\mathsf{T}\mathbf{V}\mathbf{S}^\mathsf{T}\mathbf{U}^\mathsf{T} \qquad\qquad = \mathbf{V}\mathbf{S}^\mathsf{T}\mathbf{U}^\mathsf{T}\mathbf{U}\mathbf{S}\mathbf{V}^\mathsf{T}$$

$$= \mathbf{U}(\mathbf{S}\mathbf{S}^\mathsf{T})\mathbf{U}^\mathsf{T} \qquad\qquad = \mathbf{V}(\mathbf{S}^\mathsf{T}\mathbf{S})\mathbf{V}^\mathsf{T}$$

- If $\mathbf{X}$ is an $p \times m$ matrix then $\mathbf{S}\mathbf{S}^\mathsf{T}$ is a $p \times p$ diagonal matrix with elements $S_{ii}^2 = s_i^2$

- $\mathbf{S}^\mathsf{T}\mathbf{S}$ is an $m \times m$ matrix with elements $S_{ii}^2 = s_i^2$

- $\mathbf{U}$ and $\mathbf{V}$ are matrices of eigenvectors for $\mathbf{C}$ and $\mathbf{D}$

- The eigenvalues are $\lambda_i = S_{ii}^2 = s_i^2$

Duality Revisited

- If $\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^\mathsf{T}$ then

$$\mathbf{C} = \mathbf{X}\mathbf{X}^\mathsf{T} \qquad\qquad \mathbf{D} = \mathbf{X}^\mathsf{T}\mathbf{X}$$
$$= \mathbf{U}\mathbf{S}\mathbf{V}^\mathsf{T}\mathbf{V}\mathbf{S}^\mathsf{T}\mathbf{U}^\mathsf{T} \qquad\qquad = \mathbf{V}\mathbf{S}^\mathsf{T}\mathbf{U}^\mathsf{T}\mathbf{U}\mathbf{S}\mathbf{V}^\mathsf{T}$$
$$= \mathbf{U}(\mathbf{S}\mathbf{S}^\mathsf{T})\mathbf{U}^\mathsf{T} \qquad\qquad = \mathbf{V}(\mathbf{S}^\mathsf{T}\mathbf{S})\mathbf{V}^\mathsf{T}$$

- If $\mathbf{X}$ is an $p \times m$ matrix then $\mathbf{S}\mathbf{S}^\mathsf{T}$ is a $p \times p$ diagonal matrix with elements $S_{ii}^2 = s_i^2$

- $\mathbf{S}^\mathsf{T}\mathbf{S}$ is an $m \times m$ matrix with elements $S_{ii}^2 = s_i^2$

- $\mathbf{U}$ and $\mathbf{V}$ are matrices of eigenvectors for $\mathbf{C}$ and $\mathbf{D}$

- The eigenvalues are $\lambda_i = S_{ii}^2 = s_i^2$

Duality Revisited

- If $\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^\mathsf{T}$ then

$$\mathbf{C} = \mathbf{X}\mathbf{X}^\mathsf{T} \qquad\qquad \mathbf{D} = \mathbf{X}^\mathsf{T}\mathbf{X}$$
$$= \mathbf{U}\mathbf{S}\mathbf{V}^\mathsf{T}\mathbf{V}\mathbf{S}^\mathsf{T}\mathbf{U}^\mathsf{T} \qquad\qquad = \mathbf{V}\mathbf{S}^\mathsf{T}\mathbf{U}^\mathsf{T}\mathbf{U}\mathbf{S}\mathbf{V}^\mathsf{T}$$
$$= \mathbf{U}(\mathbf{S}\mathbf{S}^\mathsf{T})\mathbf{U}^\mathsf{T} \qquad\qquad = \mathbf{V}(\mathbf{S}^\mathsf{T}\mathbf{S})\mathbf{V}^\mathsf{T}$$

- If $\mathbf{X}$ is an $p \times m$ matrix then $\mathbf{S}\mathbf{S}^\mathsf{T}$ is a $p \times p$ diagonal matrix with elements $S_{ii}^2 = s_i^2$

- $\mathbf{S}^\mathsf{T}\mathbf{S}$ is an $m \times m$ matrix with elements $S_{ii}^2 = s_i^2$

- $\mathbf{U}$ and $\mathbf{V}$ are matrices of eigenvectors for $\mathbf{C}$ and $\mathbf{D}$

- The eigenvalues are $\lambda_i = S_{ii}^2 = s_i^2$

Duality Revisited

- If $\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^{\mathsf{T}}$ then

$$\mathbf{C} = \mathbf{X}\mathbf{X}^{\mathsf{T}} \qquad\qquad \mathbf{D} = \mathbf{X}^{\mathsf{T}}\mathbf{X}$$
$$= \mathbf{U}\mathbf{S}\mathbf{V}^{\mathsf{T}}\mathbf{V}\mathbf{S}^{\mathsf{T}}\mathbf{U}^{\mathsf{T}} \qquad\qquad = \mathbf{V}\mathbf{S}^{\mathsf{T}}\mathbf{U}^{\mathsf{T}}\mathbf{U}\mathbf{S}\mathbf{V}^{\mathsf{T}}$$
$$= \mathbf{U}(\mathbf{S}\mathbf{S}^{\mathsf{T}})\mathbf{U}^{\mathsf{T}} \qquad\qquad = \mathbf{V}(\mathbf{S}^{\mathsf{T}}\mathbf{S})\mathbf{V}^{\mathsf{T}}$$

- If $\mathbf{X}$ is an $p \times m$ matrix then $\mathbf{S}\mathbf{S}^{\mathsf{T}}$ is a $p \times p$ diagonal matrix with elements $S_{ii}^2 = s_i^2$

- $\mathbf{S}^{\mathsf{T}}\mathbf{S}$ is an $m \times m$ matrix with elements $S_{ii}^2 = s_i^2$

- $\mathbf{U}$ and $\mathbf{V}$ are matrices of eigenvectors for $\mathbf{C}$ and $\mathbf{D}$

- The eigenvalues are $\lambda_i = S_{ii}^2 = s_i^2$

SS^T and $S^\mathsf{T}S$

$$S = \begin{pmatrix} s_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & s_2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & s_m & 0 & 0\cdots & 0 \end{pmatrix}$$

$$S^\mathsf{T}S = \begin{pmatrix} s_1^2 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & s_2^2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & s_m^2 & 0 & 0\cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0\cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots\cdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0\cdots & 0 \end{pmatrix}$$

$$SS^\mathsf{T} = \begin{pmatrix} s_1^2 & 0 & \cdots & 0 \\ 0 & s_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & s_m^2 \end{pmatrix}$$

SS^T and $S^\mathsf{T}S$

$$S = \begin{pmatrix} s_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & s_2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & s_m & 0 & 0\cdots & 0 \end{pmatrix}$$

$$S^\mathsf{T}S = \begin{pmatrix} s_1^2 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & s_2^2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & s_m^2 & 0 & 0\cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0\cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots\cdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0\cdots & 0 \end{pmatrix}$$

$$SS^\mathsf{T} = \begin{pmatrix} s_1^2 & 0 & \cdots & 0 \\ 0 & s_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & s_m^2 \end{pmatrix}$$

$\mathsf{SS^T}$ and $\mathsf{S^TS}$

$$\mathsf{S} = \begin{pmatrix} s_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & s_2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & s_m & 0 & 0\cdots & 0 \end{pmatrix}$$

$$\mathsf{S^TS} = \begin{pmatrix} s_1^2 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & s_2^2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & s_m^2 & 0 & 0\cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0\cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots\cdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0\cdots & 0 \end{pmatrix}$$

$$\mathsf{SS^T} = \begin{pmatrix} s_1^2 & 0 & \cdots & 0 \\ 0 & s_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & s_m^2 \end{pmatrix}$$

Having A Go

- It's really easy to verify this in MATLAB or OCTAVE

```
>> X = rand(3,2)
>> [U, S, V] = svd(X)
>> U*S*V'
>> U(:,1)'*U(:,2)
>> U'*U
>> U*U'
>> [Ua,L] = eig(X*X')
>> S*S'
```

- Test yourself!

Having A Go

- It's really easy to verify this in MATLAB or OCTAVE

```
>> X = rand(3,2)
>> [U, S, V] = svd(X)
>> U*S*V'
>> U(:,1)'*U(:,2)
>> U'*U
>> U*U'
>> [Ua,L] = eig(X*X')
>> S*S'
```

- Test yourself!

Having A Go

- It's really easy to verify this in MATLAB or OCTAVE

```
>> X = rand(3,2)
>> [U, S, V] = svd(X)
>> U*S*V'
>> U(:,1)'*U(:,2)
>> U'*U
>> U*U'
>> [Ua,L] = eig(X*X')
>> S*S'
```

- Test yourself!

Outline

1. Singular Value Decomposition

2. General Linear Mappings

3. **Linear Regression Revisited**

$$\begin{pmatrix} 0 & X \\ X^\mathsf{T} & 0 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = s \begin{pmatrix} u \\ v \end{pmatrix}$$

Linear Regression

- Given a set of data $\mathcal{D} = \{(\boldsymbol{x}_i, y_i) | k = 1, 2, \ldots, m\}$

- In linear regression we try to fit a linear model

$$f(\boldsymbol{x}|\boldsymbol{w}) = \boldsymbol{x}^\top \boldsymbol{w}$$

- Which we fit by minimising the squared error loss

$$L(\boldsymbol{w}) = \sum_{k=1}^{m} (f(\boldsymbol{x}_i|\boldsymbol{w}) - y_i)^2$$

Linear Regression

- Given a set of data $\mathcal{D} = \{(\boldsymbol{x}_i, y_i) | k = 1, 2, \ldots, m\}$

- In linear regression we try to fit a linear model

$$f(\boldsymbol{x}|\boldsymbol{w}) = \boldsymbol{x}^\mathsf{T}\boldsymbol{w}$$

- Which we fit by minimising the squared error loss

$$L(\boldsymbol{w}) = \sum_{k=1}^{m} (f(\boldsymbol{x}_i|\boldsymbol{w}) - y_i)^2$$

Linear Regression

- Given a set of data $\mathcal{D} = \{(\boldsymbol{x}_i, y_i) | k = 1, 2, ..., m\}$

- In linear regression we try to fit a linear model

$$f(\boldsymbol{x}|\boldsymbol{w}) = \boldsymbol{x}^\top \boldsymbol{w}$$

- Which we fit by minimising the squared error loss

$$L(\boldsymbol{w}) = \sum_{k=1}^{m} \left(f(\boldsymbol{x}_i|\boldsymbol{w}) - y_i\right)^2$$

Matrix Form

- In matrix from we write $L(\boldsymbol{w}) = \|\mathbf{X}\boldsymbol{w} - \boldsymbol{y}\|^2$

$$\mathbf{X} = \begin{pmatrix} \boldsymbol{x}_1^\mathsf{T} \\ \boldsymbol{x}_2^\mathsf{T} \\ \vdots \\ \boldsymbol{x}_m^\mathsf{T} \end{pmatrix} \qquad \boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2, \\ \vdots \\ y_m \end{pmatrix}$$

- Then $\boldsymbol{\nabla} L(\boldsymbol{w}^*) = 0$ implies

$$\boldsymbol{w}^* = \left(\mathbf{X}^\mathsf{T}\mathbf{X}\right)^{-1}\mathbf{X}^\mathsf{T}\boldsymbol{y} = \mathbf{X}^+\boldsymbol{y}$$

- This is known as the pseudo-inverse

Matrix Form

- In matrix from we write $L(\boldsymbol{w}) = \|\mathbf{X}\boldsymbol{w} - \boldsymbol{y}\|^2$

$$\mathbf{X} = \begin{pmatrix} \boldsymbol{x}_1^\top \\ \boldsymbol{x}_2^\top \\ \vdots \\ \boldsymbol{x}_m^\top \end{pmatrix} \qquad \boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2, \\ \vdots \\ y_m \end{pmatrix}$$

- Then $\boldsymbol{\nabla}L(\boldsymbol{w}^*) = 0$ implies

$$\boldsymbol{w}^* = \left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\boldsymbol{y} = \mathbf{X}^+\boldsymbol{y}$$

- This is known as the pseudo-inverse

Matrix Form

- In matrix from we write $L(\boldsymbol{w}) = \|\mathbf{X}\boldsymbol{w} - \boldsymbol{y}\|^2$

$$\mathbf{X} = \begin{pmatrix} \boldsymbol{x}_1^\top \\ \boldsymbol{x}_2^\top \\ \vdots \\ \boldsymbol{x}_m^\top \end{pmatrix} \qquad \boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2, \\ \vdots \\ y_m \end{pmatrix}$$

- Then $\boldsymbol{\nabla} L(\boldsymbol{w}^*) = 0$ implies

$$\boldsymbol{w}^* = \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \mathbf{X}^\top \boldsymbol{y} = \mathbf{X}^+ \boldsymbol{y}$$

- This is known as the pseudo-inverse

Using SVD

- Using $X = USV^\mathsf{T}$ then

$$X^+ = (X^\mathsf{T}X)^{-1}X^\mathsf{T}$$

$$= (VS^\mathsf{T}SV^\mathsf{T})^{-1}VS^\mathsf{T}U^\mathsf{T}$$

$$= V(S^\mathsf{T}S)^{-1}V^\mathsf{T}VS^\mathsf{T}U^\mathsf{T}$$

$$= V(S^\mathsf{T}S)^{-1}S^\mathsf{T}U^\mathsf{T} = VS^+U^\mathsf{T}$$

- If $m > p$

$$X^\mathsf{T} = \begin{pmatrix} \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid \\ & & & & & & & \\ & & & & & & & \\ \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid \end{pmatrix}, S^\mathsf{T} = \begin{pmatrix} s_1 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & s_2 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & s_3 & \cdots & 0 & & 0\cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & s_p & 0 & 0\cdots & 0 \end{pmatrix}$$

Using SVD

- Using $X = USV^\mathsf{T}$ then

$$X^+ = (X^\mathsf{T} X)^{-1} X^\mathsf{T}$$

$$= (V S^\mathsf{T} S V^\mathsf{T})^{-1} V S^\mathsf{T} U^\mathsf{T}$$

$$= V (S^\mathsf{T} S)^{-1} V^\mathsf{T} V S^\mathsf{T} U^\mathsf{T}$$

$$= V (S^\mathsf{T} S)^{-1} S^\mathsf{T} U^\mathsf{T} = V S^+ U^\mathsf{T}$$

- If $m > p$

$$X^\mathsf{T} = \begin{pmatrix} \vline & \vline & \vline & \vline & \vline & \vline & \vline \\ \end{pmatrix}, S^\mathsf{T} = \begin{pmatrix} s_1 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & s_2 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & s_3 & \cdots & 0 & & 0\cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & s_p & 0 & 0\cdots & 0 \end{pmatrix}$$

Using SVD

- Using $X = USV^\mathsf{T}$ then

$$X^+ = (X^\mathsf{T} X)^{-1} X^\mathsf{T}$$

$$= (V S^\mathsf{T} S V^\mathsf{T})^{-1} V S^\mathsf{T} U^\mathsf{T}$$

$$\textcolor{red}{= V (S^\mathsf{T} S)^{-1} V^\mathsf{T} V S^\mathsf{T} U^\mathsf{T}}$$

$$= V (S^\mathsf{T} S)^{-1} S^\mathsf{T} U^\mathsf{T} = V S^+ U^\mathsf{T}$$

- If $m > p$

$$X^\mathsf{T} = \begin{pmatrix} \Big| & \Big| & \Big| & \Big| & \Big| & \Big| & \Big| \end{pmatrix}, S^\mathsf{T} = \begin{pmatrix} s_1 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & s_2 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & s_3 & \cdots & 0 & & 0 \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & s_p & 0 & 0 \cdots & 0 \end{pmatrix}$$

Using SVD

- Using $\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^{\mathsf{T}}$ then

$$
\begin{aligned}
\mathbf{X}^{+} &= (\mathbf{X}^{\mathsf{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathsf{T}} \\
&= (\mathbf{V}\mathbf{S}^{\mathsf{T}}\mathbf{S}\mathbf{V}^{\mathsf{T}})^{-1}\mathbf{V}\mathbf{S}^{\mathsf{T}}\mathbf{U}^{\mathsf{T}} \\
&= \mathbf{V}(\mathbf{S}^{\mathsf{T}}\mathbf{S})^{-1}\mathbf{V}^{\mathsf{T}}\mathbf{V}\mathbf{S}^{\mathsf{T}}\mathbf{U}^{\mathsf{T}} \\
&= \mathbf{V}(\mathbf{S}^{\mathsf{T}}\mathbf{S})^{-1}\mathbf{S}^{\mathsf{T}}\mathbf{U}^{\mathsf{T}} = \mathbf{V}\mathbf{S}^{+}\mathbf{U}^{\mathsf{T}}
\end{aligned}
$$

- If $m > p$

$$
\mathbf{X}^{\mathsf{T}} = \left(\begin{array}{|c|c|c|c|c|c|c|c|} \hline \;\;\;\; \\ \hline \end{array}\right), \mathbf{S}^{\mathsf{T}} = \begin{pmatrix} s_1 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & s_2 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & s_3 & \cdots & 0 & & 0\cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & s_p & 0 & 0\cdots & 0 \end{pmatrix}
$$

Using SVD

- Using $X = USV^\mathsf{T}$ then

$$X^+ = (X^\mathsf{T}X)^{-1}X^\mathsf{T}$$

$$= (VS^\mathsf{T}SV^\mathsf{T})^{-1}VS^\mathsf{T}U^\mathsf{T}$$

$$= V(S^\mathsf{T}S)^{-1}V^\mathsf{T}VS^\mathsf{T}U^\mathsf{T}$$

$$= V(S^\mathsf{T}S)^{-1}S^\mathsf{T}U^\mathsf{T} = \textcolor{red}{VS^+U^\mathsf{T}}$$

- If $m > p$

$$X^\mathsf{T} = \begin{pmatrix} \vphantom{|} \\ \end{pmatrix}, S^\mathsf{T} = \begin{pmatrix} s_1 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & s_2 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & s_3 & \cdots & 0 & & 0\cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & s_p & 0 & 0\cdots & 0 \end{pmatrix}$$

Using SVD

- Using $\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^\mathsf{T}$ then

$$\begin{aligned}
\mathbf{X}^+ &= (\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}\mathbf{X}^\mathsf{T} \\
&= (\mathbf{V}\mathbf{S}^\mathsf{T}\mathbf{S}\mathbf{V}^\mathsf{T})^{-1}\mathbf{V}\mathbf{S}^\mathsf{T}\mathbf{U}^\mathsf{T} \\
&= \mathbf{V}(\mathbf{S}^\mathsf{T}\mathbf{S})^{-1}\mathbf{V}^\mathsf{T}\mathbf{V}\mathbf{S}^\mathsf{T}\mathbf{U}^\mathsf{T} \\
&= \mathbf{V}(\mathbf{S}^\mathsf{T}\mathbf{S})^{-1}\mathbf{S}^\mathsf{T}\mathbf{U}^\mathsf{T} = \mathbf{V}\mathbf{S}^+\mathbf{U}^\mathsf{T}
\end{aligned}$$

- If $m > p$

$$\mathbf{X}^\mathsf{T} = \begin{pmatrix} \Big| & \Big| & \Big| & \Big| & \Big| & \Big| & \Big| & \Big| \end{pmatrix}, \mathbf{S}^\mathsf{T} = \begin{pmatrix} s_1 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & s_2 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & s_3 & \cdots & 0 & & 0\cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & s_p & 0 & 0\cdots & 0 \end{pmatrix}$$

Using SVD

- Using $X = USV^\mathsf{T}$ then

$$X^+ = (X^\mathsf{T}X)^{-1}X^\mathsf{T}$$

$$= (VS^\mathsf{T}SV^\mathsf{T})^{-1}VS^\mathsf{T}U^\mathsf{T}$$

$$= V(S^\mathsf{T}S)^{-1}V^\mathsf{T}VS^\mathsf{T}U^\mathsf{T}$$

$$= V(S^\mathsf{T}S)^{-1}S^\mathsf{T}U^\mathsf{T} = VS^+U^\mathsf{T}$$

- If $m > p$

$$X^\mathsf{T} = \begin{pmatrix} \big| & \big| & \big| & \big| & \big| & \big| & \big| & \big| \end{pmatrix}, \quad S^\mathsf{T} = \begin{pmatrix} s_1 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & s_2 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & s_3 & \cdots & 0 & & 0 \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & s_p & 0 & 0 \cdots & 0 \end{pmatrix}$$

Pseudo-Inverse of S

$$\mathbf{S}^\mathsf{T}\mathbf{S} = \begin{pmatrix} s_1^2 & 0 & \cdots & 0 \\ 0 & s_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & s_p^2 \end{pmatrix} \qquad (\mathbf{S}^\mathsf{T}\mathbf{S})^{-1} = \begin{pmatrix} s_1^{-2} & 0 & \cdots & 0 \\ 0 & s_2^{-2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & s_p^{-2} \end{pmatrix}$$

$$\mathbf{S}^+ = (\mathbf{S}^\mathsf{T}\mathbf{S})^{-1}\mathbf{S}^\mathsf{T} = \begin{pmatrix} s_1^{-1} & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & s_2^{-1} & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & s_3^{-1} & \cdots & 0 & 0\cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & s_p^{-1} & 0 & 0\cdots & 0 \end{pmatrix}$$

Pseudo-Inverse of S

$$S^\mathsf{T} S = \begin{pmatrix} s_1^2 & 0 & \cdots & 0 \\ 0 & s_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & s_p^2 \end{pmatrix} \qquad \color{red}{\left(S^\mathsf{T} S\right)^{-1} = \begin{pmatrix} s_1^{-2} & 0 & \cdots & 0 \\ 0 & s_2^{-2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & s_p^{-2} \end{pmatrix}}$$

$$S^+ = \left(S^\mathsf{T} S\right)^{-1} S^\mathsf{T} = \begin{pmatrix} s_1^{-1} & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & s_2^{-1} & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & s_3^{-1} & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & s_p^{-1} & 0 & 0 & \cdots & 0 \end{pmatrix}$$

Pseudo-Inverse of $\mathbf{S}$

$$\mathbf{S}^\mathsf{T}\mathbf{S} = \begin{pmatrix} s_1^2 & 0 & \cdots & 0 \\ 0 & s_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & s_p^2 \end{pmatrix} \qquad \left(\mathbf{S}^\mathsf{T}\mathbf{S}\right)^{-1} = \begin{pmatrix} s_1^{-2} & 0 & \cdots & 0 \\ 0 & s_2^{-2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & s_p^{-2} \end{pmatrix}$$

$$\mathbf{S}^+ = \left(\mathbf{S}^\mathsf{T}\mathbf{S}\right)^{-1}\mathbf{S}^\mathsf{T} = \begin{pmatrix} s_1^{-1} & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & s_2^{-1} & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & s_3^{-1} & \cdots & 0 & 0\cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & s_p^{-1} & 0 & 0\cdots & 0 \end{pmatrix}$$

Pseudo-Inverse of S

$$\mathbf{S}^\mathsf{T}\mathbf{S} = \begin{pmatrix} s_1^2 & 0 & \cdots & 0 \\ 0 & s_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & s_p^2 \end{pmatrix} \qquad (\mathbf{S}^\mathsf{T}\mathbf{S})^{-1} = \begin{pmatrix} s_1^{-2} & 0 & \cdots & 0 \\ 0 & s_2^{-2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & s_p^{-2} \end{pmatrix}$$

$$\mathbf{S}^+ = (\mathbf{S}^\mathsf{T}\mathbf{S})^{-1}\mathbf{S}^\mathsf{T} = \textcolor{red}{\begin{pmatrix} s_1^{-1} & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & s_2^{-1} & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & s_3^{-1} & \cdots & 0 & 0\cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & s_p^{-1} & 0 & 0\cdots & 0 \end{pmatrix}}$$

Ill-Conditioned Data Matrix

- Recall that

$$w^* = X^+ y = V S^+ U^\top y$$

- If any of the singular values of X are small then S^+ will magnify components in that direction

- Any errors in the target y will be magnified

- This leads to poor weights

Ill-Conditioned Data Matrix

- Recall that

$$w^* = \mathbf{X}^+ y = \mathbf{V}\mathbf{S}^+\mathbf{U}^\mathsf{T} y$$

- If any of the singular values of $\mathbf{X}$ are small then $\mathbf{S}^+$ will magnify components in that direction

- Any errors in the target y will be magnified

- This leads to poor weights

Ill-Conditioned Data Matrix

- Recall that

$$\boldsymbol{w}^* = \mathbf{X}^+ \boldsymbol{y} = \mathbf{V}\mathbf{S}^+\mathbf{U}^\mathsf{T}\boldsymbol{y}$$

- If any of the singular values of $\mathbf{X}$ are small then $\mathbf{S}^+$ will magnify components in that direction

- Any errors in the target $\boldsymbol{y}$ will be magnified

- This leads to poor weights

Ill-Conditioned Data Matrix

- Recall that

$$\boldsymbol{w}^* = \mathbf{X}^+ \boldsymbol{y} = \mathbf{V}\mathbf{S}^+\mathbf{U}^\mathsf{T}\boldsymbol{y}$$

- If any of the singular values of $\mathbf{X}$ are small then $\mathbf{S}^+$ will magnify components in that direction

- Any errors in the target $\boldsymbol{y}$ will be magnified

- This leads to poor weights

Regularisation

- Consider linear regression with a regulariser

$$\mathcal{L}(\boldsymbol{w}) = \|\mathbf{X}\boldsymbol{w} - \boldsymbol{y}\|^2 + \eta\|\boldsymbol{w}\|^2$$

$$= \boldsymbol{w}^\mathsf{T}\left(\mathbf{X}^\mathsf{T}\mathbf{X} + \eta\mathbf{I}\right)\boldsymbol{w} - 2\boldsymbol{w}^\mathsf{T}\mathbf{X}^\mathsf{T}\boldsymbol{y} + \boldsymbol{y}^\mathsf{T}\boldsymbol{y}$$

- Thus

$$\boldsymbol{\nabla}\mathcal{L}(\boldsymbol{w}) = 2\left(\mathbf{X}^\mathsf{T}\mathbf{X} + \eta\mathbf{I}\right)\boldsymbol{w} - 2\mathbf{X}^\mathsf{T}\boldsymbol{y}$$

- and $\boldsymbol{\nabla}\mathcal{L}(\boldsymbol{w}^*) = 0$ gives

$$\boldsymbol{w}^* = \left(\mathbf{X}^\mathsf{T}\mathbf{X} + \eta\mathbf{I}\right)^{-1}\mathbf{X}^\mathsf{T}\boldsymbol{y}$$

Regularisation

* Consider linear regression with a regulariser

$$\mathcal{L}(\boldsymbol{w}) = \|\mathbf{X}\boldsymbol{w} - \boldsymbol{y}\|^2 + \eta\|\boldsymbol{w}\|^2$$
$$= \boldsymbol{w}^\mathsf{T}\left(\mathbf{X}^\mathsf{T}\mathbf{X} + \eta\mathbf{I}\right)\boldsymbol{w} - 2\boldsymbol{w}^\mathsf{T}\mathbf{X}^\mathsf{T}\boldsymbol{y} + \boldsymbol{y}^\mathsf{T}\boldsymbol{y}$$

* Thus

$$\boldsymbol{\nabla}\mathcal{L}(\boldsymbol{w}) = 2\left(\mathbf{X}^\mathsf{T}\mathbf{X} + \eta\mathbf{I}\right)\boldsymbol{w} - 2\mathbf{X}^\mathsf{T}\boldsymbol{y}$$

* and $\boldsymbol{\nabla}\mathcal{L}(\boldsymbol{w}^*) = 0$ gives

$$\boldsymbol{w}^* = \left(\mathbf{X}^\mathsf{T}\mathbf{X} + \eta\mathbf{I}\right)^{-1}\mathbf{X}^\mathsf{T}\boldsymbol{y}$$

Regularisation

- Consider linear regression with a regulariser

$$\mathcal{L}(\boldsymbol{w}) = \|\mathbf{X}\boldsymbol{w} - \boldsymbol{y}\|^2 + \eta\|\boldsymbol{w}\|^2$$
$$= \boldsymbol{w}^\mathsf{T}(\mathbf{X}^\mathsf{T}\mathbf{X} + \eta\mathbf{I})\boldsymbol{w} - 2\boldsymbol{w}^\mathsf{T}\mathbf{X}^\mathsf{T}\boldsymbol{y} + \boldsymbol{y}^\mathsf{T}\boldsymbol{y}$$

- Thus

$$\boldsymbol{\nabla}\mathcal{L}(\boldsymbol{w}) = 2\left(\mathbf{X}^\mathsf{T}\mathbf{X} + \eta\mathbf{I}\right)\boldsymbol{w} - 2\mathbf{X}^\mathsf{T}\boldsymbol{y}$$

- and $\boldsymbol{\nabla}\mathcal{L}(\boldsymbol{w}^*) = 0$ gives

$$\boldsymbol{w}^* = \left(\mathbf{X}^\mathsf{T}\mathbf{X} + \eta\mathbf{I}\right)^{-1}\mathbf{X}^\mathsf{T}\boldsymbol{y}$$

Regularisation Continued

- Using $X = USV^\top$

$$w^* = \left(X^\top X + \eta I\right)^{-1} X^\top y$$

$$= V\left(S^\top S + \eta I\right)^{-1} S^\top U^\top y$$

- where

$$\left(S^\top S + \eta I\right)^{-1} S^\top = \begin{pmatrix} \frac{s_1}{s_1^2+\eta} & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \frac{s_2}{s_2^2+\eta} & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & \frac{s_3}{s_3^2+\eta} & \cdots & 0 & & 0\cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{s_p}{s_p^2+\eta} & 0 & 0\cdots & 0 \end{pmatrix}$$

Regularisation Continued
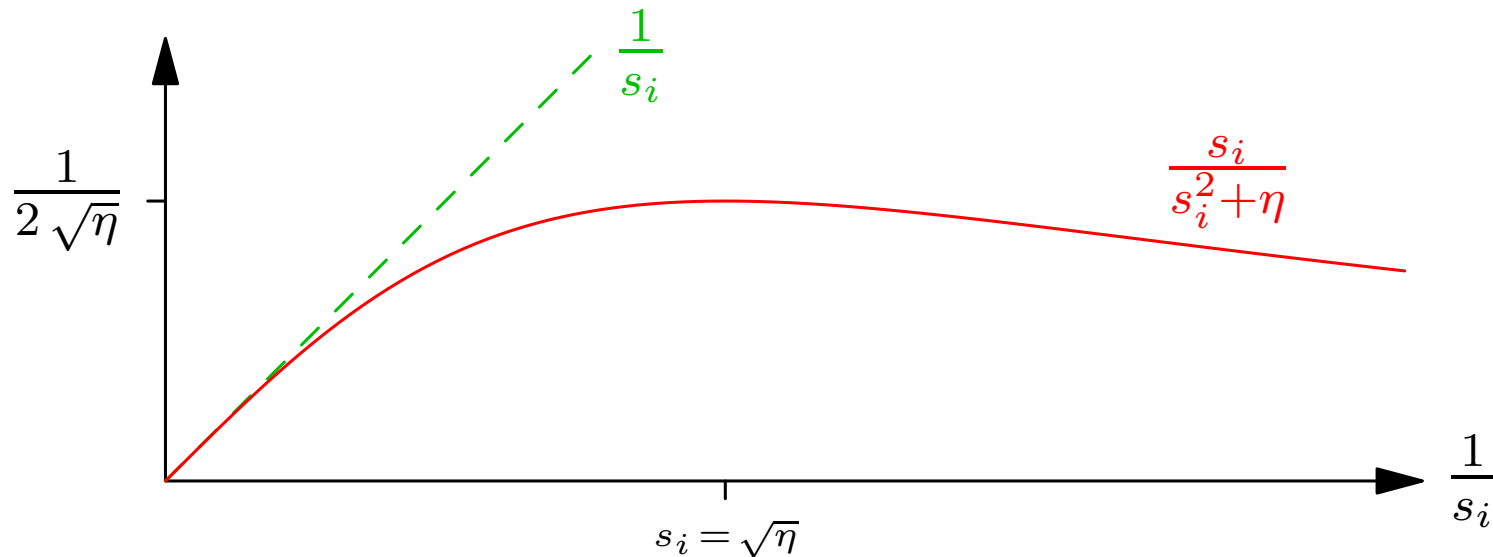
- Using $\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^\mathsf{T}$

$$\boldsymbol{w}^* = \left(\mathbf{X}^\mathsf{T}\mathbf{X} + \eta\mathbf{I}\right)^{-1}\mathbf{X}^\mathsf{T}\boldsymbol{y}$$

$$= \mathbf{V}\left(\mathbf{S}^\mathsf{T}\mathbf{S} + \eta\mathbf{I}\right)^{-1}\mathbf{S}^\mathsf{T}\mathbf{U}^\mathsf{T}\boldsymbol{y}$$

- where

$$\left(\mathbf{S}^\mathsf{T}\mathbf{S} + \eta\mathbf{I}\right)^{-1}\mathbf{S}^\mathsf{T} = \begin{pmatrix} \frac{s_1}{s_1^2+\eta} & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \frac{s_2}{s_2^2+\eta} & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & \frac{s_3}{s_3^2+\eta} & \cdots & 0 & & 0\cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{s_p}{s_p^2+\eta} & 0 & 0\cdots & 0 \end{pmatrix}$$

Regularisation Continued

- Using $X = USV^\mathsf{T}$

$$w^* = \left(X^\mathsf{T}X + \eta I\right)^{-1} X^\mathsf{T} y$$

$$= V\left(S^\mathsf{T}S + \eta I\right)^{-1} S^\mathsf{T} U^\mathsf{T} y$$

- where

$$\left(S^\mathsf{T}S + \eta I\right)^{-1} S^\mathsf{T} = \begin{pmatrix} \frac{s_1}{s_1^2+\eta} & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \frac{s_2}{s_2^2+\eta} & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & \frac{s_3}{s_3^2+\eta} & \cdots & 0 & & 0\cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{s_p}{s_p^2+\eta} & 0 & 0\cdots & 0 \end{pmatrix}$$
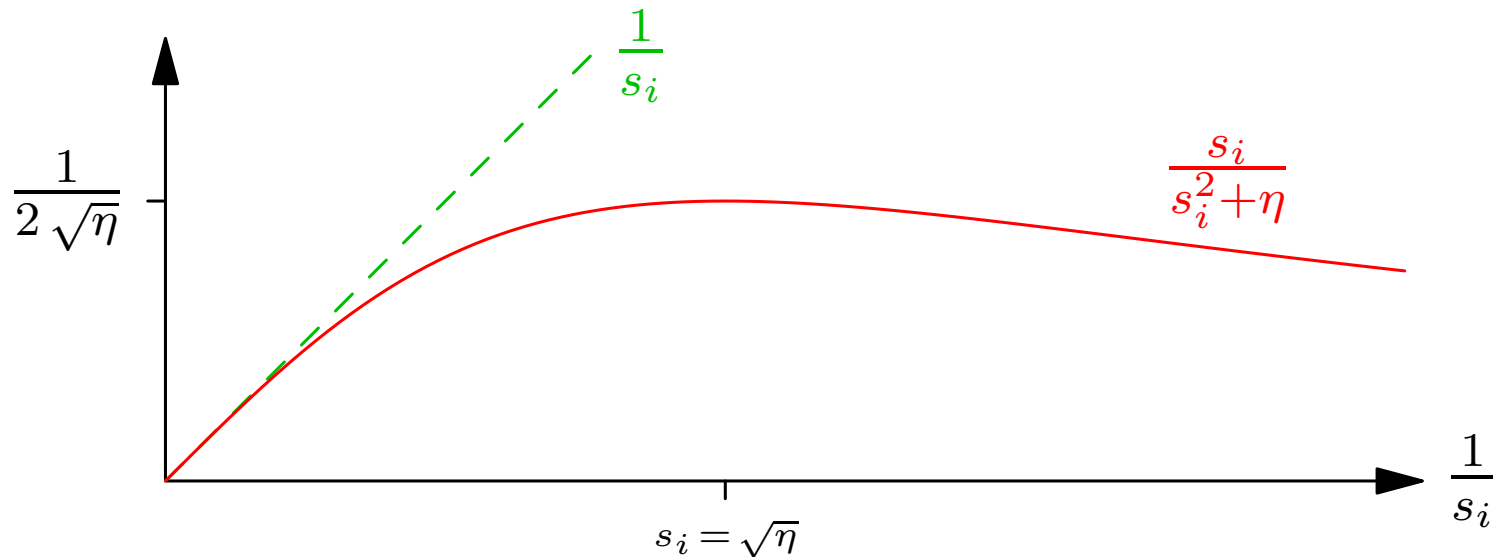
Effect of Regularisation

- Without regularisation if $s_i = 0$ the problem would be ill-posed (even S^+ does not exist since s_i^{-1} would be ill defined) and if s_i is small then S^+ is ill conditioned

- Using $\hat{S}^+ = (S^\mathsf{T} S + \eta)^{-1} S^\mathsf{T}$ instead of S^+ then

- Regularisation makes the machine much more stable (reduces the variance)

Effect of Regularisation

- Without regularisation if $s_i = 0$ the problem would be ill-posed (even S^+ does not exist since s_i^{-1} would be ill defined) and if s_i is small then S^+ is ill conditioned

- Using $\hat{S}^+ = (S^\mathsf{T} S + \eta)^{-1} S^\mathsf{T}$ instead of S^+ then

- Regularisation makes the machine much more stable (reduces the variance)
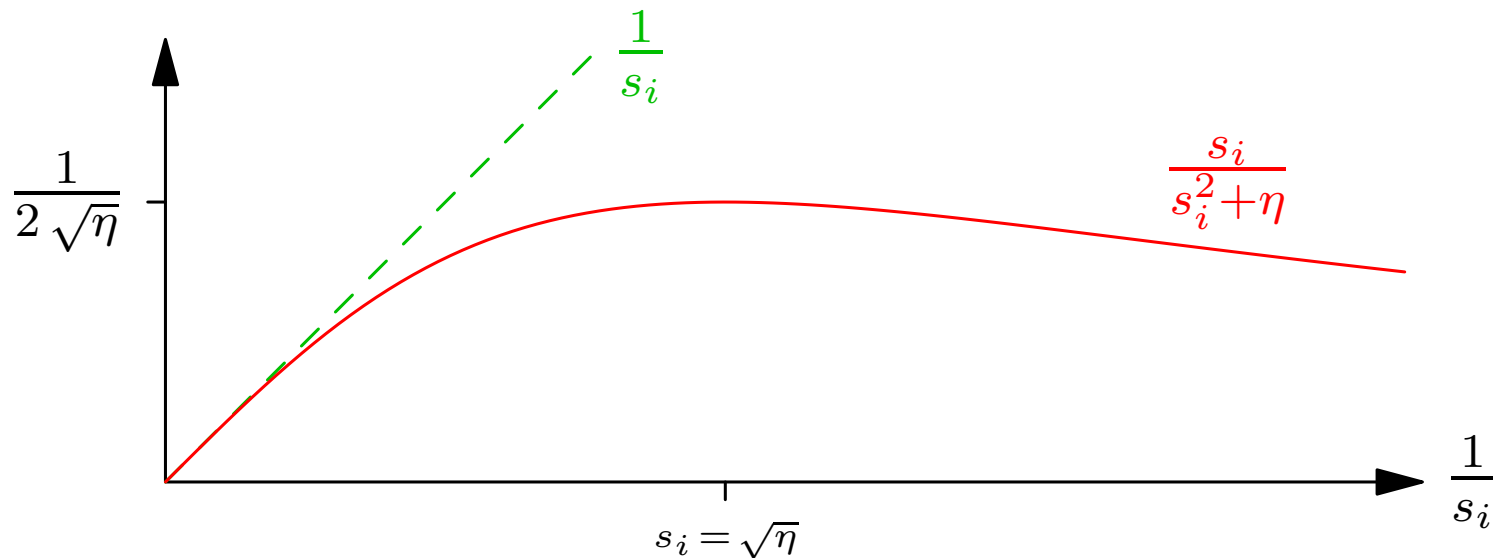
Effect of Regularisation

- Without regularisation if $s_i = 0$ the problem would be ill-posed (even $\boldsymbol{S}^+$ does not exist since s_i^{-1} would be ill defined) and if s_i is small then $\boldsymbol{S}^+$ is ill conditioned

- Using $\hat{\boldsymbol{S}}^+ = (\boldsymbol{S}^\mathsf{T}\boldsymbol{S} + \eta)^{-1}\boldsymbol{S}^\mathsf{T}$ instead of $\boldsymbol{S}^+$ then

- Regularisation makes the machine much more stable (reduces the variance)

Summary

- Any matrix can be decomposed as $\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^{\mathsf{T}}$ where

  ⋆ $\mathbf{U}$ and $\mathbf{V}$ are orthogonal (rotation matrices)
  ⋆ $\mathbf{S} = \mathrm{diag}(s_1, \ldots, s_n)$ is a diagonal matrix of positive singular values

- This describes the most general linear transform

- The transform exploits the duality between $\mathbf{X}\mathbf{X}^{\mathsf{T}}$ and $\mathbf{X}^{\mathsf{T}}\mathbf{X}$

- In linear regression the pseudo-inverse involves the reciprocal of the singular values, which can lead to poor generalisation

- Regularisation improves the conditioning of the "inverse" matrix

Summary

- Any matrix can be decomposed as $X = USV^\mathsf{T}$ where

  - U and V are orthogonal (rotation matrices)
  - $S = \mathrm{diag}(s_1, \ldots, s_n)$ is a diagonal matrix of positive singular values

- This describes the most general linear transform

- The transform exploits the duality between XX^T and $X^\mathsf{T}X$

- In linear regression the pseudo-inverse involves the reciprocal of the singular values, which can lead to poor generalisation

- Regularisation improves the conditioning of the "inverse" matrix

Summary

- Any matrix can be decomposed as $\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^\mathsf{T}$ where

  - ⋆ $\mathbf{U}$ and $\mathbf{V}$ are orthogonal (rotation matrices)
  - ⋆ $\mathbf{S} = \mathrm{diag}(s_1, \ldots, s_n)$ is a diagonal matrix of positive singular values

- This describes the most general linear transform

- The transform exploits the duality between $\mathbf{X}\mathbf{X}^\mathsf{T}$ and $\mathbf{X}^\mathsf{T}\mathbf{X}$

- In linear regression the pseudo-inverse involves the reciprocal of the singular values, which can lead to poor generalisation

- Regularisation improves the conditioning of the "inverse" matrix

Summary

- Any matrix can be decomposed as $\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^{\mathsf{T}}$ where

    ⋆ $\mathbf{U}$ and $\mathbf{V}$ are orthogonal (rotation matrices)
    ⋆ $\mathbf{S} = \operatorname{diag}(s_1, \ldots, s_n)$ is a diagonal matrix of positive singular values

- This describes the most general linear transform

- The transform exploits the duality between $\mathbf{X}\mathbf{X}^{\mathsf{T}}$ and $\mathbf{X}^{\mathsf{T}}\mathbf{X}$

- In linear regression the pseudo-inverse involves the reciprocal of the singular values, which can lead to poor generalisation

- Regularisation improves the conditioning of the "inverse" matrix

Summary

- Any matrix can be decomposed as $X = USV^\mathsf{T}$ where

  - ⋆ U and V are orthogonal (rotation matrices)
  - ⋆ $S = \mathrm{diag}(s_1, \ldots, s_n)$ is a diagonal matrix of positive singular values

- This describes the most general linear transform

- The transform exploits the duality between XX^T and $X^\mathsf{T}X$

- In linear regression the pseudo-inverse involves the reciprocal of the singular values, which can lead to poor generalisation

- Regularisation improves the conditioning of the "inverse" matrix

Summary

- Any matrix can be decomposed as $\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^\mathsf{T}$ where

  - $\star$ $\mathbf{U}$ and $\mathbf{V}$ are orthogonal (rotation matrices)
  - $\star$ $\mathbf{S} = \mathrm{diag}(s_1, \ldots, s_n)$ is a diagonal matrix of positive singular values

- This describes the most general linear transform

- The transform exploits the duality between $\mathbf{X}\mathbf{X}^\mathsf{T}$ and $\mathbf{X}^\mathsf{T}\mathbf{X}$

- In linear regression the pseudo-inverse involves the reciprocal of the singular values, which can lead to poor generalisation

- Regularisation improves the conditioning of the "inverse" matrix