# SVD, QR and Numerical/Exact Rank and eigenvalues' computation

Madhu Belur

Control & Computing group,
Electrical Engineering Dept, IIT Bombay

This talk (updated/corrected) at:

http://www.ee.iitb.ac.in/%7Ebelur/talks/

Dec 2017

# *Outline*

- Singular Value Decomposition
- Numerical and Exact Rank
- QR methods for solving $Ax = b$ and for rank determination (with tolerance)
- QR for solving $Ax = b$, for $\det(A)$, and $A^{-1}$
- Relative error and floating point
- Basics of flop count
- Eigenvalue definition and computation

Relevant reference books: Books on Matrix Computation by

- Golub & van Loan,
- David Watkins,
- Trefethen

# *SVD definition/applications*

$U \in \mathbb{R}^{n \times n}$ is called orthogonal if $U^{-1} = U^T$
Loosely speaking: using orthogonal matrices is numerically good
Orthogonal matrices are like 'rotation'.
('Reflectors' also)
View any matrix $A \in \mathbb{R}^{m \times n}$ as $A : \mathbb{R}^n \to \mathbb{R}^m$.
For $x \in \mathbb{R}^n$, we have $y \in \mathbb{R}^m$ defined as $y = Ax$.
Diagonal matrices: component-wise scaling
SVD $\equiv$ any matrix $Ax$: rotate $x$ first, then component-wise scaling, then rotation again

# *SVD: definition*

### *SVD*

**For a matrix $A \in \mathbb{R}^{m \times n}$, its Singular Value Decomposition (SVD) is defined as: $U\Sigma V := A$ with**

- **$U \in \mathbb{R}^{m \times m}$ and $V \in \mathbb{R}^{n \times n}$ are orthogonal**
- **$\Sigma \in \mathbb{R}^{m \times n}$ is diagonal with diagonal values $\sigma_i$ satisfying**

$$\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_r > 0$$

**with rank of $A := r$ and $r \leqslant \min(m, n)$**

**$\boldsymbol{\sigma_i}$ are called the singular values.**
**(After ordering), $\boldsymbol{\sigma_i}$ are unique.**
$U$ and $V$ are not unique.
$\sigma_1 =: \sigma_{\max}$, the largest singular value equals 'amplification'
'Not full rank' $\equiv r < \min(m, n)$
When $r = \min(m, n)$, (i.e. full rank), then $\sigma_{\min} := \sigma_r$ indicates 'nearness to losing rank'

$\boldsymbol{\sigma_i}$ **are called the singular values.**
**(After ordering),** $\boldsymbol{\sigma_i}$ **are unique.**
$\boldsymbol{U}$ **and** $\boldsymbol{V}$ **are not unique.**
$\boldsymbol{\sigma_1 =: \sigma_{\max}}$**, the largest singular value equals 'amplification'**
**'Not full rank'** $\equiv r < \min(m, n)$
When $r = \min(m, n)$, (i.e. full rank), then
$\sigma_{\min} := \sigma_r$ indicates 'nearness to losing rank'

**$\sigma_i$ are called the singular values.**
**(After ordering), $\sigma_i$ are unique.**
**$U$ and $V$ are not unique.**
**$\sigma_1 =: \sigma_{\max}$, the largest singular value equals 'amplification'**
**'Not full rank' $\equiv$** $r < \min(m, n)$
When $r = \min(m, n)$, (i.e. full rank), then $\sigma_{\min} := \sigma_r$ indicates 'nearness to losing rank'

$\sigma_i$ are called the singular values.
(After ordering), $\sigma_i$ are unique.
$U$ and $V$ are not unique.
$\sigma_1 =: \sigma_{\max}$, the largest singular value equals 'amplification'
'Not full rank' $\equiv r < \min(m, n)$
When $r = \min(m, n)$, (i.e. full rank), then
$\sigma_{\min} := \sigma_r$ indicates 'nearness to losing rank'

# *Nonsingular matrix*

**Generally used for square matrices only.**
**$A \in \mathbb{R}^{n \times n}$ is called nonsingular if $\det(A) \neq 0$.**
Singular $\equiv \det(A) = 0$

- Singular or nonsingular is 'layman/naive' question
- Do you live near a railway line? Yes/No
- Actually: answer is not Yes/No, but depends on 'near'
- Correct question: Do you live within 1km of a railway line?
- 50 km $\equiv$ near, for a village,
  0.1 km $\equiv$ near, for a "metro station"

Singular matrices can be made nonsingular by (as little as) 0.0000001 amount of perturbation!!

# *Nonsingular matrix*

**Generally used for square matrices only.**
**$A \in \mathbb{R}^{n\times n}$ is called nonsingular if $\det(A) \neq 0$.**
**Singular $\equiv \det(A) = 0$**

- **Singular or nonsingular is 'layman/naive' question**
- **Do you live near a railway line?** Yes/No
- Actually: answer is not Yes/No, but depends on 'near'
- Correct question: Do you live within 1km of a railway line?
- 50 km $\equiv$ near, for a village,
  0.1 km $\equiv$ near, for a "metro station"

Singular matrices can be made nonsingular by (as little as) 0.0000001 amount of perturbation!!

# Nonsingular matrix

Generally used for square matrices only.
$A \in \mathbb{R}^{n \times n}$ is called nonsingular if $\det(A) \neq 0$.
Singular $\equiv \det(A) = 0$

- Singular or nonsingular is 'layman/naive' question
- Do you live near a railway line? Yes/No
- Actually: answer is not Yes/No, but depends on 'near'
- Correct question: Do you live within 1km of a railway line?
- 50 km $\equiv$ near, for a village,
  0.1 km $\equiv$ near, for a "metro station"

Singular matrices can be made nonsingular by (as little as) 0.0000001 amount of perturbation!!

# *Nonsingular matrix*

**Generally used for square matrices only.**
**$A \in \mathbb{R}^{n\times n}$ is called nonsingular if $\det(A) \neq 0$.**
**Singular $\equiv \det(A) = 0$**

- **Singular or nonsingular is 'layman/naive' question**
- **Do you live near a railway line? Yes/No**
- **Actually: answer is not Yes/No, but depends on 'near'**
- **Correct question: Do you live within 1km of a railway line?**
- **50 km $\equiv$ near, for a village,**
  **0.1 km $\equiv$ near, for a "metro station"**

**Singular matrices can be made nonsingular by (as little as) 0.0000001 amount of perturbation!!**

# *Nonsingular matrix*

Generally used for square matrices only.
$A \in \mathbb{R}^{n \times n}$ is called nonsingular if $\det(A) \neq 0$.
Singular $\equiv \det(A) = 0$

- Singular or nonsingular is 'layman/naive' question
- Do you live near a railway line? Yes/No
- Actually: answer is not Yes/No, but depends on 'near'
- Correct question: Do you live within 1km of a railway line?
- 50 km $\equiv$ near, for a village,
  0.1 km $\equiv$ near, for a "metro station"

Singular matrices can be made nonsingular by (as little as) 0.0000001 amount of perturbation!!

# *Nonsingular matrix*

Generally used for square matrices only.
$A \in \mathbb{R}^{n \times n}$ is called nonsingular if $\det(A) \neq 0$.
Singular $\equiv \det(A) = 0$

- Singular or nonsingular is 'layman/naive' question
- Do you live near a railway line? Yes/No
- Actually: answer is not Yes/No, but depends on 'near'
- Correct question: Do you live within 1km of a railway line?
- 50 km $\equiv$ near, for a village,
  0.1 km $\equiv$ near, for a "metro station"

Singular matrices can be made nonsingular by (as little as) 0.0000001 amount of perturbation!!

# *Nonsingular matrix*

Generally used for square matrices only.
$A \in \mathbb{R}^{n\times n}$ is called nonsingular if $\det(A) \neq 0$.
Singular $\equiv \det(A) = 0$

- Singular or nonsingular is 'layman/naive' question
- Do you live near a railway line? Yes/No
- Actually: answer is not Yes/No, but depends on 'near'
- Correct question: Do you live within 1km of a railway line?
- 50 km $\equiv$ near, for a village,
  0.1 km $\equiv$ near, for a "metro station"

Singular matrices can be made nonsingular by (as little as) 0.0000001 amount of perturbation!!

# *Nonsingular matrix*

**Generally used for square matrices only.**
**$A \in \mathbb{R}^{n \times n}$ is called nonsingular if $\det(A) \neq 0$.**
**Singular $\equiv \det(A) = 0$**

- **Singular or nonsingular is 'layman/naive' question**
- **Do you live near a railway line? Yes/No**
- **Actually: answer is not Yes/No, but depends on 'near'**
- **Correct question: Do you live within 1km of a railway line?**
- **50 km $\equiv$ near, for a village,**
  **0.1 km $\equiv$ near, for a "metro station"**

**Singular matrices can be made nonsingular by (as little as) 0.0000001 amount of perturbation!!**

# *Relative distance to singular matrices*

**Consider square matrix $A \in \mathbb{R}^{n \times n}$. Following are equivalent.**

- **$A$ is nonsingular**
- **Rank $A$ is full (i.e., $r = m = n$)**
- **$\sigma_{\min} > 0$**
- **$A$ is invertible (i.e. $A^{-1}$ exists and is unique.)**
- **For each $b \in \mathbb{R}^n$, there exists an $x \in \mathbb{R}^n$ such that $Ax = b$.**
- **'Condition number of $A$' (denoted by $\kappa(A)$) is finite. ('kappa' of $A$)**

$$\kappa(A) := \frac{\sigma_1}{\sigma_n}$$

# *SVD and nearness*

**Suppose $A \in \mathbb{R}^{n \times n}$ and $\sigma_n > 0$.**
**What minimum perturbation $\Delta A$ would make $(A + \Delta A)$ singular?**
How to measure 'perturbation'? 'size' of a perturbation needed to say 'minimum'
SVD helps answer this: in the 2-norm

# *SVD and nearness*

**Suppose $A \in \mathbb{R}^{n \times n}$ and $\sigma_n > 0$.**
**What minimum perturbation $\Delta A$ would make $(A + \Delta A)$ singular?**
**How to measure 'perturbation'? 'size' of a perturbation needed to say 'minimum'**
**SVD helps answer this: in the 2-norm**

# *Norm*

**For $x \in \mathbb{R}^n$, $x = (x_1, x_2, \ldots, x_n)$, the '2-norm' is defined as $\|x\|_2 := \sqrt{x_1^2 + x_2^2 + \cdots x_n^2}$**
**Also called 'Euclidean' norm**
**Norm is a notion of 'distance' (= 'metric')**
**2-norm: most common notion of distance. There are other useful/convenient norms also.**
**$\|x\|_1$ and $\|x\|_\infty$ and $\|x\|_p$ (with $p \geqslant 1$)**
From now on: default: 2-norm: i.e. $\|x\|$ means $\|x\|_2$

- $$\sigma_{\max} = \sup_{x \in \mathbb{R}^n, x \neq 0} \frac{\|Ax\|}{\|x\|}$$
- Ratio (gain) of all possible vectors
- Actually only directions being compared
- Denominator=0 is ruled out

# Norm

For $x \in \mathbb{R}^n$, $x = (x_1, x_2, \ldots, x_n)$, the '2-norm' is defined as $\|x\|_2 := \sqrt{x_1^2 + x_2^2 + \cdots x_n^2}$
Also called 'Euclidean' norm
Norm is a notion of 'distance' (= 'metric')
2-norm: most common notion of distance. There are other useful/convenient norms also.
$\|x\|_1$ and $\|x\|_\infty$ and $\|x\|_p$ (with $p \geqslant 1$)
From now on: default: 2-norm: i.e. $\|x\|$ means $\|x\|_2$

- $$\sigma_{\max} = \sup_{x \in \mathbb{R}^n, x \neq 0} \frac{\|Ax\|}{\|x\|}$$
- Ratio (gain) of all possible vectors
- Actually only directions being compared
- Denominator=0 is ruled out

# *Matrix norms*

**For vectors $x \in \mathbb{R}^n$, $\|x\|$ indicates size/length.**
**Verrrrry small vector, verrrry small perturbation (vector), etc.**
How to put 'size' to matrices? Small matrix?

*Induced 2-norm*

$$\|A\|_2 := \sup_{x \in \mathbb{R}^n, x \neq 0} \frac{\|Ax\|}{\|x\|}$$

$\|A\|_2 = \sigma_{\max}(A)$

'induced' :≡ vector norm was used to define matrix norm
Induced 2-norm of matrix is its maximum 'amplification'
Induced 2-norm of matrix $A$ is exactly $\sigma_{\max}(A)$

# Matrix norms

For vectors $x \in \mathbb{R}^n$, $\|x\|$ indicates size/length.
Verrrrry small vector, verrrry small perturbation (vector), etc.
How to put 'size' to matrices? Small matrix?

*Induced 2-norm*

$$\|A\|_2 := \sup_{x \in \mathbb{R}^n, x \neq 0} \frac{\|Ax\|}{\|x\|}$$

$\|A\|_2 = \sigma_{\max}(A)$

'induced' :≡ vector norm was used to define matrix norm
Induced 2-norm of matrix is its maximum 'amplification'
Induced 2-norm of matrix $A$ is exactly $\sigma_{\max}(A)$

# A minimum perturbation

- Any matrix $B$ such that $\|B\| < \sigma_n$ ensures $(A + B)$ is nonsingular
- There exists (a carefully chosen) $\Delta A$ such that $\|\Delta A\| = \sigma_n$ such that $(A + \Delta A)$ is singular
- $\Delta A$ is nonunique
- In fact,

$$\frac{1}{\kappa(A)} = \min_{\det(A+\Delta A=0)} \left\{ \frac{\|\Delta A\|_2}{\|A\|_2} \right\}$$

- Ill-conditioned $\equiv \kappa(A) \gg 1$
- Well-conditioned $\equiv \kappa(A) \approx 1$
  (Note: by definition, $\kappa(A) \geqslant 1$)

# QR factorization

**Two/three variants (depending on permutation matrices)**
**Suppose $A \in \mathbb{R}^{n\times n}$. Its QR factorization is defined as $A =: QR$, with $Q$ orthogonal and $R$ upper-triangular**

- **$Q$ and $R$ are not unique.**
- **$A$ is nonsingular $\Leftrightarrow$ $R$ is nonsingular**
- **If $A$ is nonsingular, and $R$ has all diagonal elements positive, then $Q$ and $R$ are unique**

**$QR$ factorization linked to Gram-Schmidt orthogonalization**
**Permutation matrices: help in 'stability' of calculations**

# Stability

Forward and backward stability

For example, when solving for $x$ in $Ax = b$

We like: small changes in $A$ and $b$ cause small changes in $x$ ($\hat{x}$). Forward stability $x$ and $\hat{x}$ are 'close'.

Instead: Solved $\hat{x}$ is EXACT solution for a 'close by' $\hat{A}$ and $\hat{b}$: backward stability

*Algorithm stability*

Algorithm is called 'stable' if it ensures either forward or backward stability

Proofs of backward or forward stability of an algorithm are usually hard.

Even if one algorithm is bad, another good one might exist.

However, sometimes, the problem itself is

# *Stability*

Forward and backward stability

For example, when solving for $x$ in $Ax = b$

We like: small changes in $A$ and $b$ cause small changes in $x$ ($\hat{x}$). Forward stability $x$ and $\hat{x}$ are 'close'.

Instead: Solved $\hat{x}$ is EXACT solution for a 'close by' $\hat{A}$ and $\hat{b}$: backward stability

### *Algorithm stability*

Algorithm is called 'stable' if it ensures either forward or backward stability

Proofs of backward or forward stability of an algorithm are usually hard.

Even if one algorithm is bad, another good one might exist.

However, sometimes, the problem itself is

# *Stability*

Forward and backward stability

For example, when solving for $x$ in $Ax = b$

We like: small changes in $A$ and $b$ cause small changes in $x$ ($\hat{x}$). Forward stability $x$ and $\hat{x}$ are 'close'.

Instead: Solved $\hat{x}$ is EXACT solution for a 'close by' $\hat{A}$ and $\hat{b}$: backward stability

*Algorithm stability*

Algorithm is called 'stable' if it ensures either forward or backward stability

Proofs of backward or forward stability of an algorithm are usually hard.

Even if one algorithm is bad, another good one might exist.

However, sometimes, the problem itself is