Lecture 18: Introduction to Singular Value Decompositions

July 14, 2025

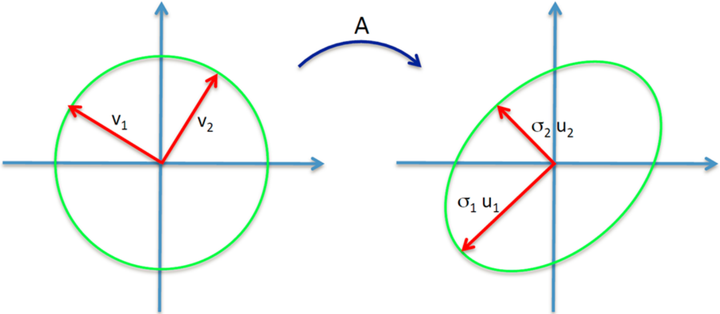
Outline

Introduction

- This lecture introduces the final decomposition called the **singular value decomposition**.
- Lecture 4 of Trefethen & Bau provides more detail, see `https://people.maths.ox.ac.uk/trefethen/text.html`.

Geometric Motivation: $AV = U\Sigma$

- The **image** of the unit hypersphere S in $\mathbb{R}^n$ under any $m \times n$ matrix transformation A is a **hyperellipse** in $\mathbb{R}^m$.
- Figure 1 shows the geometric interpretation of this transformation in $\mathbb{R}^2$.
- Both the hypersphere and hyperellipse are in $\mathbb{R}^2$ in this example.
- However, the dimensions can be any n and m, not necessarily $n = m$.



Geometric Motivation: $AV = U\Sigma$

- The factors by which the hypersphere is scaled in each of the principal semi-axes of the hyperellipse are called the **singular values** of A.
- The n singular values are denoted σ_i.
- By convention we will order them such that

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0.$$

- Notice that all the singular values are non-negative.

Geometric Motivation: $AV = U\Sigma$

- The n **left singular vectors**, u_i, of A are the unit vectors in the directions of the principal semi-axes of the ellipse.
- The n **right singular vectors**, v_i, are the unit vectors in S such that

$$Av_i = \sigma_i u_i.$$

- In other words, v_i's are the **pre-image** of u_i's under the transformation A.

Geometric Motivation: $AV = U\Sigma$ - Matrix Form

- We can write the above equation

$$Av_i = \sigma_i u_i, \qquad \text{for } i = 1, \ldots, n,$$

  in matrix form to define the **reduced SVD**.

- Pictorially, we have

$$\underbrace{\begin{bmatrix} \\ A \\ \\ \end{bmatrix}}_{A,\ m\times n} \underbrace{\begin{bmatrix} \\ v_1 & v_2 & \cdots & v_n \\ \\ \end{bmatrix}}_{V,\ n\times n}$$

$$= \underbrace{\begin{bmatrix} \\ u_1 & u_2 & \cdots & u_n \\ \\ \end{bmatrix}}_{\hat{U},\ m\times n} \underbrace{\begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{bmatrix}}_{\hat{\Sigma},\ n\times n}$$

Geometric Motivation: $AV = U\Sigma$ - Matrix Form

- The matrix $\hat{\Sigma}$ is a diagonal matrix, with the singular values of A on its diagonal.
- The matrices $\hat{U}$ and V have orthonormal columns (each is the matrix of a rotation, hence it is orthogonal).
- Note that the hat notation indicates **reduced** or **economy-sized** SVD

$$AV = \hat{U}\hat{\Sigma}.$$

- Since V is orthogonal, if we multiply by V^T on the right, we can equivalently write it as

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

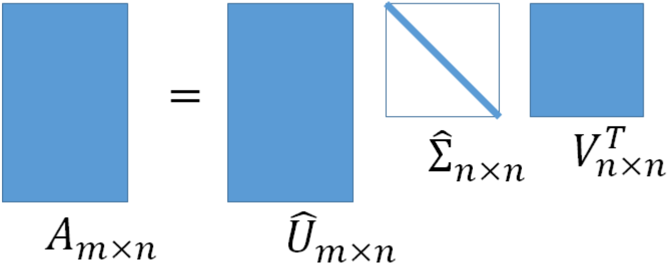
Geometric Motivation: $AV = U\Sigma$ - Matrix Form

Remarks:

1. We call the v_is **right singular vectors**, and the u_is **left singular vectors**, because of their positions in the defining equation
$$A = \underbrace{\hat{U}}_{left} \hat{\Sigma} \underbrace{V^T}_{right}.$$

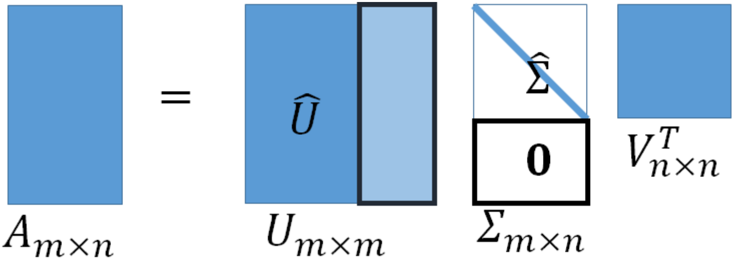
Geometric Motivation: $AV = U\Sigma$ - Matrix Form

The figure below shows this reduced SVD of A pictorially.



Geometric Motivation: $AV = U\Sigma$ - Matrix Form

- The **full SVD** is constructed in a similar way to how the full QR factorization was created from the reduced QR factorization.
- We can define a full SVD by adding $m - n$ more orthonormal columns to $\hat{U}$ to give a square, **orthogonal** U.
- Then we must also add extra empty rows to $\hat{\Sigma}$ to construct Σ.
- That is, replace $\hat{U} \to U$ and $\hat{\Sigma} \to \Sigma$ as shown in the figure below.



Geometric Motivation: $AV = U\Sigma$ - Matrix Form

- Every matrix $A \in \mathbb{R}^{m \times n}$ has a singular value decomposition.
  - We will prove this in the next lecture.
  - Furthermore, the singular values are uniquely determined.
- Also, if A is square and σ_j are distinct, then the left and right singular vectors are unique **(up to signs)**.

Geometric Motivation: $AV = U\Sigma$ - Comparison with Eigendecomposition

- The SVD is similar to the eigendecomposition we have seen previously.
- Consider the SVD vs the eigendecomposition

$$A = U\Sigma V^T \quad \text{vs} \quad A = X\Lambda X^{-1}.$$

Geometric Motivation: $AV = U\Sigma$ - Comparison with Eigendecomposition

- Both decompositions act to diagonalize a matrix.
- The SVD uses two bases: U and V, the left and right singular vectors.
- The eigendecomposition uses only one basis, the set of eigenvectors.
- The SVD always uses orthonormal vectors.
- The eigenvectors are not orthonormal **in general** (though for the real symmetric matrices we considered, they are).
- Finally, not all matrices have an eigendecomposition, but **all matrices** have an SVD, even rectangular matrices.

Properties of the SVD

- Next we will discuss some properties of the SVD.
- For the following theorems let $A \in \mathbb{R}^{m \times n}$ and $r = \#$ of non-zero singular values.

Theorem 1

$$rank(A) = r.$$

Proof.

Rank of a diagonal matrix is the number of non-zero diagonal entries. U and V are both of full rank, by definition. Hence $rank(A) = rank(\Sigma) = r$. □

Properties of the SVD

Theorem 2

$$\begin{aligned} range(A) &= span\{u_1, u_2, \ldots, u_r\}, \\ null(A) &= span\{v_{r+1}, \ldots, v_n\}. \end{aligned}$$

Properties of the SVD

Proof.

- We will not give a full proof of this theorem.
- Instead for the second property, we show that a vector in $\text{span}\{v_{r+1}, \ldots, v_n\}$ is in $\text{null}(A)$ (in other words, $\text{span}\{v_{r+1}, \ldots, v_n\} \subseteq \text{null}(A)$).
- Let $x \in \text{span}\{v_{r+1}, \ldots, v_n\}$ be arbitrary.
- Then

$$x = \sum_{i=r+1}^{n} w_i v_i \qquad \text{and so} \qquad Ax = \sum_{i=r+1}^{n} w_i (Av_i).$$

Properties of the SVD

- Observe that

$$\begin{aligned} Av_i &= U\Sigma V^T v_i \\ &= U\Sigma e_i, \end{aligned}$$

  with the last equality holding because V is an orthogonal matrix.
- But $\Sigma e_i = 0$ for $i \in [r+1, n]$ since the corresponding entries of Σ are zero.
- Therefore $Ax = 0$, so $x \in null(A)$.

□

Properties of the SVD

Lemma 3

If $A = U\Sigma V^T$*, then* $A^T A = V\Sigma^2 V^T$*.*

Proof.

We have

$$\begin{aligned} A^T A &= \left(U\Sigma V^T\right)^T U\Sigma V^T \\ &= V\Sigma^T \underbrace{U^T U}_{=I} \Sigma V^T \\ &\underbrace{=}_{\Sigma \text{ is diagonal, thus symmetric}} V\Sigma^2 V^T. \end{aligned}$$

□

Properties of the SVD

Lemma 4

Let A be a real $m \times n$ matrix, with $m \geq n$. Then $\|A^T\| = \|A\|$.

Proof.

See Lecture Notes. $\square$

Properties of the SVD

Theorem 5

$\|A\|_2 = \sigma_1$, *the largest singular value of A.*

Proof:

- Write an SVD of A:

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

- I claim that $\|\hat{U}\|_2 = \|V^T\|_2 = 1$.
- Since V is orthogonal, therefore $1 = \|V\|_2 \underbrace{=}_{\text{Lemma 4}} \|V^T\|_2$.

Properties of the SVD

- Since $\hat{U}$ has orthonormal columns, therefore

$$\begin{aligned}
\hat{U}^T\hat{U} &= I \\
\|\hat{U}^T\hat{U}\|_2 &= \|I\|_2 \\
\|\hat{U}^T\|_2\|\hat{U}\|_2 &= 1 \\
\|\hat{U}\|_2\|\hat{U}\|_2 &= 1, \text{ by Lemma 4} \\
\|\hat{U}\|_2^2 &= 1 \\
\|\hat{U}\|_2 &= 1, \text{ since } \|\hat{U}\|_2 \geq 0.
\end{aligned}$$

- This establishes the claim.
- By the claim, for any vector x, $\|Ax\|_2 = \|\Sigma x\|_2$.
- By the shape of Σ, the unit vector x' which maximizes $\|\Sigma x\|_2$ is $e_1 = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}^T$.
- Hence we have

$$\|A\|_2 = \|Ax'\|_2 = \|\Sigma x'\|_2 = \|\Sigma e_1\|_2 = \sigma_1.$$

□

Properties of the SVD

Theorem 6

$$\|A\|_2^2 = \lambda_{max}\left(A^T A\right).$$

Remark: Theorem 6 holds, even if A is not square.
Proof.

$$\begin{aligned}
\|A\|_2 &\underbrace{=}_{\text{Theorem 5}} \sigma_1 \\
\|A\|_2^2 &= \sigma_1^2 \\
&= \lambda_{\max}(\Sigma^2) \\
&\underbrace{=}_{\text{See below}} \lambda_{\max}\left(A^T A\right). \qquad (1)
\end{aligned}$$

Properties of the SVD

- Recall that, by Lemma 3, we have $A^T A = V\Sigma^2 V^T$.
- This is a similarity transformation, hence the eigenvalues of $A^T A$ equal the eigenvalues of Σ^2.
- This establishes the equality on line (1).

□

Properties of the SVD

Lemma 7

Keeping all of the above notation, $A^T \hat{U}_{m-n} = 0$*.*

Proof.

$$
\begin{aligned}
A &= \hat{U}\hat{\Sigma}V^T, \text{ so that} \\
A^T &\underbrace{=}_{\Sigma \text{ is diagonal}} V\hat{\Sigma}\hat{U}^T, \text{ so we can compute} \\
A^T\hat{U}_{m-n} &= \left(V\hat{\Sigma}\hat{U}^T\right)\hat{U}_{m-n} \\
&= V\hat{\Sigma}\underbrace{\left(\hat{U}^T\hat{U}_{m-n}\right)}_{=0} \\
&= 0,
\end{aligned}
$$

where $\hat{U}^T\hat{U}_{m-n} = 0$ holds because $\hat{U}_{m-n}$'s columns are orthogonal to $\hat{U}$'s columns. □

Properties of the SVD

Notation:

$$\begin{aligned} & \|A\|_F^2 \\ = \quad & \sum_{i,j} a_{ij}^2 \\ \underbrace{=}_{\text{See Lecture Notes}} \quad & \operatorname{tr}\left(A^T A\right), \end{aligned}$$

where $\|A\|_F$ is the **Frobenius norm**.
Recall that

$$\operatorname{tr}(A) = \sum_{i=1}^{n} a_{ii},$$

the sum of the diagonal entries of A.

Properties of the SVD

Theorem 8

$\|A\|_2 = \sigma_1$ *and* $\|A\|_F = \sqrt{\sigma_1^2 + \cdots + \sigma_r^2}$.

Proof. We have $\lambda_{\max}(A^T A) = \lambda_{\max}(\Sigma^2) = \sigma_1^2 \;\Rightarrow\; \|A\|_2 = \sigma_1$.

Properties of the SVD

Now for the Frobenius norm we have

$$
\begin{aligned}
& \|A\|_F^2 \\
= \quad & \mathrm{tr}(A^T A) \\
\underbrace{=}_{A^T A = V\Sigma^2 V^T} \quad & \mathrm{tr}\left(V\Sigma^2 V^T\right) \\
= \quad & \mathrm{tr}\left((V\Sigma)(V\Sigma)^T\right), \\
= \quad & \mathrm{tr}\left((V\Sigma)^T(V\Sigma)\right), \text{ trace identity } \mathrm{tr}(X^T Y) = \mathrm{tr}(XY^T), \\
= \quad & \mathrm{tr}\left(\Sigma V^T V\Sigma\right), \\
= \quad & \mathrm{tr}(\Sigma^2), \text{ by the orthogonality of } V, \\
= \quad & \sigma_1^2 + \cdots + \sigma_r^2.
\end{aligned}
$$

□

Properties of the SVD

Theorem 9

Non-zero singular values of A are the square roots of non-zero eigenvalues of AA^T or A^TA.

Proof.

A^TA and AA^T are similar to Σ^2.

1. We showed above (in Lemma 3) that $A^TA = V\Sigma^2V^T$.
2. Similarly,

$$\begin{aligned} AA^T &= U\Sigma V^T\left(U\Sigma V^T\right)^T \\ &= U\Sigma\underbrace{\left(V^TV\right)}_{I}\Sigma^TU^T \\ &= U\Sigma^2U^T, \text{ since } \Sigma \text{ is diagonal.} \end{aligned}$$

□

Properties of the SVD

Recall Notation: $\Lambda(A)$ is the set of eigenvalues of A.
New Notation: $\sigma(A)$ is the set of singular values of A.

Properties of the SVD

Theorem 10

If $A = A^T$, then $\sigma(A) = \{|\lambda| : \lambda \in \Lambda(A)\}$. In particular, if A is SPD then $\sigma(A) = \Lambda(A)$.

Proof.

Real symmetric matrices have orthogonal eigenvectors and real eigenvalues, so

$$A = Q\Lambda Q^T, \quad \text{with } Q \text{ orthogonal.}$$

Construct the SVD as

$$A = \underbrace{Q}_{U} \underbrace{|\Lambda|}_{\Sigma} \underbrace{\text{sign}(\Lambda)Q^T}_{V^T},$$

where $|\Lambda|$ and $\text{sign}(\Lambda)$ are diagonal matrices with entries $|\lambda_j|$ and $\text{sign}(\lambda_j)$, respectively. If desired one can also insert orthogonal permutation matrices to sort the σ's. □

Properties of the SVD

Theorem 11

The **condition number** *for* $A \in \mathbb{R}^{n \times n}$ *is* $\kappa_2(A) = \frac{\sigma_1}{\sigma_n}$.

Proof.

By the definition of κ and by Theorem 8, we have

$$\kappa_2(A) = \|A\|_2 \|A^{-1}\|_2 = \sigma_1 \|A^{-1}\|_2.$$

Since $A = U \Sigma V^T$, therefore $A^{-1} = V \Sigma^{-1} U^T$ is the SVD of A^{-1}. Therefore

$$\begin{aligned} \|A^{-1}\|_2 &= \frac{1}{\sigma_n} \\ \Rightarrow \kappa_2(A) &= \frac{\sigma_1}{\sigma_n}. \end{aligned}$$

□

Computing the SVD - 1st Attempt

- We first consider a naïve approach to computing the SVD.
- Since $A = U\Sigma V^T$ we showed above (in Lemma 3) that $A^T A = V\Sigma^2 V^T$, which is an eigendecomposition of $A^T A$!

Corollary 12

*The eigen***values** *of $A^T A$ are* **squares** *of the singular values of A.*

Corollary 13

*The eigen***vectors** *of $A^T A$ are the* **right singular** *vectors of A.*

Computing the SVD - 1st Attempt

This suggests a (naïve) method for computing the SVD:

1. Form $A^T A$ (it's symmetric and positive semi-definite, so its eigenvalues are real and non-negative),
2. Compute eigendecomposition of $A^T A = V \Lambda V^T$,
3. Compute $\Sigma = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{bmatrix}$, where $\sigma_i = \sqrt{\lambda_i}$ and $\Lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$,
4. Solve $U\Sigma = AV$ for orthogonal U (e.g., by QR factorization, as described on the next slide).

Computing the SVD - 1st Attempt

- Recovering U from the above algorithm involves (note, we already have Σ, A, V):
  - Multiply AV to get A',
  - QR factor $A' = QR$,
  - Identify $U = Q, \Sigma = R$.
- This ensures that $U = Q$ is properly orthogonal.
- Conveniently, $R = \Sigma$ will be diagonal.

Computing the SVD - 1st Attempt

- Unfortunately, this naïve method is inaccurate; the error satisfies
$$|\tilde{\sigma}_k - \sigma_k| = O\left(\frac{\epsilon \left\|A\right\|^2}{\sigma_k}\right),$$
which can be very bad for small singular values!
- (**Conceptually**, this is similar to how solving least squares by normal equations used $A^T A$.
- Effectively this "squares the condition number", therefore making it less accurate than QR factorization).
- In the next lecture we will discuss a better alternative for computing the SVD.

Computing the SVD - 1st Attempt - Example

We can find the SVD of $A = \begin{bmatrix} 0 & -\frac{1}{2} \\ 3 & 0 \\ 0 & 0 \end{bmatrix}$ in a few different ways.

Computing the SVD - 1st Attempt - Example

Method 1:

$$\begin{aligned} A^TA &= \begin{bmatrix} 0 & 3 & 0 \\ -\frac{1}{2} & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & -\frac{1}{2} \\ 3 & 0 \\ 0 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 9 & 0 \\ 0 & \frac{1}{4} \end{bmatrix} \\ &= V\Sigma^2 V^T \\ &= Q\Lambda Q^T \end{aligned}$$

Therefore $\lambda_1 = 9, \lambda_2 = \frac{1}{4}, v_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, v_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ since $Q = I$.

Therefore $\sigma_1 = 3, \sigma_2 = \frac{1}{2}$, so

$$\hat{\Sigma} = \begin{bmatrix} 3 & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \quad \text{and} \quad V = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Computing the SVD - 1st Attempt - Example

Then find U from $U\Sigma = AV$

$$\begin{bmatrix} u_1 & u_2 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & \frac{1}{2} \end{bmatrix} = \begin{bmatrix} 0 & -\frac{1}{2} \\ 3 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & -\frac{1}{2} \\ 3 & 0 \\ 0 & 0 \end{bmatrix}$$

Hence

$$\begin{aligned} 3u_1 &= \begin{bmatrix} 0 \\ 3 \\ 0 \end{bmatrix} \text{ therefore } u_1 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \\ \frac{1}{2}u_2 &= \begin{bmatrix} -\frac{1}{2} \\ 0 \\ 0 \end{bmatrix} \text{ therefore } u_2 = \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix} \end{aligned}$$

Thus $\hat{U} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \\ 0 & 0 \end{bmatrix}$.

Computing the SVD - 1st Attempt - Example

Method 2: Use AA^T instead, same idea.

Computing the SVD - 1st Attempt - Example

Method 3: Let's exploit intuition about SVD and the simple structure of this matrix.

By inspection, range$(A) = \text{span}\{u_1, u_2\}$ for $u_1 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ and $u_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, u_1 and u_2 are orthonormal. The lengths of the principal axes are 3 and $\frac{1}{2}$.

Computing the SVD - 1st Attempt - Example

Then by the definition of SVD

$$
\begin{aligned}
Av_1 &= \sigma_1 u_1 \\
\begin{bmatrix} 0 & -\frac{1}{2} \\ 3 & 0 \\ 0 & 0 \end{bmatrix} v_1 &= 3 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \\
\Rightarrow v_1 &= \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\
Av_2 &= \sigma_2 u_2 \\
\begin{bmatrix} 0 & -\frac{1}{2} \\ 3 & 0 \\ 0 & 0 \end{bmatrix} v_2 &= \frac{1}{2} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \\
\Rightarrow v_2 &= \begin{bmatrix} 0 \\ -1 \end{bmatrix}
\end{aligned}
$$

The details of solving both systems are on the following two slides.

$$
\begin{aligned}
& \left[\begin{array}{cc|c} 0 & -\frac{1}{2} & 0 \\ 3 & 0 & 3 \\ 0 & 0 & 0 \end{array}\right] \\
\sim \quad & \left[\begin{array}{cc|c} 3 & 0 & 3 \\ 0 & -\frac{1}{2} & 0 \\ 0 & 0 & 0 \end{array}\right] \begin{array}{l} R_1 \leftarrow R_2 \\ R_2 \leftarrow R_1 \\ \end{array} \\
\sim \quad & \left[\begin{array}{cc|c} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{array}\right] \begin{array}{l} R_1 \leftarrow \frac{1}{3} R_1 \\ R_2 \leftarrow -2 R_2 \\ \end{array} \\
\Rightarrow v_1 \quad = \quad & \begin{bmatrix} 1 \\ 0 \end{bmatrix}
\end{aligned}
$$

$$
\begin{aligned}
& \left[\begin{array}{cc|c} 0 & -\frac{1}{2} & \frac{1}{2} \\ 3 & 0 & 0 \\ 0 & 0 & 0 \end{array}\right] \\
\sim\ & \left[\begin{array}{cc|c} 3 & 0 & 0 \\ 0 & -\frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 \end{array}\right] \begin{matrix} R_1 \leftarrow R_2 \\ R_2 \leftarrow R_1 \\ \end{matrix} \\
\sim\ & \left[\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{array}\right] \begin{matrix} R_1 \leftarrow \frac{1}{3} R_1 \\ R_2 \leftarrow -2 R_2 \\ \end{matrix} \\
\Rightarrow v_2 =\ & \begin{bmatrix} 0 \\ -1 \end{bmatrix}
\end{aligned}
$$

Computing the SVD - 1st Attempt - Example

So $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ i.e. same solution, up to signs in U and V.