4C – Eigenanalysis & Dimensionality Reduction: Singular Value Decomposition & Applications

Derek Nowrouzezahrai
derek@cim.mcgill.ca

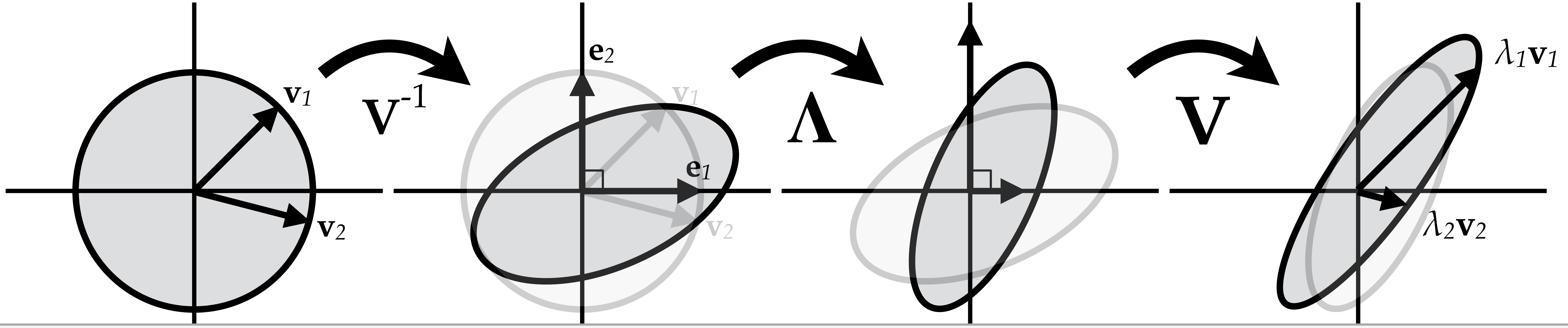
Singular Value Decomposition

Recall – Eigendecomposition

$$\mathbf{A} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{-1}$$

Transforms (with $\mathbf{V}^{-1}$) any **square** linear map ($\mathbf{A}$) into a space where it scales the canonical basis vectors (by $\boldsymbol{\Lambda}$), before transforming back to the original space (with $\mathbf{V}$)

- for arbitrary linear maps $\mathbf{A}$, the transformations $\mathbf{V}$ and $\mathbf{V}^{-1}$ may contain shears as well as rotations/reflections

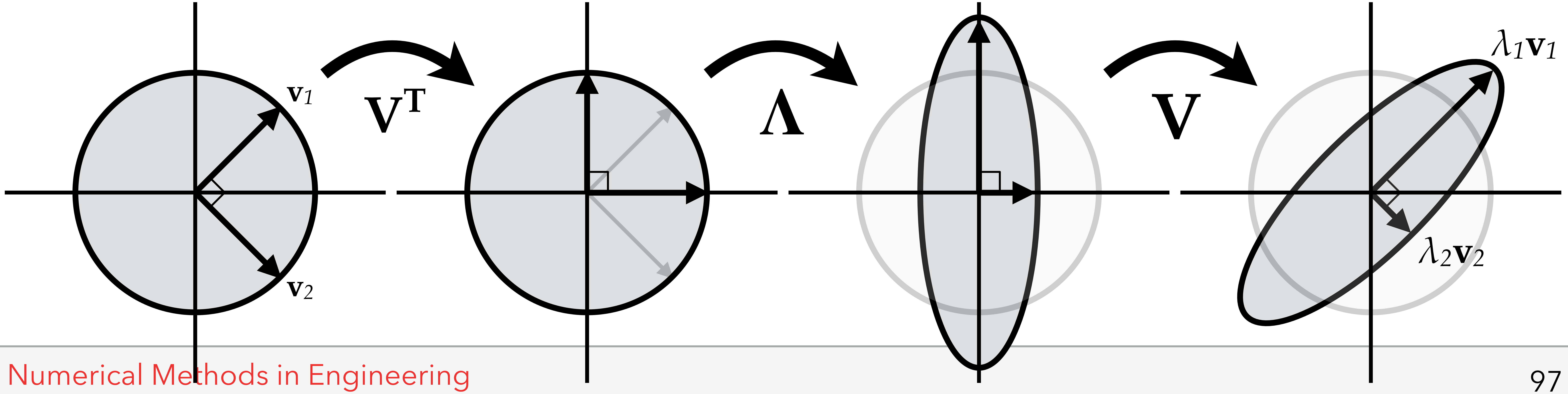


Recall – Eigendecomposition

$$\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}$$

Transforms (with $\mathbf{V}^{-1}$) any **square** linear map ($\mathbf{A}$) into a space where it scales the canonical basis vectors (by $\mathbf{\Lambda}$), before transforming back to the original space (with $\mathbf{V}$)

- for *symmetric* matrices $\mathbf{A}$, $\mathbf{V}$ and $\mathbf{V}^{-1}$ only contain a rotation



Recall – Eigendecomposition

$$\mathbf{A} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{-1}$$

The eigendecomposition is not guaranteed to exist for **an arbitrary transform** A

- $\mathbf{A}$ must be square and non-singular, among other constraints
- the power and inversion properties of eigendecompositions can be (very) useful in certain contexts

Eigenvalue problems arise naturally – and sometimes unexpectedly – in many problem domains

Singular Value Decomposition

The **Singular Value Decomposition** (SVD) is a powerful decomposition, related to the eigendecomposition but with fewer constraints and generally more applications

- we begin by deriving the decomposition,
- relating and differentiating it to the eigendecomposition, and
- discussing applications that can benefit from it

Derivation – Singular Value Decomposition

Consider a transform $\mathbf{A} \in \mathbb{R}^{m \times n}$, mapping points in $\mathbb{R}^n$ to $\mathbb{R}^m$

- this setting is already more general than the square setting

As with the eigendecomposition of square matrices, we will reason about geometric questions of the style "how does the input space in $\mathbb{R}^n$ change after applying a transformation $\mathbf{A}$?"

- how do **lengths** and **angles** change?

Derivation – Singular Value Decomposition

The length of an input vector $\mathbf{v} \in \mathbb{R}^n$ changes by a ratio of

$$R(\mathbf{v}) = \|\mathbf{A}\mathbf{v}\| \,/\, \|\mathbf{v}\|$$

and w.l.o.g. we can assume $\|\mathbf{v}\| = 1$

Since $R(\mathbf{v}) \geq 0$ we can equivalently analyze

$$[R(\mathbf{v})]^2 = \|\mathbf{A}\mathbf{v}\|^2 = \mathbf{v}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{A}\mathbf{v}$$

One interesting questions to ask:

- what are the critical points of $\mathbf{v}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{A}\mathbf{v}$ (subject to $\|\mathbf{v}\| = 1$)?
    - correspond to the vectors of maximum and minimum stretch

Derivation – Singular Value Decomposition

What are the critical points of $\mathbf{v}^\mathrm{T}\mathbf{A}^\mathrm{T}\mathbf{A}\mathbf{v}$ s.t. $\|\mathbf{v}\| = 1$?

- recall TLE: $E(\mathbf{x}) = \|\mathbf{A}\mathbf{x}\|^2/\|\mathbf{x}\|^2 \rightarrow \mathbf{A}^\mathrm{T}\mathbf{A}\mathbf{x} = [E(\mathbf{x})]\mathbf{x}$

The critical points of $\mathbf{v}^\mathrm{T}\mathbf{A}^\mathrm{T}\mathbf{A}\mathbf{v}$ s.t. $\|\mathbf{v}\| = 1$ are exactly the eigenvectors $\mathbf{v}_i$, i.e., satisfying $\mathbf{A}^\mathrm{T}\mathbf{A}\mathbf{v}_i = \lambda_i\mathbf{v}_i$

- note that the column space of $\mathbf{A}^\mathrm{T}\mathbf{A}$ spans the row space of $\mathbf{A}$
- also, as $\mathbf{A}^\mathrm{T}\mathbf{A}$ symmetric, and so its eigenvectors $\mathbf{v}_i$ form an orthogonal basis that also spans the row space of $\mathbf{A}$

Derivation – Singular Value Decomposition

Given the eigenvectors $\mathbf{v}_i$ of $\mathbf{A}^T\mathbf{A}$, define $\mathbf{u}_i = \mathbf{A}\mathbf{v}_i$

- some manipulation reveals a *parallel* eigenvalue relationship:
  $\lambda_i \mathbf{u}_i = \lambda_i \mathbf{A}\mathbf{v}_i = \mathbf{A}\,(\lambda_i \mathbf{v}_i) = \mathbf{A}\,(\mathbf{A}^T\mathbf{A}\mathbf{v}_i) = (\mathbf{A}\mathbf{A}^T)\,\mathbf{A}\mathbf{v}_i = (\mathbf{A}\mathbf{A}^T)\mathbf{u}_i$
- and so each $\mathbf{u}_i = \mathbf{A}\mathbf{v}_i$ is an eigenvector of $\mathbf{A}\mathbf{A}^T$
  - it is easy to show that $\| \mathbf{u}_i \| = \sqrt{\lambda_i}\|\mathbf{v}_i\|$ here, but we further normalize $\mathbf{u}_i$ so that $\| \mathbf{u}_i \| = 1$
- we similarly observe – as before – that the column space of $\mathbf{A}\mathbf{A}^T$ now spans the column space of $\mathbf{A}$
- $\mathbf{A}\mathbf{A}^T$ is also symmetric, and so its eigenvectors $\mathbf{u}_i$ form an orthogonal basis that similarly span the column space of $\mathbf{A}$

Derivation – Singular Value Decomposition

[<u>HALF-WAY THROUGH</u>: RECAP… AND TAKE A DEEP BREATH]

$$(\mathbf{A}^{\mathrm{T}}\mathbf{A})\,\mathbf{v}_i = \lambda_i\,\mathbf{v}_i \quad \text{with } \mathbf{v}_i \in \mathbb{R}^n$$
$$(\mathbf{A}\mathbf{A}^{\mathrm{T}})\,\mathbf{u}_i = \lambda_i\,\mathbf{u}_i \quad \text{with } \mathbf{u}_i \in \mathbb{R}^m$$

- the orthogonal $\mathbf{v}_i$ span the row space of $\mathbf{A}$
- the orthogonal $\mathbf{u}_i$ span the column space of $\mathbf{A}$
- the symmetric $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ and $\mathbf{A}\mathbf{A}^{\mathrm{T}}$ are *at least* positive semidefinite
  - with $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ and $\mathbf{A}\mathbf{A}^{\mathrm{T}}$ can't <u>both</u> have full rank (for $m \neq n$)
  - e.g., if $m > n$ then $\mathbf{A}^{\mathrm{T}}\mathbf{A} \in \mathbb{R}^{n \times n}$ is the smaller square system and there are *at most* n non-zero λ_i – and vice-versa for $m < n$

Derivation – Singular Value Decomposition

Let k be the number of strictly positive (i.e., non-zero) eigenvalues λ_i, we form matrices $\bar{\mathbf{U}} \in \mathbb{R}^{m \times k}$ and $\bar{\mathbf{V}} \in \mathbb{R}^{n \times k}$ with the eigenvectors $\mathbf{u}_i \in \mathbb{R}^m$ and $\mathbf{v}_i \in \mathbb{R}^n$ as their columns

- let's generalize the diagonalization of $\mathbf{A}$ that we saw earlier with square $\mathbf{A}$ to the non-square setting:

$$
\begin{aligned}
(\bar{\mathbf{U}}^{\mathrm{T}}\mathbf{A}\bar{\mathbf{V}})\mathbf{e}_i &= \bar{\mathbf{U}}^{\mathrm{T}}\mathbf{A}\mathbf{v}_i = (1/\lambda_i)\,\bar{\mathbf{U}}^{\mathrm{T}}\mathbf{A}(\lambda_i\mathbf{v}_i) = (1/\lambda_i)\,\bar{\mathbf{U}}^{\mathrm{T}}\mathbf{A}(\mathbf{A}^{\mathrm{T}}\mathbf{A}\mathbf{v}_i) \\
&= (1/\sqrt{\lambda_i})\,\bar{\mathbf{U}}^{\mathrm{T}}\mathbf{A}\mathbf{A}^{\mathrm{T}}\mathbf{u}_i = \sqrt{\lambda_i}\,\bar{\mathbf{U}}^{\mathrm{T}}\mathbf{u}_i \\
&= \sqrt{\lambda_i}\,\mathbf{e}_i
\end{aligned}
$$

Derivation – Singular Value Decomposition

Defining $\bar{\boldsymbol{\Sigma}} = \mathrm{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_k})$ we arrive at the diagonalization of $\mathbf{A}$ as

$$\bar{\mathbf{U}}^{\mathrm{T}} \mathbf{A} \bar{\mathbf{V}} = \bar{\boldsymbol{\Sigma}}$$

From here, we can almost re-arrange and solve for $\mathbf{A}$ in factorized form – according to $\bar{\mathbf{U}} \in \mathbb{R}^{m \times k}$ and $\bar{\mathbf{V}} \in \mathbb{R}^{n \times k}$ – however these two matrices are not square-orthogonal

- we need to add m - k columns to $\bar{\mathbf{U}}$ and n - k columns to $\bar{\mathbf{V}}$ before we can invert (i.e., transpose) them to isolate $\mathbf{A}$

Derivation – Singular Value Decomposition

We augment $\bar{\mathbf{U}} \in \mathbb{R}^{m \times k}$ with orthogonal column vectors that lie in the *null space* of $\mathbf{AA}^\mathrm{T}$, satisfying $(\mathbf{AA}^\mathrm{T})\,\mathbf{u}_i = \mathbf{0}$

- we similarly augment $\bar{\mathbf{V}}$ with vectors in the *null space* of $\mathbf{A}^\mathrm{T}\mathbf{A}$

After forming the extended $\mathbf{U} \in \mathbb{R}^{m\times m}$ and $\mathbf{V} \in \mathbb{R}^{n\times n}$ matrices, we will have $(\mathbf{U}^\mathrm{T}\mathbf{AV})\mathbf{e}_i = \mathbf{0}$ and/or $\mathbf{e}_i^\mathrm{T}\,(\mathbf{U}^\mathrm{T}\mathbf{AV}) = \mathbf{0}^\mathrm{T}$ for $i > k$

If both $\mathbf{AA}^\mathrm{T}$ and $\mathbf{A}^\mathrm{T}\mathbf{A}$ are rank deficient

If only one of $\mathbf{AA}^\mathrm{T}$ or $\mathbf{A}^\mathrm{T}\mathbf{A}$ is rank deficient

For completeness, we extend $\bar{\boldsymbol{\Sigma}}$ with 0-padding for all diagonal entries past the i^{th} diagonal entry, forming $\boldsymbol{\Sigma}$

Derivation – Singular Value Decomposition

Isolating $\mathbf{A}$ yields its singular value decomposition:

$$\mathbf{A} = \mathbf{U}\,\boldsymbol{\Sigma}\,\mathbf{V}^{\mathrm{T}}$$

- the columns of $\mathbf{U}$ are called *left singular vectors* and the columns of $\mathbf{V}$ are called the *right singular vectors*
- the columns of $\mathbf{U}$ that correspond to non-zero singular values σ_i span the column space of $\mathbf{A}$
- the columns of $\mathbf{V}$ that correspond to non-zero singular values σ_i span the row space of $\mathbf{A}$

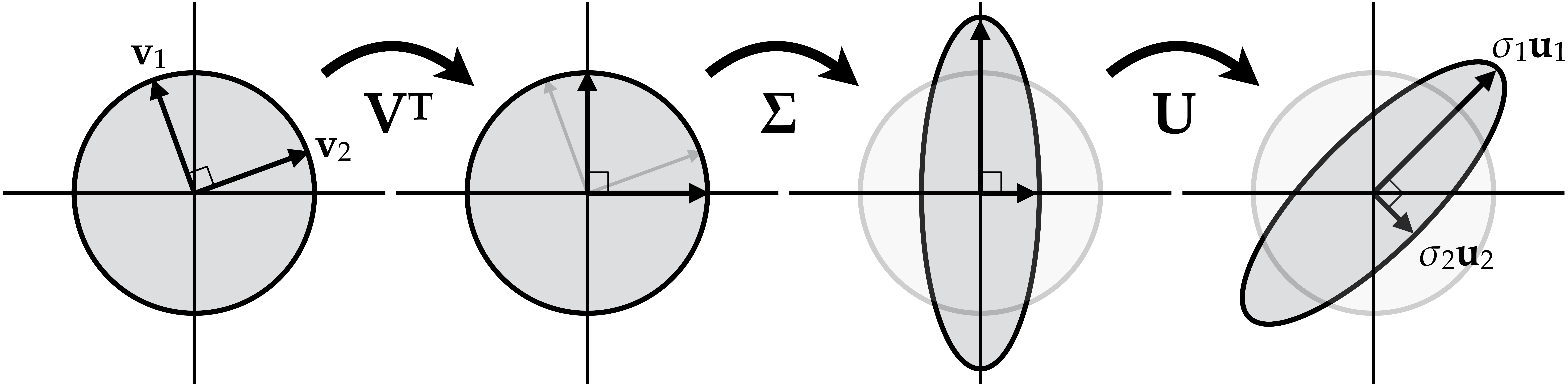
SVD for Wide vs. Tall Matrices

SVD for Wide vs. Tall Matrices

$$\underset{m \times n}{\begin{bmatrix} \\ \quad \mathbf{A} \quad \\ \\ \end{bmatrix}} = \underbrace{\underset{m \times m}{\begin{bmatrix} | & & | \\ \mathbf{u}_1 & \cdots & \mathbf{u}_m \\ | & & | \end{bmatrix}}}_{\mathbf{U}} \underbrace{\underset{m \times n}{\left[\begin{array}{ccc:c} \sigma_1 & & & \\ & \ddots & & \mathbf{0} \\ & & \sigma_k & \end{array}\right]}}_{\mathbf{\Sigma}} \underbrace{\underset{n \times n}{\begin{bmatrix} \text{---} & \mathbf{v}_1 & \text{---} \\ & \vdots & \\ \text{---} & \mathbf{v}_m & \text{---} \end{bmatrix}}}_{\mathbf{V}^{\mathrm{T}}}$$

SVD – Geometric Interpretation

As with eigendecompositions, we can visualize the SVD in terms of the geometric transforms in its composition



- note that this visualization is only representative for square $\mathbf{A}$, since the dimensionality of input and output spaces differ for for $m \neq n$ – i.e., $\mathbf{U}$ and $\mathbf{V}$ are m- and n-dimensional rotations

SVD vs. Eigendecomposition

Eigendecomposition and SVD are closely related

- indeed, we use two eigendecompositions to derive the SVD

1. Existence

- the eigendecomposition does not exist for every matrix
    - matrices must be square, among other requirements (see *the spectral theorem*), for the eigendecomposition to exist
- the SVD exists for **every** matrix (see *the fundamental theorem of linear algebra*)

SVD vs. Eigendecomposition

Eigendecomposition and SVD are closely related

- indeed, we use two eigendecompositions to derive the SVD

2. Structure

- diagonalizing transforms in an eigendecomposition – $\mathbf{V}$ and $\mathbf{V}^{-1}$ – must be inverses of each other, but may not be isometric
- eigenvalues can generally take on any complex value ($\mathbb{C}$)
- diagonalizing transformations $\mathbf{U}$ and $\mathbf{V}^T$ in the SVD are not generally related to each other, however they are isometric
- singular values are all real and non-negative

SVD vs. Eigendecomposition

Eigendecomposition and SVD are closely related

- indeed, we use two eigendecompositions to derive the SVD

3. Conditions for Equivalence

- the eigendecomposition and SVD of a matrix are only equivalent under a very restrictive setting: when $\mathbf{A}$ is *symmetric positive semidefinite*
  - recall – when $\mathbf{A}$ is symmetric and real, its eigenvalues are real and its eigenvectors are orthogonal

$$\mathbf{A} = \mathbf{U}\,\mathbf{\Sigma}\,\mathbf{V}^T \qquad \mathbf{A}^T\mathbf{A} = (\mathbf{U}\,\mathbf{\Sigma}\,\mathbf{V}^T)^T\,\mathbf{U}\,\mathbf{\Sigma}\,\mathbf{V}^T = (\mathbf{V}\,\mathbf{\Sigma}^T\,\mathbf{U}^T)\,\mathbf{U}\,\mathbf{\Sigma}\,\mathbf{V}^T = \mathbf{V}\,\mathbf{\Sigma}^2\,\mathbf{V}^T$$