Introduction to Linear Algebra

Final Review Class

Wang Tianyu[1]

Fall, 2019

Outline

Content

10' True / False
- look carefully at questions on slides

38' General Question & Calculation & Simple Justification
- focus on EVD and SVD

12' Proof

Metric

Metric, denoted as $d(x,y)$, is a generalization of the concept "distance for an arbitrary set.

Constraints on Metric:

1. $d(x,y) \geq 0$
2. $d(x,y) = 0 \leftrightarrow x = y$
3. $d(x,y) = d(y,x)$
4. $d(y,z) \leq d(x,y) + d(x,z)$

So a set $\mathcal{S}$ containing those x, y, z together with a valid metric d is called a **metric space**.

Convergence of Point Sequence

A sequence of points $\{a_n\}$ in a set $\mathcal{S}$ is said to **converge** to $a \in \mathcal{S}$ if

$$\lim_{n\to\infty} d(a, a_n) = 0$$

From the definition, it can be easily seen that this concept is associated with the choice of metric.

Cauchy Sequence

A sequence $\{a_n\}$ in a metric space $\mathcal{S}$ is said to be a **Cauchy sequence** if for any $\epsilon > 0$ there exists an N such that

$$d(a_m, a_n) < \epsilon \quad \text{whenever} \quad m, n \geq N$$

- Any convergent sequence is Cauchy sequence (Proof using triangular inequality);
- **Not** every Cauchy sequence is convergent.

Norm

Norm, denoted as $||x||$ is a generalization of the concept "length" for an arbitrary vector space.

Constraints on Norm:

- $||x|| \geq 0$
- $||ax|| = |a|||x||$
- $||x + y|| \leq ||x|| + ||y||$

So a vector space with a valid norm is called a **normed vector space**.

Special Norms

- p-norm

$$||\mathbf{v}||_p = \left(\sum_{i=1}^{n} |v_i|^p \right)^{1/p}$$

- Frobenius Norm

$$||\mathbf{A}||_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2}$$

Compatible & Sub-multiplicative

A matrix norm on $\mathbb{R}^{n\times n}$ is said to be **compatible** to another norm for $\mathbb{R}$ if

$$\forall \mathbf{A} \in \mathbb{R}^{n\times n}, \mathbf{v} \in \mathbb{R}^{n}, \quad ||\mathbf{Av}|| \leq ||\mathbf{A}|| \cdot ||\mathbf{v}||$$

A matrix norm on $\mathbb{R}^{n\times n}$ is said to be **sub-multiplicative** if

$$\forall \mathbf{A}, \mathbf{B} \in \mathbb{R}^{n\times n}, \quad ||\mathbf{AB}|| \leq ||\mathbf{A}|| \cdot ||\mathbf{B}||$$

Operator Norm

The matrix norm $||\cdot||_o : \mathbf{R}^{n\times n} \to \mathbf{R}$,

$$||\mathbf{A}||_o = \max_{\hat{\mathbf{x}}} ||\mathbf{A}\hat{\mathbf{x}}||$$

is known as the **operator norm** induced by the vector norm $||\cdot||$ on $\mathbf{R}^n$.

- $||\mathbf{A}||_1$ = max sum over some column in $\mathbf{A}$
- $||\mathbf{A}||_\infty$ = max sum over some row in $\mathbf{A}$
- $||\mathbf{A}||_2$ = max singular value in $\mathbf{A}$

Inner Product

Inner product, denoted as $\langle u, v\rangle$, is a generalization of the concept dot product for an arbitrary set.

Constraints on Inner Product:

- $\langle \mathbf{u}, \mathbf{v}\rangle = \langle \mathbf{v}, \mathbf{u}\rangle$
- $\langle \mathbf{u}+\mathbf{v}, \mathbf{w}\rangle = \langle \mathbf{u}, \mathbf{w}\rangle + \langle \mathbf{v}, \mathbf{w}\rangle$
- $\langle \alpha\mathbf{u}, \mathbf{v}\rangle = \alpha\langle \mathbf{u}, \mathbf{v}\rangle = \langle \mathbf{u}, \alpha\mathbf{v}\rangle$
- $\langle \mathbf{u}, \mathbf{u}\rangle \geq 0 \;\text{ and }\; \langle \mathbf{u}, \mathbf{u}\rangle = 0 \text{ iff } \mathbf{u} = \mathbf{0}$

So a vector space with a valid inner product is called a **inner product space**.

Redefined Terms

Basic terms for inner product space:

- **length**: $||\mathbf{v}|| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$;
- **distance**: $d(\mathbf{u}, \mathbf{v}) = ||\mathbf{u} - \mathbf{v}|| = \sqrt{\langle \mathbf{u} - \mathbf{v}, \mathbf{u} - \mathbf{v} \rangle}$;
- **orthogonal**: $\langle \mathbf{u}, \mathbf{v} \rangle = 0$;
- **angle**: $\cos\theta = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{||\mathbf{u}|| \cdot ||\mathbf{v}||}$

Based on these redefined terms, theorems like *Cauchy-Schwarz inequality*, *Triangle inequality* and *Parallelogram law* are still satisfied.

Some Notes

1. Inner product of a vector space is **NOT** unique.
2. Every inner product space is a metric space as well as a normed space.
3. $\mathcal{L}_p$-norm for $p \neq 2$ does not correspond to any inner product. In such cases, *Pythagorean law* will not hold.
4. An inner product space $\mathcal{H}$ is called a **Hilbert space** if every Cauchy sequence in $\mathcal{H}$ converges to an element of $\mathcal{H}$ with respect to the induced norm.

Orthogonal Set

A set of vectors in an inner product space is called an **orthogonal set** if any two different vectors in it are orthogonal. And if all vectors in this set are unit, then it is also called **orthonormal set**.

Parsevals Theorem

If in an inner product space, $\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n$ is an orthonormal basis, then

$$||\mathbf{v}||^2 = \sum_{i=1}^{n} c_i^2 \quad \text{where} \quad \mathbf{v} = \sum_{i=1}^{n} c_i \mathbf{u}_i$$

Orthogonal Matrix

An $n \times n$ matrix $\mathbf{Q}$ is said to be an **orthogonal matrix** if its columns form an orthonormal set.

$n \times n$ Orthogonal matrices have several properties:

- $\mathbf{Q}^{\mathrm{T}}\mathbf{Q} = \mathbf{I}$
- $\mathbf{Q}^{\mathrm{T}} = \mathbf{Q}^{-1}$
- $\langle \mathbf{Qx}, \mathbf{Qy} \rangle = \langle \mathbf{x}, \mathbf{y} \rangle$
- $||\mathbf{Qx}|| = ||\mathbf{x}||$
- $\mathbf{Q}^{\mathrm{T}}$ is also orthogonal

Projection

The **vector projection** of a vector $\mathbf{y}$ onto another vector $\mathbf{u}$ is defined as

$$\mathrm{proj}_{\mathbf{u}}\mathbf{y} = ||\mathbf{y}|| \cos\theta \frac{\mathbf{u}}{||\mathbf{u}||} = \frac{\langle \mathbf{u}, \mathbf{y} \rangle}{\langle \mathbf{u}, \mathbf{u} \rangle}\mathbf{u},$$

and the vector $(\mathbf{y} - \mathrm{proj}_{\mathbf{u}}\mathbf{y})$ is called the **vector component** of $\mathbf{y}$ orthogonal to $\mathbf{u}$.

Orthogonality between Vector & Space

A vector $\mathbf{x}$ in a inner product space $\mathcal{V}$ is orthogonal to a subspace of $\mathcal{V}$ if it it orthogonal to any vector in that space.

For a vector $\mathbf{x}$ in $\mathcal{V}$ and a subspace $\mathcal{W}$ of $\mathcal{V}$, we can **uniquely** write

$$\mathbf{x} = \mathbf{x}^{\parallel} + \mathbf{x}^{\perp}$$

where $\mathbf{x}^{\parallel}$ is in $\mathcal{W}$ and $\mathbf{x}^{\perp}$ is orthogonal to $\mathcal{W}$.

For an orthonormal basis $\mathcal{B} = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_m\}$,

$$\mathbf{x}^{\parallel} = \mathrm{proj}_{\mathcal{W}}(\mathbf{x}) = \langle \mathbf{u}_1, \mathbf{x} \rangle \mathbf{u}_1 + \langle \mathbf{u}_2, \mathbf{x} \rangle \mathbf{u}_2 + \cdots + \langle \mathbf{u}_m, \mathbf{x} \rangle \mathbf{u}_m$$

Orthogonality between Spaces

For two subspace $\mathcal{X}$, $\mathcal{Y}$ of an inner product space $\mathcal{V}$, if any vector in $\mathcal{X}$ is orthogonal to any vector in $\mathcal{Y}$, then we write

$$\mathcal{X} \perp \mathcal{Y}$$

For a subspace $\mathcal{W}$ of an inner product space $\mathcal{V}$, the set of all vectors in $\mathcal{V}$ which is orthogonal to $\mathcal{W}$ is defined as its **orthogonal complement**.

1. The orthogonal complement $\mathcal{W}^{\perp}$ of $\mathcal{W}$ is a subspace of $\mathcal{V}$.
2. $\mathcal{W}^{\perp} \cap \mathcal{W} = \{\mathbf{0}\}$.
3. $\mathrm{dim}\mathcal{W} + \mathrm{dim}\mathcal{W}^{\perp} = n$.
4. $(\mathcal{W}^{\perp})^{\perp} = \mathcal{W}$

$$\mathrm{null}(\mathbf{A}) = (\mathrm{col}(\mathbf{A}^{\mathrm{T}}))^{\perp} \quad \text{and} \quad \mathrm{null}(\mathbf{A}^{\mathrm{T}}) = (\mathrm{col}(\mathbf{A}))^{\perp}$$

Direct Sum

The direct sum of two subspaces $\mathcal{X}$, $\mathcal{Y}$ of $\mathcal{W}$ is the set of vectors in $\mathcal{W}$ which can be **uniquely** represented as sum of a vector in $\mathcal{X}$ and a vector in $\mathcal{Y}$.

If $\mathcal{S}$ is a subspace of $\mathcal{W}$, then

$$\mathcal{W} = \mathcal{S} \oplus \mathcal{S}^{\perp}$$

Gram-Schmidt QR Factorization

If $\mathbf{A}$ is an $m \times n$ matrix of rank n, then $\mathbf{A}$ can be factored into a product

$$\mathbf{A} = \mathbf{QR}$$

where $\mathbf{Q}$ is an $m \times n$ matrix with orthonormal column vectors and $\mathbf{R}$ is an upper triangular $n \times n$ matrix whose diagonal entries are all positive.

QR factorization is often used to solve the linear least squares problem and is the basis for a particular eigenvalue algorithm, the QR algorithm.

Motivation

The vector reaches the **steady-state** regardless of the initial state, that is, $\exists t_0 \geq 0 \; \forall t \geq t_0, \mathbf{A}\mathbf{w}_t = \mathbf{w}_t$ for any choice of $\mathbf{w}_0$.

In general, if a linear transformation is represented by an $n \times n$ matrix $\mathbf{A}$ and we can find a nonzero vector $\mathbf{x}$ so that $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$, for some scalar λ, then, for this transformation, $\mathbf{x}$ is a natural choice to use as a basis vector for $\mathbb{R}^n$.

For an $n \times n$ matrix $\mathbf{A}$, a scalar λ is said to be an **eigenvalue** of $\mathbf{A}$ if there exists a nonzero vector $\mathbf{x}$ such that

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$$

The vector $\mathbf{x}$ is said to be an **eigenvector** corresponding to λ.

Key Concepts

- **Eigenspace**: the subspace null$(\mathbf{A} - \lambda\mathbf{I})$
- **Characteristic Polynomial**: $p(\lambda) = \det(\mathbf{A} - \lambda\mathbf{I})$
- **Algebraic Multiplicity** & **Geometric Multiplicity**: the degree of a root in the characteristic polynomial & the dimension of the eigenspace ($\geq$)
- **Complex Vector Space**: the one in which the scalars are complex numbers ($\mathbb{R}^n$ is not a subspace of $\mathbb{C}^n$)

If λ is an eigenvalue of a **real** $n \times n$ matrix $\mathbf{A}$, and if x is a eigenvector belonging to λ, then $\bar{\lambda}$ is also an eigenvalue of $\mathbf{A}$, and $\bar{\mathbf{x}}$ is a corresponding eigenvector.

Some Properties of Eigens

Assume for a $n \times n$ square matrix $\mathbf{A}$,

1. Eigenvectors with distinct eigenvalues are linearly independent.
2. The number of non-zero eigenvalues is equal to the rank.
3. If λ is an eigenvalue of $\mathbf{A}$, then
   - $\alpha\lambda, \alpha \neq 0$ is an eigenvalue of $\alpha\mathbf{A}$;
   - $\lambda^s, s \in \mathbb{Z}^+$ is an eigenvalue of $\mathbf{A}^s$;
   - $p(\lambda)$ is an eigenvalue of $p(\mathbf{A})$, where $p(\cdot)$ is a polynomial function;
   - λ is an eigenvalue of transpose matrix $\mathbf{A}^{\mathrm{T}}$;
   - if $\mathbf{A}$ is invertible, λ^{-1} is an eigenvalue of $\mathbf{A}^{-1}$.
4. if $\mathbf{A}$ is diagonal or triangular, the eigenvalues are the diagonal elements.

Similarity Transformation

For two $n \times n$ matrices $\mathbf{A}$ and $\mathbf{B}$, if there exists a nonsingular matrix $\mathbf{P}$ such that $\mathbf{B} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}$, then $\mathbf{B}$ is **similar** to $\mathbf{A}$. $\mathbf{A}$ and $\mathbf{B}$ are similar matrices then.

$$
\begin{tikzcd}
{[\mathbf{v}]_{\mathcal{A}}} \arrow[r, "\mathbf{A}"] & {[L(\mathbf{v})]_{\mathcal{A}}} \arrow[d, "\mathbf{P}^{-1}"] \\
{[\mathbf{v}]_{\mathcal{B}}} \arrow[u, "\mathbf{P}"] \arrow[r, "\mathbf{B}"'] & {[L(\mathbf{v})]_{\mathcal{B}}}
\end{tikzcd}
$$

Similarity Invariants

Some properties preserved by a similarity transformation:

1. the same eigenvalues
2. the same determinant
3. the same trace
4. the same rank
5. the same nullity
6. the same invertibility
7. the same eigenspace dimension corresponding to some λ

Diagonal

Now we can approach a special case of the similarity transformation, namely diagonalization.

A matrix $\mathbf{A}_{n\times n}$ is **diagonalizable** if there is an invertible matrix $\mathbf{P}$ and a diagonal matrix $\mathbf{D}$ such that $\mathbf{D} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}$.

1. Diagonalizable $\leftrightarrow$ n linearly independent eigenvectors
2. Diagonalizable $\rightarrow$ columns of $\mathbf{P}$ are the n eigenvectors; diagonal elements of $\mathbf{D}$ are corresponding eigenvalues
3. Diagonalizing matrix $\mathbf{P}$ is not unique.

In this sense, $\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-\mathbf{1}}$ and $\mathbf{A}^k = \mathbf{P}\mathbf{D}^k\mathbf{P}^{-1}$. We can easily calculate $e^{\mathbf{A}} = \mathbf{I} + \mathbf{A} + \frac{1}{2!}\mathbf{A}^2 + \frac{1}{3!}\mathbf{A}^3 + \cdots = \mathbf{P}e^{\mathbf{D}}\mathbf{P}^{-1}$ now!

Real Matrix with Complex Eigenvalues

1. Consider the real matrix $\mathbf{C} = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$ where $b \geq 0$. If a and b are not both zero, then $\mathbf{C} = \begin{bmatrix} r & 0 \\ 0 & r \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ where $\theta = \arg(\lambda_1)$, $r = \mathrm{mod}(\lambda_1)$.
2. For a real 2×2 matrix $\mathbf{C}$ with complex eigenvalues $\lambda = a \pm b\mathrm{i}, b > 0$, $\mathbf{P} = [\mathrm{Re}(\mathbf{x})\ \mathrm{Im}(\mathbf{x})]$ is invertible and $\mathbf{A} = \mathbf{P} \begin{bmatrix} a & -b \\ b & a \end{bmatrix} \mathbf{P}^{-1}$ where $\mathbf{x}$ is an eigenvector corresponding to $a - b\mathrm{i}$.

Combining the two facts above gives us, for every real 2×2 matrix $\mathbf{C}$ with complex eigenvalues, $\mathbf{C} = \mathbf{P}\mathbf{S}\mathbf{R}_\theta\mathbf{P}^{-1}$.

Complex Notation

For $\mathbf{u} = [u_1, \cdots, u_n]^{\mathrm{T}}$ and $\mathbf{v} = [v_1, \cdots, v_n]^{\mathrm{T}}$ in $\mathbb{C}^n$,

- $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^{\mathrm{T}} \bar{\mathbf{v}} = \bar{\mathbf{v}}^{\mathrm{T}} \mathbf{u}$
- $||\mathbf{v}|| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle} = \sqrt{|v_1|^2 + \cdots + |v_n|^2}$, $|v_i| = \mathrm{mod}(v_i)$

If $\{\mathbf{u}_1, \cdots, \mathbf{u}_n\}$ is an orthonormal basis and $\mathbf{z} = \sum_{i=1}^{n} \alpha_i \mathbf{u}_i$, then $\alpha_i = \langle \mathbf{z}, \mathbf{u}_i \rangle$ and $||\mathbf{z}||^2 = \sum_{i=1}^{n} \alpha_i \bar{\alpha}_i$.

Hermitian Matrix - Symmetric Matrix

A matrix $\mathbf{A}$ is said to be **Hermitian** if $\mathbf{A} = \mathbf{A}^{\mathrm{H}}$. Hermitian matrices can be viewed as the complex analogue of symmetric real matrices. Some nice properties hold for Hermitian matrices,

1. The eigenvalues are all real. In other sense, the main diagonal entries are all real.
2. The eigenvectors belonging to distinct eigenvalues are orthogonal.
3. $\mathbf{A} + \mathbf{A}^{\mathrm{H}}$, $\mathbf{A}\mathbf{A}^{\mathrm{H}}$ and $\mathbf{A}^{\mathrm{H}}\mathbf{A}$ are all Hermitian.
4. $\mathbf{A}^k$ is Hermitian for all $k = 1, 2, \cdots$; if $\mathbf{A}$ is invertible, then $\mathbf{A}^{-1}$ is also Hermitian.
5. If $\mathbf{A}$, $\mathbf{B}$ are Hermitian, then $\alpha\mathbf{A} + \beta\mathbf{B}$ is Hermitian for all real scalars α, β.

Unitary Matrix - Orthogonal Matrix

An $n \times n$ matrix $\mathbf{U}$ is said to be **unitary** if its columns are orthonormal in $\mathbb{C}^n$, *i.e.*, $\mathbf{U}^{\mathrm{H}}\mathbf{U} = \mathbf{I}$. It follows that $\mathbf{U}^{-1} = \mathbf{U}^{\mathrm{H}}$.

If the eigenvalues of a Hermitian matrix $\mathbf{A}$ are **distinct**, then there exists a unitary matrix $\mathbf{U}$ such that $\mathbf{D} = \mathbf{U}^{-1}\mathbf{A}\mathbf{U} = \mathbf{U}^{\mathrm{H}}\mathbf{A}\mathbf{U}$.

For each $n \times n$ matrix $\mathbf{A}$, there exists a unitary matrix $\mathbf{U}$ such that $\mathbf{R} = \mathbf{U}^{\mathrm{H}}\mathbf{A}\mathbf{U}$ is upper triangular. The factorization $\mathbf{A} = \mathbf{U}\mathbf{R}\mathbf{U}^{\mathrm{H}}$ is **Schur Decomposition** of $\mathbf{A}$.

Hermitian matrix always has a unitary matrix $\mathbf{U}$ that diagonalizes it.

Normal

If $\mathbf{A}$ can be factored into $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{U}^{\mathrm{H}}$ where $\mathbf{U}$ is unitary and $\mathbf{D}$ is diagonal, then $\mathbf{A}\mathbf{A}^{\mathrm{H}} = \mathbf{A}^{\mathrm{H}}\mathbf{A}$. A matrix that satisfies this is called **normal** (**if and only if**).

1. A normal matrix is **Hermitian** if and only if all its eigenvalues are real.
2. If $\mathbf{A}$ and $\mathbf{B}$ are normal with $\mathbf{AB} = \mathbf{BA}$, then both $\mathbf{AB}$ and $\mathbf{A} + \mathbf{B}$ are also normal. Moreover, there exists a unitary matrix $\mathbf{U}$ such that $\mathbf{UAU}^H$ and $\mathbf{UBU}^H$, which is called **simultaneously diagonalizable**.

Background

For a $m \times n$ matrix,

1. $\mathbf{A}$ and $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ have the same null space.
2. $\mathbf{A}$ and $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ have the same row space.
3. $\mathbf{A}^{\mathrm{T}}$ and $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ have the same column space.
4. $\mathbf{A}$ and $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ have the same rank

If $\mathbf{A}$ is an $m \times n$ matrix, then

1. $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ is orthogonally diagonalizable.
2. The eigenvalues of $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ are nonnegative.

Square roots of eigenvalues of $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ are called the **singular values** of $\mathbf{A}$.

Singular Value Decomposition

If $\mathbf{A}$ is an $m \times n$ of rank k, then $\mathbf{A}$ can be factored as

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

$$= \left[\begin{array}{ccc|ccc} \mathbf{u}_1 & \cdots & \mathbf{u}_k & \mathbf{u}_{k+1} & \cdots & \mathbf{u}_m \end{array}\right] \left[\begin{array}{ccc|c} \sigma_1 & \cdots & 0 & \\ \vdots & \ddots & \vdots & \mathbf{0}_{k\times(n-k)} \\ 0 & \cdots & \sigma_k & \\ \hline & \mathbf{0}_{(m-k)\times k} & & \mathbf{0}_{(m-k)\times(n-k)} \end{array}\right] \left[\begin{array}{c} \mathbf{v}_1^{\mathrm{T}} \\ \vdots \\ \mathbf{v}_k^{\mathrm{T}} \\ \hline \mathbf{v}_{k+1}^{\mathrm{T}} \\ \vdots \\ \mathbf{v}_n^{\mathrm{T}} \end{array}\right]$$

in which $\mathbf{U}$, $\boldsymbol{\Sigma}$ and $\mathbf{V}$ are matrices of size $m \times m$, $m \times n$, and $n \times n$, respectively.

Singular Value Decomposition

- The nonzero diagonal entries of $\mathbf{\Sigma}$ are nonzero singular values of $\mathbf{A}$,

$$\sigma_1 = \sqrt{\lambda_1}, \quad \sigma_2 = \sqrt{\lambda_2}, \ldots, \quad \sigma_k = \sqrt{\lambda_k}$$

where λ_i are the nonzero eigenvalues of $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ in order of decreasing size, so

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_k > 0$$

- Vector $\mathbf{u}_j$ is defined as the normalized image of $\mathbf{v}_j$ under $\mathbf{A}$

$$\mathbf{u}_j = \frac{\mathbf{A}\mathbf{v}_j}{\|\mathbf{A}\mathbf{v}_j\|} = \frac{1}{\sigma_j}\mathbf{A}\mathbf{v}_j \qquad \text{for } j = 1, 2, \ldots, k$$

- The set $\{\mathbf{u}_{k+1}, \ldots, \mathbf{u}_m\}$ is an extension set of $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ so that

$$\{\mathbf{u}_1, \ldots \mathbf{u}_k, \mathbf{u}_{k+1}, \ldots, \mathbf{u}_m\}$$

forms an orthonormal basis for $\mathbb{R}^m$.

Singular Value Decomposition

For every linear map $T : \mathbb{K}^n \to \mathbb{K}^m$ we can find orthonormal bases of them such that T maps the i-th basis vector of $\mathbb{K}^n$ to a non-negative multiple of the i-th basis vector of $\mathbb{K}^m$, and sends the left-over basis vectors to zero. With respect to these bases, the map T is represented by a diagonal matrix with non-negative real diagonal entries.

Reduced Singular Value Decomposition

For matrix $\mathbf{A}$ whose rank $k < n$, it is often sufficient to consider $\mathbf{A} = \mathbf{U}_k \mathbf{\Sigma}_k \mathbf{V}_k^{\mathrm{T}}$, where $\mathbf{U}_k = [\mathbf{u}_1, \cdots, \mathbf{u}_k]$, $\mathbf{V}_k = [\mathbf{v}_1, \cdots, \mathbf{v}_k]$, $\mathbf{\Sigma}_k$ is the diagonal matrix containing nonzero singular values of $\mathbf{A}$.

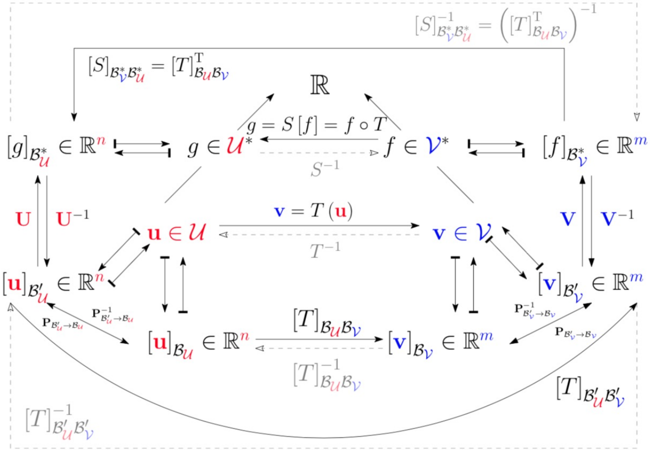
Instead of storing the whole matrix $\mathbf{A}$, we store the σ_j's, $\mathbf{u}_j$'s and $\mathbf{v}_j$'s up to rank k.

With SVD, we can diagonalize a rectangular system (not restricted to square system as EVD does),

$$\mathbf{A}\mathbf{x} = \mathbf{b}, \ \mathbf{A} \in \mathbb{R}^{m \times n}$$

$$\mathbf{\Sigma}(\mathbf{V}^{\mathrm{T}}\mathbf{x}) = (\mathbf{U}^{\mathrm{T}}\mathbf{b})$$

General Picture



Generalized Inverse

For an $m \times n$ matrix $\mathbf{A}$ of rank k, the generalized inverse $\mathbf{A}^+$ is defined as

$$\mathbf{A}^+ = \mathbf{V}\boldsymbol{\Sigma}^+\mathbf{U}^\mathrm{T} = \mathbf{V}_k\boldsymbol{\Sigma}_k^{-1}\mathbf{U}_k^\mathrm{T}$$

If $\mathbf{A}$ has linearly independent columns, then $\mathbf{A}^\mathrm{T}\mathbf{A}$ is invertible and,

$$\mathbf{A}^+ = (\mathbf{A}^\mathrm{T}\mathbf{A})^{-1}\mathbf{A}^\mathrm{T}$$

Motivation

A quadratic form in n variables is a function $f : \mathbb{R}^n \to \mathbb{R}$ such that $f(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}$ where $\mathbf{A}$ is a symmetric $n \times n$ matrix.

To find the matrix associated with some quadratic form:

1. coefficients of squared terms correspond to the diagonal elements;
2. coefficients of cross-product terms are averaged among elements at corresponding indices.

For every quadratic form $f(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}$, the substitution $\mathbf{x} = \mathbf{Q}\mathbf{y}$ results in $\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} = \mathbf{y}^{\mathrm{T}}\mathbf{D}\mathbf{y} = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2$, where $\mathbf{Q}$ is the orthogonal matrix such that $\mathbf{Q}^{\mathrm{T}}\mathbf{A}\mathbf{Q} = \mathbf{D}$.

Key Concepts

For $n \times n$ real symmetric matrix $\mathbf{A}$, the quadratic form is,

1. Positive definite: $f(\mathbf{x}) > 0$ for all $\mathbf{x} \neq \mathbf{0}$ (all of the eigenvalues are positive);
2. Negative definite: $f(\mathbf{x}) < 0$ for all $\mathbf{x} \neq \mathbf{0}$ (all of the eigenvalues are negative);
3. Positive semidefinite: $f(\mathbf{x}) \geq 0$ for all $\mathbf{x}$ (all of the eigenvalues are nonnegative);
4. Negative semidefinite: $f(\mathbf{x}) \leq 0$ for all $\mathbf{x}$ (all of the eigenvalues are nonpositive);
5. Indefinite: $f(\mathbf{x})$ alters between positive and negative (both positive and negative eigenvalues);

Application

For $f(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}$, subject to $||\mathbf{x}|| = 1$, we have

1. $f(\mathbf{x})$ is bounded between the minimum and maximum eigenvalues;
2. The minimum value is attained when $\mathbf{x}$ is a unit eigenvector corresponding to minimum eigenvalue;
3. The maximum value is attained when $\mathbf{x}$ is a unit eigenvector corresponding to maximum eigenvalue.

Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is three times differentiable, and f has a critical point at $\mathbf{a}$. If $\mathbf{H}$ is positive definite at $\mathbf{a}$, then $f(a)$ is a local minimum for f.