SI231b: Matrix Computations

Lecture 3: Basic Concepts (Part 2)

Yue Qiu

qiuyue@shanghaitech.edu.cn

School of Information Science and Technology

ShanghaiTech University

Sept. 15, 2020

Basic Concepts: Part 2

- subspaces, span
- linear independence, basis
- range and null spaces
- dimension of subspaces, rank
- inner product, orthogonality
- matrix products, computational complexity

Obvious interpretation of subspaces in the visual spaces $\mathbb{R}^2$ and $\mathbb{R}^3$

- flat surfaces passing through the origin

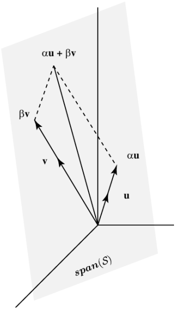
Spanning sets

- for a set of vectors $\mathcal{S} = \{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_r\}$, the subspaces

$$span(\mathcal{S}) = \{\alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \cdots + \alpha_r \mathbf{v}_r\}$$

  generated by forming all linear combinations of vectors from $\mathcal{S}$ is called the space spanned by $\mathcal{S}$

- If $\mathcal{V}$ is a vector space such that $\mathcal{V} = span(\mathcal{S})$, we say $\mathcal{S}$ is a spanning set for $\mathcal{V}$.



- $\mathcal{S} = \{\mathbf{u}, \mathbf{v}\}$ is a spanning set of the indicated plan in the left figure
- The unit vectors

$$\left\{ \mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \mathbf{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}$$

spans $\mathbb{R}^3$

If $\mathcal{X}$ and $\mathcal{Y}$ are subspaces of a vector space $\mathcal{V}$,

- $\mathcal{X} \cap \mathcal{Y}$ is also a subspace
- $\mathcal{X} \cup \mathcal{Y}$ need not be a subspace

If $\mathcal{X}$ and $\mathcal{Y}$ are subspaces of a vector space $\mathcal{V}$, define the sum of two subspaces by

$$\mathcal{X} + \mathcal{Y} = \{\mathbf{x} + \mathbf{y} | \mathbf{x} \in \mathcal{X} \text{ and } \mathbf{y} \in \mathcal{Y}\}$$

then

- the sum $\mathcal{X} + \mathcal{Y}$ is again a subspace of $\mathcal{V}$
- if $\mathcal{S}_X$, $\mathcal{S}_Y$ spans $\mathcal{X}$ and $\mathcal{Y}$, then $\mathcal{S}_X \cup \mathcal{S}_Y$ spans $\mathcal{X} + \mathcal{Y}$

Examples

- If $\mathcal{X} \subset \mathbb{R}^2$ and $\mathcal{Y} \subset \mathbb{R}^2$ are subspaces defined by two different lines through the origin, then $\mathcal{X} + \mathcal{Y} = \mathbb{R}^2$
- If $\mathcal{X}$ is a subspace represents a plane passing through the origin in $\mathbb{R}^3$ and $\mathcal{Y}$ is a subspace defined by the line through the origin that is perpendicular to $\mathcal{X}$, $\mathcal{X} + \mathcal{Y} = \mathbb{R}^3$

Question: any span is a subspace, can any subspace be written as a span?

Theorem

Let $\mathcal{S}$ be a subspace of $\mathbb{R}^m$. There exists a positive integer n and a collection of vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathcal{S}$ such that $\mathcal{S} = \mathrm{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$.

- Implication: we can always represent a subspace by a span

Linear Independence

A collection of vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^m$ is said to be linearly independent if

$$\sum_{i=1}^{n} \alpha_i \mathbf{a}_i \neq \mathbf{0}, \quad \text{for all } \boldsymbol{\alpha} \in \mathbb{R}^n \text{ with } \boldsymbol{\alpha} \neq \mathbf{0};$$

- an equivalent way of defining linear dependence: $\{\mathbf{a}_1, \ldots, \mathbf{a}_n\} \subset \mathbb{R}^m$ is a linearly dependent vector set if there exists $\boldsymbol{\alpha} \in \mathbb{R}^m$, $\boldsymbol{\alpha} \neq \mathbf{0}$, such that

Some known facts:

- if $\{\mathbf{a}_1, \ldots \mathbf{a}_n\} \subset \mathbb{R}^m$ is linearly independent, then any $\mathbf{a}_j$ *cannot* be a linear combination of the other $\mathbf{a}_i$'s; i.e., $\mathbf{a}_j \neq \sum_{i \neq j} \alpha_i \mathbf{a}_i$ for any α_i's.
- if $\{\mathbf{a}_1, \ldots \mathbf{a}_n\} \subset \mathbb{R}^m$ is linearly dependent, then *there exists* an $\mathbf{a}_j$ such that $\mathbf{a}_j$ is a linear combination of the other $\mathbf{a}_i$'s; i.e., $\mathbf{a}_j = \sum_{i \neq j} \alpha_i \mathbf{a}_i$ for some α_i's.
- if $\{\mathbf{a}_1, \ldots \mathbf{a}_n\} \subset \mathbb{R}^m$ is linearly independent, then $n \leq m$ must hold.
- let $\{\mathbf{a}_1, \ldots, \mathbf{a}_m\} \subset \mathbb{R}^m$ be a linearly independent vector set. Suppose $\mathbf{y} \in \mathrm{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$. Then the coefficient $\boldsymbol{\alpha}$ for the representation

$$\mathbf{y} = \sum_{i=1}^{n} \alpha_i \mathbf{a}_i$$

is unique; i.e., there does *not* exist a $\boldsymbol{\beta} \in \mathbb{R}^n$, $\boldsymbol{\beta} \neq \boldsymbol{\alpha}$, such that $\mathbf{y} = \sum_{i=1}^{n} \beta_i \mathbf{a}_i$.

Linear Independence

Let $\{\mathbf{a}_1, \ldots \mathbf{a}_n\} \subset \mathbb{R}^m$, and denote $\{i_1, \ldots, i_k\} \subseteq \{1, \ldots, n\}$ as an index subset with $k \leq n$ and $i_j \neq i_l$ for all $j \neq l$.

A vector subset $\{\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_k}\}$ is called a maximal linear independent subset of $\{\mathbf{a}_1, \ldots \mathbf{a}_n\}$ if

1. $\{\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_k}\}$ is linear independent;
2. $\{\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_k}\}$ is not contained by any other linearly independent subset of $\{\mathbf{a}_1, \ldots \mathbf{a}_n\}$.

- physical meaning: find a set of non-redundant vectors from $\{\mathbf{a}_1, \ldots \mathbf{a}_n\}$

Linear Independence

- Example:

$$\mathbf{a}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{a}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{a}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad \mathbf{a}_4 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}.$$

The linear independent subets of $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4\}$ are

$$\{\mathbf{a}_1\}, \{\mathbf{a}_2\}, \{\mathbf{a}_3\}, \{\mathbf{a}_4\},$$

$$\{\mathbf{a}_1, \mathbf{a}_2\}, \{\mathbf{a}_1, \mathbf{a}_3\}, \{\mathbf{a}_1, \mathbf{a}_4\}, \{\mathbf{a}_2, \mathbf{a}_3\}, \{\mathbf{a}_2, \mathbf{a}_4\}, \{\mathbf{a}_3, \mathbf{a}_4\},$$

$$\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}, \quad \{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_4\}, \quad \{\mathbf{a}_1, \mathbf{a}_3, \mathbf{a}_4\}.$$

But the maximal linear independent subsets are

$$\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}, \quad \{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_4\}, \quad \{\mathbf{a}_1, \mathbf{a}_3, \mathbf{a}_4\}.$$

Linear Independence

Facts:

- $\{\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_k}\}$ is a maximal linear independent subset of $\{\mathbf{a}_1, \ldots \mathbf{a}_n\}$
  - if and only if $\{\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_k}, \mathbf{a}_j\}$ is linear dependent for any $j \in \{1, \ldots, n\} \setminus \{i_1, \ldots, i_k\}$
- if $\{\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_k}\}$ is a maximal linearly independent subset of $\{\mathbf{a}_1, \ldots \mathbf{a}_n\}$, then

$$\mathrm{span}\{\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_k}\} = \mathrm{span}\{\mathbf{a}_1, \ldots \mathbf{a}_n\}.$$

- key concepts for defining the basis of a subspace

Let $\mathcal{S} \subseteq \mathbb{R}^m$ be a nontrivial subspace. A vector set $\{\mathbf{b}_1, \ldots, \mathbf{b}_k\} \subset \mathbb{R}^m$ is called a basis for $\mathcal{S}$ if $\{\mathbf{b}_1, \ldots, \mathbf{b}_k\}$ is linearly independent and

$$\mathcal{S} = \operatorname{span}\{\mathbf{b}_1, \ldots, \mathbf{b}_k\}.$$

- Examples:

  let $\{\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_k}\}$ be a maximal linearly independent vector subset of $\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$. Then, $\{\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_k}\}$ is a basis for $\operatorname{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$.

Some facts:

- we may have more than one basis for $\mathcal{S}$
- all bases for $\mathcal{S}$ have the same number of elements; i.e., if $\{\mathbf{b}_1, \ldots, \mathbf{b}_k\}$ and $\{\mathbf{c}_1, \ldots, \mathbf{c}_l\}$ are bases for $\mathcal{S}$, then $k = l$

Range Spaces

1. The range of a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ denoted by $\mathcal{R}(\mathbf{A})$, is defined to be the subspace of $\mathbb{R}^m$ generated by the range of $\mathbf{Ax}$

$$\mathcal{R}(\mathbf{A}) = \{\mathbf{y} \in \mathbb{R}^m \mid \mathbf{y} = \mathbf{Ax}, \ \mathbf{x} \in \mathbb{R}^n\} \subset \mathbb{R}^m$$

   - also called column space

2. The range of $\mathbf{A}^T$ is the subspace of $\mathbb{R}^n$ defined by

$$\mathcal{R}(\mathbf{A}^T) = \left\{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{x} = \mathbf{A}^T\mathbf{y}, \ \mathbf{y} \in \mathbb{R}^m\right\} \subset \mathbb{R}^n$$

   - also called row space

3. $\mathcal{R}(\mathbf{A})$ is the set of all "images" of vectors $\mathbf{x} \in \mathbb{R}^n$ under transformation by $\mathbf{A}$, sometimes $\mathcal{R}(\mathbf{A})$ is called the image space of $\mathbf{A}$.

Null Spaces

1. The null space of a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ denoted by $\mathcal{N}(\mathbf{A})$, is defined to be the subspace of $\mathbb{R}^n$ with

$$\mathcal{N}(\mathbf{A}) = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{A}\mathbf{x} = \mathbf{0}\} \subset \mathbb{R}^n$$

   - $\mathcal{N}(\mathbf{A})$ is simply the set of all solutions to the homogeneous system $\mathbf{A}\mathbf{x} = \mathbf{0}$.

2. Similarly, the nullspace of $\mathbf{A}^T$, i.e., $\mathcal{N}(\mathbf{A}^T)$

$$\mathcal{N}(\mathbf{A}^T) = \left\{\mathbf{y} \in \mathbb{R}^m \mid \mathbf{A}^T\mathbf{y} = \mathbf{0}\right\} \subset \mathbb{R}^m$$

   - also called left-hand nullspace of $\mathbf{A}$ since it is the set of all solutions to the left-hand homogeneous system $\mathbf{y}^T\mathbf{A} = \mathbf{0}^T$

The dimension of a nontrivial subspace $\mathcal{S}$ is defined as the number of elements of a basis for $\mathcal{S}$.

- the dimension of of the trivial subspace $\{\mathbf{0}\}$ is defined as 0.
- $\dim \mathcal{S}$ will be used as the notation for denoting the dimension of $\mathcal{S}$
- physical meaning: effective degrees of freedom of the subspace
- examples:
  - $\dim \mathbb{R}^m = m$
  - if k is the number of maximal linearly independent vectors of $\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$, then $\dim \operatorname{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\} = k$.

Properties:

- let $\mathcal{S}_1, \mathcal{S}_2 \subseteq \mathbb{R}^m$ be subspaces. If $\mathcal{S}_1 \subseteq \mathcal{S}_2$, then $\dim \mathcal{S}_1 \leq \dim \mathcal{S}_2$.
- let $\mathcal{S}_1, \mathcal{S}_2 \subseteq \mathbb{R}^m$ be subspaces. If $\mathcal{S}_1 \subseteq \mathcal{S}_2$ and $\dim \mathcal{S}_1 = \dim \mathcal{S}_2$, then $\mathcal{S}_1 = \mathcal{S}_2$.
- let $\mathcal{S} \subseteq \mathbb{R}^m$ be a subspace. Then

$$\dim \mathcal{S} = m \Longleftrightarrow \mathcal{S} = \mathbb{R}^m.$$

- let $\mathcal{S}_1, \mathcal{S}_2 \subseteq \mathbb{R}^m$ be subspaces. We have $\dim(\mathcal{S}_1 + \mathcal{S}_2) \leq \dim \mathcal{S}_1 + \dim \mathcal{S}_2$.
  - as a more advanced result, we also have

$$\dim(\mathcal{S}_1 + \mathcal{S}_2) = \dim \mathcal{S}_1 + \dim \mathcal{S}_2 - \dim(\mathcal{S}_1 \cap \mathcal{S}_2).$$

Rank

The rank of a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$, denoted by $\text{rank}(\mathbf{A})$, is defined as the number of elements of a maximal linearly independent subset of $\{\mathbf{a}_1, \dots, \mathbf{a}_n\}$.

- $\text{rank}(\mathbf{A})$ is the maximum number of linearly independent columns of $\mathbf{A}$
- $\dim \mathcal{R}(\mathbf{A}) = \text{rank}(\mathbf{A})$ by definition

Facts:

- $\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{A}^T)$, i.e., the rank of $\mathbf{A}$ is also the maximum number of linearly independent rows of $\mathbf{A}$
- $\text{rank}(\mathbf{A} + \mathbf{B}) \leq \text{rank}(\mathbf{A}) + \text{rank}(\mathbf{B})$
- $\text{rank}(\mathbf{AB}) \leq \min\{\text{rank}(\mathbf{A}), \text{rank}(\mathbf{B})\}$.
    - Equality holds when $\mathbf{A}$ and $\mathbf{B}$ are full rank.

Rank

- $\mathbf{A} \in \mathbb{R}^{m\times n}$ is said to have
  - full column rank if the columns of $\mathbf{A}$ are linearly independent (more precisely, the collection of *all* columns of $\mathbf{A}$ is linearly independent)
    - $\mathbf{A} \in \mathbb{R}^{m\times n}$ being of full column rank $\Longleftrightarrow m \geq n, \text{rank}(\mathbf{A}) = n$
  - full row rank if the rows of $\mathbf{A}$ are linearly independent
    - $\mathbf{A} \in \mathbb{R}^{m\times n}$ being of full row rank $\Longleftrightarrow m \leq n, \text{rank}(\mathbf{A}) = m$
  - full rank if $\text{rank}(\mathbf{A}) = \min\{m, n\}$; i.e., it has either full column rank or full row rank
  - rank deficient if $\text{rank}(\mathbf{A}) < \min\{m,\ n\}$

Invertible Matrices

A square matrix $\mathbf{A}$ is said to be nonsingular or invertible if

- $\mathbf{A}$ is full rank
- all the columns of $\mathbf{A}$ are linear independent
- $\mathbf{Ax} = \mathbf{0} \Longleftrightarrow \mathbf{x} = \mathbf{0}$
- alternatively, we say $\mathbf{A}$ is singular if $\mathbf{Ax} = \mathbf{0}$ for some $\mathbf{x} \neq \mathbf{0}$.

The inverse of an invertible $\mathbf{A}$, denoted by $\mathbf{A}^{-1}$, is a square matrix that satisfies

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}.$$

Invertible Matrices

Facts (for a nonsingular $\mathbf{A}$):

- $\mathbf{A}^{-1}$ always exists and is unique (or there are no two inverses of $\mathbf{A}$)
- $\mathbf{A}^{-1}$ is nonsingular
- $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$
- $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$
- $(\mathbf{A}\mathbf{B})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$, where $\mathbf{A}, \mathbf{B}$ are square and nonsingular
- $(\mathbf{A}^T)^{-1} = (\mathbf{A}^{-1})^T$
  - as a shorthand notation, we will denote $\mathbf{A}^{-T} = (\mathbf{A}^T)^{-1}$

The inner product of two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ is defined as

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^{n} y_i x_i = \mathbf{y}^T \mathbf{x}.$$

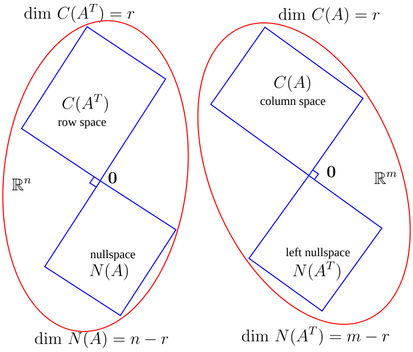
- $\mathbf{x}, \mathbf{y}$ are said to be orthogonal to each other if $\langle \mathbf{x}, \mathbf{y} \rangle = 0$
- $\mathbf{x}, \mathbf{y}$ are said to be parallel if $\mathbf{x} = \alpha \mathbf{y}$ for some α

The angle between two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ is defined as

$$\theta = \arccos\left(\frac{\mathbf{y}^T \mathbf{x}}{\|\mathbf{x}\|_2 \|\mathbf{y}\|_2} \right).$$

- $\mathbf{x}, \mathbf{y}$ are orthogonal if $\theta = \pi/2$
- $\mathbf{x}, \mathbf{y}$ are parallel if $\theta = 0$ or $\theta = \pi$

Orthogonality of Four Fundamental Subspaces



- $\mathcal{R}(\mathbf{A}) \perp \mathcal{N}(\mathbf{A}^T)$
- $\mathcal{R}(\mathbf{A}^T) \perp \mathcal{N}(\mathbf{A})$
- Details will follow in the later part

Cauchy-Schwartz inequality:

$$|\mathbf{x}^T\mathbf{y}| \leq \|\mathbf{x}\|_2\|\mathbf{y}\|_2.$$

Also, the above equality holds if and only if $\mathbf{x} = \alpha\mathbf{y}$ for some $\alpha \in \mathbb{R}$.

- Proof: suppose $\mathbf{y} \neq \mathbf{0}$; the case of $\mathbf{y} = \mathbf{0}$ is trivial. For any $\alpha \in \mathbb{R}$,

$$0 \leq \|\mathbf{x} - \alpha\mathbf{y}\|_2^2 = (\mathbf{x} - \alpha\mathbf{y})^T(\mathbf{x} - \alpha\mathbf{y}) = \|\mathbf{x}\|_2^2 - 2\alpha\mathbf{x}^T\mathbf{y} + \alpha^2\|\mathbf{y}\|_2^2. \qquad (*)$$

Also, the equality above holds if and only if $\mathbf{x} = \beta\mathbf{y}$ for some β. Let

$$f(\alpha) = \|\mathbf{x}\|_2^2 - 2\alpha\mathbf{x}^T\mathbf{y} + \alpha^2\|\mathbf{y}\|_2^2.$$

The function f is minimized when $\alpha = (\mathbf{x}^T\mathbf{y})/\|\mathbf{y}\|_2^2$. Plugging this α back to $(*)$ leads to the desired result.

Hölder inequality:

$$|\mathbf{x}^T\mathbf{y}| \leq \|\mathbf{x}\|_p\|\mathbf{y}\|_q,$$

for any p, q such that $1/p + 1/q = 1,\ p \geq 1$.

- examples:
  - $(p, q) = (2, 2)$: Cauchy-Schwartz inequality
  - $(p, q) = (1, \infty)$: $|\mathbf{x}^T\mathbf{y}| \leq \|\mathbf{x}\|_1\|\mathbf{y}\|_\infty$.

    This can be easily verified to be true:

$$|\mathbf{x}^T\mathbf{y}| \leq \sum_{i=1}^n |x_i y_i| \leq \max_j |y_j| \left(\sum_{i=1}^n |x_i|\right) = \|\mathbf{x}\|_1\|\mathbf{y}\|_\infty.$$

Let $\mathbf{A} \in \mathbb{R}^{m \times k}$, $\mathbf{B} \in \mathbb{R}^{k \times n}$, and consider

$$\mathbf{C} = \mathbf{A}\mathbf{B}.$$

- column representation:

$$\mathbf{c}_i = \mathbf{A}\mathbf{b}_i, \quad i = 1, \ldots, n$$

- inner-product representation: redefine $\mathbf{a}_i \in \mathbb{R}^k$ as the ith row of $\mathbf{A}$.

$$\mathbf{A}\mathbf{B} = \begin{bmatrix} \mathbf{a}_1^T \\ \vdots \\ \mathbf{a}_m^T \end{bmatrix} \begin{bmatrix} \mathbf{b}_1 & \cdots & \mathbf{b}_n \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1^T\mathbf{b}_1 & \cdots & \mathbf{a}_1^T\mathbf{b}_n \\ \vdots & & \vdots \\ \mathbf{a}_m^T\mathbf{b}_1 & \cdots & \mathbf{b}_m^T\mathbf{b}_n \end{bmatrix}$$

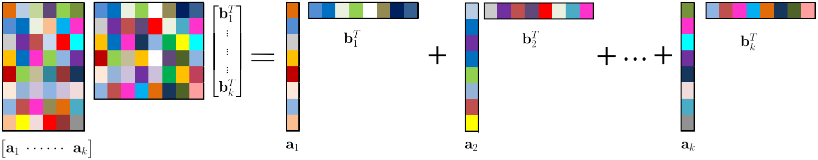
Thus,

$$c_{ij} = \mathbf{a}_i^T\mathbf{b}_j, \quad \text{for any } i, j.$$

Matrix Product Representations

- outer-product representation: redefine $\mathbf{b}_i \in \mathbb{R}^k$ as the ith row of $\mathbf{B}$. Thus,

$$\mathbf{C} = \mathbf{AB} = \sum_{i=1}^{k} \mathbf{a}_i \mathbf{b}_i^T$$



Matrix Product Representations

- The matrix of the form $\mathbf{X} = \mathbf{a}\mathbf{b}^T$ for some $\mathbf{a}, \mathbf{b}$ is called a rank-one outer product.
  - It can be verified that $\text{rank}(\mathbf{X}) \leq 1$, and $\text{rank}(\mathbf{X}) = 1$ if $\mathbf{a} \neq \mathbf{0}, \mathbf{b} \neq \mathbf{0}$.
- the outer-product representation

$$\mathbf{C} = \sum_{i=1}^{k} \mathbf{a}_i \mathbf{b}_i^T$$

  is a sum of k rank-one outer products.
- does it mean that $\text{rank}(\mathbf{C}) = k$?
  - $\text{rank}(\mathbf{C}) \leq \sum_{i=1}^{k} \text{rank}(\mathbf{a}_i \mathbf{b}_i^T) \leq k$ is true [1]
  - but the above equality is generally not attained; e.g., $k = 2$, $\mathbf{a}_1 = \mathbf{a}_2$, $\mathbf{b}_1 = -\mathbf{b}_2$ leads to $\mathbf{C} = \mathbf{0}$
  - $\text{rank}(\mathbf{C}) = k$ only when $\mathbf{A}$ and $\mathbf{B}$ are full rank (take home exam)

[1] use the rank inequality $\text{rank}(\mathbf{A} + \mathbf{B}) \leq \text{rank}(\mathbf{A}) + \text{rank}(\mathbf{B})$.

Sometimes it may be useful to manipulate matrices in a block form.

- let $\mathbf{A} \in \mathbb{R}^{m\times n}$, $\mathbf{x} \in \mathbb{R}^n$. By partitioning

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix}$$

  where $\mathbf{A}_1 \in \mathbb{R}^{m\times n_1}, \mathbf{A}_2 \in \mathbb{R}^{m\times n_2}, \mathbf{x}_1 \in \mathbb{R}^{n_1}, \mathbf{x}_2 \in \mathbb{R}^{n_2}$, with $n_1 + n_2 = n$, we can write

$$\mathbf{A}\mathbf{x} = \mathbf{A}_1\mathbf{x}_1 + \mathbf{A}_2\mathbf{x}_2$$

- similarly, by partitioning

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix},$$

  we can write

$$\mathbf{A}\mathbf{x} = \begin{bmatrix} \mathbf{A}_{11}\mathbf{x}_1 + \mathbf{A}_{12}\mathbf{x}_2 \\ \mathbf{A}_{21}\mathbf{x}_1 + \mathbf{A}_{22}\mathbf{x}_2 \end{bmatrix}$$

- consider $\mathbf{AB}$. By an appropriate partitioning,

$$\mathbf{AB} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 \end{bmatrix} \begin{bmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \end{bmatrix} = \mathbf{A}_1\mathbf{B}_1 + \mathbf{A}_2\mathbf{B}_2$$

- similarly, by an appropriate partitioning,

$$\mathbf{AB} = \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \end{bmatrix} \begin{bmatrix} \mathbf{B}_1 & \mathbf{B}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{A}_1\mathbf{B}_1 & \mathbf{A}_1\mathbf{B}_2 \\ \mathbf{A}_2\mathbf{B}_1 & \mathbf{A}_2\mathbf{B}_2 \end{bmatrix}$$

- we showcase two-block partitioning only, but the same manipulations apply to multi-block partitioning like

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \cdots & \mathbf{A}_{1q} \\ \vdots & & \vdots \\ \mathbf{A}_{p1} & \cdots & \mathbf{A}_{pq} \end{bmatrix}$$

Extension to $\mathbb{C}^n$

- all the concepts described above apply to the complex case
- we only need to replace every "$\mathbb{R}$" with "$\mathbb{C}$", and every "T" with "H"; e.g.,
  - $$\text{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\} = \{\mathbf{y} \in \mathbb{C}^m \mid \mathbf{y} = \textstyle\sum_{i=1}^n \alpha_i \mathbf{a}_i,\ \boldsymbol{\alpha} \in \mathbb{C}^n\},$$
  - $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{y}^H \mathbf{x}$;
  - $\|\mathbf{x}\|_2 = \sqrt{\mathbf{x}^H \mathbf{x}}$, and so forth.

Extension to $\mathbb{C}^{m\times n}$

- the concepts also apply to the matrix case
  - e.g., we may write

$$\mathrm{span}\{\mathbf{A}_1,\ldots,\mathbf{A}_k\} = \{\mathbf{Y}\in\mathbb{R}^{m\times n} \mid \mathbf{Y} = \textstyle\sum_{i=1}^k \alpha_i\mathbf{A}_i,\ \boldsymbol{\alpha}\in\mathbb{R}^k\}.$$

  - sometimes it is more convenient to *vectorize* $\mathbf{X}$ as a vector $\mathbf{x}\in\mathbb{R}^{mn}$, and use the same treatment as in the $\mathbb{R}^n$ case
  - inner product for $\mathbb{R}^{m\times n}$:

$$\langle\mathbf{X},\mathbf{Y}\rangle = \textstyle\sum_{i=1}^m\sum_{j=1}^n x_{ij}y_{ij} = \mathrm{tr}\big(\mathbf{Y}^T\mathbf{X}\big),$$

  - the matrix version of the Euclidean norm is called the Frobenius norm:

$$\|\mathbf{X}\|_F = \sqrt{\textstyle\sum_{i=1}^m\sum_{j=1}^n |x_{ij}|^2} = \sqrt{\mathrm{tr}\big(\mathbf{X}^T\mathbf{X}\big)}$$

- extension to $\mathbb{C}^{m\times n}$ is just as straightforward as in that to $\mathbb{C}^n$

Complexities of Matrix Computations

- every vector/matrix operation such as $\mathbf{x} + \mathbf{y}$, $\mathbf{y}^T\mathbf{x}$, $\mathbf{A}\mathbf{x}$, $\ldots$ incurs computational costs, and they cost more as the vector and matrix sizes get bigger
- we typically look at floating point arithmetic operations, such as add, subtract, multiply, and divide

- flops: one flop means one floating point arithmetic operation.
- flops count of some standard vector/matrix operations: $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{B} \in \mathbb{R}^{n \times p}$,
  - $\mathbf{x} + \mathbf{y}$: n adds, so n flops
  - $\mathbf{y}^T \mathbf{x}$: n multiplies and $n - 1$ adds, so $2n - 1$ flops
  - $\mathbf{Ax}$: m inner products, so $m(2n - 1)$ flops
  - $\mathbf{AB}$: do "$\mathbf{Ax}$" above p times, so $pm(2n - 1)$ flops

- we are often interested in the *order* of the complexity
- big $\mathcal{O}$ notation: given two functions $f(n), g(n)$, the notation

$$f(n) = \mathcal{O}(g(n))$$

  means that there exists a constant $C > 0$ and n_0 such that $|f(n)| \leq C|g(n)|$ for all $n \geq n_0$.
- big $\mathcal{O}$ complexities of standard vector/matrix operations:
  - $\mathbf{x} + \mathbf{y}$: $\mathcal{O}(n)$ flops
  - $\mathbf{y}^T\mathbf{x}$: $\mathcal{O}(n)$ flops
  - $\mathbf{Ax}$: $\mathcal{O}(mn)$ flops
  - $\mathbf{AB}$: $\mathcal{O}(mnp)$ flops

- Discussion: flop counts do not always translate into the actual efficiency of the execution of an algorithm
- things like pipelining, FPGA, parallel computing (multiple GPUs, multiple servers, cloud computing), etc., can make the story different.
- flop counts also ignore memory usage and other overheads...
- that said, we need at least a crude measure of the computational cost of an algorithm, and counting the flops serves that purpose.

How to Save Computations

- computational complexities depend much on how we design and write an algorithm
- generally, it is about
  - top-down, analysis-guided, designs:
    - seen in class, research papers
    - looks elegant
  - facts are
    - usually *not* taught much in class
    - not commonplace in papers
    - subtly depends on your problem at hand
    - a bunch of small differences can make a big difference, say in actual running time
- here we give several, but by no means all, tips for saving computations

- apply matrix operations wisely
- Example: try this on Matlab

```
>> A=randn(5000,2);
>> B=randn(2,10000);
>> C=randn(10000,10000);
>>
>> tic; D= A*B*C; toc
Elapsed time is 1.334183 seconds.
>> tic; D= (A*B)*C; toc     % ask Matlab to do AB first
Elapsed time is 1.205725 seconds.
>> tic; D= A*(B*C); toc     % ask Matlab to do BC first
Elapsed time is 0.067979 seconds.
```

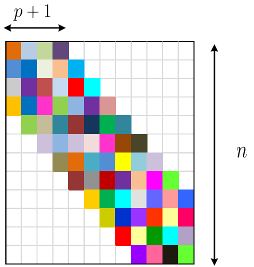
- let us analyze the complexities in the last example
  - $\mathbf{A} \in \mathbb{R}^{m \times n}, \mathbf{B} \in \mathbb{R}^{n \times p}, \mathbf{C} \in \mathbb{R}^{p \times p}$, with $n \ll \min\{m, p\}$.
  - We want to compute $\mathbf{D} = \mathbf{ABC}$.
  - if we compute $\mathbf{AB}$ first, and then $\mathbf{D} = (\mathbf{AB})\mathbf{C}$, the flop count will be
    $$\mathcal{O}(mnp) + \mathcal{O}(mp^2) = \mathcal{O}(m(n+p)p) \approx \mathcal{O}(mp^2)$$
  - if we compute $\mathbf{BC}$ first, and then $\mathbf{D} = \mathbf{A}(\mathbf{BC})$, the flop count will be
    $$\mathcal{O}(np^2) + \mathcal{O}(mnp) = \mathcal{O}((m+p)np).$$
  - the 2nd option is preferable if n is much smaller than m, p

How to Save Computations

- use structures, if available
- example: let $\mathbf{A} \in \mathbb{R}^{n \times n}$ and suppose that

  $$a_{ij} = 0 \text{ for all } i, j \text{ such that } |i - j| > p,$$

  for some integer $p > 0$.
  - such a structured $\mathbf{A}$ is called banded matrix
  - if we don't use structures, computing $\mathbf{Ax}$ requires $\mathcal{O}(n^2)$



  - if we use the banded + sparsity[1] structures, we can compute $\mathbf{Ax}$ with $\mathcal{O}(pn)$
  - different problems may have different fancy/advanced structures to be exploited

[1] a vector or matrix is said to be sparse if it contains many zeros