Optimization Theory and Algorithm	**Lecture 5 - 05/11/2021**

Lecture 5

Lecturer:Xiangyu Chang *Scribe: Xiangyu Chang*

Edited by: Xiangyu Chang

1 Vector Space

Definition 1 *Suppose a set V satisfies, for any $\mathbf{x}, \mathbf{y} \in V$ implies $a \cdot \mathbf{x} + b \cdot \mathbf{y} \in V$, where "+" and "·" are the addition and scalar-multiplication which are defined on V. The we call the set V as a vector space, and any $\mathbf{x} \in V$ is a vector of V.*

Example 1 *The examples of Vector Space:*

- $\mathbb{R}^2$ *and* $\mathbb{R}^n$ *are vector spaces.*
- $D = \left\{ A \in \mathbb{R}^{3\times 3} : A = \begin{bmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{bmatrix} \right\}$
- $\mathbb{R}^2, \mathbb{R}$ *and* $\{0\}$ *are the sub-spaces of* $\mathbb{R}^3$.
- *The Column Space of $A = (\mathbf{a}_1, \dots, \mathbf{a}_n), \mathbf{a}_i \in \mathbb{R}^m$, $C(A) = \{\mathbf{b} \in \mathbb{R}^m : \mathbf{b} = \sum_{i=1}^n x_i \mathbf{a}_i, x_i \in \mathbb{R}\}$. Thus, actually $\mathbf{b}$ is the linear combination of the column vectors of A. The matrix form can be written as $C(A) = \{\mathbf{b} \in \mathbb{R}^m : \mathbf{b} = A\mathbf{x}\}$.*
- *The Null Space of A: $N(A) = \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} = 0\}$.*
- $\mathcal{F} = \{f | f(x) = \mathbf{a}^\top \mathbf{x} + b, \mathbf{x}, \mathbf{a} \in \mathbb{R}^n\}$ *is a vector space??*

Definition 2 *If $\sum_{i=1}^n x_i \mathbf{a}_i = 0$ implies that $x_1 = x_2 = \cdots = x_n = 0$, then the vectors $a_i, i = 1, \dots, n$ are linearly independent.*

Note that $A = (\mathbf{a}_1, \dots, \mathbf{a}_n)$ and $N(A) = 0$ means that vectors $a_i, i = 1, \dots, n$ are linearly independent.

Definition 3 *The basis of a vector space V is a set of vectors $v_1, \dots, v_d$ satisfies:*

- $v_1, \dots, v_d$ *are linearly independent.*
- $span\{v_1, \dots, v_d\} = \{\mathbf{b} : \mathbf{b} = \sum_{i=1}^d x_i v_d, x_i \in \mathbb{R}\} = V$.

Then we say the dimensionality of V is d, denoted as $dim(V) = d$.

Example 2 *The four fundamental sub-spaces of $A \in \mathbb{R}^{m\times n}$ are the column space $C(A)$, null space $N(A)$, the row space $R(A) = C(A^\top)$ and the left-null space $N(A^\top) = \{\mathbf{x} : A^\top \mathbf{x} = \mathbf{x}^\top A = 0\}$. We can see that $C(A) \subset \mathbb{R}^m, N(A^\top) \subset \mathbb{R}^m, R(A) \subset \mathbb{R}^n, R(A) \subset \mathbb{R}^n$, and $dim(C(A)) = rank(A) = r, dim(N(A^\top) = m - r$ and $dim(R(A)) = r, dim(N(A)) = n - r$ (see Figure 1).*

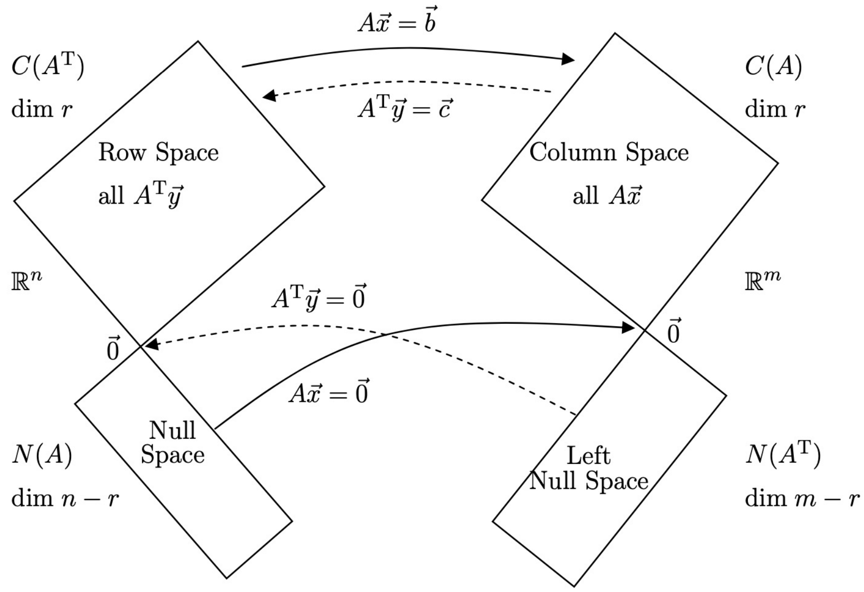

Figure 1: The four fundamental sub-spaces of A
.

2 Vector Norm

Vector Norm:

Definition 4 *The norm of a vector $v \in \mathbb{R}^n$ is a function $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}$ satisfies:*

- $\|v\| \geq 0$ *and* $\|v\| = 0$ *if and only if* $v = 0$.
- $\|\alpha v\| = |\alpha|\|v\|$ *for any* $\alpha \in \mathbb{R}$.
- $\|v + u\| \leq \|v\| + \|u\|$

Example 3 *We demonstrate some norm examples.*

- ℓ_p*-norm* $1 \leq p < \infty$*:* $\|\mathbf{x}\|_p = \left(\sum_{i=1}^n |x_i|^p\right)^{1/p}$.
- ℓ_∞*-norm:* $\|\mathbf{x}\|_\infty = \max_i |x_i|$.
- ℓ_0*-norm:* $\|\mathbf{x}\|_0$ *is the number of nonzero elements of* $\mathbf{x}$.
- **Q:** *is ℓ_0-norm a vector norm??*

Theorem 1

$$\|\mathbf{x}\|_\infty \leq \|\mathbf{x}\|_1 \leq n\|\mathbf{x}\|_\infty \tag{1}$$
$$\|\mathbf{x}\|_\infty \leq \|\mathbf{x}\|_2 \leq \sqrt{n}\|\mathbf{x}\|_\infty \tag{2}$$
$$\|\mathbf{x}\|_2 \leq \|\mathbf{x}\|_1 \leq \sqrt{n}\|\mathbf{x}\|_2 \tag{3}$$
$$\|\mathbf{x}\|_p \leq \|\mathbf{x}\|_q \leq n^{\frac{1}{q}-\frac{1}{p}}\|\mathbf{x}\|_p, p \geq q > 1. \tag{4}$$

Proof 1 *Sample Proof: Let* $\mathbf{v} = (v_1, \ldots, v_n)^\top$, *where* $v_i = |x_i|/x_i$ *if* $x_i \neq 0$. *Thus,* $|v_i| = 1$ *and* $|x_i| = v_i x_i$. *Then*

$$\|\mathbf{x}\|_1 = \sum_i |x_i| = \sum v_i x_i = \mathbf{v}^\top \mathbf{x} \leq \|\mathbf{v}\|_2 \|\mathbf{x}\|_2 = \sqrt{n}\|\mathbf{x}\|_1, \tag{5}$$

where the last inequality comes from the Cauchy inequality (6). *The geometric interpretation can be found in Figure 2.*

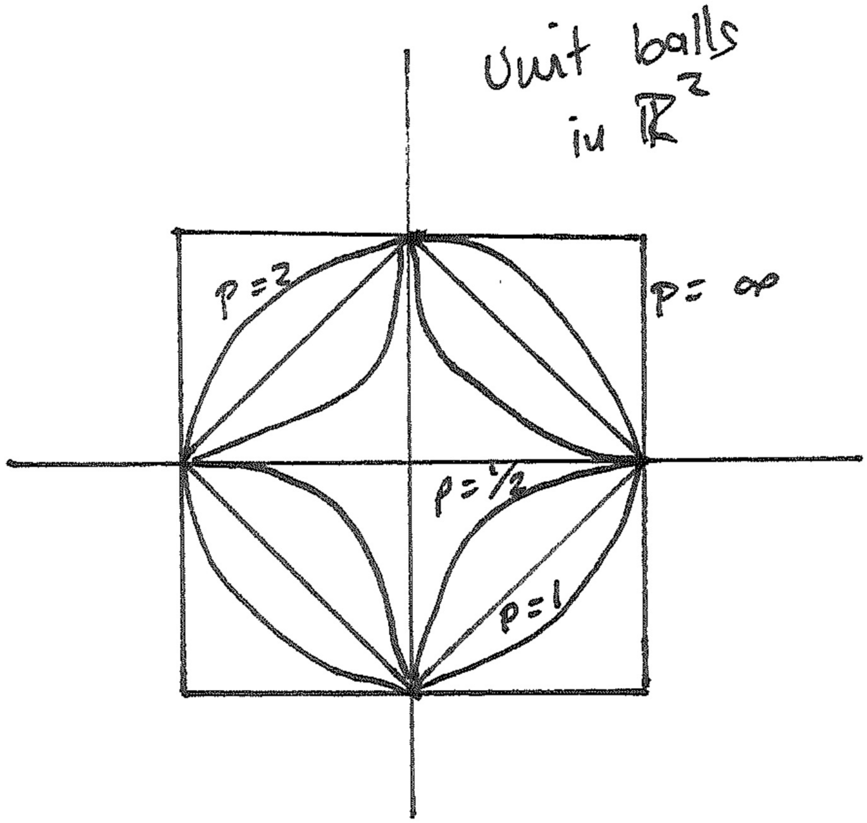

Figure 2: The balls of unit norm in $\mathbb{R}^2$

.

Definition 5 *We define the inner product of* $v, u \in \mathbb{R}^n$ *is* $\langle v, u\rangle = v^\top u$. *Then the* ℓ_2*-norm is the norm with respect to the inter product in* $\mathbb{R}^n$, *that is* $\|v\|_2^2 = v^\top v = \langle v, v\rangle$.

Theorem 2 *(Pythagorean Theorem)*

$$\|u + v\|_2^2 = \|u\|_2^2 + \|v\|_2^2,$$

if $v \perp u$, *namely* $\langle u, v\rangle = 0$.

Theorem 3 *(Cauchy Inequality)*

$$|\langle u, v\rangle| \leq \|u\|_2 \|v\|_2. \tag{6}$$

Based on the Cauchy inequality, we can define the angle between two vectors is $\cos(u, v) = \frac{\langle u,v \rangle}{\|u\|_2 \|v\|_2}$. This can be seen as the similarity of two vectors.

Q: How to project a vector **a** on **b**?

Theorem 4 *(Hölder Inequality)*

$$|\langle u, v \rangle| \leq \|u\|_p \|v\|_q, \tag{7}$$

where $p, q \geq 1$ *and* $\frac{1}{p} + \frac{1}{q} = 1$.

3 Matrix Norm

Matrix Norm:

Definition 6 *The norm of a matrix* $A \in \mathbb{R}^{m \times n}$ *is a function* $\| \cdot \| : \mathbb{R}^{m \times n} \to \mathbb{R}$ *satisfies:*

- $\|A\| \geq 0$ *and* $\|A\| = 0$ *if and only if* $A = 0$.
- $\|\alpha A\| = |\alpha| \|A\|$ *for any* $\alpha \in \mathbb{R}$.
- $\|A + B\| \leq \|A\| + \|B\|$
- $\|A \cdot B\| \leq \|A\| \cdot \|B\|$

Definition 7 *A matrix norm and a vector norm are compatible if*

$$\|A\mathbf{x}\| \leq \|A\| \|\mathbf{x}\|. \tag{8}$$

3.1 Vector-based Norms

For a give matrix $A \in \mathbb{R}^{m \times n}$, consider the vector $vec(A) \in \mathbb{R}^{mn}$ (the columns of A stacked on top of one another), and apply the standard vector p-norm, then implies

- $\|A\|_1 = \sum_{i=1}^m \sum_{j=1}^n |a_{ij}|$;
- $\|A\|_\infty = \max_{ij} |a_{ij}|$;
- $\|A\|_2 = \sqrt{\sum_{i=1}^m \sum_{j=1}^n a_{ij}^2}$. The vector-based ℓ_2 matrix norm is commonly called a *Frobenius* norm and denoted as $\|A\|_F$.

Let us give a sample proof to guarantee that the vector-based norms are the matrix norms.

Proof 2 *Let us prove that* $\|A\|_1$ *is a matrix norm.*

$$\|AB\|_1 = \sum_{i,j} |(AB)_{ij}| = \sum_{i,j} |\sum_{k=1} a_{ik} b_{kj}| \tag{9}$$

$$\leq \sum_{i,j} \sum_{k=1} |a_{ik} b_{kj}| \leq \sum_{i,j} \sum_{k=1} |a_{ik}| |b_{kj}| \tag{10}$$

$$= \|A\|_1 \|B\|_1. \tag{11}$$

You can use the same trick to justify the compatibility of Frobenius norm.

Theorem 5 *The Frobenious norm of matrix A is*

$$\|A\|_F^2 = tr(A^\top A), \tag{12}$$

where $tr(A) = \sum_i a_{ii}$ is the trace of any symmetric matrix.

Prove it by your self.

Theorem 6 *Suppose that U and V are orthogonal matrices, namely $U^\top U = UU^\top = I$, then*

$$\|UAV\|_F = \|A\|_F. \tag{13}$$

3.2 Induced Matrix Norms

Definition 8 *Given any vector norm, the induced matrix norm is give by*

$$\|A\|_{p,q} = \sup_{\mathbf{x}\neq 0} \frac{\|A\mathbf{x}\|_p}{\|\mathbf{x}\|_q} = \sup_{\|\mathbf{x}\|_q=1} \|A\mathbf{x}\|_p. \tag{14}$$

We use a simple notation for $\|A\|_{p,p} = \|A\|_p$.

You can check that these norms are automatically compatible with the vector norm that produced them.

Example 4 *Let us give some examples of the induced matrix norms.*

- $\|A\|_1 = \max_j \sum_i |a_{ij}|$*, it is the largest column sum.*
- $\|A\|_\infty = \max_i \sum_j |a_{ij}|$*, it is the largest row sum.*
- $\|A\|_2 = \max_i \sigma_i$*, where σ_i is the largest singular value.*

Proof 3 *Let us give a sample proof.*

$$\begin{aligned}
\|A\mathbf{x}\|_1 &= \sum_i |\sum_j a_{ij}x_j| \leq \sum_i \sum_j |a_{ij}||x_j| &(15)\\
&= \sum_j \left(\sum_i |a_{ij}|\right) \cdot |x_j| \leq \sum_j \left(\max_k \sum_i |a_{ik}|\right) \cdot |x_j| &(16)\\
&= \left(\max_k \sum_i |a_{ik}|\right) \cdot \sum_j |x_j| = \left(\max_k \sum_i |a_{ik}|\right) \cdot \|\mathbf{x}\|_1. &(17)
\end{aligned}$$

Thus, based on the definition of induced norm we have $\|A\|_1 \leq \max_k \sum_i |a_{ik}|$. Fourther, suppose that $k_0 = \arg\max_k \sum_i |a_{ik}|$, and take $\mathbf{x} = e_{k_0}$, then $\|A\|_1 = \sum_i |a_{ik_0}| = \max_k \sum_i |a_{ik}|$.

- Matrix Inner production: $A, B \in \mathbb{R}^{m\times n}$, then $\langle A, B\rangle = tr(AB^\top) = \sum_i \sum_j a_{ij}b_{ji}$.
- So, $\|A\|_F^2 = \langle A.B\rangle$.
- Cauchy Inequality:

$$|\langle A, B\rangle| \leq \|A\|_F \|B\|_F. \tag{18}$$

3.3 singular-value-based Matrix Norms

For any matrix A with the singular value decomposition form $A = U\Sigma V^\top$, then we can define the following singular-value-based matrix norms as:

- Spectral Radius: $\rho(A) = \|A\|_2 = \max_i \sigma_i$, where σ_i is the ith singular value of A.
- $\|A\|_F = \sqrt{\sum_i \sigma_i^2}$.
- $\|A\|_* = \sum_i \sigma_i$, this is called *nuclear norm*.
- $\|A\|_\infty = \max_i \sigma_i$, the same as the spectral radius.

Theorem 7 *Suppose that $\|A\|$ is a matrix norm, then*

$$\rho(A) \leq \|A\|. \tag{19}$$

3.4 Singular Value Decomposition

Theorem 8 *Any matrix $A \in \mathbb{R}^{m\times n}$ can be factors as*

$$A = U\Sigma V^\top, \tag{20}$$

where $U^\top U = V^\top V = I$ and Σ is a diagonal matrix with σ_i on the diagonal.

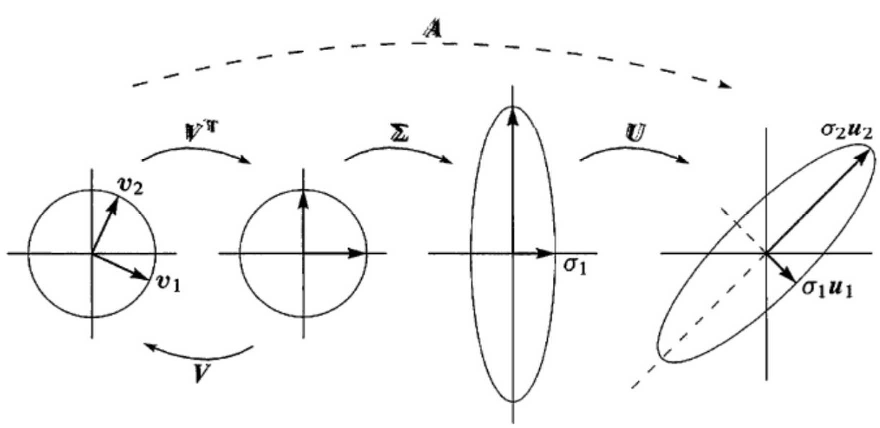

Figure 3: Geometric Interpretation of SVD

Remark 1 *This theorem is not very rigorous. Actually, we need to show U and V implicitly.*

- *Full SVD: $U \in \mathbb{R}^{m^2}, V \in \mathbb{R}^{n^2}$ and $\Sigma \in \mathbb{R}^{m\times n}$.*
- *condensed SVD: $U \in \mathbb{R}^{m\times r}, V^\top \in \mathbb{R}^{r\times n}$ and $\Sigma \in \mathbb{R}^{r\times r}$, where $r = rank(A)$.*
- *Thin SVD: $U \in \mathbb{R}^{m\times r}, V^\top \in \mathbb{R}^{n\times n}$ and $\Sigma \in \mathbb{R}^{r\times n}$.*
- *Thin SVD: $U \in \mathbb{R}^{m\times m}, V^\top \in \mathbb{R}^{r\times n}$ and $\Sigma \in \mathbb{R}^{m\times r}$.*
- *In this note, we use the condensed SVD. Then $A^\top A = V^\top \Sigma^2 V$ and $AA^\top = U^\top \Sigma^2 U$.*

- *Then we can compute the* U*,* V *and* Σ *by the eigenvalue decomposition of the symetric matrix* $A^\top A$ *and* $AA^\top$*. This is not Unique!!!*
- *Singular value decomposition of* A *is*
$$A = \sum_{i=1}^{r} \sigma_i u_i v_i^\top. \tag{21}$$
- *Geometric Interpretation of SVD (see Figure 3).*
- *Pseudo-inverse:* $A = U\Sigma V^\top$ *then* $A^+ = V\Sigma^{-1}U^\top$*.*
- *Let us consider the LS problem. The solution is* $\mathbf{x}^* = (A^\top A)^{-1}A^\top \mathbf{b} = V\Sigma^{-1}U^\top \mathbf{b}$*. Then* A^+ *has the similar behavior of* A^{-1}*.*

References