MATRIX ANALYSIS

1. NOTATION AND BASIC FACTS

↪ We denote by $\mathcal{M}_{n,d}(\mathbb{R})$ the space of all $n \times d$ matrices with real coefficients. For a matrix $A \in \mathcal{M}_{n,d}(\mathbb{R})$ and integer $k \leq n$ and $l \leq d$, we let $A_{\lceil k,l \rceil} \in \mathcal{M}_{k,l}(\mathbb{R})$ denote the matrix A restricted to its first k rows and first l columns. The i,j entry of A will be denoted by $A(i,j)$ or $A^{(i,j)}$.

We assume that the reader is familiar with elementary matrix analysis, including, in particular the fact that symmetric matrices are diagonalizable in an orthonormal basis, i.e., if $A \in \mathcal{M}_{d,d}(\mathbb{R})$ is a symmetric matrix (whose space is denoted by $\mathcal{S}_d$), there exists an orthogonal matrix $U \in \mathcal{O}_d$ (i.e., satisfying $U^T U = U U^T = \mathrm{Id}(\mathbb{R}^d)$) and a diagonal matrix $D \in \mathcal{M}_{d,d}(\mathbb{R})$ such that:

$$A = U D U^T. \tag{2.1}$$

The identity $AU = UD$ then implies that the columns of U form an orthonormal basis of eigenvectors of A.

↪ If $A \in \mathcal{S}_d^+$ is positive semi-definite (i.e., $u^T A u \geq 0$ for all $u \in \mathbb{R}^d$), the entries of D in the decomposition $A = UDU^T$ are non-negative, and one can define the matrix square root of A as $S = U D^{\odot 1/2} U^T$ where $D^{\odot 1/2}$ is the diagonal matrix formed taking the square roots of all coefficients of D. We will use the notation $S = A^{1/2}$. Note that $D^{1/2} = D^{\odot 1/2}$ if D is diagonal and positive semi-definite.

If $A \in \mathcal{S}_d^{++}$ is positive definite (i.e., A is positive semi-definite and $u^T A u = 0$ implies $u = 0$) and B is positive semi-definite, both being $d \times d$ matrices, the generalized eigenvalue problem associated with A and B consists in finding a diagonal matrix D and a matrix U such that $BU = AUD$ and $U^T A U = \mathrm{Id}(\mathbb{R}^d)$. Letting $\tilde{U} = A^{1/2} U$, the problem is equivalent to solving $A^{-1/2} B A^{-1/2} \tilde{U} = \tilde{U} D$ with $\tilde{U}^T \tilde{U} = \mathrm{Id}(\mathbb{R}^d)$, i.e., finding the eigenvalue decomposition of the symmetric positive-definite matrix $A^{-1/2} B A^{-1/2}$.

If $A \in \mathcal{M}_{n,d}(\mathbb{R})$, it can be decomposed as:

$$A = U D V^T, \tag{2.2}$$

where $U \in \mathcal{O}_n(\mathbb{R})$ and $V \in \mathcal{O}_d(\mathbb{R})$ are orthogonal matrices and $D \in \mathcal{M}_{n,d}(\mathbb{R})$ is diagonal (i.e., such that $D(i,j) = 0$ whenever $i \neq j$) with non-negative diagonal coefficients. These coefficients are called the singular values of A, and the procedure is called a <u>singular value decomposition</u> (SVD) of A. An equivalent formulation is that there exists orthonormal bases $u_1, \dots, u_n$ of $\mathbb{R}^n$ and $v_1, \dots, v_d$ of $\mathbb{R}^d$ (forming the columns of U and V) such that:

[...]

$$A v_i = \lambda_i u_i \,, \tag{2.3}$$

for $i \leq \min(n,d)$, where $\lambda_1, \dots, \lambda_{\min(n,d)}$ are the singular values. Of course, if A is square and symmetric positive semi-definite, an eigenvalue decomposition of A is also a SVD (and the singular values coincide with the eigenvalues). More generally, if $A = UDV^T$, then $AA^T = UDD^TU^T$ and $A^TA = VD^TDV^T$ are eigenvalue decompositions of AA^T and A^TA. Singular values are uniquely defined, up to regarding. However, the matrices U and V are not unique up to column reordering in general.

If $m = \min(n,d)$, then, forming the matrices $\tilde{U} = U_{\lceil n,m \rceil}$ (resp. $\tilde{V} = V_{\lceil d,m \rceil}$) by removing from U (resp. V) its last $n-m$ (resp. $d-m$) columns, and $\tilde{D} = D_{\lceil m,m \rceil}$ by removing from D its $n-m$ rows and $d-m$ columns, one has:

$$A = \tilde{U}\tilde{D}\tilde{V}^T \,, \tag{2.4}$$

with $\tilde{U}, \tilde{D}$ and $\tilde{V}$ having respectively size $n \times m$, $m \times m$, and $m \times d$, $\tilde{U}^T\tilde{U} = \tilde{V}^T\tilde{V} = \mathrm{Id}(\mathbb{R}^m)$ and $\tilde{D}$ diagonal with non-negative coefficients. This representation provides a <u>reduced</u> SVD of A and one can create a full SVD from a reduced one by completing the missing rows of $\tilde{U}$ and $\tilde{V}$ to form orthogonal matrices, and by adding the required number of zeros to the matrix $\tilde{D}$.

---

2. The Trace Inequality

▫ <u>Theorem[1]</u> (Von Neumann): Let $A, B \in \mathcal{M}_{n,d}(\mathbb{R})$ have singular values $(\lambda_1, \dots, \lambda_m)$ and $(\mu_1, \dots, \mu_m)$, respectively, where $m = \min(n,d)$. Assume that these eigenvalues are listed in decreasing order so that $\lambda_1 \geq \dots \geq \lambda_m$ and $\mu_1 \geq \dots \geq \mu_m$. Then:

$$\mathrm{tr}(A^TB) \leq \sum_{i=1}^{m} \lambda_i \mu_i \,. \tag{2.5}$$

Moreover, if $\mathrm{tr}(A^TB) = \sum_{i=1}^m \lambda_i \mu_i$, then exist $n \times n$ and $d \times d$ orthogonal matrices U and V such that U^TAV and U^TBV are both diagonal, i.e., one can find SVDs of A and B in the same bases of $\mathbb{R}^n$ and $\mathbb{R}^d$.

*) <u>Remark[1]</u>: Note that, since the singular values of $-A$ and of A coincide, the Von Neumann theorem implies:

$$\left| \mathrm{tr}(A^TB) \right| \leq \sum_{i=1}^{m} \lambda_i \mu_i \,, \tag{2.6}$$

for all matrices A and B, with equality if either A and B or $-A$ and B have an SVD using the same basis.

◇

*) <u>Remark[2]</u>: The proof of the Von Neumann theorem can be found in [Leon Mirsky, 79(4): 303-306, 1975].

◇

(3.) APPLICATIONS

↳ Let p and d be integers with $p \leq d$. Let $A \in \mathcal{S}_d(\mathbb{R})$, $B \in \mathcal{S}_p(\mathbb{R})$ be symmetric matrices. We consider the following optimization problem : maximize, over matrices $U \in \mathcal{M}_{d,p}(\mathbb{R})$ such that $U^TU = Id(\mathbb{R}^p)$, the function:

$$F(U) = tr(U^TAUB) = tr(AUBU^T). \tag{2.7}$$

We first note that the singular values of UBU^T, which is $d \times d$, are the same as the eigenvalues of B completed with zeros. Letting $\lambda_1 \geq \dots \geq \lambda_d$ be the eigenvalues of A and $\mu_1 \geq \dots \geq \mu_p$ those of B, we therefore have, from THEOREM 1 :

$$F(U) \leq \sum_{i=1}^{p} \lambda_i \mu_i . \tag{2.8}$$

Introduce the eigenvalue decompositions of A and B in the form $A = V\Lambda V^T$ and $B = WMW^T$. For $F(U)$ to be equal to its upper-bound, we know that we must arrange UBU^T to take the form :

$$UBU^T = V\begin{pmatrix} M & 0 \\ 0 & 0 \end{pmatrix}V^T . \tag{2.9}$$

Use, as before, the notation $V_{\lceil d p \rceil}$ to denote the matrix formed with the p first columns of V. Take $U = V_{\lceil d p \rceil}W^T$, which satisfies $U^TU = Id(\mathbb{R}^p)$. We then have :

$$V_{\lceil d p \rceil}W^TBWV_{\lceil d p \rceil}^T = V_{\lceil d p \rceil}MV_{\lceil d p \rceil}^T = V\begin{pmatrix} M & 0 \\ 0 & 0 \end{pmatrix}V^T, \tag{2.10}$$

which shows that U is optimal. We summarize this discussion in the next theorem.

■ THEOREM 2 : Let $A \in \mathcal{S}_d(\mathbb{R})$ and $B \in \mathcal{S}_p(\mathbb{R})$ be symmetric matrices, with $p \leq d$. Let eigenvalue decompositions of A and B be given by $A = V\Lambda V^T$ and $B = WMW^T$, where the diagonal elements of Λ (resp. M) are $\lambda_1 \geq \dots \geq \lambda_d$ (resp. $\mu_1 \geq \dots \geq \mu_p$). Define $F(U) = tr(AUBU^T)$, for $U \in \mathcal{M}_{d,p}(\mathbb{R})$. Then :

$$\max\left\{F(U) : U^TU = Id(\mathbb{R}^p)\right\} = \sum_{i=1}^{p} \lambda_i \mu_i . \tag{2.11}$$

The maximum is attained at $U = V_{\lceil d,p \rceil}W^T$.

◆ COROLLARY 1 : Let $A \in \mathcal{S}_d(\mathbb{R})$ be a symmetric matrix with eigenvalues $\lambda_1 \geq \dots \geq \lambda_d$. For $p \leq d$, let $\mu_1 \geq \dots \geq \mu_p > 0$ and define :

$$F(e_1, \dots, e_p) = \sum_{i=1}^{p} \mu_i e_i^T A e_i . \tag{2.12}$$

Then, the maximum of F over all orthonormal families $e_1, \dots, e_p$ in $\mathbb{R}^d$ is $\sum_{i=1}^{p} \lambda_i \mu_i$ and is attained when $e_1, \dots, e_p$ are eigenvectors of A with eigenvalues $\lambda_1, \dots, \lambda_p$. The minimum of F over all orthonormal families $e_1, \dots, e_p$ in $\mathbb{R}^d$ is $\sum_{i=1}^{p} \lambda_{d-i+1} \mu_i$ and is attained when $e_1, \dots, e_p$ are eigenvectors of A with eigenvalues $\lambda_d, \dots, \lambda_{d-p+1}$.

*) REMARK[3] : Note that the COROLLARY[1] applies the THEOREM[2] with $B = \mathrm{diag}(\mu_1, \dots, \mu_p)$. The statement about the maximum is just a special case of THEOREM[2], with $B = \mathrm{diag}(\mu_1, \dots, \mu_p)$, noting that the i^{th} diagonal element of $U^T A U$ is $e_i^T A e_i$ where e_i is the i^{th} column of U. The statement about the minimum is deduced by replacing A by $-A$.

*) REMARK[4] : Applying COROLLARY[1] with $p = 1$, we retrieve the elementary result that $\lambda_1 = \max\{u^T A u : |u| = 1\}$ and $\lambda_d = \min\{u^T A u : |u| = \}$.

■ THEOREM[3] (Rayleigh) : Let $A \in \mathcal{M}_{d,d}(\mathbb{R})$ be a symmetric matrix with eigenvalues $\lambda_1 \geq \dots \geq \lambda_d$. Then :

$$\lambda_k = \max_{V : \dim(V) = k} \min\{u^T A u, u \in V, |u| = 1\} = \min_{V : \dim(V) = k} \max\{u^T A u, u \in V, |u| = 1\}, \tag{2.13}$$

where the min and max are taken over linear subspaces of $\mathbb{R}^d$.

---

④ SOME MATRIX NORMS

↳ The operator norm of a matrix $A \in \mathcal{M}_{n,d}(\mathbb{R})$, is defined as :

$$|A|_{op} := \max\{|Ax| : x \in \mathbb{R}^d, |x| = 1\}. \tag{2.14}$$

It is equal to the square root of the largest eigenvalue of $A^T A$, i.e., to the largest singular value of A.

↳ The Fröbenius norm of a matrix $A \in \mathcal{M}_{n,d}(\mathbb{R})$ is defined as :

$$|A|_F := \sqrt{\mathrm{tr}(A^T A)} = \sqrt{\sum_{i,j=1}^{d} [A(i,j)]^2} = \sqrt{\sum_{k=1}^{m} \sigma_k^2}, \tag{2.15}$$

where $\sigma_1, \dots, \sigma_m$ are the singular values of A, and $m = \min(n, d)$.

↳ The nuclear norm of a matrix $A \in \mathcal{M}_{n,d}(\mathbb{R})$ is defined as :

$$|A|_* := \sum_{k=1}^{d} \sigma_k. \tag{2.16}$$

One can prove that this is a norm using an equivalent definition, provided by the following proposition.

■ PROPOSITION[1] : Let A be an $n\times d$ matrix. Then:

$$|A|_* = \max\{tr(UAV^T) : U \in M_{n,n} \wedge U^TU = Id,\ V \in M_{d,d} \wedge V^TV = Id\}. \tag{2.17}$$

↳ The fact that $|A|_*$ is a norm, for which the only non-trivial fact was the triangular inequality, now is an easy consequence of the PROPOSITION[1], because the maximum of the sum of two functions is always less than the sum of their maximums. More precisely, we have:

$$\begin{aligned} |A+B|_* &= \max\{tr(UAV^T) + tr(UBV^T) : U^TU = Id,\ V^TV = Id\} \\ &\leq \max\{tr(UAV^T) : U^TU = Id,\ V^TV = Id\} + \max\{tr(UBV^T) : U^TU = Id,\ V^TV = Id\} \\ &= |A|_* + |B|_*. \end{aligned} \tag{2.18}$$

The nuclear norm is also called the Ky Fan norm of order d. Ky Fan norms of order k (for $1 \leq k \leq d$) associate to a matrix A the quantity:

$$|A|_{(k)} = \lambda_1 + \dots + \lambda_k = \sum_{i=1}^{k} \lambda_i, \tag{2.19}$$

i.e., the sum of its k largest singular values. One has the following proposition.

■ PROPOSITION[2] : The Ky Fan norms satisfy the triangular inequality.

↳ We refer to [Rajendra Bhatia, *Matrix Analysis*, vol. 169, 2013] for more examples of matrix norms, including, in particular those provided by taking p^{th} powers in Ky Fan's norms, defining:

$$|A|_{(k,p)} = (\lambda_1^p + \dots + \lambda_k^p)^{1/p} = \left(\sum_{i=1}^{k} \lambda_i^p\right)^{1/p}. \tag{2.20}$$