

SCRIPT OF A TALK ON THE KY FAN AND SCHATTEN P NORMS

VISHVAS VASUKI

CONTENTS

| | |
|--|---|
| 1. Introduction | 2 |
| 1.1. The theme | 2 |
| 1.2. The intention | 2 |
| 1.3. The proofs | 2 |
| 1.4. Outline | 2 |
| 1.5. Notation | 2 |
| 1.6. Things to leave on the whiteboard | 2 |
| 2. The norms | 2 |
| 2.1. The intention | 2 |
| 2.2. Vector and matrix norms | 3 |
| 2.2.1. Absolute norm | 3 |
| 2.3. Ky Fan (p,k) norms | 3 |
| 2.4. Schatten p norms | 3 |
| 2.4.1. Frobenius (Hilbert-Schmidt, Euclidian) norm | 3 |
| Optional info | 3 |
| 2.4.2. Trace (Nuclear) norm | 3 |
| 2.5. Convexity | 3 |
| 3. Stochastic matrices | 3 |
| 3.1. The intention | 3 |
| 3.2. Bistochastic matrix | 3 |
| 3.3. Doubly substochastic matrix Q | 3 |
| 4. Ideas about Majorization | 3 |
| 4.1. The intention | 3 |
| 4.2. Strong majorization | 3 |
| 4.2.1. Differing notations | 3 |
| 4.3. Connection with stochastic matrices | 3 |
| 4.4. Weak majorization | 3 |
| 4.4.1. Connection with stochastic matrices | 4 |
| 4.4.2. Weak Majorization preserved under convex increasing fn | 4 |
| 5. Singular value Σ properties | 4 |
| 5.1. Motivation | 4 |
| 5.2. Unitary invariance | 4 |
| 5.3. Trace of A | 4 |
| 5.4. Effect of row or column deletion | 4 |
| 5.5. The inequality $\prod_{i=1}^k \sigma_i(X_k^* A Y_k) \leq \prod_{i=1}^k \sigma_i(A)$ | 4 |
| 5.6. The inequality $\sum_{i=1}^k \lambda_i(A) ^p \leq \sum_{i=1}^k \sigma_i(A)^p$ | 4 |

| | |
|--|---|
| 5.7. Polar decomposition | 5 |
| 5.8. The inequality $\sum_{i=1}^k \sigma_i(AB)^p \leq \sum_{i=1}^k \sigma_i(A)^p \sigma_i(B)^p$ | 5 |
| 5.9. k partial isometry | 5 |
| 5.10. An identity about partial sum of top k singular values | 5 |
| 5.11. The inequality $\sum_{i=1}^k \sigma_i(A+B) \leq \sum_{i=1}^k \sigma_i(A) + \sum_{i=1}^k \sigma_i(B)$ | 5 |
| 6. Proofs of normness | 5 |
| 6.1. Unitary invariance of norms | 5 |
| 6.1.1. Symmetric gauge fn g | 5 |
| 6.2. Ky Fan (p, k) norms | 6 |
| 6.2.1. Schatten p norms | 6 |
| 6.2.2. The general case | 6 |
| 6.2.3. Matrix normness | 6 |
| 7. Interesting ideas about Doubly Substochastic matrix Q | 6 |
| References | 7 |

1. INTRODUCTION

1.1. **The theme.** We will introduce the Nuclear norm, the Schatten p norms, the Ky Fan norms, and some of their properties. In particular, we will show that these norms are indeed matrix norms, that they are convex functions of the matrix entries and that they are unitarily invariant.

1.2. **The intention.** One begins to understand new mathematics only when one can explain it in terms of simpler and familiar mathematics. The intention is to explain these properties in terms of elementary linear algebra, to the extent possible in 1.5 hours. We will get a feel for a good variety of ideas which go into these proofs.

1.3. **The proofs.** The proofs and ideas presented here come from [1] and [2].

There are over 20 theorems and lemmas involved, and it is a joy to read through them all. We will see proofs of important claims in detail. But, in the interests of time, we will omit proofs to other subsidiary claims.

Also, though I have done the algebra, I have not yet satisfactorily tried to view various claims from the geometric perspective.

1.4. **Outline.** First we will define some norms, consider some ideas about stochastic matrices, majorization and σ inequalities, then we will show the normness of Ky Fan norms and Schatten p norms.

1.5. **Notation.** Unless specified otherwise, any matrix is assumed to be $m \times n$. Let $q = \min\{m, n\}$. (a_i) represents a vector whose i th element is a_i . $(a_i) \downarrow$ represents that in this vector, all elements are arranged in descending order.

1.6. **Things to leave on the whiteboard.** The outline. The definition of norms, matrix norms, Ky Fan and Schatten norms, Birkhoff's theorem, majorization.

2. THE NORMS

2.1. **The intention.** We will now define some norms of interest.

2.2. Vector and matrix norms. Any vector norm obeys non negativity, positive definiteness, homogeneity / scalability ($\|cv\| = c\|v\|$) and triangle inequality.

In addition to this, matrix norms obey submultiplicativity: $\|AB\| \leq \|A\| \|B\|$.

2.2.1. Absolute norm. Absolute norm: $\|a\| = \||a|\|$. Absoluteness iff monotonicity ($|x| \leq |y| \implies \|x\| \leq \|y\|$).

2.3. Ky Fan (p,k) norms. Take σ_i in descending order. $\|A\|_{p,k} = (\sum_{i=1}^k \sigma_i^p)^{1/p}$ for $p \geq 1$: p norm to top k σ .

2.4. Schatten p norms. Apply p norm to singular values. Special case of Ky Fan norm: $\|A\|_{p,q} = \|A\|_{S_p} = (\sum \sigma_i^p)^{1/p}$.

2.4.1. Frobenius (Hilbert-Schmidt, Euclidian) norm. $\|A\|_{S_2} = \|A\|_F$.
Optional info. $(\sum a_{i,j}^2)^{\frac{1}{2}} = (\sum \|a_j\|^2)^{\frac{1}{2}} = (tr A^* A)^{\frac{1}{2}} = (tr A A^*)^{\frac{1}{2}} = (tr \Sigma^* \Sigma)^{\frac{1}{2}} = \|A\|_F$. So, based on matrix inner product: $\langle A, B \rangle = tr(B^* A)$.

2.4.2. Trace (Nuclear) norm. $\|A\|_{S_1} = \|A\|_{tr} = \sum \sigma_i = tr((A^* A)^{1/2})$.

2.5. Convexity. Convexity of vector norms is a direct consequence of their homogeneity and congruence with the Δ inequality.

3. STOCHASTIC MATRICES

3.1. The intention. This is an important idea we will need in proving important ideas about majorization, and in our proofs of normness. We will introduce the ideas, and will return to the proofs later if we still have time in the end.

3.2. Bistochastic matrix. S is bistochastic if $S \geq 0$, every row and column sums to 1.

For permutation matrix P, PS or SP also stochastic.

(Birkhoff): $\{S\}$ = set of finite convex combinations of permutation matrices P_i .

3.3. Doubly substochastic matrix Q. Q is doubly Substochastic if $Q \geq 0$, every row and column sums to at most 1.

4. IDEAS ABOUT MAJORIZATION

4.1. The intention. Majorization is an important idea we will need in deriving singular value inequalities, and in our proofs of normness. We will introduce the ideas, and will return to the proofs later if we still have time in the end.

4.2. Strong majorization. Take $a, b \in C^m$, rearrange in descending order to get the vectors $(a_{[i]}), (b_{[i]})$, and in ascending order to get $(a_{(i)}), (b_{(i)})$.

$a \preceq b$ (b majorizes a) if $\sum_{i=1}^m a_i = \sum_{i=1}^m b_i, \sum_{i=1}^k a_{[i]} \leq \sum_{i=1}^k b_{[i]} \forall k$.

4.2.1. Differing notations. One can get an \equiv notion from using ascending order and saying a majorizes b. Note: [1] and [2] differ slightly in the way they define majorization notations.

4.3. Connection with stochastic matrices. b majorizes a iff \exists doubly stochastic S such that $a = Sb$.

4.4. Weak majorization. Weak majorization (\succeq) is like strong majorization, except the $\sum_{i=1}^m a_i = \sum_{i=1}^m b_i$ condition is omitted.

4.4.1. *Connection with stochastic matrices.* b weakly majorizes $a \geq 0$ iff \exists doubly substochastic Q : $a = Qb$.

b weakly majorizes a iff \exists doubly stochastic S : $a \leq Sb$.

4.4.2. *Weak Majorization preserved under convex increasing fn.* Take convex increasing scalar fn f . Let $f(a)$ denote the vector $(f(a_i))$. If b weakly majorizes a , then $f(b)$ weakly majorizes $f(a)$.

Proof sketch(3.3.8 in [2]):

As b weakly majorizes a ($a \preceq b$), for some doubly stochastic Q , $a \leq Qb$.

Due to monotonicity of f , $f(a) \leq f(Qb)$. So, $f(a) \preceq f(Qb)$.

By Birkhoff's thm, for $\alpha_i \geq 0$, $\sum \alpha_i = 1$, $f(Qb) = f((\sum \alpha_i P_i)b) \leq \sum \alpha_i f((P_i)b)$ (using the convexity of f) $= \sum \alpha_i P_i f(b) = Qf(b)$. So, $f(Qb) \preceq f(b)$. Hence, we have that $f(a) \preceq f(Qb) \preceq f(b)$. \square

A corollary: If $0 \leq a, b$ are vectors with entries in descending order, if $\prod_{i=1}^k a_i \leq \prod_{i=1}^k b_i$, and if g is such that $g(e^x)$ is convex increasing, then $g(b)$ weakly majorizes (\succeq) $g(a)$.

Proof sketch: Taking log on both sides, we get $\log a \preceq \log b$. Then, using $f(x) = g(e^x)$ we get $g(a) \preceq g(b)$. In this sketch, we will ignore cases where $a_{i>k} = 0$.

5. SINGULAR VALUE Σ PROPERTIES

5.1. **Motivation.** Singular value inequalities are important in proving the vector and matrix normness of Ky Fan norms.

5.2. **Unitary invariance.** $\Sigma \in R^{mn}$ always, so $\Sigma = \Sigma^*$. $\Sigma = U^*AV$: so, Σ is unitary invariant: $\sigma_i(A) = \sigma_i(Q_1AQ_2)$.

5.3. **Trace of A.** $tr(A) = \sum_i a_{i,i}$. A linear map: $tr(kA + lB) = k tr(A) + l tr(B)$. $tr(AB) = tr(BA)$. Similarity invariant: $tr(P^{-1}AP) = tr(APP^{-1}) = tr(A)$. So, $tr(A) = tr(QTQ^*) = \sum \lambda_i$.

5.4. **Effect of row or column deletion.** A_r : A with r rows or cols deleted; $\sigma_k(A) \geq \sigma_k(A_r) \geq \sigma_{k+r}(A)$. (Proof only for 2 hr talk.)

Proof sketch: (3.3.1 [2]) Prove for $r=1$, get general case thence. Suppose sth col is deleted: For upper bound, use $\sigma_k(A_1) = \max_{S \subset C^n, dim(S)=k} \min_{x \in S} \|Ax\|_2$ with extra cond: $x \perp e_s$; for lower bound use $\sigma_k(A_1) = \min_{S \subset C^n, dim(S)=n-k} \max_{x \in S} \|Ax\|_2$ with extra cond. For row deletion, consider A^* .

5.5. **The inequality** $\prod_{i=1}^k \sigma_i(X_k^*AY_k) \leq \prod_{i=1}^k \sigma_i(A)$. X_k, Y_k have \perp columns.

Proof sketch:

By block mult, for any arbit sq orth X, Y , $S_k = X_k^*AY_k$ is the upper left submatrix of $S = X^*AY$.

So, $\sigma_i(X_k^*AY_k) = \sigma_i(S_k) \leq \sigma_i(S) = \sigma_i(X^*AY) = \sigma_i(A)$. So, $|\det(X_k^*AY_k)| = \prod_{i=1}^k \sigma_i(X_k^*AY_k) \leq \prod_{i=1}^k \sigma_i(A)$.

5.6. **The inequality** $\sum_{i=1}^k |\lambda_i(A)|^p \leq \sum_{i=1}^k \sigma_i(A)^p$. **Proof sketch:**

For square A , take $A = QTQ^*$, $T = Q^*AQ$, $(\lambda_i) \downarrow$; and take the k -principal submatrix T_k . By block multiplication, we get: $T_k = Q_k^*AQ_k$.

So, $|\det(T_k)| = \prod_{i=1}^k |\lambda_i(A)| = |\det(Q_k^*AQ_k)| \leq \prod_{i=1}^k \sigma_i(A)$ using a lemma proved earlier. Equality holds for $k=m$. (3.3.2 [2])

By majorization theorems, $|tr(A)| \leq \sum_{i=1}^q |\lambda_i(A)| \leq \sum_{i=1}^q \sigma_i(A)$. Also, for any $p \geq 1$, $\sum_{i=1}^k |\lambda_i(A)|^p \leq \sum_{i=1}^k \sigma_i(A)^p$.

5.7. Polar decomposition. $m \leq n$: take SVD $A = U[\Sigma \ 0][V_1 \ V_2]^* = U\Sigma V_1^*$, $P^2 = AA^* = U\Sigma^2 U^*$: +ve semidefinite; take $P = U\Sigma U^*$: Hermitian +ve semidefinite. So, $A = U\Sigma V_1^* = PUV_1^* = PY$, where Y has orthonormal rows.

So, if $m \geq n$: $A = YQ$ for Hermitian +ve semidefinite Q, Y with orth columns: apply thm to A^* .

5.8. The inequality $\sum_{i=1}^k \sigma_i(AB)^p \leq \sum_{i=1}^k \sigma_i(A)^p \sigma_i(B)^p$. **Proof sketch:** $A \in C^{m \times p}$, $B \in C^{p \times n}$. Take $AB = U\Sigma V^*$; $U_k^* ABV_k = \Sigma_k$. Take polar decomposition of $BV_k = X_k Q$; $Q^2 = V_k^* B^* B V_k$.

Thus, $det(Q^2) \leq \prod_{i=1}^k \sigma_i(B^* B) = \prod_{i=1}^k \sigma_i(B)^2$, and $det(Q) = \prod_{i=1}^k \sigma_i(B)$. So, $\prod_{i=1}^k \sigma_i(AB) = |det(U_k^* ABV_k)| = |det(U_k^* AX_k) det(Q)| \leq \prod_{i=1}^k \sigma_i(A) \sigma_i(B)$ using a lemma proved earlier. (3.3.4 [2])

By majorization theorems, for $p \geq 1$, $\sum_{i=1}^k \sigma_i(AB)^p \leq \sum_{i=1}^k \sigma_i(A)^p \sigma_i(B)^p$.

5.9. k partial isometry. $A = U\Sigma V^*$ with $\Sigma = \begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix}$.

5.10. An identity about partial sum of top k singular values. $\sum_{i=1}^k \sigma_i(A) = \max \{ |tr(X^* AY)| : X^* X = Y^* Y = I, X \in C^{m \times k}, Y \in C^{n \times k} \}$
 $= \max \{ |tr(AC)| : C \in C^{n \times m} \text{ is rank k partial isometry} \}$.

Proof sketch: We can get C from Y, X by taking $C = YX^*$. Then, $\sigma_i(C^* C) = \sigma_i(XX^*) = \sigma_i(X^* X) = 1$, and we confirm that C is a rank k partial isometry.

We can also get Y, X from C: use SVD: $C = U\Sigma V^* = U_k V_k^*$.

$|tr(AC)| = | \sum_{i=1}^q \lambda_i(AC) | \leq \sum_{i=1}^q |\lambda_i(AC)|$
 $\leq \sum_{i=1}^q \sigma_i(AC) \leq \sum_{i=1}^q \sigma_i(A) \sigma_i(C) = \sum_{i=1}^k \sigma_i(A)$ using properties shown earlier.
 Take $A = U\Sigma V^*$, then for rank k isometry $C = V \hat{I}_k U^*$, $tr(AC) = tr(U\Sigma \hat{I}_k U^*) = \sum_{i=1}^k \sigma_i(A)$.

5.11. The inequality $\sum_{i=1}^k \sigma_i(A+B) \leq \sum_{i=1}^k \sigma_i(A) + \sum_{i=1}^k \sigma_i(B)$. **Proof sketch** (3.4.3 [2]) :

Let C be some rank k partial isometry.

$\sum_{i=1}^k \sigma_i(A+B) = \max \{ |tr((A+B)C)| \} \leq \max \{ |tr(AC)| + |tr(BC)| \}$
 $\leq \max \{ |tr(AC)| \} + \max \{ |tr(BC)| \} = RHS$.

6. PROOFS OF NORMNESS

6.1. Unitary invariance of norms. If $\|\cdot\|$ unitary invariant, by SVD, $\|A\| = \|\Sigma\|$.

6.1.1. Symmetric gauge fn g. $g : C^q \rightarrow R^+$ is a vector norm on C^q , and is also an absolute norm, and is permutation invariant: $g(Px) = g(x)$. So, it is a fn on a set rather than a sequence.

Given a unitarily invariant $\|\cdot\| : g(x) = \|X\| : X = diag(x)$ is symm gauge. The $diag(x)$ function returns a $m \times n$ matrix whose diagonal elements correspond to the elements of x.

Proof sketch(3.5.18 [2]): Normness of g comes from normness of $\|\cdot\|$. Permutation invariance and absoluteness come from unitary invariance of $\|\cdot\|$. \square

Given symmetric gauge g, $\|X\| = g(\sigma)$ is unitary invariant matrix norm.

Proof sketch(3.5.18 [2]): Unitary invariance of $\|\cdot\|$ follows from unitary invariance of Σ . As g is vector norm, we get +ve definiteness, non negativity, homogeneity. We next prove Δ inequality.

g is absolute, so it is monotone. $\sigma(A+B)$ weakly majorized by $\sigma(A) + \sigma(B)$, so $\sigma(A+B) \leq S[\sigma(A) + \sigma(B)]$ for some doubly stochastic S . So, using monotonicity of g , $g(\sigma(A+B)) \leq g(S(\sigma(A) + \sigma(B))) = g(\sum_i \alpha_i P_i(\sigma(A) + \sigma(B)))$ (using Birkhoff's thm) $\leq \sum \alpha_i (g(P_i \sigma(A)) + g(P_i \sigma(B)))$ (using homogeneity and Δ inequality of g) $\leq g(\sigma(A)) + g(\sigma(B)) = \|A\| + \|B\|$, using permutation invariance of g .

6.2. Ky Fan (p, k) norms. We have already seen Δ inequality for (1, k) norm.

6.2.1. Schatten p norms. Normness of the Ky Fan (p, q) norms follows easily from the fact that p norms are symmetric gauge functions. So, we are mostly done with them.

6.2.2. The general case. The following is optional, and will be presented in case there is time left:

We now show the normness of general (p, k) norms by showing that their dependence on $\sigma(A)$ is a symmetric gauge.

Consider the symmetric gauge $\|x\|_{p,k}$ corresponding to $\|A\|_{p,k}$. We want to show that this is indeed a symmetric gauge. Only Δ inequality is nontrivial. : **Proof sketch:** Take a, b in absolute descending order to get the vectors $a' = (a_{[i]}), b' = (b_{[i]})$.

We can show by induction on k that $\sum_{i=1}^k (|a'| + |b'|)_i = \sum_{i=1}^k (|a_{[i]}| + |b_{[i]}|) \geq \sum_{i=1}^k (|a| + |b|)_{[i]} \geq \sum_{i=1}^k |a + b|_{[i]}$. So, $|a'| + |b'|$ weakly majorizes $a + b$.

So, by weak majorization theorems, for $p \geq 1$: $\sum_{i=1}^k (|a'| + |b'|)_i^p = \sum_{i=1}^k (|a_{[i]}| + |b_{[i]}|)^p \geq \sum_{i=1}^k |a + b|_{[i]}^p$.

So, raising both sides to $1/p$, we have that $\| |a'| + |b'| \|_{p,k} \geq \|a + b\|_{p,k}$.

Now, consider the vectors $a'_{(k)}$ and $b'_{(k)}$, which are a' and b' with $a'_i = b'_i = 0 \forall i > k$. By the Δ inequality of the p norm, we have that $\| |a'| + |b'| \|_{p,k} = \| |a'_{(k)}| + |b'_{(k)}| \|_p \leq \| |a'_{(k)}| \|_p + \| |b'_{(k)}| \|_p = \|a'\|_{p,k} + \|b'\|_{p,k} = \|a\|_{p,k} + \|b\|_{p,k}$. \square

6.2.3. Matrix normness. Matrix normness is seen using inequalities seen earlier: $\sum_{i=1}^k \sigma_i(AB)^p \leq \sum_{i=1}^k \sigma_i(A)^p \sigma_i(B)^p \leq \sum_{i=1}^k \sigma_i(A)^p \sum_{i=1}^k \sigma_i(B)^p$.

7. INTERESTING IDEAS ABOUT DOUBLY SUBSTOCHASTIC MATRIX Q

This is unlikely to be covered.

Q is dbl substochastic iff B has dbl stochastic dilation S : make deficiency vectors d_r, d_c ; get difference matrices $D_r = \text{diag}(d_r), D_c = \text{diag}(d_c)$; get $S = \begin{pmatrix} Q & D_r \\ D_c^T & Q^T \end{pmatrix}$.

$\{Q\}$ equivalent to set of convex combos of partial permutation matrices: Dilate Q to S , get finite convex combo of P_i , take the convex combo of principal submatrices.

$Q \in C^{n \times n}$ is dbl substochastic iff \exists dbl stochastic $S \in C^{n \times n}$ with $A \geq B$: Take any Q , get finite convex combo of partial permutation matrices; alter each to get permutation matrix; their convex combo is S .

REFERENCES

- [1] Horn and Johnson. *Matrix Analysis*. CUP, 1985.
- [2] Horn and Johnson. *Topics in Matrix Analysis*. CUP, 1991.