Charles Zheng EE 378b HW 3

1. a. Since the space of unit vectors is compact, there exists x^0 such that $||M||_2 = ||Mx^0||_2$. For the same reason, there exist unit vectors x°, y° such that

$$\langle x^\circ, My^\circ \rangle = \max_{||x||=||y||=1} \langle x, My \rangle$$

Letting $y^0 = Mx^0/||M||_2$, we have $||y||_2 = 1$. Therefore

$$||M||_2 = (y^0)^* M x^0 \leq \max_{||y||=||x||=1} \langle y, Mx \rangle$$

Also, by Cauchy-Schwarz we have

$$\max_{||x||=||y||=1} \langle x, My \rangle = (x^\circ)^* M y^\circ \leq ||x^\circ||_2 ||My^\circ||_2 \leq ||My^\circ||_2 \leq \max_{||y||=1} ||My||_2 = ||M||_2$$

Having shown that

$$||M||_2 \leq \max_{||x||=||y||=1} \langle x, My \rangle \leq ||M||_2$$

we conclude that the definitions are equivalent

b. Let the SVD of M be written $M = UDV^T$ where $D = diag(\sigma_1, \ldots, \sigma_n)$. Let v_1 be the first column of V, then $||v_1||_2 = 1$ and

$$||Mv_1||_2 = ||UDV^T v_1||_2 = ||UDe_1||_2 = ||De_1||_2 = \sigma_1$$

Hence

$$\sigma_1 = ||Mv_1||_2 \leq \max_{||x||=1} ||Mx||_2 = ||M||_2$$

Meanwhile for any unit vector x, defining $y = V^T x$ we have $||y||_2 \leq 1$. Then

$$\max_{||x||=1} ||Mx||_2 = \max_{||x||=1} ||UDV^T x||_2 \leq \max_{||x||=1} ||DV^T x||_2 \leq \max_{||y||=1} ||Dy||_2$$

But defining $a_i = y_i^2$,

$$\max_{||y||=1} ||Dy||_2^2 = \max_{||y||=1} \sum_{i=1}^n \sigma_i^2 y_i^2 = \max_{\sum a_i = 1, a_i \geq 0} \sum_{i=1}^n \sigma_i^2 a_i$$

is maximized by $a = e_1$, hence $\max_{||y||=1} ||Dy||_2 = \sigma_1$.

Having shown that

$$\sigma_1 \leq ||M||_2 \leq \sigma_1$$

we conclude that the two definitions are equivalent.

2. a. From 1a we have

$$||M^*||_2 = \max_{||x||=||y||=1} \langle x, M^* y\rangle = \max_{||x||=||y||=1} \langle y, Mx\rangle = ||M||_2$$

b. We have

$$||AB||_2 = \max_{||x||=1} ||ABx||_2 = \max_{y=Bx \text{ for some } ||x||=1} ||Ay||_2$$

Meanwhile, if $y = Bx$, and $||x||_2 = 1$, we have $||y||_2 \leq ||B||_2$. Therefore the set $\{y : y = Bx \text{ for some } x \text{ such that } ||x|| = 1\}$ is contained in the set $\{y : ||y||_2 \leq ||B||_2\}$. Hence

$$\max_{y=Bx \text{ for some } ||x||=1} ||Ay||_2 \leq \max_{||y||=||B||_2} ||Ay||_2 = ||A||_2||B||_2$$

3. i.

$$||aM||_2 = \max_{||x||=1} ||aMx||_2 = \max_{||x||=1} |a|||M_2x||_2 = |a| \max_{||x||=1} ||Mx||_2 = a||M||_2$$

ii.

$$||A+B||_2 = \max_{||x||=1} ||Ax+Bx||_2 \leq \max_{||x||=1} ||Ax||_2+||Bx||_2 \leq \max_{||x||=1} ||Ax||_2+\max_{||x||=1} ||Bx||_2 = ||A||_2+||B||_2$$

iii. Proof of contrapositive: If $M \neq 0$, then some column M_i is nonzero. But then $||Me_i||_2 = ||M_i||_2 > 0$, so $||M||_2 > 0$.

4. a. Let m_i denote the columns of M^*.

$$||M||_2 = ||M^*||_2 = \max_{||x||=1} M^*x = \left\| \sum_{i=1}^{m} x_i m_i \right\| \leq \sum_{i=1}^{m} x_i ||m_i||_2$$

Now let μ be the vector defined by $\mu_i = ||m_i||_2$. By Hölder's inequality, we have

$$||M||_2 \leq \max_{||x||_2=1} \langle x, \mu\rangle \leq \max_{||x||_2=1} ||x||_1||\mu||_\infty$$

But $\max_{||x||_2=1} ||x||_1 = \sqrt{m}$ and $||\mu||_\infty = \max_i ||m_i||_2$. Hence

$$||M||_2 \leq \sqrt{m} \max_i ||m_i||_2$$

as needed.

We see that the upper bound is tight by taking $M = 1_m 1_n^T$, in which case $m_1 = \ldots = m_m = 1_n$

$$||M||_2 = ||M\frac{1}{\sqrt{n}}1_n||_2 = ||\sqrt{n}1_m||_2 = \sqrt{m}||1_n||_2 = \sqrt{m}\max_i ||m_i||_2$$

b. Using the fact that $||x||_2 \geq \frac{1}{\sqrt{m}}||x||_1$ for any vector $x \in \mathbb{R}^m$, we have

$$||M||_2 \geq ||M\frac{1}{\sqrt{n}}1_n||_2 = ||\mu||_2 \geq \frac{1}{\sqrt{m}}||\mu||_1 = \frac{1}{\sqrt{mn}}\sum_{i=1}^{m}|m_i^* 1_n|$$

where μ is the m-vector with $m_i = \frac{1}{\sqrt{n}}m_i^* 1_n$. Hence the upper bound is proved.

To show that the upper bound is tight, take $M = 1_m 1_n^T$, in which case

$$||M||_2 = \sqrt{mn} = \frac{1}{\sqrt{mn}}mn = \frac{1}{\sqrt{mn}}\sum_{i=1}^{m}|m_i^* 1_n|$$

5. a. We have

$$||M||_F = \sqrt{\sum_i \sum_j M_{ij}^2} = \sqrt{\mathrm{tr}M^T M} = \sqrt{\mathrm{tr}M^T M}$$

Without loss of generality assume $n \leq m$. Since the eigenvalues of $M^T M$ are $\sigma_1^2, \ldots, \sigma_n^2$, we have

$$\mathrm{tr}M^T M = \sum_{i=1}^{n}\sigma_i^2$$

hence dropping the assumption that $n \leq m$, we have

$$||M||_F = \sqrt{\sum_{i=1}^{\min(n,m)}\sigma_i^2}$$

b. Only $\sigma_1, \ldots, \sigma_r$ are nonzero, where $r = \mathrm{rank}(M)$. Letting $d = (\sigma_1, \ldots, \sigma_r)$, we have $||M||_F = ||d||_2$. But we know that for all vectors $x \in \mathbb{R}^r$,

$$||x||_\infty \leq ||x||_2 = \sqrt{\sum_{i=1}^{r}x_i^2} \leq \sqrt{\sum_{i=1}^{r}||x||_\infty^2} = \sqrt{r}||x||_\infty$$

Hence

$$||M||_2 = \sigma_1 = ||d||_\infty \leq ||d||_2 = ||M||_F = ||d||_2 \leq \sqrt{r}||d||_\infty = \sqrt{r}||M||_2$$

as needed.