By: Alex Levis alevis@g.harvard.edu

In this note, we expand some arguments in this post to prove Weyl's inequality, a result bounding the distance between singular values of two matrices (or alternatively bounding the distance between eigenvalues of two symmetric matrices).

Review of singular value decomposition

Let $A \in \mathbb{R}^{m\times n}$ be an arbitrary $m \times n$ real matrix. Let $r = \text{rank}(A)$, and the singular value decomposition of A be given by $A = U\Sigma V^T = \sum_{i=1}^r \sigma_i u_i v_i^T$, where $U = [u_1 \cdots u_m] \in \mathbb{R}^{m\times m}$, $V = [v_1 \cdots v_n] \in \mathbb{R}^{n\times n}$ are orthogonal, and $\Sigma \in \mathbb{R}^{m\times n}$ a rectangular diagonal matrix, i.e., the (i,j)-th entry of Σ is 0, for $i \neq j$, and the (i,i)-th entry of Σ is $\sigma_i \geq 0$, for $i = 1,\ldots,\min(m,n)$ — we can arrange the singular values of A such that $\sigma_1 \geq \cdots \geq \sigma_r > 0 = \sigma_{r+1} = \cdots = \sigma_{\min(m,n)}$. Recall that $v_1,\ldots,v_n$ are eigenvectors of symmetric matrix A^TA with corresponding eigenvalues $\sigma_1^2,\ldots,\sigma_r^2,0,\ldots,0$, and $v_1,\ldots,v_r$ form an orthonormal basis for $\mathcal{C}(A^TA) = \mathcal{R}(A) \subseteq \mathbb{R}^n$. Similarly, $u_1,\ldots,u_m$ are eigenvectors of symmetric matrix AA^T with corresponding eigenvalues $\sigma_1^2,\ldots,\sigma_r^2,0,\ldots,0$, and $u_1,\ldots,u_r$ form an orthonormal basis for $\mathcal{C}(AA^T) = \mathcal{C}(A) \subseteq \mathbb{R}^m$.

For matrix A, we defined the operator norm as

$$\|A\|_{\text{op}} = \sup_{x\in\mathbb{R}^n\setminus\{\mathbf{0}\}} \frac{\|Ax\|}{\|x\|} = \sup_{x\in S^{n-1}} \|Ax\|,$$

where $S^{n-1} = \{x \in \mathbb{R}^n : \|x\| = 1\}$ is the unit sphere in $\mathbb{R}^n$. We showed that $\|A\|_{\text{op}} = \sigma_1$, using a result from homework:

$$\|A\|_{\text{op}}^2 = \sup_{x\in S^{n-1}} x^T(A^TA)x = \sigma_1^2,$$

since σ_1^2 is the maximum eigenvalue of symmetric matrix A^TA.

Variational characterizations of singular values

Let $\mathcal{V}_k^{(p)} = \{V \subseteq \mathbb{R}^p \,|\, V \text{ a subspace}, \dim(V) = k\}$ be the class of all k-dimensional subspaces of $\mathbb{R}^p$, for any $p,k \in \mathbb{N}$ with $k \leq p$.

Theorem (Fischer-Courant)**.** *With $A \in \mathbb{R}^{m\times n}$ given as above, and $k \in \{1,\ldots,\min(m,n)\}$,*

$$\sigma_k = \sup\left\{\inf_{x\in V\cap S^{n-1}} \|Ax\| : V \in \mathcal{V}_k^{(n)}\right\} = \inf\left\{\sup_{x\in V\cap S^{n-1}} \|Ax\| : V \in \mathcal{V}_{n-k+1}^{(n)}\right\}.$$

Proof. For the first equality, let $V \in \mathcal{V}_k^{(n)}$ be arbitrary, and note that $V \cap \mathscr{L}(v_k,\ldots,v_n) \neq \{\mathbf{0}\}$, since otherwise we could form a set of $k+(n-k+1) = n+1$ linearly independent vectors in $\mathbb{R}^n$ (recall result from first homework). Thus, we can take $z \in V \cap \mathscr{L}(v_k,\ldots,v_n)$ such that $\|z\| = 1$ (take any non-zero vector and normalize it), so there are constants $a_k,\ldots,a_n \in \mathbb{R}$ such that

$$z = \sum_{i=k}^n a_i v_i,$$

and $1 = \|z\|^2 = \sum_{i=k}^n a_i^2\|v_i\|^2 = \sum_{i=k}^n a_i^2$. Observe that

$$Az = \sum_{i=1}^{\min(m,n)} \sum_{j=k}^n a_j\sigma_i u_i v_i^T v_j = \sum_{i=k}^{\min(m,n)} a_i\sigma_i u_i.$$

Hence, as σ_k is greater than all subsequent singular values,

$$\inf_{x\in V\cap S^{n-1}} \|Ax\| \leq \sqrt{\|Az\|^2} = \sqrt{\sum_{i=k}^{\min(m,n)} a_i^2\sigma_i^2} \leq \sqrt{\sigma_k^2 \sum_{i=k}^n a_i^2} = \sqrt{\sigma_k^2} = \sigma_k,$$

showing that $\sigma_k \geq \sup\left\{\inf_{x\in V\cap S^{n-1}} \|Ax\| : V \in \mathcal{V}_k^{(n)}\right\}$.

Conversely, set $V = \mathscr{L}(v_1, \ldots, v_k) \in \mathcal{V}_k^{(n)}$, and let $x \in V \cap S^{n-1}$ be an arbitrary unit vector in this subspace. Then there are constants $b_1, \ldots, b_k \in \mathbb{R}$ such that

$$x = \sum_{i=1}^{k} b_i v_i, \implies Ax = \sum_{i=1}^{\min(m,n)} \sum_{j=1}^{k} b_j \sigma_i u_i v_i^T v_j = \sum_{i=1}^{k} b_i \sigma_i u_i,$$

and $1 = \|x\|^2 = \sum_{i=1}^{k} b_i^2 \|v_i\|^2 = \sum_{i=1}^{k} b_i^2$. Thus,

$$\|Ax\| = \sqrt{\sum_{i=1}^{k} b_i^2 \sigma_i^2} \geq \sigma_k \sqrt{\sum_{i=1}^{k} b_i^2} = \sigma_k,$$

so $\inf_{x \in V \cap S^{n-1}} \|Ax\| \geq \sigma_k$. But this infimum is attained at $x = v_k$, as $Av_k = \sigma_k u_k$ has norm σ_k. This shows that $\sigma_k = \inf_{x \in V \cap S^{n-1}} \|Ax\|$, so $\sigma_k \leq \sup \left\{ \inf_{x \in V \cap S^{n-1}} \|Ax\| : V \in \mathcal{V}_k^{(n)} \right\}$, establishing the first equality.

The second equality is shown in an analogous fashion, but we will elaborate for completeness. Let $V \in \mathcal{V}_{n-k+1}^{(n)}$ be arbitrary, and note that $V \cap \mathscr{L}(v_1, \ldots, v_k) \neq \{\mathbf{0}\}$, since otherwise we could form a set of $k + (n-k+1) = n+1$ linearly independent vectors in $\mathbb{R}^n$. Thus, we can take $z \in V \cap \mathscr{L}(v_1, \ldots, v_k)$ such that $\|z\| = 1$, so there are constants $a_1, \ldots, a_k \in \mathbb{R}$ such that

$$z = \sum_{i=1}^{k} a_i v_i,$$

and $1 = \|z\|^2 = \sum_{i=1}^{k} a_i^2 \|v_i\|^2 = \sum_{i=1}^{k} a_i^2$. Observe that

$$Az = \sum_{i=1}^{\min(m,n)} \sum_{j=1}^{k} a_j \sigma_i u_i v_i^T v_j = \sum_{i=1}^{k} a_i \sigma_i u_i.$$

Hence, as σ_k is less than all previous singular values,

$$\sup_{x \in V \cap S^{n-1}} \|Ax\| \geq \sqrt{\|Az\|^2} = \sqrt{\sum_{i=1}^{k} a_i^2 \sigma_i^2} \geq \sqrt{\sigma_k^2 \sum_{i=1}^{k} a_i^2} = \sqrt{\sigma_k^2} = \sigma_k,$$

showing that $\sigma_k \leq \inf \left\{ \sup_{x \in V \cap S^{n-1}} \|Ax\| : V \in \mathcal{V}_{n-k+1}^{(n)} \right\}$.

Conversely, set $V = \mathscr{L}(v_k, \ldots, v_n) \in \mathcal{V}_{n-k+1}^{(n)}$, and let $x \in V \cap S^{n-1}$ be an arbitrary unit vector in this subspace. Then there are constants $b_k, \ldots, b_n \in \mathbb{R}$ such that

$$x = \sum_{i=k}^{n} b_i v_i, \implies Ax = \sum_{i=1}^{\min(m,n)} \sum_{j=k}^{n} b_j \sigma_i u_i v_i^T v_j = \sum_{i=k}^{\min(m,n)} b_i \sigma_i u_i,$$

and $1 = \|x\|^2 = \sum_{i=k}^{n} b_i^2 \|v_i\|^2 = \sum_{i=k}^{n} b_i^2$. Thus,

$$\|Ax\| = \sqrt{\sum_{i=k}^{\min(m,n)} b_i^2 \sigma_i^2} \leq \sigma_k \sqrt{\sum_{i=k}^{n} b_i^2} = \sigma_k,$$

so $\sup_{x \in V \cap S^{n-1}} \|Ax\| \leq \sigma_k$. But this supremum is attained at $x = v_k$, as $Av_k = \sigma_k u_k$ has norm σ_k. This shows that $\sigma_k = \sup_{x \in V \cap S^{n-1}} \|Ax\|$, so $\sigma_k \geq \inf \left\{ \sup_{x \in V \cap S^{n-1}} \|Ax\| : V \in \mathcal{V}_{n-k+1}^{(n)} \right\}$, as desired. $\square$

Singular values of sums of matrices

We will now apply the Fischer-Courant characterization of singular values to relate the singular values of a matrix sum to those of its summands. For arbitrary matrix $A \in \mathbb{R}^{m\times n}$, we know that singular values are uniquely defined as the square root of the eigenvalues of the symmetric positive semi-definite matrix A^TA (or equivalently AA^T), so we can define the function $\sigma_k : \mathbb{R}^{m\times n} \to [0,\infty)$, for $k = 1,\dots,\min(m,n)$, such that $\sigma_1(A) \geq \sigma_2(A) \geq \cdots \geq \sigma_{\min(m,n)}(A)$ are the ordered singular values of A for any $A \in \mathbb{R}^{m\times n}$. Note that $\sigma_1(A) = \|A\|_{\mathrm{op}}$, and for $k = 1,\dots,\min(m,n)$, $\sigma_k(A) = \sigma_k(-A)$, since the identical matrices $A^TA = (-A)^T(-A)$ share the same eigenvalues.

Theorem. *Let A, $B \in \mathbb{R}^{m\times n}$ be any two real matrices of the same dimension, then*

$$\sigma_{k+\ell-1}(A+B) \leq \sigma_k(A) + \sigma_\ell(B),$$

for any $k,\ell \in \{1,\dots,\min(m,n)\}$ such that $k+\ell-1 \leq \min(m,n)$.

Proof. By Fischer-Courant,

$$\sigma_{k+\ell-1}(A+B) = \inf\left\{\sup_{x\in V\cap S^{n-1}} \|Ax+Bx\| : V \in \mathcal{V}^{(n)}_{n-k-\ell+2}\right\}, \tag{1}$$

and we can write similar formulae for $\sigma_k(A)$ and $\sigma_\ell(B)$. In the proof of Fischer-Courant, we showed that the infimum over subspaces is exactly attained, so we can take $V_A \in \mathcal{V}^{(n)}_{n-k+1}$ and $V_B \in \mathcal{V}^{(n)}_{n-\ell+1}$ such that $\sigma_k(A) = \sup_{x\in V_A\cap S^{n-1}}\|Ax\|$ and $\sigma_\ell(B) = \sup_{x\in V_B\cap S^{n-1}}\|Bx\|$. Let $W = V_A \cap V_B$, and note that

$$\dim(W) = \dim(V_A) + \dim(V_B) - \dim(V_A+V_B) \geq (n-k+1) + (n-\ell+1) - n = n-k-\ell+2.$$

Moreover, by the choice of V_A and V_B,

$$\sup_{x\in W\cap S^{n-1}} \|Ax+Bx\| \leq \sup_{x\in W\cap S^{n-1}} (\|Ax\| + \|Bx\|) \leq \sup_{x\in V_A\cap S^{n-1}} \|Ax\| + \sup_{x\in V_B\cap S^{n-1}} \|Bx\| = \sigma_k(A) + \sigma_\ell(B).$$

We could achieve the same bound by replacing W with any $V \in \mathcal{V}^{(n)}_{n-k-\ell+2}$ such that $V \subseteq W$, as then $\sup_{x\in V\cap S^{n-1}}\|Ax+Bx\| \leq \sup_{x\in W\cap S^{n-1}}\|Ax+Bx\|$. Hence,

$$\sigma_k(A) + \sigma_\ell(B) \geq \sigma_{k+\ell-1}(A+B),$$

by formula (1). □

Corollary. *The functions $\sigma_k : \mathbb{R}^{m\times n} \to [0,\infty)$ are Lipschitz continuous with respect to the operator norm, with Lipschitz constant 1. In other words, for any $A, B \in \mathbb{R}^{m\times n}$ and $k \in \{1,\dots,\min(m,n)\}$,*

$$|\sigma_k(A) - \sigma_k(B)| \leq \|A-B\|_{\mathrm{op}}.$$

Proof. Applying the above theorem with $\ell = 1$, we find

$$\sigma_k(A) = \sigma_k(B + (A-B)) \leq \sigma_k(B) + \sigma_1(A-B),$$

and similarly

$$\sigma_k(B) = \sigma_k(A + (B-A)) \leq \sigma_k(A) + \sigma_1(B-A).$$

Noting that $\sigma_1(A-B) = \sigma_1(B-A) = \|A-B\|_{\mathrm{op}}$, we obtain

$$|\sigma_k(A) - \sigma_k(B)| \leq \|A-B\|_{\mathrm{op}}.$$

□

Analogous results for symmetric matrices

We have seen that any real symmetric matrix $A \in \mathbb{R}^{n\times n}$ has spectral decomposition $V\Lambda V^T = \sum_{i=1}^n \lambda_i v_i v_i^T$, where $V = [v_1 \cdots v_n] \in \mathbb{R}^{n\times n}$ is orthogonal, $\Lambda = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$, and the columns $v_1, \dots, v_n$ are eigenvectors of A with corresponding eigenvalues $\lambda_1, \dots, \lambda_n \in \mathbb{R}$. Since for A symmetric, $A^TA = AA^T = A^2$, which has eigenvalues $\lambda_1^2, \dots, \lambda_n^2$, it is clear that the eigenvalues and singular values of a symmetric matrix are related: if λ is an eigenvalue of A, then $|\lambda| = \sigma_j(A)$ for some $j \in \{1, \dots, n\}$. Given this close relationship between spectral and singular value decompositions, we might expect all of the above results for singular values of matrices to have analogs concerning the eigenvalues of symmetric matrices. This is indeed the case, and we give the details of this below.

In general, let $\mathrm{Sym}_n \subseteq \mathbb{R}^{n\times n}$ denote the space of $n \times n$ real symmetric matrices (remember this is a vector space, i.e., linear combinations of symmetric matrices are symmetric), and let $\lambda_k : \mathrm{Sym}_n \to \mathbb{R}$, $k = 1, \dots, n$ be such that $\lambda_1(A) \geq \cdots \geq \lambda_n(A)$ are the ordered eigenvalues of any symmetric matrix A. Recall from homework that $\lambda_1(A) = \sup_{x \in S^{n-1}} x^TAx$ and $\lambda_n(A) = \inf_{x\in S^{n-1}} x^TAx$. This motivates that we can replace the role of σ_k with λ_k, orthogonal matrix U with V, and $\|Ax\|$ with x^TAx in all the above arguments, to get analogous results — for the two theorems below, check that the logic flows through seamlessly!

Theorem (Fischer-Courant)**.** *For any $A \in \mathrm{Sym}_n$, and $k \in \{1, \dots, n\}$,*

$$\lambda_k(A) = \sup\left\{\inf_{x\in V\cap S^{n-1}} x^TAx : V \in \mathcal{V}_k^{(n)}\right\} = \inf\left\{\sup_{x\in V\cap S^{n-1}} x^TAx : V \in \mathcal{V}_{n-k+1}^{(n)}\right\}.$$

Theorem. *Let A, $B \in \mathrm{Sym}_n$ be any two real symmetric matrices of the same dimension, then*

$$\lambda_{k+\ell-1}(A+B) \leq \lambda_k(A) + \lambda_\ell(B),$$

for any $k, \ell \in \{1, \dots, n\}$ such that $k + \ell - 1 \leq n$.

Corollary (Weyl's Inequality)**.** *For any $A, B \in \mathrm{Sym}_n$,*

$$|\lambda_k(A) - \lambda_k(B)| \leq \|A - B\|_{\mathrm{op}},$$

for all $k \in \{1, \dots, n\}$.

Proof. As before, applying the above theorem with $\ell = 1$, we find

$$\lambda_k(A) = \lambda_k(B + (A - B)) \leq \lambda_k(B) + \lambda_1(A - B),$$

and similarly

$$\lambda_k(B) = \lambda_k(A + (B - A)) \leq \lambda_k(A) + \lambda_1(B - A).$$

Next, note that for any $C \in \mathrm{Sym}_n$, $\sigma_1(C) = \max(|\lambda_1(C)|, |\lambda_n(C)|)$, by the correspondence between eigenvalues and singular values of symmetric matrices. Therefore,

$$\lambda_1(A - B) \leq \sigma_1(A - B) = \|A - B\|_{\mathrm{op}},$$

and

$$\lambda_1(B - A) \leq \sigma_1(B - A) = \sigma_1(A - B) = \|A - B\|_{\mathrm{op}}.$$

Hence,

$$|\lambda_k(A) - \lambda_k(B)| \leq \|A - B\|_{\mathrm{op}}.$$

□