THE MIN-MAX PRINCIPLE

Let A be a symmetric $n \times n$ matrix. The eigenvalues are real and hence we can order them $\lambda_1(A) \le \lambda_2(A) \le \cdots \le \lambda_n(A)$ where we count them in their multiplicity. E.g., if A is 4×4 and the eigenvalues are $1, 2, 3$ and the multiplicity of 2 is two then we write $1 \le 2 \le 2 \le 3$. We denote the corresponding eigenvector by $\vec{u}_1, \dots, \vec{u}_n$ and we choose them to be an orthonormal basis of $\mathbb{R}^n$. The following is a convenient way of writing the matrix A

$$A = \sum_{k=1}^{n} \lambda_k(A) \vec{u}_k \vec{u}_k^T \ . \tag{1}$$

You are used to compute the eigenvalues using the characteristic polynomial. For symmetric matrices there is another way to achieve that, which is already implicit in the proof that any symmetric matrix can be diagonalized. Here is an interesting theorem.

Theorem 0.1. *Let A be a symmetric matrix. For any vector $\vec{x}$ with norm one, i.e., on the unit sphere, consider the function*

$$f(\vec{x}) = \vec{x}^T A \vec{x} \ .$$

Then the minimum of this function on the sphere is attained, i.e., there exists a vector $\vec{x}_1$ such $f(\vec{x}_1) \le f(\vec{x})$ for all $\vec{x}$ on the unit sphere. Moreover the vector $\vec{x}_1$ is an eigenvector and the value $f(\vec{x}_1)$ is the corresponding eigenvalue.

Proof. Using (1) we have

$$\vec{x}^T A \vec{x} = \sum_{k=1}^{n} \lambda_k(A) \vec{x}^T \vec{u}_k \vec{u}_k^T \vec{x} = \sum_{k=1}^{n} \lambda_k(A) (\vec{u}_k^T \vec{x})^2 \ .$$

If we replace all $\lambda_k(A)$ be $\lambda_1(A)$ which is the smallest one we get

$$\vec{x}^T A \vec{x} \ge \lambda_1(A) \|\vec{x}\|^2 \ .$$

If we choose $\vec{x} = \vec{u}_1$ we get equality. □

Recall that whenever a subspace V is invariant under A, then its orthogonal complement is also invariant under A. This is because A is symmetric. *Thus, how do we find the second eigenvalue? Just minimize the function over all vector $\vec{x}$ that are perpendicular to the vector $\vec{x}_1$! Continuing this way we recover all the eigenvalues* Of course, in general we cannot find the vector $\vec{x}_1$ precisely and therefore, we know the lowest eigenvalue only approximately. Nevertheless this point of view is very useful.

As we said, computing eigenvalues is not easy and here is a first variational principle that allows in principle to compute the eigenvalues within some accuracy. This principle goes back a long way, Cauchy knew much of what follows as did Henri Poincaré, Curant and Hilbert who called it the min-max principle. Here is the description: Let $V \subset \mathbb{R}^n$ be a k-dimensional subspace. Then we can compute (in principle)

$$\max_{\vec{x} \in V, \|\vec{x}\|=1} \vec{x}^T A \vec{x} \ .$$

This is a function on the set of sub-spaces having dimension k. E.g., if we choose V to be the span of the first k eigenvectors, then

$$\max_{\vec{x}\in V, \|\vec{x}\|=1} \vec{x}^T A\vec{x} = \lambda_k(A) \ .$$

The problem, however, is that we do not know the subspace associated with the eigenvectors. Next we vary the subspace, keeping the dimension k fixed and set

$$\mu_k(A) = \min_{V\subset\mathbb{R}^n \ , \dim V = k} \max_{\vec{x}\in V, \|\vec{x}\|=1} \vec{x}^T A\vec{x} \ .$$

Theorem 0.2 (Min-Max)**.** *We have that*

$$\mu_k(A) = \lambda_k(A) \ , k = 1, 2, \ldots, n \ .$$

Proof. We reduce the problem to proving two inequalities. First we show that $\mu_k(A) \le \lambda_k(A)$. For this we pick as a subspace V_k the space spanned by the vector $\vec{u}_1, \ldots, \vec{u}_k$. This space has obviously the dimension k. Pick any vector $\vec{x} \in V_k$ which means that we can write it as

$$\vec{x} = \sum_{j=1}^{k} c_j \vec{u}_j$$

and if we assume that $\vec{x}$ is a unit vector we have that

$$1 = \vec{x}^T\vec{x} = \sum_{i,j=1}^{k} c_i c_j \vec{u}_i^T \vec{u}_j = \sum_{j=1}^{k} c_j^2 \ ,$$

and similarly

$$\vec{x}^T A\vec{x} = \sum_{j=1}^{k}(A)\lambda_j c_j^2 \le \lambda_k \sum_{j=1}^{k} c_j^2 = \lambda_k(A)$$

because of the ordering of the eigenvalues. Hence

$$\mu_k(A) \le \max_{\vec{u}\in V_k, \|\vec{u}\|=1} \vec{u}^T A\vec{u} \le \lambda_k(A) \ .$$

To show the converse, namely that $\lambda_k(A) \le \mu_k(A)$. We pick any k-dimension subspace $V \subset \mathbb{R}^n$. We would like to find a vector $\vec{x} \in V$ of the form

$$\vec{x} = \sum_{j=k}^{n} c_k \vec{u}_j$$

where $\sum_{j=k}^{n} c_k^2 = 1$. The reason for this is that

$$\vec{x}^T A\vec{x} = \sum_{j=k}^{n} \lambda_j(A) c_j^2 \ge \lambda_k(A) \ .$$

Since this inequality holds for any subspace V we also have

$$\mu_k(A) = \min_{V, \dim V = k} \max_{\vec{x}\in V, \|x\|=1} \vec{x}^T A\vec{x} \ge \lambda_k(A)$$

which would together with the previous inequality prove the theorem. It remains to show that such a vector $\vec{x}$ exists. To this end consider the space W_k spanned by the vectors $\vec{u}_k, \ldots, \vec{u}_n$. This space has dimension $n-k+1$. If $V \cap W_k = \{0\}$ then picking a basis in V together with the basis for W_k we get $n+1$ linearly independent vectors in $\mathbb{R}^n$ which cannot be. Hence the exists a vector $x \in V \cap W_k$. $\square$

Here is another companion theorem.

Theorem 0.3 (Max-Min). *We have that*

$$\lambda_{n-k+1}(A) = \max_{V, \dim V = k} \min_{\vec{x} \in V, \|x\|=1} \vec{x}^T A \vec{x} \ .$$

Proof. Again, we give the right side a name, $\nu_{n-k+1}(A)$. Picking V to be the space spanned by the vectors $\vec{u}_{n-k+1}, \dots, \vec{u}_n$ we get that

$$\min_{\vec{x} \in V, \|x\|=1} \vec{x}^T A \vec{x} = \lambda_{n-k+1}(A) \ .$$

and hence $\nu_{n-k+1}(A) \geq \lambda_{n-k+1}(A)$. For the opposite inequality we pick an arbitrary subspace V of dimension k and note that the space U_k spanned by the vectors $\vec{u}_1, \dots, \vec{u}_{n-k+1}$ is not disjoint with V, as a dimensional argument shows. Thus, there exists $\vec{x} \in V$ of the form

$$\vec{x} = \sum_{j=1}^{n-k+1} c_j \vec{u}_j \ ,$$

with $\sum_{j=1}^{n-k+1} c_k^2 = 1$. Hence

$$\vec{x}^T A \vec{x} = \sum_{j=1}^{n-k+1} \lambda_j(A) (\vec{x}^T \vec{u}_j)^2 \leq \lambda_{n-k+1}(A) \ .$$

Hence $\mu_{n-k+1}(A) \leq \lambda_{n-k+1}(A)$ which finishes the proof. □

All what we just said works also for Hermitean matrices.

Exercise 1: Reformulate and prove the Min-Max Theorem and Max-Min Theorem for Hermitan matrices.

Exercise 2: Prove the following

Theorem 0.4. *Given two hermitean $n \times n$ matrices A and B, assume that for all $\vec{z} \in \mathbb{C}^n$, $\vec{z}^* A \vec{z} \leq \vec{z}^* B \vec{z}$. Then the corresponding eigenvalues satisfy*

$$\lambda_i(A) \leq \lambda_i(B) \ , i = 1, \dots, n \ .$$

Here is another useful version of this principle. We state is for hermitean matrices since in this case the eigenvalues are real and the proof is the same as the one for symmetric matrices. All you have to recall is that the dot product is now replaced by the inner product

$$\langle \vec{z}, \vec{w} \rangle = \sum_{i=1}^n \overline{z}_i w_i \ .$$

Theorem 0.5. *Let A be an $n \times n$ hermitean matrix, i.e., $A^* = A$ and denote by $\lambda_1 \leq \lambda_2, \leq \dots \leq \lambda_n$ the eigenvalues and by $\vec{x}_1, \dots, \vec{x}_n$ the corresponding eigenvectors. For $k \leq n$ pick any k orthonormal vectors $\vec{u}_1, \dots, \vec{u}_k$ and form the matrix*

$$B_{ij} = \langle \vec{u}_i, A \vec{u}_j \rangle \ .$$

Denote by $\mu_1, \dots, \mu_k$ the eigenvalues of the matrix B arranged in increasing order. Then

$$\lambda_1 \leq \mu_1 \ , \ \lambda_2 \leq \mu_2 \ , \ \dots \ , \ \lambda_k \leq \mu_k \ .$$

This version of the min-max principle is especially useful. All one does is to come up with a set of orthonormal trial vectors, computes the matrix B and its eigenvalues. These values provide upper bounds to the actual eigenvalues of A.

Proof. First we make a little observation. Suppose that we are given k orthonormal vectors $\vec{v}_j$ and form the matrix $B_{ij} = \vec{v}_i^* A \vec{v}_j$. We have the eigenvalues $\mu_1, \dots, \mu_k$ and the corresponding eigenvectors $\vec{U}_1, \dots, \vec{U}_k$. Note that this vectors are in $\mathbb{C}^k$ and not in $\mathbb{C}^n$. We may write B in the form

$$C^* B C = D$$

where

$$C = \left[\vec{U}_1, \dots, \vec{U}_k\right]$$

is a unitary matrix and D has the μ_js in the diagonal. Hence for $\ell \le k$

$$\mu_\ell = \sum_{ij} C^*_{\ell i} B_{ij} C_{j\ell} = \vec{w}_\ell^* A \vec{w}_\ell$$

where $\vec{w}_\ell = \sum_i \vec{v}_i C_{i\ell} \in C^n$. We also not that $\vec{w}_{\ell'}^* A \vec{w}_\ell = 0$ if $\ell' \ne \ell$. Consider the subspace of $\mathbb{C}^n$ spanned by the vectors $\vec{w}_1, \dots, \vec{w}_\ell$. This is an ℓ dimensional subspace V of $\mathbb{C}^n$. Pick any vector $\vec{w}$ in this subspace V and write it as $\vec{w} = \sum_p d_p \vec{w}_p$. We compute

$$\vec{w}^* A \vec{w} = \sum_{p=1}^{\ell} |d_p|^2 \mu_p \le \mu_\ell$$

and

$$\max_{\vec{w} \in V, \|\vec{w}\|=1} \vec{w}^* A \vec{w} = \mu_\ell \ ,$$

just pick $\vec{w} = \vec{w}_\ell$. Hence, by the previous theorem $\lambda_\ell(A) \le \mu_\ell$. Thus, $\lambda_\ell(A) \le \mu_\ell, \ell = 1, \dots, k$. □

Another nice theorem is the following, attributed to Henri Poincaré but already known to Cauchy.

Theorem 0.6. *Let A be any hermitean $n \times n$ matrix with eigenvalues $\lambda_1 \le \lambda_2 \le \cdots \lambda_n$. Consider the matrix B which one gets by removing the column j and the row j from A and denote the eigenvalues by $\mu_1 \le \mu_2, \le \cdots \mu_{n-1}$. Then*

$$\lambda_1 \le \mu_1 \le \lambda_2 \le \mu_2 \le \cdots \le \mu_{n-1} \le \lambda_n \ .$$

In other words, the eigenvalues of A and B interlace.

Proof. We may assume that B is obtained from A by removing the last row and the last column. The matrix B being $(n-1) \times (n-1)$ has eigenvectors $\vec{v}_1, \dots, \vec{v}_{n-1} \in \mathbb{R}^{n-1}$. We may add a 0 in the last entry and in this way they become vectors $\vec{w}_1, \dots, \vec{w}_{n-1} \in \mathbb{R}^n$. For $1 \le k \le n-1$ consider the subspace V spanned by $\vec{v}_1, \dots \vec{v}_k$ and note that using the Min-Max Theorem

$$\lambda_k(B) = \max_{\vec{x} \in V, \|x\|=1} \vec{x}^T B \vec{x} = \max_{\vec{x} \in W, \|x\|=1} \vec{x}^T A \vec{x} \ge \lambda_k(A) \ ,$$

where W is the space spanned by the vectors $\vec{w}_1, \dots, \vec{w}_k$. It remains to show that $\lambda_k(B) \le \lambda_{k+1}(A)$ and for this us the Max-Min Theorem. We take for V the space spanned by the vectors $\vec{v}_{n-k}, \ldots, \vec{v}_{n-1}$ which, you recall, are the eigenvectors of B. Adding the zero entry at the bottom of these vectors gets us the vectors $\vec{w}_{n-k}, \ldots, \vec{w}_{n-1}$. We have that

$$\mu_{n-k}(B) = \min_{\vec{x} \in V, \|x\|=1} \vec{x}^T B \vec{x} = \min_{\vec{x} \in W, \|x\|=1} \vec{x}^T A \vec{x}$$

and since

$$\lambda_{n-k+1} = \max_{V, \dim V = k} \min_{\vec{x} \in V, \|x\|=1} \vec{x}^T A \vec{x}$$

it follows that

$$\mu_{n-k}(B) \leq \lambda_{n-k+1}(A) \ , k = 1, \ldots, (n-1) \ .$$

This proves the Theorem. □