THE MIN-MAX PRINCIPLE

Let A be a symmetric $n \times n$ matrix. The eigenvalues are really and hence we can order them $\lambda_1(A) \leq \lambda_2(A) \leq \cdots \leq \lambda_n(A)$ 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 $1 \leq 2 \leq 2 \leq 3$. We denote the corresponding eigenvector by $\vec{x}_1, \ldots, \vec{x}_n$ and we choose them to be an orthonormal basis of $\mathbb{R}^n$.

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 $f(\vec{x}_1) \leq f(\vec{x})$ 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. The function $f(\vec{x})$ is a differentiable function on the sphere. Instead of looking at the function f consider the function

$$\frac{\vec{x}^T A \vec{x}}{\vec{x}^T \vec{x}}$$

which is defined for all $\vec{x} \neq 0$. Obviously both the numerator and the denominator are differentiable. Any differentiable function is continuous and any continuous function on the sphere attains its maximum and minimum. Denote the point where the minimum is attained by $\vec{x}_1$, $|\vec{x}_1\| = 1$. Take any vector $\vec{y} \in R^n$ and consider the function

$$g(\varepsilon) = \frac{(\vec{x}_1 + \varepsilon \vec{y})^T A (\vec{x}_1 + \varepsilon \vec{y})}{\|\vec{x}_1 + \varepsilon \vec{y}\|^2} \ .$$

This function is differentiable and satisfies $g(\varepsilon) \geq g(0)$. Hence, the derivative at $\varepsilon = 0$ has to vanish and we compute the derivative to be

$$\frac{2\vec{y}^T A \vec{x}_1}{\|\vec{x}_1\|^2} - \frac{(\vec{x}_1^T A \vec{x}_1) 2\vec{y}^T \vec{x}_1}{\|\vec{x}_1\|^4} = 0$$

Since $\|\vec{x}_1\| = 1$ we have that

$$\vec{y}^T A \vec{x}_1 - f(\vec{x}_1) 2\vec{y}^T \vec{x}_1 = 0$$

or

$$\vec{y}^T [A\vec{x}_1 - f(\vec{x}_1)\vec{x}_1] = 0$$

and since $\vec{y}$ is arbitrary

$$A\vec{x}_1 = f(\vec{x}_1)\vec{x}_1$$

and we are done.

Here is another argument more in line with what you have learned in Calculus III. Assuming the a minimizer exists, we use Lagrange multipliers to minimize $\vec{x}^T A \vec{x}$ given the constraint

$\|\vec{x}\|^2 = 1$. we have that $\nabla \vec{x}^T A\vec{x} = 2A\vec{x}$, $\nabla \|\vec{x}|^2 = 2\vec{x}$ so that at the minimum $A\vec{x} = \lambda\vec{x}$. Thus the minimizing vector must be an eigenvector. □

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 to Courant and Hilbert and is called 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)

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

Next we vary the subspace, keeping the dimension k fixed and set

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

Theorem 0.2. *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 $\mu_k(A) \leq \lambda_k(A)$. For this we pick as a subspace V_k the space spanned by the vector $\vec{x}_1, \ldots, \vec{x}_k$. This space has obviously the dimension k. Pick any vector $\vec{u} \in V_k$ which means that we can write it as

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

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

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

and similarly

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

because of the ordering of the eigenvalues. Hence

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

To show the converse We pick any k-dimension subspace $V \subset \mathbb{R}^n$. The space $W_k \subset \mathbb{R}^n$ spanned by $\vec{x}_k, \vec{x}_{k+1}, \ldots, \vec{x}_n$ is a $(n-k+1)$ dimensional subspace. Hence the subspace $V \cap W_k$ has a dimension at least one. Otherwise, if $V \cap W_k = \{\vec{0}\}$ then any basis of V together with the vectors $\vec{x}_k, \ldots, \vec{x}_n$ would yield $n+1$ linearly independent vectors in $\mathbb{R}^n$ which cannot be. Hence there exists a nonzero vector $\vec{u} \in V \cap W_k$ of length one. We can write $\vec{u} = \sum_{j=k}^{n} c_j \vec{x}_j$ and hence

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

Hence we have shown that for any subspace $V \subset \mathbb{R}^n$ whose dimension is k

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

and therefore

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

which finishes the proof. □

Corollary 0.3. *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*

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

The proof is an exercise.

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

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

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

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

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 $\mu_1, \ldots, \mu_k$ and the corresponding eigenvectors $\vec{U}_1, \ldots, \vec{U}_k$. Note that this vectors are in $\mathbb{C}^k$ and not in $\mathbb{C}^n$. We may write B in the form

$$C^* BC = D$$

where

$$C = \left[\vec{U}_1, \ldots, U_k\right]$$

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

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

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 $\ell' \neq \ell$. Consider the subspace of $\mathbb{C}^n$ spanned by the vectors $\vec{w}_1, \ldots, \vec{w}_\ell$. 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

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

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 $\lambda_\ell(A) \leq \mu_\ell$. Thus, $\lambda_\ell(A) \leq \mu_\ell, \ell = 1, \ldots, k$. □

Another nice theorem is the following, attributed to Henri Poincaré.

Theorem 0.5. *Let A be any hermitean $n \times n$ matrix with eigenvalues $\lambda_1 \leq \lambda_2 \leq \cdots \lambda_n$. Consider the matrix B which one gets by removing the column j and the row j from A and denote the eigenvalues by $\mu_1 \leq \mu_2, \leq \cdots \mu_{n-1}$. Then*

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

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

Proof. By adding a constant multiple the identity matrix to A we may assume that the eigenvalue of A are all strictly positive. To simplify the notation we assume that B is created by removing the last column and last row from A. The matrix B, being a sub matrix of A satisfies the assumption of the previous theorem and hand we have that $\lambda_i \leq \mu_i \ , i = 1, \ldots, n-1$. Hence the values $\mu_i > 0, i = 1, \ldots, n-1$. Next we add a column and a row consisting of zeros to B which turns B into an $n \times n$ matrix $\tilde{B}$ that has 0 as an additional eigenvalue. Thus the eigenvalues of $\tilde{B}$ as $0 \leq \mu_1 \leq \mu_2 \leq \ldots \mu_{n-1}$. Take the subspace V spanned by the first ℓ eigenvectors of A. By the Theorem 0.2 we have that $0 \leq \lambda_1$ and $\mu_{\ell+1} \leq \lambda_\ell$. Since ℓ is arbitrary this proves the theorem.

□