and
\[
\sum_{j=1}^{n} \lambda_j |y_j|^2 \geq \lambda_1 \sum_{j=1}^{n} |y_j|^2 = \lambda_1 \|y\|^2
\]
therefore
\[
\lambda_1 \leq R(x) \leq \lambda_n \quad \text{for all } x \neq 0
\]

Equalities are attained since \( R(U(e_1)) = 1 \) and \( R(U(e_n)) = n \). Going to coordinates \( x \) minimum is attained for \( x = Ue_1 = u_1 = \text{eigenvector corresponding to } \lambda_1 \) since \( R(u_1) = R(U(e_1)) = \lambda_1 \), and for \( x = Ue_n = u_n = \text{eigenvector corresponding to } \lambda_n \), maximum is attained since \( R(u_n) = R(U(e_n)) = \lambda_n \). This proves:

**Theorem 39.** If \( A \) is a self-adjoint matrix then
\[
\max \frac{\langle x, Ax \rangle}{\|x\|^2} = \lambda_n \text{ the max eigenvalue of } A, \text{ attained for } x = u_n
\]
and
\[
\min \frac{\langle x, Ax \rangle}{\|x\|^2} = \lambda_1 \text{ the min eigenvalue of } A, \text{ attained for } x = u_1
\]

As an important consequence in numerical calculations: the maximum eigenvalue of \( A \) can be found by solving a maximization problem, and the minimum eigenvalue - by a minimization problem.

### 4.3. The minimax principle.
Reducing the dimension of \( A \) we can find all the eigenvalues, one by one. Consider the eigenvalues (24) of \( A \) and the corresponding eigenvectors \( u_1, \ldots, u_n \) which form an orthonormal basis:
\[
F^n = \bigoplus_{j=1}^{n} Fu_j.
\]

We saw that \( \max R(x) = \lambda_n = R(u_n) \). The subspace \( Sp(u_n) \) and its orthogonal complement \( Sp(u_n)^\perp = \bigoplus_{j=1}^{n-1} Sp(u_j) \) are invariant under \( A \).

Consider \( A \) as a linear transformation of the \( n-1 \) dimensional vector space \( Sp(u_n)^\perp \) to itself: its eigenvalues are \( \lambda_1, \ldots, \lambda_{n-1} \), the largest being \( \lambda_{n-1} \). We reduced the dimension!

Using Theorem 39 for \( A \) as a linear transformation on the vector space \( Sp(u_n)^\perp \) it follows that
\[
(25) \quad \max_{x \in Sp(u_n)^\perp} R(x) = \lambda_{n-1} \text{ is attained for } x = u_{n-1}
\]
The statement \( x \in Sp(u_n)^\perp \) can be formulated as the constraint \( \langle x, u_n \rangle = 0 \):
\[
\max_{x: \langle x, u_n \rangle = 0} R(x) = \lambda_{n-1}
\]

We can do even better: we can obtain \( \lambda_{n-1} \) \emph{without knowing} \( u_n \) or \( \lambda_n \).

To achieve this, subject \( x \) to \emph{any} constraint: \( \langle x, z \rangle = 0 \) for some \( z \neq 0 \).

It is easier to see what happens in coordinates \( y = U^*x \) in which \( A \) is diagonal. The constraint \( \langle x, z \rangle = 0 \) is equivalent to \( \langle y, w \rangle = 0 \) where \( w = Uz \) is some nonzero vector.
Step I. We have

$$\max_{y : \langle y, w \rangle = 0} R_U(y) \geq \lambda_{n-1} \quad \text{for all } w \neq 0$$

since there is some nonzero vector $y$ belonging to both the $n - 1$ dimensional subspace $\{y : \langle y, w \rangle = 0\}$ and the two dimensional subspace $F e_{n-1} \oplus F e_n$. (Such a vector is easy to find: $y = (0, \ldots, 0, y_{n-1}, y_n)^T$ with $\langle y, w \rangle = 0$; if $w_n \neq 0$ take $y_{n-1} = 1$ and $y_n = -w_{n-1}/w_n$, and if $w_n = 0$ take $y_{n-1} = 0, y_n = 1$). Using formula (23)

$$R_U(y) = \frac{\lambda_{n-1} |y_{n-1}|^2 + \lambda_n |y_n|^2}{|y_{n-1}|^2 + |y_n|^2} \geq \frac{\lambda_{n-1} |y_{n-1}|^2 + \lambda_{n-1} |y_n|^2}{|y_{n-1}|^2 + |y_n|^2} = \lambda_{n-1}$$

proving (26).

Step II. Inequality (26) implies that

$$\min_{w \neq 0} \max_{y : \langle y, w \rangle = 0} R_U(y) \geq \lambda_{n-1}$$

Step III. We now show that equality is attained in (28) for special $w$.

For $w = e_n$ we have, by (25),

$$\max_{y : \langle y, e_n \rangle = 0} R_U(y) = \lambda_{n-1} \quad \text{attained for } y = e_n$$

hence in (28) there is equality

$$\min_{w \neq 0} \max_{y : \langle y, w \rangle = 0} R_U(y) = \lambda_{n-1}$$

In a similar way it is shown that $\lambda_{n-2}$ is obtained by a minimum-maximum process, but with two constraints:

$$\min_{w_1, w_2 \neq 0} \max_{\langle y, w_1 \rangle = 0, \langle y, w_2 \rangle = 0} R_U(y) = \lambda_{n-2}$$

Indeed, consider a nonzero vector $\tilde{y} = (0, \ldots, 0, y_{n-2}, y_{n-1}, y_n)^T$ satisfying $\langle \tilde{y}, w_1 \rangle = 0$ and $\langle \tilde{y}, w_2 \rangle = 0$. Then in formula (23)

$$R_U(\tilde{y}) = \frac{\lambda_{n-2} |y_{n-2}|^2 + \lambda_{n-1} |y_{n-1}|^2 + \lambda_n |y_n|^2}{|y_{n-2}|^2 + |y_{n-1}|^2 + |y_n|^2} \geq \frac{\lambda_{n-2} |y_{n-2}|^2 + \lambda_{n-2} |y_{n-1}|^2 + \lambda_{n-2} |y_n|^2}{|y_{n-2}|^2 + |y_{n-1}|^2 + |y_n|^2} = \lambda_{n-2}$$

which shows that

$$\max_{\langle y, w_1 \rangle = 0, \langle y, w_2 \rangle = 0} R_U(y) \geq \lambda_{n-2}$$

Since for $w_1 = e_n$ and $w_2 = e_{n-1}$ we have equality in (30), and this implies (29).
Step by step, adding one extra constraint, the minimax procedure yields the next largest eigenvalue.

Going back to the variable $x$ it is found that:

**Theorem 40. The minimax principle**

Let $A$ be a self-adjoint matrix, with its eigenvalues numbered in an increasing sequence:

$$\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$$

corresponding to the eigenvectors $v_1, \ldots, v_n$.

Then its Rayleigh’s quotient

$$R(x) = \frac{\langle x, Ax \rangle}{\|x\|^2}$$

satisfies

$$\max_{x \neq 0} R(x) = \lambda_n$$

$$\min_{x \neq 0} \max_{\langle x, z \rangle = 0} R(x) = \lambda_{n-1}$$

$$\min_{z_1, z_2 \neq 0} \max_{\langle x, z_1 \rangle = 0, \langle x, z_2 \rangle = 0} R(x) = \lambda_{n-2}$$

and in general

$$\min_{z_1, \ldots, z_k \neq 0} \max_{\langle x, z_1 \rangle = 0, \ldots, \langle x, z_k \rangle = 0} R(x) = \lambda_{n-k}, \quad k = 1, 2, \ldots, n - 1$$

**Remark.** Sometimes the minimax principle is formulated as

$$\min_{V_j} \max_{x \in V_j} R(x) = \lambda_j, \quad j = 1, \ldots, n$$

where $V_j$ denotes an arbitrary subspace of dimension $j$.

The two formulations are equivalent since the set

$$V_{n-k} = \{x \mid \langle x, z_1 \rangle = 0, \ldots, \langle x, z_k \rangle = 0\}$$

is a vector space of dimension $n - k$ if $z_1, \ldots, z_k$ are linearly independent.

A similar construction starting with the lowest eigenvalue produces:

**Theorem 41. The maximin principle**

Under the assumptions of Theorem 40

$$\min_{x \neq 0} R(x) = \lambda_1$$

$$\max_{z \neq 0} \min_{\langle x, z \rangle = 0} R(x) = \lambda_2$$
and in general

$$\max_{\mathbf{z}_1, \ldots, \mathbf{z}_k \neq \mathbf{0}} \min_{\langle \mathbf{x}, \mathbf{z}_1 \rangle = 0} R(\mathbf{x}) = \lambda_{k+1}, \quad k = 1, 2, \ldots, n - 1$$

$$\cdots$$

$$\langle \mathbf{x}, \mathbf{z}_k \rangle = 0$$

4.4. The minimax principle for the generalized eigenvalue problem. Suppose $\lambda_1 \leq \lambda_1 \leq \ldots \leq \lambda_n$ are eigenvalues for the problem

$$A \mathbf{v} = \lambda B \mathbf{v}, \quad A \text{ symmetric, } B \text{ positive definite}$$

(31)

It was seen in §3.8 that if $S = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$ is the matrix whose columns are the generalized eigenvectors of the problem (31), then both matrices $A$ and $B$ are diagonalized using a congruence transformation: $S^T AS = \Lambda$ and $S^T BS = I$.

Defining

$$R(\mathbf{x}) = \frac{\langle \mathbf{x}, A\mathbf{x} \rangle}{\langle \mathbf{x}, B\mathbf{x} \rangle}$$

it is found that in coordinates $\mathbf{x} = S\mathbf{y}$:

$$R(\mathbf{x}) = R(S\mathbf{y}) = \frac{\langle S\mathbf{y}, A S \mathbf{y} \rangle}{\langle S\mathbf{y}, B S \mathbf{y} \rangle} = \frac{\langle \mathbf{y}, S^T A S \mathbf{y} \rangle}{\langle \mathbf{y}, S^T B S \mathbf{y} \rangle} = \frac{\lambda_1 y_1^2 + \ldots + \lambda_n y_n^2}{y_1^2 + \ldots + y_n^2}$$

and therefore

$$\max R(\mathbf{x}) = \lambda_n, \quad \min R(\mathbf{x}) = \lambda_1$$
5. Singular Value Decomposition

5.1. Rectangular matrices.

For rectangular matrices $M$ the notions of eigenvalue/vector cannot be defined. However, the products $MM^*$ and/or $M^*M$ (which are square, even self-adjoint, and even positive semi-definite matrices) carry a lot of information about $M$. The first result is that they have the same nonzero eigenvalues:

**Proposition 42.** Let $M$ be an $m \times n$ matrix. The matrices $MM^*$ and $M^*M$ are positive semi-definite. Moreover, they have the same nonzero eigenvalues (with the same multiplicity).

Moreover, let $\lambda_1, \ldots, \lambda_r$ be the positive eigenvalues. If $M^*Mv_j = \lambda_j v_j$ with $\lambda_j > 0$ and $v_1, \ldots, v_r$ an orthonormal set, then $MM^*u_j = \lambda_j v_j$ for $u_j = \frac{1}{\sqrt{\lambda_j}}Mv_j$ and $u_1, \ldots, u_r$ is an orthonormal set.

**Proof.** The two matrices obviously self-adjoint and are positive semi-definite since $\langle x, M^*Mx \rangle = \langle Mx, Mx \rangle \geq 0$ and $\langle x, MM^*x \rangle = \langle M^*x, M^*x \rangle \geq 0$.

Let $v_1, \ldots, v_n$ be an orthonormal set of eigenvectors of $M^*M$, the first $r$ corresponding to nonzero eigenvalues: $M^*Mv_j = \lambda_j v_j$ with $\lambda_j > 0$, for $j = 1, \ldots, r$ and $M^*Mv_j = 0$ for $j > r$.

Applying $M$ we discover that $MM^*Mv_j = \lambda_j Mv_j$ with $\lambda_j > 0$, for $j = 1, \ldots, r$ and $MM^*Mv_j = 0$ for $j > r$ which would mean that $Mv_j$ are eigenvectors to $MM^*$ corresponding to the eigenvalue $\lambda_j$ provided we ensure that $Mv_j \neq 0$. This is certainly true for all $j = 1, \ldots, r$, since $\|Mv_j\|^2 = \langle Mv_j, Mv_j \rangle = \langle v_j, M^*Mv_j \rangle = \langle v_j, \lambda_j v_j \rangle = \lambda_j \neq 0$ for $j \leq r$.

On the other hand, note that all $Mv_1, \ldots, Mv_r$ are mutually orthogonal, since $\langle Mv_j, Mv_i \rangle = \langle v_j, M^*Mv_i \rangle = \lambda_i \delta_{ij}$ so $Mv_j \perp Mv_i$ for all $i \neq j \leq r$, and $\|Mv_j\|^2 = \lambda_j$. Therefore, all the nonzero eigenvalues of $M^*M$ are also eigenvalues for $MM^*$, and with the same multiplicity, and with corresponding unit eigenvectors $u_j = \frac{1}{\sqrt{\lambda_j}}Mv_j$, $j = 1, \ldots, r$.

The same argument can be applied replacing $M$ by $M^*$, showing that indeed, $MM^*$ and $M^*M$ have the same nonzero eigenvalues and with the same multiplicity. $\Box$

Another result is that the Euclidian norm of $M$ can be read from the maximal eigenvalue of $MM^*$ (or $M^*M$):

$$\sum_{i,j} |M_{ij}|^2 = \text{the max eigenvalue of } M^*M$$

(We will discuss more later.)

5.2. The SVD theorem. Let $M$ be an $m \times n$ matrix. We are going to bring it as close to a diagonal form as humanly possible. Namely, we are going to write it as $M = U\Sigma V^*$ where $\Sigma$ is a diagonal $m \times n$ matrix, and $U$