We introduce the following theorem which will be used later.

Theorem 1 (Courant-Fischer min-max theorem [HJ10, Theorem 4.2.11])**.** *Let* $\mathbf{A}$ *be a* $n \times n$ *symmetric matrix with eigenvalues* $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$*. Then we have*

$$\max_{S:\dim(S)=k} \min_{\mathbf{x}\in S;\|\mathbf{x}\|_2=1} \mathbf{x}^T\mathbf{A}\mathbf{x} = \lambda_k \tag{1}$$

$$\min_{S:\dim(S)=n-k+1} \max_{\mathbf{x}\in S;\|\mathbf{x}\|_2=1} \mathbf{x}^T\mathbf{A}\mathbf{x} = \lambda_k \tag{2}$$

for all $k \in [n]$ *and* S *ranges over all subspaces of* $\mathbb{R}^n$*.*

Example 2. For the case $k = 1$, the theorem boils down to the following:

$$\max_{\mathbf{x}\in\mathbb{R}^n;\|\mathbf{x}\|_2=1} \mathbf{x}^T\mathbf{A}\mathbf{x} = \lambda_1.$$

1 Matrix Norms

We work with the vector spaces $\mathbb{R}^n$ and $\mathbb{C}^n$ where $\mathbb{R}$ and $\mathbb{C}$ are set of real and complex numbers.

Definition 3. For a vector space $\mathcal{V}$, a norm is defined as a function $\|.\| : \mathcal{V} \to \mathbb{R}$, satisfying following properties for $x \in \mathcal{V}$:

- Non-negativity: $\|\mathbf{x}\| \geq 0$
- Positive: $\|\mathbf{x}\| = 0$ if and only if $\mathbf{x} = 0$
- Homogeneous: $\|c\mathbf{x}\| = |c| \, \|\mathbf{x}\|$ for $c \in \mathbb{F}$
- Triangle inequality: $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$

Example 4. For $p \geq 1$, ℓ_p norm of a vector $x \in \mathbb{R}^n$ is defined as:

$$\|\mathbf{x}\|_p = (\sum_{i=1}^{n} |x_i|^p)^{1/p}.$$

Following are some examples:

- ℓ_1-norm: $\|\mathbf{x}\|_1 = \sum_{i=1}^n |x_i|$
- ℓ_2-norm: $\|\mathbf{x}\|_p = \sqrt{(\sum_{i=1}^n |x_i|^2)}$
- ℓ_∞-norm: $\|\mathbf{x}\|_\infty = \max_{i\in\{1,\ldots,n\}} |x_i|$

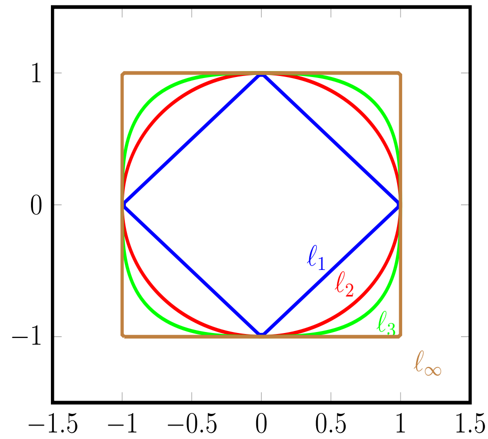

Figure 1: ℓ_p unit balls for $\ell_p = \ell_1, \ell_2, \ell_3, \ell_\infty$

Figure 1 shows the ℓ_p unit balls for $\ell_p = \ell_1, \ell_2, \ell_3, \ell_\infty$, centered at origin, which is defined as:

$$B_p = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\|_p \leq 1\}.$$

Note that since ℓ_p norm satisfies triangle inequality, the ℓ_p unit ball is convex. The unit ball expands with p from a rotated square for ℓ_1 norm to a square for ℓ_∞ norm.

The definition is valid for a general vector space over a field $\mathbb{F}$. Since the space of all $n \times n$ matrices, denoted by $\mathcal{M}_n$, is also a vector space, we can similarly talk about matrix norms. Sometimes matrix norms are called to those norms satisfying so called submultiplicative property (for instance [HJ10]):

$$\|\mathbf{XY}\| \leq \|\mathbf{X}\| \|\mathbf{Y}\|.$$

Let's continue by some common examples of matrix norms. Some common examples are ℓ_1 norm or ℓ_∞ (max) norm. ℓ_1 norm of a matrix A is given by:

$$\|\mathbf{A}\|_1 = \sum_{i,j \in [n]} |a_{ij}|.$$

Similarly, ℓ_∞ norm of a matrix $\mathbf{A}$ is given by:

$$\|\mathbf{A}\|_\infty = \max_{i,j \in [n]} |a_{ij}|.$$

Note that ℓ_∞ norm is not submultiplicative.

1.1 Operator Norm

A normed vector space can induce a norm on the space of linear transformations:

$$\|\mathbf{A}\| = \max_{\|\mathbf{x}\|=1} \|\mathbf{Ax}\|.$$

The norm is called the operator norm of the matrix. From this definition, it can be seen that:

$$\|\mathbf{Ax}\| \leq \|\mathbf{A}\| \|\mathbf{x}\|. \tag{3}$$

Given each vector norm, a norm can be induced on $\mathcal{M}_n$. For instance ℓ_p norm for $p \geq 1$ induces a norm on $\mathcal{M}_n$. In general we define the following[1]:

$$\|\mathbf{A}\|_{p \to q} = \max_{\|\mathbf{x}\|_p = 1} \|\mathbf{A}\mathbf{x}\|_q .$$

Spectral radius of a matrix is defined as follows.

Definition 5 (Spectral radius)**.** The specral radius $\rho(\mathbf{A})$ of a matrix $\mathbf{A} \in \mathcal{M}_n$ is defined as:

$$\rho(\mathbf{A}) = \max\{|\lambda| : \lambda \text{ is an eigenvalue of } \mathbf{A}\}.$$

The operators norms, and in general submultiplicative matrix norms, of a matrix are all bounded from below by its spectral radius.

Theorem 6. *The (submultiplicative) matrix norm $\|\cdot\|$ of an arbitrary matrix $\mathbf{A}$ is lower bounded by its spectral radius:*

$$\|\mathbf{A}\| \geq \rho(\mathbf{A}).$$

Proof. Suppose that $\rho(\mathbf{A}) = |\lambda_k|$ and $\mathbf{x}$ is a non-zero eigenvector corresponding to λ_1. For a general submultiplicative norm, define $\mathbf{V} = [\mathbf{x}, \ldots, \mathbf{x}] \in \mathbb{R}^{n \times n}$. Then from submultiplicative property we have:

$$\|\mathbf{A}\mathbf{V}\| \leq \|\mathbf{A}\| \|\mathbf{V}\| .$$

Moreover $\mathbf{A}\mathbf{V} = [\mathbf{A}\mathbf{x}, \ldots, \mathbf{A}\mathbf{x}] = \lambda_k \mathbf{V}$ and $\|\lambda_k \mathbf{V}\| = |\lambda_k| \|\mathbf{V}\|$. Therefore $\|\mathbf{A}\| \geq |\lambda_k| = \rho(\mathbf{A})$. □

The proof for operator norms, easily follows from:

$$|\lambda_k| \|\mathbf{x}\| = \|\lambda_k \mathbf{x}\| = \|\mathbf{A}\mathbf{x}\| \leq \|\mathbf{A}\| \|\mathbf{x}\| .$$

where the last inequality is the result of (3). Therefore $\|\mathbf{A}\| \geq \rho(\mathbf{A})$.

Example 7. Consider the case where $p = q = 1$:

$$\|\mathbf{A}\|_{1 \to 1} = \max_{\|\mathbf{x}\|_1 = 1} \|\mathbf{A}\mathbf{x}\|_1 .$$

For all $\mathbf{x} \in \mathbb{R}^n$ with $\|\mathbf{x}\|_1 = 1$, we have:

$$\begin{aligned}
\|\mathbf{A}\mathbf{x}\|_1 &= \sum_i \left| \sum_j a_{ij} x_j \right| \leq \sum_i \sum_j |x_j| |a_{ij}| \\
&= \sum_j |x_j| \sum_i |a_{ij}| \leq \max_j \sum_i |a_{ij}| = \max_j \|\mathbf{c}_j\|_1
\end{aligned}$$

where $\mathbf{c}_j$ is the column j of the matrix. Therefore $\max_j \|\mathbf{c}_j\|_1$ is an upper bound for this norm. It is achievable by choosing the canonical basis $\mathbf{e}_{j^*}$ such that $\mathbf{c}_{j^*} = \max_j \|\mathbf{c}_j\|_1$. Hence:

$$\|\mathbf{A}\|_{1 \to 1} = \max_j \|\mathbf{c}_j\|_1 .$$

[1] The notation is chosen as such to avoid confusion with entrywise norms.

Example 8. Let's consider the norm on linear transformations induced by ℓ_∞ norm:

$$\|\mathbf{A}\|_{\infty\to\infty} = \max_{\|\mathbf{x}\|_\infty=1} \|\mathbf{A}\mathbf{x}\|_\infty .$$

For all $\mathbf{x} \in \mathbb{R}^n$ with $\|\mathbf{x}\|_\infty = 1$, we have:

$$\begin{aligned}\|\mathbf{A}\mathbf{x}\|_\infty &= \max_i \left|\sum_j a_{ij}x_j\right| \leq \max_i \sum_j |x_j|\,|a_{ij}| \\ &= \max_i \sum_j |a_{ij}| = \max_i \|\mathbf{r}_i\|_1\end{aligned}$$

where $\mathbf{r}_i$ is the row i of the matrix. This upper bound is achievable by choosing vector of ones $\mathbf{1}_{n\times 1}$. Hence:

$$\|\mathbf{A}\|_{\infty\to\infty} = \max_i \|\mathbf{r}_i\|_1 .$$

Example 9. (*spectral norm*) Particular case of ℓ_2 norm is called *spectral norm*. Sometimes the spectral norm is directly defined as :

$$\|\mathbf{A}\|_{2\to 2} = \max\{\sqrt{\lambda} : \lambda \text{ is an eigenvalue of } \mathbf{A}^T\mathbf{A}\}.$$

Note that $\mathbf{A}^T\mathbf{A}$ is positive semidefinite and all its eigenvalues are nonnegative. Normal matrices could be diagonalized and they are the most general class of matrices that can be orthogonally diagonalized. In this case we have $\mathbf{A} = \mathbf{U}^T\Lambda\mathbf{U}$. If $\|\mathbf{x}\|_2 = 1$, we have:

$$\begin{aligned}\|\mathbf{A}\mathbf{x}\|_2 &= \left\|\mathbf{U}^T\mathbf{\Lambda}\mathbf{U}\mathbf{x}\right\|_2 \leq \|\mathbf{\Lambda}\|_2 \|\mathbf{x}\|_2 \\ &= \max\{|\lambda| : \lambda \text{ is an eigenvalue of } \mathbf{A}\}.\end{aligned}$$

This implies that the spectral norm of normal matrices are equal to the maximum eigenvalue in absolute value which is the spectral radius $\rho(\mathbf{A})$:

$$\rho(\mathbf{A}) = \|\mathbf{A}\|_{2\to 2} .$$

In general this is not true and the spectral radius may even not be a norm (see Exercise 1). Take the following matrix as example.

$$\mathbf{A} = \begin{bmatrix} 4 & 2 & 1 \\ 2 & 5 & 1 \\ 1 & 8 & 8 \end{bmatrix}$$

See that $\|\mathbf{A}\mathbf{x}\|_{2\to 2} \approx 12$ and $\rho(\mathbf{A}) \approx 10$.

Finally, note that a real matrix $\mathbf{A}$ having positive eigenvalues does not imply $\mathbf{x}^T\mathbf{A}\mathbf{x} > 0$ for all $\mathbf{x} \in \mathbb{R}^n$. Consider $\mathbf{A} = \begin{bmatrix} 0.1 & 1 \\ 0 & 0.1 \end{bmatrix}$. The eigenvalues are both 0.1 and positive, however choosing $\mathbf{x} = [-1, 1]^T$ gives $\mathbf{x}^T\mathbf{A}\mathbf{x} = -0.8 < 0$. Being symmetric is essential in the definition of positive definite and similarly nonnegative definite matrices.

References

[HJ10] Roger A. Horn and Charles R. Johnson. *Matrix analysis*. Cambridge Univ. Press, Cambridge, 2010.