Math. A2, Lesson 9.
Linear transformations, eigenvalues–eigenvectors

Let V, W be vector spaces. A map $T : V \to W$ is called *linear transformation* if $T(\mathbf{u} + \mathbf{v}) = T\mathbf{u} + T\mathbf{v}$ and $T(c\mathbf{u}) = cT\mathbf{u}$ for any $\mathbf{u}, \mathbf{v} \in V$ and $c \in \mathbb{R}$. Consequently, $T\mathbf{0} = \mathbf{0}$ and the linear transformation of a straight line is a straight line, etc.

From now on, we will confine ourselves to finite dimensional real inner product spaces. Linear operations between these spaces can be described by matrices of real entries. To stress that the elements of the Euclidean space are finite dimensional vectors, we will use bold-face lower-case letters, further vectors are treated as column-vectors. The inner product of the vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ is therefore written with matrix multiplication, like $\mathbf{x}^T\mathbf{y}$, where ^{T} stands for the transposition, hence $\mathbf{x}^T$ is a row-vector. Matrices will be denoted by bold-face upper-case letters. An $m \times n$ matrix $\mathbf{A} = (a_{ij})$ of real entries a_{ij}'s corresponds to an $\mathbb{R}^n \to \mathbb{R}^m$ linear transformation. Its transpose, $\mathbf{A}^T$, is an $n \times m$ matrix. An $n \times n$ matrix is called quadratic and it maps $\mathbb{R}^n$ into itself. The identity matrix is denoted by $\mathbf{I}$ or $\mathbf{I}_n$ if we want to refer to its size.

Note that for an $m \times n$ matrix $\mathbf{A}$, its range is

$$\mathcal{R}(\mathbf{A}) = \mathrm{lin}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$$

where $\mathbf{a}_1, \ldots, \mathbf{a}_n$ are the column vectors of $\mathbf{A}$ for which fact the notation $\mathbf{A} = (\mathbf{a}_1, \ldots, \mathbf{a}_n)$ will be used. The *rank* of $\mathbf{A}$ is the dimension of its range: and it is also equal to the maximum number of linearly independent rows of $\mathbf{A}$; trivially, $\mathrm{rank}(\mathbf{A}) \leq \min\{m, n\}$.

The *kernel* (null space) of $\mathbf{A}$ is

$$Ker(\mathbf{A}) = \{\mathbf{x} : \mathbf{A}\mathbf{x} = \mathbf{0}\}.$$

Both $\mathcal{R}(\mathbf{A})$ and $Ker(\mathbf{A})$ are subspaces in $\mathbb{R}^m$ and $\mathbb{R}^n$ respectively, and their dimension sum to n.

The quadratic matrix $\mathbf{A}$ is *symmetric* if $\mathbf{A} = \mathbf{A}^T$ and *orthogonal* if $\mathbf{A}\mathbf{A}^T = \mathbf{I}$. A 2×2 orthogonal matrix corresponds to a plane rotation by angle α, and its matrix is

$$\begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}.$$

The $n \times n$ matrix $\mathbf{A}$ has an inverse if and only if its determinant, $|\mathbf{A}| \neq 0$, and its inverse is denoted by $\mathbf{A}^{-1}$. In this case, the linear transformation corresponding to $\mathbf{A}^{-1}$ undoes the effect of the $\mathbb{R}^n \to \mathbb{R}^n$ transformation corresponding to $\mathbf{A}$, i.e. $\mathbf{A}^{-1}\mathbf{y} = \mathbf{x}$ if and only if $\mathbf{A}\mathbf{x} = \mathbf{y}$ for any $\mathbf{y} \in \mathbb{R}^n$. It is important that in case of an invertible (*regular*) matrix $\mathbf{A}$, the *range* (or image space) of $\mathbf{A}$ – denoted by $\mathcal{R}(\mathbf{A})$ – is the whole $\mathbb{R}^n$, and in exchange, the kernel of $\mathbf{A}$ (the subspace of vectors that are mapped into the zero vector by $\mathbf{A}$) consists of the only $\mathbf{0}$.

An orthogonal matrix $\mathbf{A}$ is always regular and $\mathbf{A}^{-1} = \mathbf{A}^T$; further its rows (or columns) constitute a complete orthonormal set in $\mathbb{R}^n$. Let k $(1 \leq k < n)$ be an integer; an $n \times k$ matrix $\mathbf{A}$ is called *suborthogonal* if its columns form (a not complete) orthonormal set in $\mathbb{R}^n$. For such an $\mathbf{A}$, the relation $\mathbf{A}^T\mathbf{A} = \mathbf{I}_k$ holds, but $\mathbf{A}\mathbf{A}^T \neq \mathbf{I}_n$. In fact, the $n \times n$ matrix $\mathbf{P} = \mathbf{A}\mathbf{A}^T$ is symmetric and

idempotent ($\mathbf{P}^2 = \mathbf{P}$), hence, it corresponds to the orthogonal projection onto $\mathcal{R}(\mathbf{A})$. The *trace* of the $n \times n$ matrix $\mathbf{A}$ is

$$\mathrm{tr}(\mathbf{A}) = \sum_{i=1}^{n} a_{ii}.$$

How the above matrix–matrix and matrix–scalar functions will look like if the underlying matrix is a product? If $\mathbf{A}$ and $\mathbf{B}$ can be multiplied together ($\mathbf{A}$ is $m \times n$ and $\mathbf{B}$ is $n \times k$ type), then their product corresponds to the succession of linear operations $\mathbf{B}$ and $\mathbf{A}$ in this order, therefore

$$(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$$

and if $\mathbf{A}$ and $\mathbf{B}$ are regular $n \times n$ matrices, then so is $\mathbf{AB}$, and

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}.$$

Further, $(\mathbf{A}^{-1})^T = (\mathbf{A}^T)^{-1}$, and vice versa. If $\mathbf{A}$ and $\mathbf{B}$ are $n \times n$ matrices, then

$$|\mathbf{AB}| = |\mathbf{A}| \cdot |\mathbf{B}|.$$

Therefore, the determinant of the product of several matrices of the same size does not depend on the succession of the matrices, however, the matrix multiplication is usually not commutative. The trace is commutative in the following sense: if $\mathbf{A}$ is an $n \times k$ and $\mathbf{B}$ is a $k \times n$ matrix, then

$$\mathrm{tr}(\mathbf{AB}) = \mathrm{tr}(\mathbf{BA}).$$

For several factors, the trace is accordingly, cyclically commutative:

$$\mathrm{tr}(\mathbf{A}_1\mathbf{A}_2 \dots \mathbf{A}_n) = \mathrm{tr}(\mathbf{A}_2 \dots \mathbf{A}_n\mathbf{A}_1) = \dots = \mathrm{tr}(\mathbf{A}_n\mathbf{A}_1 \dots \mathbf{A}_{n-1})$$

when, of course, the sizes of the factors are such that the successive multiplications in $\mathbf{A}_1 \dots \mathbf{A}_n$ can be performed and the number of rows in $\mathbf{A}_1$ is equal to the number of columns in $\mathbf{A}_n$. Further,

$$\mathrm{rank}(\mathbf{AB}) \le \min\{\mathrm{rank}(\mathbf{A}), \mathrm{rank}(\mathbf{B})\},$$

consequently, the rank cannot be increased in course of matrix multiplications.

Given an $n \times n$ symmetric real matrix $\mathbf{A}$, the quadratic form in the variables $x_1, \dots, x_n$ is the homogeneous quadratic function of these variables:

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_ix_j = \mathbf{x}^T\mathbf{A}\mathbf{x},$$

where $\mathbf{x} = (x_1, \dots, x_n)^T$, hence the matrix multiplication results in a scalar. The possible signs of a quadratic form (with different $\mathbf{x}$'s) characterize the underlying matrix. Accordingly, they fall into exactly one of the following categories.

Definition: Let $\mathbf{A}$ be $n \times n$ symmetric real matrix.

- $\mathbf{A}$ is positive (negative) definite if $\mathbf{x}^T\mathbf{A}\mathbf{x} > 0$ ($\mathbf{x}^T\mathbf{A}\mathbf{x} < 0$), $\forall \mathbf{x} \neq \mathbf{0}$.
- $\mathbf{A}$ is positive (negative) semidefinite if $\mathbf{x}^T\mathbf{A}\mathbf{x} \ge 0$ ($\mathbf{x}^T\mathbf{A}\mathbf{x} \le 0$), $\forall \mathbf{x} \in \mathbb{R}^n$, and $\mathbf{x}^T\mathbf{A}\mathbf{x} = \mathbf{0}$ for at least one $\mathbf{x} \neq \mathbf{0}$.

- $\mathbf{A}$ is indefinite if $\mathbf{x}^T\mathbf{A}\mathbf{x}$ takes on both positive and negative values (with different, non-zero $\mathbf{x}$'s).

The positive and negative definite matrices are all regular, whereas the positive and negative semidefinite ones are singular. The indefinite matrices can be either regular or singular. To more easily characterize the definiteness of symmetric matrices, we will use their eigenvalues.

The notion of an eigenvalue and eigenvector is introduced: λ is an eigenvalue of the $n \times n$ real matrix $\mathbf{A}$ with corresponding eigenvector $\mathbf{u} \neq \mathbf{0}$ if $\mathbf{A}\mathbf{u} = \lambda\mathbf{u}$. If $\mathbf{u}$ is an eigenvector of $\mathbf{A}$, it is easy to see that for $c \neq 0$, $c\mathbf{u}$ is also an eigenvector with the same eigenvalue. Therefore, it is better to speak about *eigen-direction*s instead of eigenvectors; or else, we will consider specially normalized, e.g. unit-norm eigenvectors, when only the orientation is bivalent. It is well known that an $n \times n$ matrix $\mathbf{A}$ has exactly n eigenvalues (with multiplicities) which are (possibly complex) roots of the characteristic polynomial $|\mathbf{A} - \lambda\mathbf{I}|$. Knowing the eigenvalues, the corresponding eigenvectors are obtained by solving the system of linear equations $(\mathbf{A} - \lambda\mathbf{I})\mathbf{u} = \mathbf{0}$ which must have a non-trivial solution due to the choice of λ. In fact, there are infinitely many solutions (in case of single eigenvalues they are constant multiples of each other). An eigenvector corresponding to a complex eigenvalue must also have complex coordinates, but in case of our main interest (the symmetric matrices) this cannot occur.

If the $n \times n$ matrix $\mathbf{A}$ has all distinct eigenvalues, then the corresponding eigenvectors are linearly independent, and choosing them as a new basis, our matrix becomes diagonal. Let $\mathbf{U}$ contain these eigenvectors in its columns ($\mathbf{U}$ is a regular matrix and it is the transformation matrix from the standard basis to the new basis formed by the eigenvectors). Then the matrix

$$\mathbf{U}^{-1}\mathbf{A}\mathbf{U}$$

becomes diagonal. When there are multiple eigenvalues, the matrix is diagonalizable only if to any multiple eigenvalue as many linearly independent eigenvectors correspond as the multiplicity of this eigenvalue. These are the so-called *normal matrices*. The quadratic real matrix $\mathbf{A}$ is called normal if $\mathbf{A}\mathbf{A}^T = \mathbf{A}^T\mathbf{A}$. Among real matrices, only the symmetric, anti-symmetric ($\mathbf{A}^T = -\mathbf{A}$), and orthogonal matrices are normal. Normal matrices have the following important spectral property: to their eigenvalues there corresponds an orthonormal set of eigenvectors; choosing this as a new basis, the matrix becomes *diagonal* (all the off-diagonal entries are zeros).

Hilbert–Schmidt theorem: The $n \times n$ symmetric, real matrix $\mathbf{A}$ has real eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$ (with multiplicities), and the corresponding eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_n$ can be chosen such that they constitute a complete orthonormal set in $\mathbb{R}^n$.

This so-called Spectral Decomposition theorem implies the following SD of the $n \times n$ symmetric matrix $\mathbf{A}$:

$$\mathbf{A} = \sum_{i=1}^{n} \lambda_i \mathbf{u}_i \mathbf{u}_i^T = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T, \tag{1}$$

where $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ is the diagonal matrix containing the eigenvalues – called *spectrum* – in its main diagonal, while $\mathbf{U} = (\mathbf{u}_1, \ldots, \mathbf{u}_n)$ is the orthogonal matrix containing the corresponding eigenvectors of $\mathbf{A}$ in its columns in the

order of the eigenvalues. Of course, permuting the eigenvalues in the main diagonal of $\mathbf{\Lambda}$, and the columns of $\mathbf{U}$ accordingly, will lead to the same SD, however – if not otherwise stated – we will enumerate the real eigenvalues in non-increasing order. About the uniqueness of the above SD we can state the following: the unit-norm eigenvector corresponding to a single eigenvalue is unique (up to orientation), whereas to an eigenvalue with multiplicity m there corresponds a unique m-dimensional so-called *eigen-subspace* within which any orthonormal set can be chosen for the corresponding eigenvectors.

It is easy to verify that for the eigenvalues of the symmetric matrix $\mathbf{A}$

$$\sum_{i=1}^{n} \lambda_i = \mathrm{tr}(\mathbf{A}) \quad \text{and} \quad \prod_{i=1}^{n} \lambda_i = |\mathbf{A}|$$

hold. Therefore $\mathbf{A}$ is singular if and only if it has a 0 eigenvalue, and

$$r = \mathrm{rank}(\mathbf{A}) = \mathrm{rank}(\mathbf{\Lambda}) = |\{i \,:\, \lambda_i \neq 0\}|;$$

moreover, $\mathcal{R}(\mathbf{A}) = \mathrm{lin}\{\mathbf{u}_i \,:\, \lambda_i \neq 0\}$. Therefore, the SD of $\mathbf{A}$ simplifies to

$$\sum_{i=1}^{r} \lambda_i \mathbf{u}_i \mathbf{u}_i^T.$$

Its spectrum also determines the definiteness of $\mathbf{A}$ in the following manner.

Proposition: Let $\mathbf{A}$ be $n \times n$ symmetric real matrix.

- $\mathbf{A}$ is positive (negative) definite if and only if all of its eigenvalues are positive (negative).
- $\mathbf{A}$ is positive (negative) semidefinite if and only if all of its eigenvalues are nonnegative (nonpositive), and its spectrum includes the zero.
- $\mathbf{A}$ is indefinite if its spectrum contains at least one positive and one negative eigenvalue.

The matrix of an orthogonal projection $\mathbf{P}_F$ onto the r-dimensional subspace $F \subset \mathbb{R}^n$ has the following SD (only the $r < n$ case is of importance, since in the $r = n$ case $\mathbf{P}_F = \mathbf{I}_n$):

$$\mathbf{P}_F = \sum_{i=1}^{r} \mathbf{u}_i \mathbf{u}_i^T = \mathbf{A}\mathbf{A}^T,$$

where $\mathbf{u}_1, \ldots, \mathbf{u}_r$ is any orthonormal set in F which is the eigen-subspace corresponding to the eigenvalue 1 of multiplicity r. Note that the eigenspace corresponding to the other eigenvalue 0 of multiplicity $n - r$ is the orthogonal complementary subspace $F^\perp$ of F in $\mathbb{R}^n$, but it has no importance, as only the eigenvectors in the first r columns of $\mathbf{U}$ enter into the above SD of $\mathbf{P}_F$. With the notation $\mathbf{A} = (\mathbf{u}_1, \ldots, \mathbf{u}_r)$, the SD of $\mathbf{P}_F$ simplifies to $\mathbf{A}\mathbf{A}^T$, indicating that $\mathbf{A}$ is a suborthogonal matrix. The matrix of *reflection* through the subspace F is $2\mathbf{P}_F - \mathbf{I}$.