EE6421 Statistical Signal Processing

Appendix. Review of Linear Algebra

Let $\mathbf{A}$ be an $m \times n$ matrix, with $m \leq n$. Let

$$\mathbf{A} = \begin{bmatrix} \boldsymbol{a}_1 \\ \boldsymbol{a}_2 \\ \vdots \\ \boldsymbol{a}_m \end{bmatrix} \tag{1}$$

where $\boldsymbol{a}_i$ $(i = 1, 2, \ldots, m)$ is a $1 \times n$ matrix (or row vector).

The k $(k \leq m)$ row vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_k$ are said to be linearly independent if whenever the following relation holds:

$$b_1 \boldsymbol{a}_1 + b_2 \boldsymbol{a}_2 + \cdots + b_k \boldsymbol{a}_k = \mathbf{0} \tag{2}$$

we must have $b_1 = b_2 = \cdots = b_k = 0$.

The row rank of $\mathbf{A}$ is defined as the maximum number of its linearly independent row vectors.

An alternative definition of rank is:

The rank of $\mathbf{A}$ is defined as the greatest integer d such that the determinant of some $d \times d$ submatrix of $\mathbf{A}$ is a nonzero constant.

Linearly independent column vectors can be similarly defined.

For a given matrix $\mathbf{B}$ of any size, the maximum number of its linearly independent row vectors is always equal to its linearly independent column vectors.

$\mathbf{A}$ is said to be of full row rank if $\operatorname{rank}(\mathbf{A}) = m$, and of full column rank if $\operatorname{rank}(\mathbf{A}) = n$.

An $m \times n$ matrix can be both of full row rank and of full column rank only if it is a square matrix ($m = n$).

Let $\mathbf{A}$ and $\mathbf{B}$ be two matrices of compatible sizes. Then

$$\text{rank}\,(\mathbf{AB}) \leq \text{ rank}\,(\mathbf{A});\ \text{rank}\,(\mathbf{AB}) \leq \text{ rank}\,(\mathbf{B}) \qquad (3)$$

$$\text{rank}\,(\mathbf{A} + \mathbf{B}) \leq \text{ rank}\,(\mathbf{A}) + \text{ rank}\,(\mathbf{B})$$

For a square matrix, we can define its power. $\mathbf{A}^0 = I$, where I is an identity matrix of the same size as $\mathbf{A}$. $\mathbf{A}^1 = \mathbf{A}$. $\mathbf{A}^2 = \mathbf{AA}$. In general, $\mathbf{A}^{k+l} = \mathbf{A}^k\mathbf{A}^l$ for any nonnegative integers k, l.

The square matrix $\mathbf{A}$ is said to be a *projection* matrix if $\mathbf{A}^2 = \mathbf{A}$.

The matrix $\mathbf{A}$ is said to have a *null space* if there exists at least one nonzero column vector $\boldsymbol{x}$ such that $\mathbf{A}\boldsymbol{x} = \mathbf{0}$. We say that $\boldsymbol{x}$ is orthogonal to the rows of $\mathbf{A}$. Suppose that $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_l$ are a set of linearly independent column vectors such that

$$\mathbf{A}\boldsymbol{x}_i = \mathbf{0}, \quad i = 1, \ldots, l \tag{4}$$

and if for any column vector $\mathbf{b}$ satisfying

$$\mathbf{A}\mathbf{b} = \mathbf{0} \tag{5}$$

we must have

$$\mathbf{b} = e_1\boldsymbol{x}_1 + \cdots + e_l\boldsymbol{x}_l \tag{6}$$

for some constants $e_1, \ldots, e_l$. In such a case, the null space of $\mathbf{A}$ is generated by $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_l$, and hence its dimension equals l.

The vector $\boldsymbol{y}$ is said to be in the *range* of $\mathbf{A}$ if it can be expressed as $\mathbf{A}\boldsymbol{x} = \boldsymbol{y}$ for some $\boldsymbol{x}$.

One of the most powerful concepts in matrix algebra is eigenanalysis. Letting $\mathbf{A}$ be a square matrix, the nonzero vector $\boldsymbol{v}$ is said to be an *eigenvector* of $\mathbf{A}$ if it satisfies

$$\mathbf{A}\boldsymbol{v} = \lambda\boldsymbol{v} \tag{7}$$

where λ is the scalar termed the *eigenvalue* associated with $\boldsymbol{v}$.

To find all the eigenvalues of $\mathbf{A}$, we can just find all the roots of $\det(\mathbf{A} - \lambda I) = 0$. Suppose λ_1 is the eigenvalue of $\mathbf{A}$, then the corresponding eigenvector $\boldsymbol{v}_1$ can be obtained by searching a vector in the null space of the matrix $\mathbf{A} - \lambda_1 I$.

A square matrix $\mathbf{A}$ of dimension $n \times n$ will have n linearly independent eigenvectors if all its n eigenvalues are distinct. Otherwise, $\mathbf{A}$ may or may not have n linearly independent eigenvectors.

A *symmetric* matrix equals its transpose ($\mathbf{A}^t = A$). Correlation matrices, which have the form $\mathbf{A} = \mathcal{E}[\boldsymbol{x}\boldsymbol{x}^t]$, are symmetric.

- The eigenvalues of a symmetric matrix are real.
- A symmetric matrix of dimension $n \times n$ will always have n linearly independent eigenvectors, which may or may not be orthogonal, but can be made as orthogonal.
- The eigenvectors associated with distinct eigenvalues of a symmetric matrix are orthogonal.
- If $\mathbf{A}$ is positive definite (semidefinite), all of its eigenvalues are positive (nonnegative).
- If $\mathbf{A}$ is positive definite, all of its principal minors are positive.

Define a matrix $\mathbf{V}$ having its columns to be the eigenvectors of the matrix $\mathbf{A}$.

$$\mathbf{V} = [\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n] \tag{8}$$

If $\mathbf{A}$ is symmetric, we can always choose its n eigenvectors as orthogonal, implying that $\mathbf{V}$ is an orthogonal matrix. $\mathbf{A}$ then can be expressed as

$$\begin{aligned} \mathbf{A} &= \mathbf{V}\Lambda\mathbf{V}^t \\ &= \sum_{i=1}^{n} \lambda_i \boldsymbol{v}_i \boldsymbol{v}_i^t \end{aligned} \tag{9}$$

where Λ is the diagonal matrix $\operatorname{diag}[\lambda_1, \lambda_2, \ldots, \lambda_n]$.

Furthermore, if $\mathbf{A}$ is nonsingular, we have

$$\mathbf{A}^{-1} = \sum_{i=1}^{n} \frac{1}{\lambda_i} \boldsymbol{v}_i \boldsymbol{v}_i^t \tag{10}$$

For a positive definite matrix $\mathbf{A}$, if $\mathbf{A} = \mathbf{V}\Lambda\mathbf{V}^t$ is its eigenvalue decomposition, where Λ is the diagonal matrix defined earlier, we define $\sqrt{\mathbf{A}}$ as $\sqrt{\mathbf{A}} = \mathbf{V}\sqrt{\Lambda}\mathbf{V}^t$ where

$$\sqrt{\Lambda} = \text{ diag} \left[\sqrt{\lambda_1}, \sqrt{\lambda_2}, \ldots, \sqrt{\lambda_n}\right]$$

.

Singular value decomposition (SVD)

Letting $\mathbf{A}$ be an $m \times n$ matrix consisting of complex entries, it can be expressed by

$$\mathbf{A} = \mathbf{U}D\mathbf{V}' \tag{11}$$

where D is the diagonal matrix $\operatorname{diag}[\sigma_1, \ldots, \sigma_k]$ and where $\mathbf{U}(m \times k)$ and $\mathbf{V}'(k \times n)$ are matrices satisfying $\mathbf{U}'\mathbf{U} = I_k$ and $\mathbf{V}'\mathbf{V} = I_k$.

Letting $\boldsymbol{u}_i$ and $\boldsymbol{v}_i$ denote the ith columns of $\mathbf{U}$ and $\mathbf{V}$ respectively, then

$$\mathbf{A}\boldsymbol{v}_i = \sigma_i \boldsymbol{u}_i \quad \boldsymbol{u}_i'\mathbf{A} = \sigma_i \boldsymbol{v}_i' \tag{12}$$

The scalars σ_i are termed the singular values of the matrix $\mathbf{A}$, whereas $\boldsymbol{u}_i$ and $\boldsymbol{v}_i$ are termed the left and right singular vectors of $\mathbf{A}$, respectively.

The number of nonzero singular values equals $k \leq \min(m, n)$. Notice that $k =$ rank $(\mathbf{A})$.

The SVD is related to eigenanalysis by

$$\mathbf{A}'\mathbf{A}\boldsymbol{v}_i = \sigma_i^2 \boldsymbol{v}_i \quad \mathbf{A}\mathbf{A}'\boldsymbol{u}_i = \sigma_i^2 \boldsymbol{u}_i \tag{13}$$

When $\mathbf{A}$ is not square, $\mathbf{A}'\mathbf{A}$ and $\mathbf{A}\mathbf{A}'$ cannot be both of full rank. Similar to eigenanalysis, $\mathbf{A}$ can also be expressed as

$$\mathbf{A} = \sum_{i=1}^{k} \sigma_i \boldsymbol{u}_i' \boldsymbol{v}_i \tag{14}$$

SVD can be used to define the *pseudoinverse* of a rectangular matrix. Assuming that rank $(A) = \min(m, n)$, the pseudoinverse $\mathbf{A}^{\sim 1}$ satisfies either $\mathbf{A}\mathbf{A}^{\sim 1} = I_m$, or $\mathbf{A}^{\sim 1}\mathbf{A} = I_n$ according to which dimension of the matrix is the smallest.

$$\mathbf{A}^{\sim 1} = \sum_{i=1}^{k} \frac{1}{\sigma_i} \boldsymbol{v}_i' \boldsymbol{u}_i \tag{15}$$

or in matrix terms $\mathbf{A}^{\sim 1} = \mathbf{V} D^{-1} \mathbf{U}'$. Finally, the pseudoinverse can be defined more directly by either $\mathbf{A}^{\sim 1} = \mathbf{A}'(\mathbf{A}\mathbf{A}')^{-1}$ or $\mathbf{A}^{\sim 1} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$.

Inequalities

The Cauchy-Schwarz inequality can be used to simplify optimization problems and provide explicit solutions. For two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ it asserts that

$$(\boldsymbol{y}^T\boldsymbol{x})^2 \leq (\boldsymbol{y}^T\boldsymbol{y})(\boldsymbol{x}^T\boldsymbol{x})$$

with equality if and only if $\boldsymbol{y} = c\,\boldsymbol{x}$ for c an arbitrary constant.

As applied to (complex) functions $g(x)$ and $g(x)$, it takes the form

$$\left|\int g(x)h(x)dx\right|^2 \leq \int |g(x)|^2 dx \int |h(x)|^2 dx$$

with equality if and only if $g(x) = c\,h^*(x)$ for c an arbitrary complex constant.