CISC 271 Class 6

Spectral Decomposition and Positive [Semi–]Definite Matrices

Text Correspondence: §6.3

Main Concepts:

- *Spectral theorem: diagonalization uses matrix transpose*
- *Positive Definite: each eigenvalue is greater than zero*
- *Positive Semi-Definite: each eigenvalue is at least zero*
- *Quadratic form for data matrix M: $\vec{u}^T M \vec{u}$*
- *Covariance matrix: positive definite matrix from statistics*

Sample Problem, Machine Inference: For a set of data vectors, what vector basis makes the data statistically independent?

The eigenvectors of a diagonalizable matrix $A \in \mathbb{R}^{n\times n}$ are a basis for the vector space $\mathbb{R}^n$. This means that any vector $\vec{u} \in \mathbb{R}^n$ can be represented as a linear sum of the eigenvectors $\vec{v}_j$, written as

$$\vec{u} = \alpha_1\vec{v}_1 + \alpha_2\vec{v}_2 + \cdots \alpha_n\vec{v}_n = \sum_{j=1}^{n} \alpha_j\vec{v}_j \tag{6.1}$$

6.1 The Spectral Decomposition

A fundamental theorem in linear algebra, which we will use often in this course, is the ***Spectral Theorem:***

Every normal matrix A has the factorization

$$A = Q\Lambda Q^T$$

where Q is an orthogonal matrix in which the columns are eigenvectors of A, and Λ is a diagonal matrix in which the entries are the eigenvalues of A.

This has an important corollary, the ***Spectral Theorem for Symmetric Matrices:***

> *Every real symmetric matrix B has the factorization*
>
> $$B = Q\Lambda Q^T$$
>
> *where Q is an orthogonal matrix in which the columns are eigenvectors of B, and Λ is a real diagonal matrix in which the entries are the eigenvalues of B.*

The difference between these two theorems is that a real-valued symmetric matrix has real eigenvalues and eigenvectors, whereas an asymmetric normal matrix might have complex eigenvalues and eigenvectors.

We can now consider Equation 6.1 for a symmetric matrix (or other normal matrix). The eigenvectors are a basis, and more importantly their unit forms are an *orthonormal* basis. We can write Equation 6.1 concisely as

$$\vec{u} = Q\vec{\alpha} \tag{6.2}$$

Because Q is an orthogonal matrix, its inverse Q^{-1} is its transpose Q^T. We can therefore solve Equation 6.2 by pre-multiplying both sides by $Q^{-1} = Q^T$, so

$$Q^T\vec{u} = \vec{\alpha} \tag{6.3}$$

Consider expanding Equation 6.3 by rows. This gives

$$\begin{array}{rcl} \vec{q}_1^T\vec{u} & = & \alpha_1 \\ \vec{q}_2^T\vec{u} & = & \alpha_2 \\ & \vdots & \\ \vec{q}_n^T\vec{u} & = & \alpha_n \end{array} \tag{6.4}$$

For an orthonormal basis, particularly for one that arises from a symmetric matrix, we can solve for the “weights” of the basis by using the dot-product form of each line in Equation 6.4 as

$$\alpha_k = \vec{q}_k \cdot \vec{u} \tag{6.5}$$

This property of orthonormal basis vectors is one that we will use in studying large data sets by principal-components analysis.

6.2 Positive-Definite and Positive-Semidefinite Matrices

One kind of symmetric matrix that is frequently encountered has every eigenvalue is greater than zero, which can be written as $\lambda_j > 0$. These are of interest to us for at least two reasons:

- Such a matrix arises in many application domains
- It helps us to understand a generalization of eigenvalues to non-square matrices

It is helpful to recall some useful abbreviations for saying that a matrix has eigenvalues that are all positive, or that are all negative. Assuming that $B = B^T$is real, and that $\vec{u} \in \mathbb{R}^n$ is a non-zero vector $\vec{u} \neq \vec{0}$, these symbols and quadratic forms are equivalent:

Symbol	**Name**	**Eigenvalues**	**Quadratic Form**
$B \succ 0$	Positive definite	$\forall_j \lambda_j > 0$	$\vec{u}^T B \vec{u} > 0$
$B \succeq 0$	Positive semidefinite	$\forall_j \lambda_j \geq 0$	$\vec{u}^T B \vec{u} \geq 0$
$B \prec 0$	Negative definite	$\forall_j \lambda_j < 0$	$\vec{u}^T B \vec{u} < 0$
$B \preceq 0$	Negative semidefinite	$\forall_j \lambda_j \leq 0$	$\vec{u}^T B \vec{u} \leq 0$
	Indefinite	$(\exists_i \lambda_i > 0) \wedge (\exists_j \lambda_j < 0)$	

Of these abbreviations, we will often use $B \succ 0$ and $B \succeq 0$. We can easily investigate some relationships between eigenvalues, eigenvectors, and quadratic forms for a real symmetric matrix.

6.3 Quadratic Form of a Symmetric Matrix

For a real symmetric matrix $B \in \mathbb{R}^{m \times m}$, we know from the Spectral Theorem that B can be decomposed into a product of factor matrices that have real entries, as

$$B = Q \Lambda Q^T \tag{6.6}$$

Because B is diagonalizable, an eigenvector basis is the orthonormal vectors $\vec{q}_j$ that are columns of the matrix Q in Equation 6.6. Any non-zero vector $\vec{u} \in \mathbb{R}^m$ can be written in this basis as a linear combination

$$\vec{u} = \alpha_1 \vec{q}_1 + \alpha_2 \vec{q}_2 + \cdots + \alpha_m \vec{q}_m \tag{6.7}$$

We assumed that $\vec{u} \neq \vec{0}$, so at least one scalar α_j in Equation 6.7 must be non-zero.

Because each $\vec{q}_j$ is an eigenvector of B, we know that $B\vec{q}_j = \lambda_j \vec{q}_j$. Because the vectors $\vec{q}_j$ are an orthonormal basis for $\mathbb{R}^m$, we know that if $i \neq j$ then $\vec{q}_i^T \vec{q}_j = 0$ and that $\vec{q}_j^T \vec{q}_j = 1$. Using

Equation 6.7 and these properties, we can first expand the product $B\vec{u}$ and then the quadratic form $\vec{u}^T B\vec{u}$ as

$$\begin{aligned} B\vec{u} &= \lambda_1\alpha_1\vec{q}_1 + \lambda_2\alpha_2\vec{q}_2 + \cdots + \lambda_m\alpha_m\vec{q}_m \\ \vec{u}^T B\vec{u} &= \lambda_1\alpha_1^2(\vec{q}_1^T\vec{q}_1) + \lambda_2\alpha_2^2(\vec{q}_2^T\vec{q}_2) + \cdots + \lambda_m\alpha_m^2(\vec{q}_m^T\vec{q}_m) \\ &= \lambda_1\alpha_1^2 + \lambda_2\alpha_2^2 + \cdots + \lambda_m\alpha_m^2 \end{aligned} \tag{6.8}$$

Because each scalar value $\alpha_j^2 \geq 0$, and because $\vec{u}$ is an arbitrary non-zero vector in $\mathbb{R}^m$, we can deduce that the quadratic form $\vec{u}^T B\vec{u} > 0$ if and only if each $\lambda_j > 0$, which we abbreviate as $B \succ 0$. We can likewise reason for $B \succeq 0$.

6.4 Example: Product of a Full-Rank Matrix and its Transpose

Suppose that we are given a matrix $A \in \mathbb{R}^{m\times n}$ that has column vectors that are linearly independent. If we compute the RREF of A, we will find that the nullspace of A is empty: for every non-zero vector $\vec{u} \neq \vec{0}$ we know that it transforms to a non-zero data vector $\vec{y} \neq \vec{0}$, so

$$[\vec{y} = A\vec{u}] \quad \Rightarrow \quad [A\vec{u} \neq \vec{0}] \quad \Rightarrow \quad [\|A\vec{u}\| > 0]$$

One way to write $\|\vec{y}\|^2$ is to use the dot product of $\vec{y}$ with itself. Using this fact, and replacing the dot product with the vector transpose, we have

$$\begin{aligned} \|\vec{y}\|^2 &= \vec{y}\cdot\vec{y} \\ &= \vec{y}^T\vec{y} \\ &= [A\vec{u}]^T[A\vec{u}] \\ &= \vec{u}^T[A^TA]\vec{u} > 0 \end{aligned} \tag{6.9}$$

The matrix A^TA is symmetric, by construction. The inequality of Equation 6.9 implies that the eigenvalues of the A^TA must be greater than zero; it is easy to reason this out using proof by contradiction, substituting the unit eigenvector for a negative eigenvalue into Equation 6.9. The final term of Equation 6.9, which for a general square matrix M is written as $\vec{u}^T M\vec{u}$, is called the *quadratic form* and has applications in the analysis of polynomials of several variables.

The reasoning we have used so far in this class gives us two important results:

- If A has linearly independent columns, then A^TA is symmetric positive definite
- If, for a symmetric matrix B and every non-zero vector $\vec{u} \neq \vec{0}$ we have $\vec{u}^T B\vec{u} > 0$, then B is positive definite

An important application of symmetric positive [semi–]definite matrices is in statistical analysis of data. We will not explore such statistical analysis in detail. We can, however, use some of the basic ideas to inform us as we explore our data.

6.5 Statistics of Vectors: Means and Variance

The *mean* of a vector $\vec{a}$ is defined as the sum of the entries divided by the number of entries. This is also called the arithmetic mean, or the average, but we will simply use the word "mean".

The mean is a scalar value that is commonly written as the vector symbol with a plain bar overtop. Using this notation, we will define the mean of a vector $\vec{x} \in \mathbb{R}^m$ as

$$\bar{x} \stackrel{\text{def}}{=} \frac{1}{m} \sum_{i=1}^{m} x_i \tag{6.10}$$

The variance of a set of samples is a measure of how the samples vary from their mean value. To find the sample variance, we begin by subtracting the mean value to find the *zero-mean* vector

$$\vec{m} = \begin{bmatrix} x_1 - \bar{x} \\ x_2 - \bar{x} \\ \vdots \\ x_m - \bar{x} \end{bmatrix} = \vec{x} - \bar{x}\vec{1} \tag{6.11}$$

where $\vec{1} \in \mathbb{R}^m$ is the "ones" column vector, which has each entry equal to unity.

The *sample* variance of the data in $\vec{x}$ is the sum of the squares differences from the mean $\bar{x}$, divided by the degrees of freedom in the difference vector. This sometimes written as a function, $\text{var}(\vec{x})$, and sometimes as σ^2 or as s^2. We will define the sample variance as

$$\text{var}(\vec{x}) = \frac{1}{m-1} \sum_{j=1}^{n} (x_j - \bar{x})^2 = \frac{(\vec{x} - \bar{x}\vec{1}) \cdot (\vec{x} - \bar{x}\vec{1})}{m-1} = \frac{\vec{m} \cdot \vec{m}}{m-1} = \frac{\vec{m}^T \vec{m}}{m-1} \tag{6.12}$$

We can use examples to calculate means and variances. Suppose that we have two data vectors

$$\vec{x}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \qquad \vec{x}_2 = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} \tag{6.13}$$

Using Equation 6.10, the means of the data in Example 6.13 are $\bar{x}_1 = 2$ and $\bar{x}_2 = 5$. the zero-mean vectors for these data vectors are

$$\vec{m}_1 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \qquad \vec{m}_2 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \tag{6.14}$$

We can see, in Example 6.14, that the zero-mean versions are the same for distinct data vectors. The reason for this is that we have "lost" one degree of freedom for each zero-meran, which is in the mean value $\bar{x}_j$.

To better understand variances, suppose that we have two data vectors

$$\vec{x}_3 = \begin{bmatrix} 1 \\ -1 \\ 3 \end{bmatrix} \qquad \vec{x}_4 = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} \tag{6.15}$$

The variances of the data vectors in Example 6.15 can be calculated as $\text{var}(\vec{x}_3) = 4$ and $\text{var}(\vec{x}_4) = 1$. We would say that the data in vector $\vec{x}_3$ vary more than do the data in vector $\vec{x}_4$.

6.6 Example: Covariance Matrix in Statistics

A concept that is related to variance is how much two sets of samples differ from each other. The process is similar: mean-correct each data set and take the inner product. If the zero-mean vector for the data $\vec{x}_1$ is $\vec{m}_1$ and the zero-mean vector for the data $\vec{x}_2$ is $\vec{m}_2$, then the *covariance* of $\vec{x}_1$ and $\vec{x}_2$ is

$$\text{cov}(\vec{x}_1, \vec{x}_2) = \frac{(\vec{x}_1 - \bar{x}_1\vec{1}) \cdot (\vec{x}_2 - \bar{x}_2\vec{1})}{m-1} = \frac{\vec{m}_1 \cdot \vec{m}_2}{m-1} = \frac{\vec{m}_1^T \vec{m}_2}{m-1} \tag{6.16}$$

The dot product is commutative, so for any $\vec{u}$ and $\vec{v}$ in the same vector space,

$$^\circ\, \text{cov}(\vec{u}, \vec{v}) = \text{cov}(\vec{v}, \vec{u})$$

Consider writing the variance of $\vec{x}_1$ as c_{11}, the covariance of $\vec{x}_1$ and $\vec{x}_2$ as c_{12}, and so on. These terms can be gathered into a *variance-covariance* matrix

$$B \stackrel{\text{def}}{=} \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} \tag{6.17}$$

Because the covariance is commutative, the matrix C in Equation 6.17 is symmetric.

Another way to write the covariance matrix for two vectors is to combine Equation 6.12 and Equation 6.16. We can write Equation 6.17 as

$$
\begin{aligned}
B &= \begin{bmatrix} \mathrm{var}(\vec{x}_1, \vec{x}_1) & \mathrm{cov}(\vec{x}_1, \vec{x}_2) \\ \mathrm{cov}(\vec{x}_2, \vec{x}_1) & \mathrm{var}(\vec{x}_2, \vec{x}_2) \end{bmatrix} \\
&= \frac{1}{m-1} \begin{bmatrix} \vec{m}_1^T \vec{m}_1 & \vec{m}_1^T \vec{m}_2 \\ \vec{m}_2^T \vec{m}_1 & \vec{m}_2^T \vec{m}_2 \end{bmatrix} \\
&= \frac{1}{m-1} \begin{bmatrix} \vec{m}_1^T \\ \vec{m}_2^T \end{bmatrix} \begin{bmatrix} \vec{m}_1 & \vec{m}_2 \end{bmatrix} \\
&= \frac{1}{m-1} M^T M
\end{aligned}
\tag{6.18}
$$

Because the matrix B in Equation 6.18 is constructed from a matrix M, the matrix B is symmetric and positive semidefinite. If the zero-mean vectors that are the columns of M are linearly independent, then the covariance matrix B is symmetric and positive definite.

We will use the covariance matrix later in this course, when we perform principal components analysis on potentially very large sets of data.

6.7 Example: Linear Elastic Structures

One example of a symmetric positive definite matrix arises when we try to model either a "uniform" physical material or a simple robotic device. Let us look at how springs can be used to model a material, and what the equations looks like.

A *linear* spring is one that follows Hooke's Law, which says that a displacement of a spring produces a force that is a constant multiple of the displacement. The usual terminology is:

Symbol	**Meaning**
x	Displacement of the spring from the zero position
f	Force that the spring exerts at the displacement x
k	The stiffness constant of the spring

Hooke's Law, stated using these symbols, is

$$f = kx \tag{6.19}$$

where the energy, or work, performed is force times distance or $E = fx$. An illustration of a linear spring, a displacement, and a force are given in Figure 6.1.

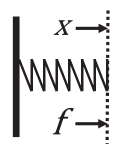

Figure 6.1: A scalar Hookean spring, where x is the displacement and f is the spring force.

What if there are multiple springs acting in a plane? The displacements will be x_1 in one axis and x_2 in the other axis, with resulting forces f_1 and f_2 respectively. This is illustrated in Figure 6.2.

The mathematical model of a planar spring system is a vector version of Equation 6.19, where the displacement is $\vec{x}$ and the resulting force is $\vec{f}$. These are related by the stiffness matrix, K, as

$$\vec{f} = K\vec{x} \tag{6.20}$$

A basic result in materials science is the Maxwell-Betti Law, which for our purposes states that the stiffness matrix K is symmetric[1].

[1] The law is usually stated and proved using stress tensors, which are higher-dimensional extensions of matrices.

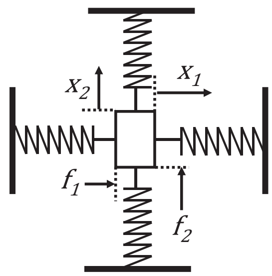

Figure 6.2: A set of springs acting on a body in the plane, where the body can only translate and cannot rotate. Displacements x_1 and x_2, acting together, result in forces f_1 and f_2.

The energy is still the product of the force and the displacement. Energy is a scalar and the other terms are vectors, so the energy is usually be written as the dot product $E = \vec{f} \cdot \vec{x}$. Using the commutative property of the dot product, and using the vector transpose to replace the dot product, we find that

$$\begin{aligned} E &= \vec{f} \cdot \vec{x} \\ &= \vec{x} \cdot \vec{f} \\ &= \vec{x}^T \vec{f} \end{aligned} \tag{6.21}$$

When we substitute Equation 6.20 into Equation 6.21 we get

$$\begin{aligned} E &= \vec{x}^T \vec{f} \\ &= \vec{x}^T K \vec{x} \end{aligned} \tag{6.22}$$

Energy cannot be destroyed, so the value E in Equation 6.22 must be greater than zero. This implies that the eigenvalues of the stiffness matrix K must be greater than zero; it is easy to reason this out using proof by contradiction, substituting the unit eigenvector for a negative eigenvalue into Equation 6.22.

This implies that the stiffness matrix K is symmetric positive definite.

End of Extra Notes