Chapter 23. Change of Basis

Let \( V \) be a finite-dimensional vector space. We have seen that any two bases of \( V \) have the same number of elements, and we defined \( \dim V = n \) if the bases of \( V \) have \( n \) elements. In this chapter we begin with the question:

Can we describe all bases of \( V \)?

One response is to parameterize all bases, just as we parameterize the points on a circle. You can start with any reference point \( p \) on the circle, and parameterize the other points \( p' \) on the circle in terms of the unique angle from \( p \) to \( p' \). Similarly, you can start with any reference basis \( B \) of \( V \) and parameterize the other bases \( B' \) in terms of a unique invertible matrix taking \( B \) to \( B' \). More precisely, we have:

**The Change of Basis Theorem:** If \( \{v_1, \ldots, v_n\} \) and \( \{u_1, \ldots, u_n\} \) are two bases of \( V \) then there is a unique matrix \( B = [b_{ij}] \) such that

\[
   u_j = \sum_{i=1}^{n} b_{ij} v_i, \quad \text{for all} \quad 1 \leq j \leq n,
\]

and this matrix \( B \) is invertible.

**Proof:** Since \( \{v_1, \ldots, v_n\} \) is a basis of \( V \), each vector in \( V \) can be written uniquely as a linear combination of the \( v_i \). Hence there exist unique coefficients \( b_{ij} \) such that

\[
   u_j = \sum_{i=1}^{n} b_{ij} v_i, \quad \text{for all} \quad 1 \leq j \leq n.
\]

On the other hand, since \( \{u_1, \ldots, u_n\} \) is a basis of \( V \) each vector in \( V \) can be written uniquely as a linear combination of the \( u_i \) so there exist coefficients \( c_{ij} \) such that

\[
   v_j = \sum_{i=1}^{n} c_{ij} u_i, \quad \text{for all} \quad 1 \leq j \leq n.
\]

Thus we have two matrices \( B = [b_{ij}] \) and \( C = [c_{ij}] \). Now we have

\[
   u_j = \sum_{k=1}^{n} b_{kj} v_k = \sum_{k=1}^{n} b_{kj} \left( \sum_{i=1}^{n} c_{ik} u_i \right) = \sum_{i=1}^{n} \left( \sum_{k=1}^{n} c_{ik} b_{kj} \right) u_i.
\]

In the last equality we have switched summands in order to gather the coefficients of \( u_i \). Since the \( u_i \)'s are also linearly independent, the coefficients on both sides of the previous equation must agree, so we have

\[
   \sum_{k=1}^{n} c_{ik} b_{kj} = \begin{cases} 
   1 & \text{if } i = j \\
   0 & \text{if } i \neq j.
   \end{cases}
\]

This means the matrices \( B = [b_{ij}] \) and \( C = [c_{ij}] \) satisfy \( CB = I \), so \( B \) is invertible and \( B^{-1} = C \). □

Prop. 1 says that, given one reference basis \( \{v_1, \ldots, v_n\} \), we have a 1-1 correspondence (i.e. a parametrization)

\[
   \{\text{bases of } V\} \leftrightarrow \{\text{invertible matrices}\}
\]
in which the basis \( \{u_1, \ldots, u_n\} \) corresponds to the matrix \( B \) in the Change of Basis Theorem.

For this parametrization to be helpful, we must thoroughly understand invertible matrices. The invertible matrices can be characterized in several ways; we list five of them.

**The Invertible Matrix Theorem:** For an \( n \times n \) matrix \( B \), the following are equivalent.

1. \( \det B \neq 0 \);
2. \( B \) is invertible;
3. \( \ker B = \{0\} \);
4. \( \text{im} \) \( B = \mathbb{R}^n \);
5. The columns of \( B \) are a basis of \( \mathbb{R}^n \).

**Proof:** The equivalence \( 1 \Leftrightarrow 2 \) follows from Cramer’s formula (chap. 18)

\[
B^{-1} = \frac{1}{\det B} [(-1)^{i+j} \det B_{ij}]^T,
\]

where \( B_{ij} \) is obtained from \( B \) by deleting \( \text{row}_i(B) \) and \( \text{col}_j(B) \).

For \( 2 \Rightarrow 3 \), assume \( B \) is invertible. If \( v \in \ker B \) then \( Bv = 0 \). Applying \( B^{-1} \), we get \( v = B^{-1}Bv = B^{-1}0 = 0 \), so \( \ker B = \{0\} \).

For \( 3 \Rightarrow 4 \), assume \( \ker B = \{0\} \). By the Kernel-Image Theorem we have

\[
\dim(\ker B) + \dim(\text{im} B) = \dim \mathbb{R}^n = n.
\]

Since \( \dim(\ker B) = 0 \) we have \( \dim(\text{im} B) = n \). Now we have \( \text{im} B = \mathbb{R}^n \) by Prop. 1 of chap. 22.

For \( 4 \Rightarrow 5 \), assume \( \text{im} B = \mathbb{R}^n \). Since \( \text{im} B \) is the span of the column vectors \( Be_i \), it follows that the \( n \) columns of \( B \) span \( \mathbb{R}^n \). By the Goldilocks principle, the columns of \( B \) are a basis of \( \mathbb{R}^n \).

To complete the chain of equivalences, we prove \( 5 \Rightarrow 2 \). Let \( u_j = Be_j \) be the columns of \( B \), so that

\[
u_j = \sum_{i=1}^n b_{ij}e_i, \quad \text{for all } j = 1, \ldots, n.
\]

Applying Prop. 1 to the bases \( \{e_1, \ldots, e_n\} \) and \( \{u_1, \ldots, u_n\} \) shows that \( B \) is invertible. \( \blacksquare \)

At the beginning of this course we diagonalized many \( 2 \times 2 \) matrices \( A \). In all cases the point was to convert the standard basis \( \{e_1, e_2\} \) to a basis \( \{u, v\} \) of eigenvectors for \( A \). This worked most simply in the case that the eigenvalues \( \lambda, \mu \) of \( A \) were distinct and real. In two dimensions, we did not need much theoretical information because simple computations showed that everything worked.

In higher dimensions, the calculations are not so simple. However, the Invertible Matrix Theorem implies that diagonalization works the same way in all dimensions, as shown in:
The Diagonalization Theorem: Let $A$ be an $n \times n$ matrix with distinct real eigenvalues $\lambda_1, \ldots, \lambda_n$. Let $v_1, \ldots, v_n$ be corresponding eigenvectors for $A$, so that each $v_i$ is nonzero and $Av_i = \lambda_i v_i$. Then the matrix $B = [v_1 \ v_2 \ \cdots \ v_n]$, whose columns are the $v_i$, is invertible and

$$B^{-1}AB = \begin{bmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \cdots & \lambda_{n-1} \\ 0 & 0 & \cdots & 0 & \lambda_n \end{bmatrix}.$$ 

Proof: We first prove that $\{v_1, \ldots, v_n\}$ is linearly independent. Suppose, if possible, that $\{v_1, \ldots, v_n\}$ is linearly dependent. By Lemma 1 in chap. 21, there is an index $j$ such that $v_j$ is a linear combination of $\{v_1, \ldots, v_{j-1}\}$. Choose the smallest such $j$, so that $\{v_1, \ldots, v_{j-1}\}$ is linearly independent and there are scalars $c_1, \ldots, c_{j-1}$ such that

$$v_j = c_1 v_1 + \cdots + c_{j-1} v_{j-1}. \tag{1}$$

Applying $A$ to both sides of (1) and remembering that $Av_i = \lambda_i v_i$ for each $i$, we get

$$\lambda_j v_j = c_1 \lambda_1 v_1 + \cdots + c_{j-1} \lambda_{j-1} v_{j-1}. \tag{2}$$

But we can also multiply both sides of (1) by $\lambda_j$, to get

$$\lambda_j v_j = c_1 \lambda_j v_1 + \cdots + c_{j-1} \lambda_j v_{j-1}. \tag{3}$$

Subtracting equation (2) from (3), we get

$$0 = c_1 (\lambda_1 - \lambda_j) v_1 + \cdots + c_{j-1} (\lambda_{j-1} - \lambda_j) v_{j-1}.$$ 

Since $\{v_1, \ldots, v_{j-1}\}$ are linearly independent and $\lambda_i \neq \lambda_j$ for $i \neq j$, it follows that $c_1 = \cdots = c_{j-1} = 0$, so $v_j = 0$. This is a contradiction, since eigenvectors are nonzero. Therefore $\{v_1, \ldots, v_n\}$ are linearly independent, as claimed.

By the Goldilocks principle, if it follows that $\{v_1, \ldots, v_n\}$ is a basis of $\mathbb{R}^n$. By the Invertible Matrix Theorem, the matrix $B$, whose columns are $\{v_1, \ldots, v_n\}$, is invertible.

The last part of the proof should look familiar: Column $j$ of $B^{-1}AB$ is

$$B^{-1}AB e_j = B^{-1}A(v_j) = B^{-1} (v_j) = \lambda_j B^{-1} v_j = \lambda_j e_j.$$ 

This means $B^{-1}AB$ is the diagonal matrix above. ■

The Diagonalization Theorem is just one aspect of a larger principle: Suppose you have a linear function

$$L : V \rightarrow V,$$

where $V$ is a vector space of dimension $n$. To make $L$ explicit you choose a basis, say $\{v_1, \ldots, v_n\}$. Then $L$ is determined by its effect on these basis vectors. Namely there are scalars $a_{ij}$ such that

$$L(v_j) = \sum_{i=1}^{n} a_{ij} v_i,$$
so \( L \) is described by the matrix \( A_L = [a_{ij}] \), using the basis \( \{v_1, \ldots, v_n\} \).

But no basis is intrinsically superior to any other. You may also express \( L \) with respect to another basis \( \{u_1, \ldots, u_n\} \) of \( V \). This means there are scalars \( a'_{ij} \) such that

\[
L(u_j) = \sum_{i=1}^{n} a'_{ij} u_i,
\]

so \( L \) is also described by the matrix \( A'_L = [a'_{ij}] \).

The basic question is: What is the relation between the matrices \( A_L \) and \( A'_L \)? The answer is:

\[
B^{-1}A_L B = A'_L,
\]

where \( B = [b_{ij}] \) is the change-of-basis matrix expressing the \( u \)-basis in terms of the \( v \)-basis:

\[
u_j = \sum_{i=1}^{n} b_{ij} v_i.
\]

This is proved quite similarly to the Change of Basis Theorem, see Exercise 23.1.
Exercise 23.1 Let $L : V \to V$ be a linear function, and let $\{v_1, \ldots, v_n\}$ and $\{u_1, \ldots, u_n\}$ be two bases of $V$. Then there are scalars $a_{ij}$ and $a'_{ij}$ such that

$$L(v_j) = \sum_{i=1}^{n} a_{ij} v_i, \quad L(u_j) = \sum_{i=1}^{n} a'_{ij} u_i.$$  

Prove that the matrices $A_L = [a_{ij}]$, $A'_L = [a'_{ij}]$ are related by

$$B^{-1} A_L B = A'_L,$$

where $B$ is the matrix in the change of basis theorem.

[Hint: Show that $AB = BA'$ by computing $L(u_j)$ in two different ways and reversing the sum, as in the proof of the Change of Basis Theorem.]

Exercise 23.2 Suppose $A$ is an $m \times n$ matrix, where $m < n$. Show that the following are equivalent.

1. Some $m \times m$ minor of $A$ is nonzero.
2. $\dim(\ker A) = n - m$.
3. $\im A = \mathbb{R}^m$.
4. The columns of $A$ span $\mathbb{R}^m$.

Exercise 23.3 Suppose $A$ is an $m \times n$ matrix, where $m > n$. Show that the following are equivalent.

1. Some $m \times m$ minor of $A$ is nonzero.
2. $\ker A = \{0\}$.
3. $\dim(\im A) = n$.
4. The columns of $A$ are linearly independent

Exercise 23.4 Let $A$ be an $n \times n$ matrix with distinct real eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. Assume that $|\lambda_1| > |\lambda_i|$ for all $i \neq 1$. Prove that for any $v \in V$ the limit

$$\lim_{n \to \infty} \frac{1}{\lambda_1^n} A^n v$$

lies on the $\lambda_1$-eigenline for $A$. (This is called the “Power Method” for finding the $\lambda_1$-eigenvector. In chap. 6 we used this method to approximate Fibonacci numbers.)