MA110: Lecture 10

Saurav Bhaumik
Department of Mathematics
IIT Bombay

Spring 2025

Matrix of a Linear Transformation

General Case: Let V and W be vector subspaces of dimension n and m, and let $T : V \to W$ be a linear map. Suppose $E := (\mathbf{v}_1, \ldots, \mathbf{v}_n)$ and $F := (\mathbf{w}_1, \ldots, \mathbf{w}_m)$ are ordered bases of V and W respectively. Then the matrix of T with respect to E and F is the $m \times n$ matrix $A = (a_{jk})$ determined by

$$T(\mathbf{v}_k) = \sum_{j=1}^{m} a_{jk}\mathbf{w}_j \quad \text{for } k = 1, \ldots, n.$$

This matrix is denoted by $\boxed{\mathbf{M}_F^E(T)}$. Note that the kth column of this matrix corresponds to the coefficients of $T(v_k)$ when expressed as a linear combination of $\mathbf{w}_1, \ldots, \mathbf{w}_m$. Also, for $\mathbf{v} = \alpha_1\mathbf{v}_1 + \cdots + \alpha_n\mathbf{v}_n \in V$, $\mathbf{w} = \beta_1\mathbf{w}_1 + \cdots + \beta_m\mathbf{w}_m \in W$,

$$T(\mathbf{v}) = \mathbf{w} \iff \mathbf{A}\begin{bmatrix}\alpha_1 \\ \vdots \\ \alpha_n\end{bmatrix} = \begin{bmatrix}\beta_1 \\ \vdots \\ \beta_m\end{bmatrix}.$$

Composites of Linear Transformation

Suppose U is another vector space of dimension p and $D = (\mathbf{u}_1, \ldots, \mathbf{u}_p)$ is an ordered basis of U. If $\mathbf{S} : U \to V$ is a linear map, then we can form the composite of S and T, i.e.,

$$T \circ S : U \to W$$

Now if $B = (b_{k\ell})$ is the $n \times p$ matrix of S w.r.t. D and E, i.e., $B = \mathbf{M}_E^D(S)$, then for $\ell = 1, \ldots, p$,

$$S(u_\ell) = \sum_{k=1}^{n} b_{k\ell}\mathbf{v}_k \quad \text{and hence} \quad (T \circ S)(u_\ell) = \sum_{k=1}^{n} b_{k\ell} T(\mathbf{v}_k);$$

substituting for $T(\mathbf{v}_k)$ from the earlier equation, we see that

$$(T \circ S)(u_\ell) = \sum_{k=1}^{n} b_{k\ell} \sum_{j=1}^{m} a_{jk}\mathbf{w}_j = \sum_{j=1}^{m} \left(\sum_{k=1}^{n} a_{jk} b_{k\ell} \right) \mathbf{w}_j$$

Thus we see that $\boxed{\mathbf{M}_F^D(T \circ S) = AB = \mathbf{M}_F^E(T)\mathbf{M}_E^D(S).}$

Remark: Let V be a subspace of $\mathbb{R}^{n\times 1}$, W be a subspace of $\mathbb{R}^{m\times 1}$, and let $T : V \to W$ be a linear map. Two important subspaces associated with T are as follows.

(i) $\mathcal{N}(T) := \{\mathbf{x} \in V : T(\mathbf{x}) = \mathbf{0}\}$, called the **null space** of T,

(ii) $\mathcal{I}(T) := \{T(\mathbf{x}) : \mathbf{x} \in V\}$, called the **image space** of T.

We note that
a linear map T is one-one $\iff \mathcal{N}(T) = \{\mathbf{0}\}$, and
a linear map T is onto $\iff \mathcal{I}(T) = W$.

Further, if $V := \mathbb{R}^{n\times 1}$, $W := \mathbb{R}^{m\times 1}$, and $\mathbf{A} \in \mathbb{R}^{m\times n}$, then

$$\begin{aligned}\mathcal{N}(T_{\mathbf{A}}) &= \{\mathbf{x} \in \mathbb{R}^{n\times 1} : \mathbf{A}\mathbf{x} = \mathbf{0}\} = \mathcal{N}(\mathbf{A}),\\ \mathcal{I}(T_{\mathbf{A}}) &= \{\mathbf{A}\mathbf{x} : \mathbf{x} \in \mathbb{R}^{n\times 1}\} = \mathcal{C}(\mathbf{A}).\end{aligned}$$

The last equality follows by noting that if $\mathbf{A} = \begin{bmatrix}\mathbf{c}_1 & \cdots & \mathbf{c}_n\end{bmatrix}$, then $\mathbf{A}\mathbf{x} = x_1\mathbf{c}_1 + \cdots + x_n\mathbf{c}_n$ for $\mathbf{x} := \begin{bmatrix}x_1 & \cdots & x_n\end{bmatrix} \in \mathbb{R}^{n\times 1}$.

Example

Let $\mathbf{A} := \begin{bmatrix} 1 & -1 \\ -1 & 2 \\ 0 & 1 \end{bmatrix} \in \mathbb{R}^{3\times 2}$. Then $T_{\mathbf{A}} : \mathbb{R}^{2\times 1} \to \mathbb{R}^{3\times 1}$.

In fact, $T_{\mathbf{A}} : \begin{bmatrix} x_1 & x_2 \end{bmatrix}^{\mathsf{T}} \longmapsto \begin{bmatrix} x_1 - x_2 & -x_1 + 2x_2 & x_2 \end{bmatrix}^{\mathsf{T}}$ for all $\begin{bmatrix} x_1 & x_2 \end{bmatrix}^{\mathsf{T}} \in \mathbb{R}^{2\times 1}$. Clearly, $\mathcal{N}(T_{\mathbf{A}}) = \{\mathbf{0}\}$. Also,

$$\mathcal{I}(T_{\mathbf{A}}) = \left\{ \begin{bmatrix} y_1 & y_2 & y_3 \end{bmatrix}^{\mathsf{T}} \in \mathbb{R}^{3\times 1} : y_1 + y_2 - y_3 = 0 \right\}.$$

To see this, note that $(x_1 - x_2) + (-x_1 + 2x_2) - x_2 = 0$ for all $\begin{bmatrix} x_1 & x_2 \end{bmatrix}^{\mathsf{T}} \in \mathbb{R}^{2\times 1}$, and if $\begin{bmatrix} y_1 & y_2 & y_3 \end{bmatrix}^{\mathsf{T}} \in \mathbb{R}^{3\times 1}$ satisfies $y_1 + y_2 - y_3 = 0$, then we may let $x_1 := y_1 + y_3$, $x_2 := y_3$, so that $x_1 - x_2 = y_1$, $-x_1 + 2x_2 = y_2$ and $x_2 = y_3$, that is, $T_{\mathbf{A}}(\begin{bmatrix} x_1 & x_2 \end{bmatrix}^{\mathsf{T}}) = \begin{bmatrix} y_1 & y_2 & y_3 \end{bmatrix}^{\mathsf{T}}$.

Note: $\mathcal{I}(T_{\mathbf{A}})$ is a plane through the origin $\begin{bmatrix} 0 & 0 & 0 \end{bmatrix}^{\mathsf{T}}$ in $\mathbb{R}^{3\times 1}$.

Complex Numbers

In our development of matrix theory, we have so far used real numbers as scalars, and we have considered matrices whose entries are real numbers. It turns out that all the concepts extend easily when the scalars allowed to be **complex numbers** instead of real numbers. We shall assume familiarity with the basics of complex numbers and just recall a few properties. The set of all complex numbers is denoted by $\mathbb{C}$. Let $z \in \mathbb{C}$. Then $z = x + iy$ for unique $x, y \in \mathbb{R}$. We call x the **real part** of z, and denote it by $\Re(z)$. Also, y is called the **imaginary part** of z, and it is denoted by $\Im(z)$. The complex number $x - iy$ is called the **conjugate** of $z = x + iy$, and it is denoted by $\bar{z}$. Let $a, b, c, d \in \mathbb{R}$. The addition and multiplication of complex numbers $a + ib$, $c + id$ is defined by

$$\begin{aligned}(a+ib)+(c+id) &= (a+c)+i(b+d)\\ (a+ib)(c+id) &= (ac-bd)+i(ad+bc).\end{aligned}$$

Recall also that the **absolute value** of a complex number $z = x + iy \in \mathbb{C}$ is defined by by

$$|z| := \sqrt{z\overline{z}} = \sqrt{x^2 + y^2}.$$

Note that the **triangle inequality**

$$|z_1 + z_2| \leq |z_1| + |z_2|$$

holds for all $z_1, z_2 \in \mathbb{C}$. In terms of the norm

$$\|(x, y)\| := \sqrt{x^2 + y^2} \quad \text{of a vector } (x, y) \in \mathbb{R}^2,$$

this corresponds to the inequality

$$\|(x_1, y_1) + (x_2, y_2)\| \leq \|(x_1, y_1)\| + \|(x_2, y_2)\|$$

for all $(x_1, y_1), (x_2, y_2) \in \mathbb{R}^2$. Also, note that

$$\max\{|\Re(z)|, |\Im(z)|\} \ \leq \ |z| \ \leq \ |\Re(z)|+|\Im(z)| \quad \text{for all } z \in \mathbb{C}.$$

We shall use complex numbers as scalars and consider matrices whose entries are complex numbers.

The set of all $m \times n$ matrices with entries in $\mathbb{C}$ is denoted by $\mathbb{C}^{m\times n}$. In particular, $\mathbb{C}^{1\times n}$ is the set of all row vectors of length n, while $\mathbb{C}^{m\times 1}$ is the set of all column vectors of length m.

For $\mathbf{A} := [a_{jk}] \in \mathbb{C}^{m\times n}$, define $\mathbf{A}^* := [\overline{a_{kj}}]$. Then $\mathbf{A}^* \in \mathbb{C}^{n\times m}$. It is called the **conjugate transpose** or the **adjoint** of $\mathbf{A}$. Note: $(\alpha\mathbf{A} + \beta\mathbf{B})^* = \overline{\alpha}\mathbf{A}^* + \overline{\beta}\mathbf{B}^*$ for $\mathbf{A}, \mathbf{B} \in \mathbb{C}^{m\times n}$ and $\alpha, \beta \in \mathbb{C}$. In case $m = n$, then $(\mathbf{AB})^* = \mathbf{B}^*\mathbf{A}^*$.

- A square matrix $\mathbf{A} = [a_{jk}]$ is called **Hermitian** or **self-adjoint** if $\mathbf{A}^* = \mathbf{A}$, that is, if $a_{jk} = \overline{a_{kj}}$ for all j, k.
- A square matrix $\mathbf{A} = [a_{jk}]$ is called **skew-Hermitian** or **skew self-adjoint** if $a_{jk} = -\overline{a}_{kj}$ for all j, k.

Note: Every diagonal entry of a self-adjoint matrix is real since $a_{jj} = \overline{a_{jj}} \implies a_{jj} \in \mathbb{R}$ for $j = 1, \ldots, n$. On the other hand, the real part of every diagonal entry of a skew self-adjoint matrix

Note: If $\mathbf{x} := \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}^{\mathsf{T}} \in \mathbb{C}^{n\times 1}$ is a column vector, then $\mathbf{x}^* = \begin{bmatrix} \overline{x_1} & \cdots & \overline{x_n} \end{bmatrix} \in \mathbb{C}^{1\times n}$ is a row vector, and $\mathbf{x}^*\mathbf{x} = |x_1|^2 + \cdots + |x_n|^2$. It follows that $\mathbf{x}^*\mathbf{x} = 0 \iff \mathbf{x} = \mathbf{0}$.

A matrix $\mathbf{A} \in \mathbb{C}^{m\times n}$ defines a linear transformation from $\mathbb{C}^{n\times 1}$ to $\mathbb{C}^{m\times 1}$, and every linear transformation from $\mathbb{C}^{n\times 1}$ to $\mathbb{C}^{m\times 1}$ can be represented by a matrix $\mathbf{A} \in \mathbb{C}^{m\times n}$ (with respect to an ordered basis for $\mathbb{C}^{n\times 1}$ and an ordered basis for $\mathbb{C}^{m\times 1}$).

Similarly, we can consider vector subspaces of $\mathbb{C}^{n\times 1}$, and the concepts of linear dependence of vectors and of the span of a subset carry over to $\mathbb{C}^{n\times 1}$. The Fundamental Theorem for Linear Systems remains valid for matrices with complex entries.

Having thus completed our discussion of solution of a linear system, we shall turn to solution of an 'eigenvalue problem' associated with a matrix. In this development, the role of complex numbers will turn out to be important.

Matrix Eigenvalue Problem

The German word 'eigen' means 'belonging to itself'. The eigenvalue problem for a matrix consists of finding a nonzero vector which is sent to a scalar multiple of itself by the linear transformation defined by the matrix.

Eigenvalue problems come up frequently in many engineering branches, quantum mechanics, physical chemistry, biology, and even in economics and psychology.

Please refer to Section 8.2 of Kreyszig's book for applications of eigenvalue problems to stretching of elastic membranes, to vibrating mass-spring systems, to Markov processes and to population control models.

In the development that follows, we shall use either real numbers or complex numbers as scalars. To facilitate a general discussion which applies to both types of scalars, we shall write $\mathbb{K}$ to mean either $\mathbb{R}$ or $\mathbb{C}$. When we want to switch to a special treatment valid for only the real scalars, or only for the complex scalars, we shall specify $\mathbb{K} := \mathbb{R}$ or $\mathbb{K} := \mathbb{C}$.

Definition

Let $\mathbf{A}$ *be an* $n \times n$ *matrix with entries in* $\mathbb{K}$*, that is, let* $\mathbf{A} \in \mathbb{K}^{n\times n}$*. A scalar* $\lambda \in \mathbb{K}$ *is called an* **eigenvalue** *of* $\mathbf{A}$ *if*

$$\mathbf{A}\mathbf{x} = \lambda\,\mathbf{x} \quad \textit{for some } \mathbf{x} \in \mathbb{K}^{n\times 1} \textit{ with } \mathbf{x} \neq \mathbf{0}.$$

Any nonzero vector $\mathbf{x} \in \mathbb{K}^{n\times 1}$ *satisfying* $\mathbf{A}\mathbf{x} = \lambda\,\mathbf{x}$ *is called an* **eigenvector** *of* $\mathbf{A}$ *corresponding to the eigenvalue* λ*. Further,*

$$\{\mathbf{x} \in \mathbb{K}^{n\times 1} : \mathbf{A}\mathbf{x} = \lambda\,\mathbf{x}\} = \mathcal{N}(\mathbf{A} - \lambda\mathbf{I}).$$

is called the **eigenspace** *of* $\mathbf{A}$ *associated with* λ*.*

How to find eigenvalues

Let $\mathbf{A} = [a_{jk}] \in \mathbb{K}^{n\times n}$ and let $\lambda \in \mathbb{K}$. Clearly,

$$\begin{aligned}\lambda \text{ is an eigenvalue of } \mathbf{A} &\iff \mathcal{N}(\mathbf{A} - \lambda\mathbf{I}) \neq \{\mathbf{0}\}\\ &\iff \operatorname{rank}(\mathbf{A} - \lambda\mathbf{I}) < n\\ &\iff \det(\mathbf{A} - \lambda\mathbf{I}) = 0.\end{aligned}$$

The last condition suggests that we consider the polynomial

$$p_{\mathbf{A}}(t) := \det(\mathbf{A} - t\,\mathbf{I}) = \det\begin{bmatrix} a_{11} - t & \cdots & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{n1} & \cdots & a_{nn} - t \end{bmatrix}.$$

This is called the **characteristic polynomial** of $\mathbf{A}$. It is a polynomial of degree n with coefficients in $\mathbb{K}$ and for $\lambda \in \mathbb{K}$,

$$\lambda \text{ is an eigenvalue of } \mathbf{A} \iff \lambda \text{ is a root of } p_{\mathbf{A}}, \text{ i.e., } p_{\mathbf{A}}(\lambda) = 0.$$

In particular, an $n \times n$ matrix with entries in $\mathbb{K}$ has at most n eigenvalues in $\mathbb{K}$.

Algebraic and Geometric Multiplicities

Definition

Let $\mathbf{A} = [a_{jk}] \in \mathbb{K}^{n\times n}$ *and let* $\lambda \in \mathbb{K}$ *be an eigenvalue of* $\mathbf{A}$.

- *The* **algebraic multiplicity** *of* λ *(as an eigenvalue of* $\mathbf{A}$*) is the order* m *of the root* λ *of* $p_{\mathbf{A}}(t)$*, i.e.,* m *is the largest positive integer such that* $(t-\lambda)^m$ *divides* $p_{\mathbf{A}}(t)$.
- *The* **geometric multiplicity** *of* λ *(as an eigenvalue of* $\mathbf{A}$*) is the dimension of its eigenspace, i.e.,* $\dim \mathcal{N}(\mathbf{A}-\lambda\mathbf{I})$.

Observe that if $\lambda \in \mathbb{K}$ be an eigenvalue of $\mathbf{A}$, then

geometric multiplicity of $\lambda = \text{nullity}(\mathbf{A}-\lambda\mathbf{I}) = n - \text{rank}(\mathbf{A}-\lambda\mathbf{I})$.

This can be calculated by solving the homogeneous system $(\mathbf{A}-\lambda\mathbf{I})\mathbf{x} = \mathbf{0}$ using, for instance, Gaussian elimination. In fact, GEM and back substitution will also give the *basic solutions*, i.e., a set of eigenvectors of $\mathbf{A}$ which forms a basis of the eigenspace of $\mathbf{A}$ associated to λ.

Examples

(i) Let $\mathbf{A} = \text{diag}(a_1, \ldots, a_n)$, i.e., let $\mathbf{A}$ be a diagonal matrix with diagonal entries $a_1, \ldots, a_n$ in that order. Clearly

$$p_{\mathbf{A}}(t) = (a_1 - t)(a_2 - t) \cdots (a_n - t).$$

Thus the eigenvalues of $\mathbf{A}$ are precisely $a_1, \ldots, a_n$. Note that this can also be seen directly since $\mathbf{A}\mathbf{e}_k = a_k \mathbf{e}_k$ for each $k = 1, \ldots, n$, where $\mathbf{e}_1, \ldots, \mathbf{e}_n$ are the basic column vectors in $\mathbb{K}^{n \times 1}$. Observe that in this case the algebraic multiplicity of each eigenvalue is equal to its geometric multiplicity. Indeed, if $\lambda \in \{a_1, \ldots, a_n\}$, then the algebraic multiplicity of λ equals

$$m := \text{the number of } i \in \{1, \ldots, n\} \text{ such that } a_i = \lambda.$$

To find the geometric multiplicity of λ, consider $\mathbf{A} - \lambda \mathbf{I}$ and note that this is a diagonal matrix with exactly m rows of zeros and $n - m$ nonzero rows. So $\text{rank}(\mathbf{A} - \lambda \mathbf{I}) = n - m$ and hence

$$\text{geometric multiplicity of } \lambda = \text{nullity}(\mathbf{A} - \lambda \mathbf{I}) = m.$$

Examples Contd.

On the other hand, let $\lambda \in \mathbb{K}$ and $\lambda \notin \{a_1, \ldots, a_n\}$. Then $(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0} \iff (a_1 - \lambda)x_1 = \cdots = (a_n - \lambda)x_n = 0$. Thus the linear system $(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}$ has only the zero solution, and so λ is not an eigenvalue of $\mathbf{A}$.

If $\lambda \in \mathbb{K}$ appears g times on the diagonal of the matrix $\mathbf{A}$, then the geometric multiplicity of the eigenvalue λ is equal to g since the number of zero rows in an REF of $\mathbf{A} - \lambda\mathbf{I}$ is g.

(ii) Let $\mathbf{A}$ be an upper triangular matrix with diagonal entries $a_{11}, \ldots, a_{nn}$. Again, the characteristic polynomials factors as

$$p_{\mathbf{A}}(t) = (a_{11} - t)(a_{22} - t) \cdots (a_{nn} - t).$$

So the eigenvalues of $\mathbf{A}$ are precisely $a_{11}, \ldots, a_{nn}$. The algebraic multiplicities can be found as in the previous example. However, they may not always coincide with the corresponding geometric multipliciities.

Examples Contd.

For example, consider the 2×2 matrix $\mathbf{A} := \begin{bmatrix} 3 & 1 \\ 0 & 3 \end{bmatrix}$. Clearly 3 is the only eigenvalue of $\mathbf{A}$ and its algebraic multiplicity is 2. On the other hand, the homogeneous linear system $(\mathbf{A} - 3\mathbf{I})\mathbf{x} = \mathbf{0}$ comprises of the equations $x_2 = 0$ and $0 = 0$. So the eigenspace of $\mathbf{A}$ associated with the eigenvalue 3 has $[1\ 0]^{\mathsf{T}}$ as its basis. Thus the geometric multiplicity of the eigenvalue 3 of $\mathbf{A}$ is 1.

Let $\mathbf{A}$ be an upper triangular matrix with diagonal entries $a_{11}, \dots, a_{nn}$. Now a_{11} is an eigenvalue of $\mathbf{A}$ since $\mathbf{A}\mathbf{e}_1 = a_{11}\mathbf{e}_1$. In fact, each $a_{11}, \dots, a_{nn}$ is an eigenvalue of $\mathbf{A}$. To see this, let $\lambda \in \{a_{11}, \dots, a_{nn}\}$, and let k be the smallest positive integer such that $a_{kk} = \lambda$. Then $\mathbf{A} - \lambda\mathbf{I}$ is an upper triangular matrix with the kth diagonal entry zero and earlier diagonal entries nonzero, and so in an REF of $\mathbf{A} - \lambda\mathbf{I}$, the kth column cannot be pivotal. Hence $\text{rank}(\mathbf{A} - \lambda\mathbf{I}) < n$. This shows that λ is an eigenvalue of $\mathbf{A}$; eigenvectors of $\mathbf{A}$ corresponding to λ can be found by backward substitution.
On the other hand, if $\lambda \notin \{a_{11}, \dots, a_{nn}\}$, then backward substitution shows that the linear system $(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}$ has only the zero solution, and so λ is not an eigenvalue of $\mathbf{A}$.

(iii) Let $\mathbf{A}$ be a lower triangular matrix.Then $\mathbf{A}^\mathsf{T}$ is an upper triangular matrix with the same diagonal elements as $\mathbf{A}$. Also, $\operatorname{rank}(\mathbf{A} - \lambda \mathbf{I}) = \operatorname{rank}(\mathbf{A}^\mathsf{T} - \lambda \mathbf{I})$ for any $\lambda \in \mathbb{K}$. It follows that each diagonal element of $\mathbf{A}$ is an eigenvalue of $\mathbf{A}$, and no other $\lambda \in \mathbb{K}$ is an eigenvalue of $\mathbf{A}$. Eigenvectors of $\mathbf{A}$ corresponding to λ can be found by forward substitution.

In general, solving an eigenvalue problem $\mathbf{A}\mathbf{x} = \lambda\,\mathbf{x}$ is much harder than finding solutions of a linear system $\mathbf{A}\mathbf{x} = \mathbf{b}$. In the latter case, the matrix $\mathbf{A}$ and the right side $\mathbf{b}$ are given. On the other hand, in the eigenvalue problem, the 'unknown' vector $\mathbf{x}$ appears on both sides of the equation, and additionally, an 'unknown' scalar λ appears on the right side.

We need to find an eigenvalue λ of $\mathbf{A}$ and a corresponding eigenvector $\mathbf{x}$ of $\mathbf{A}$ simultaneously. It is tough, but if one of them is known beforehand, then the other can be found easily. Suppose a scalar λ is known to be an eigenvalue of $\mathbf{A}$. Then all eigenvectors of $\mathbf{A}$ corresponding to λ can be obtained by finding the general solution of the homogeneous linear system $(\mathbf{A} - \lambda\mathbf{I})\,\mathbf{x} = \mathbf{0}$. Next, suppose a nonzero vector $\mathbf{x}$ is known to be an eigenvector of $\mathbf{A}$. Then one only needs to calculate $\mathbf{A}\mathbf{x}$ and observe that it is a scalar multiple of $\mathbf{x}$. This scalar is the corresponding eigenvalue of $\mathbf{A}$.

Examples

(i) Let $\mathbf{A} := \begin{bmatrix} -2 & 2 & -3 \\ 2 & 1 & -6 \\ -1 & -2 & 0 \end{bmatrix}$, and suppose somehow we know that $\lambda := -3$ is an eigenvalue of $\mathbf{A}$. Let

$$\mathbf{B} := \mathbf{A} - (-3)\mathbf{I} = \mathbf{A} + 3\mathbf{I} = \begin{bmatrix} 1 & 2 & -3 \\ 2 & 4 & -6 \\ -1 & -2 & 3 \end{bmatrix}.$$

By EROs, we can transform $\mathbf{B}$ to $\mathbf{B}' := \begin{bmatrix} 1 & 2 & -3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$. Now the solution space of $\mathbf{B}'\mathbf{x} = \mathbf{0}$ is $\{\mathbf{x} \in \mathbb{R}^{3\times 1} : x_1 + 2x_2 - 3x_3 = 0\}$, which is also the solution space of $\mathbf{Bx} := (\mathbf{A} + 3\mathbf{I})\mathbf{x} = \mathbf{0}$, the basic solutions being $\mathbf{s}_2 := \begin{bmatrix} -2 & 1 & 0 \end{bmatrix}^{\mathsf{T}}$ and $\mathbf{s}_3 := \begin{bmatrix} 3 & 0 & 1 \end{bmatrix}^{\mathsf{T}}$. Thus the eigenvectors of $\mathbf{A}$ corresponding to the eigenvalue $\lambda = -3$ are the nonzero linear combinations of $\mathbf{s}_2$ and $\mathbf{s}_3$. The geometric multiplicity of the eigenvalue -3 is equal to 2.

(ii) Let $\mathbf{A} := \frac{1}{10}\begin{bmatrix} 29 & 6 & -1 \\ 29 & 16 & -11 \\ 25 & 10 & 15 \end{bmatrix}$, and suppose somehow we know that $\mathbf{x} := \begin{bmatrix} 1 & -1 & 3 \end{bmatrix}^{\mathsf{T}}$ is an eigenvector of $\mathbf{A}$. We easily find that $\mathbf{Ax} = 2\begin{bmatrix} 1 & -1 & 3 \end{bmatrix}^{\mathsf{T}}$. Hence 2 is the corresponding eigenvalue of $\mathbf{A}$.

We saw that eigenvalue problems for diagonal matrices are the easiest to solve. We wonder when a nondiagonal matrix would 'behave like a diagonal matrix'. To make this precise, we introduce the following notion.

Let $\mathbf{A}, \mathbf{B} \in \mathbb{K}^{n\times n}$. We say that $\mathbf{A}$ is **similar** to $\mathbf{B}$ (over $\mathbb{K}$) if there is an invertible $\mathbf{P} \in \mathbb{K}^{n\times n}$ such that $\mathbf{B} = \mathbf{P}^{-1}\mathbf{AP}$, that is, $\mathbf{AP} = \mathbf{PB}$. In this case, we write $\mathbf{A} \sim \mathbf{B}$.

Similarity of Square Matrices

Definition

Let $\mathbf{A}, \mathbf{B} \in \mathbb{K}^{n\times n}$. *We say that* $\mathbf{A}$ *is* **similar** *to* $\mathbf{B}$ *(over* $\mathbb{K}$*) if there is an invertible* $\mathbf{P} \in \mathbb{K}^{n\times n}$ *such that* $\mathbf{B} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}$, *that is,* $\mathbf{A}\mathbf{P} = \mathbf{P}\mathbf{B}$. *In this case, we write* $\mathbf{A} \sim \mathbf{B}$.

One can easily check (i) $\mathbf{A} \sim \mathbf{A}$, (ii) if $\mathbf{A} \sim \mathbf{B}$ then $\mathbf{B} \sim \mathbf{A}$, and (iii) if $\mathbf{A} \sim \mathbf{B}$ and $\mathbf{B} \sim \mathbf{C}$, then $\mathbf{A} \sim \mathbf{C}$.

Examples:

(i) Let $\mathbf{A} := \begin{bmatrix} 4 & -3 \\ 2 & -1 \end{bmatrix}$. It is easily seen that $\mathbf{P} := \begin{bmatrix} 3 & 1 \\ 2 & 1 \end{bmatrix}$ is invertible and . $\mathbf{P}^{-1} = \begin{bmatrix} 1 & -1 \\ -2 & 3 \end{bmatrix}$. Hence $\mathbf{A}$ is similar to

$$\mathbf{B} := \begin{bmatrix} 1 & -1 \\ -2 & 3 \end{bmatrix} \begin{bmatrix} 4 & -3 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} 3 & 1 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ -2 & 3 \end{bmatrix} \begin{bmatrix} 6 & 1 \\ 4 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix},$$

which is a diagonal matrix.

More Examples and a Characterization of Similarity

(ii) Let $\mathbf{A} \in \mathbb{K}^{n\times n}$. Then $\mathbf{A} \sim \mathbf{I} \iff \mathbf{A} = \mathbf{I}$.

(iii) Let $\mathbf{A} \in \mathbb{K}^{n\times n}$, and let $\mathbf{E}$ be $n\times n$ an elementary matrix. Then $\mathbf{B} := \mathbf{EAE}^{-1}$ is similar to $\mathbf{A}$. Note: $\mathbf{EA}$ is obtained from $\mathbf{A}$ by an elementary row operation on $\mathbf{A}$, and $\mathbf{B}$ is obtained from $\mathbf{EA}$ by the 'reverse column operation' on $\mathbf{EA}$.

Similarity of matrices has the following characterization.

Proposition

Let $\mathbf{A}, \mathbf{B} \in \mathbb{K}^{n\times n}$. Then $\mathbf{A} \sim \mathbf{B}$ if and only if there is an ordered basis E for $\mathbb{K}^{n\times 1}$ such that $\mathbf{B}$ is the matrix of the linear transformation $T_{\mathbf{A}} : \mathbb{K}^{n\times 1} \to \mathbb{K}^{n\times 1}$ with respect to E.

In fact, $\mathbf{B} = \mathbf{P}^{-1}\mathbf{AP}$ if and only if the columns of $\mathbf{P}$ form an ordered basis, say E, for $\mathbb{K}^{n\times 1}$ and $\mathbf{B} = \mathbf{M}_E^E(T_{\mathbf{A}})$.

A proof of this proposition will be outlined later.

Diagonalizability

Definition

A matrix $\mathbf{A} \in \mathbb{K}^{n\times n}$ *is called* **diagonalizable** *(over* $\mathbb{K}$*) if* $\mathbf{A}$ *is similar to a diagonal matrix (over* $\mathbb{K}$*).*

Here is a useful characterization of diagonalizability.

Proposition

A matrix $\mathbf{A} \in \mathbb{K}^{n\times n}$ is diagonalizable if and only if there is a basis for $\mathbb{K}^{n\times 1}$ consisting of eigenvectors of $\mathbf{A}$. In fact,

$$
\begin{aligned}
& \mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{D}, \text{ where } \mathbf{P}, \mathbf{D} \in \mathbb{K}^{n\times n} \text{ are of the form} \\
& \mathbf{P} = \begin{bmatrix} \mathbf{x}_1 & \cdots & \mathbf{x}_n \end{bmatrix} \text{ and } \mathbf{D} = \operatorname{diag}(\lambda_1, \ldots, \lambda_n) \\
\iff \quad & \{\mathbf{x}_1, \ldots, \mathbf{x}_n\} \text{ is a basis for } \mathbb{K}^{n\times 1} \text{ and} \\
& \mathbf{A}\mathbf{x}_k = \lambda_k \mathbf{x}_k \text{ for } k = 1, \ldots, n.
\end{aligned}
$$

We shall see a proof of this result as well as many examples of diagonalizable and non-diagonalizable matrices later.