Conversation 37: Diagonalization

Winfried Just
Department of Mathematics, Ohio University

MATH3200: Applied Linear Algebra

A mystery matrix

Bob: Here is a practice problem that I found online:

Let $\vec{\mathbf{x}}_1 = \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix} \quad \vec{\mathbf{x}}_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \quad \vec{\mathbf{x}}_3 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$

Find a 3×3 matrix $\mathbf{A}$ such that the corresponding transformation $L_{\mathbf{A}} : \mathbb{R}^3 \to \mathbb{R}^3$:

- stretches every vector $\vec{\mathbf{x}}$ on the line $span(\vec{\mathbf{x}}_1)$ by a factor of 2,
- flips every vector $\vec{\mathbf{x}}$ on the line $span(\vec{\mathbf{x}}_2)$ to $-\vec{\mathbf{x}}$,
- and maps every vector on the line $span(\vec{\mathbf{x}}_3)$ to the origin $\vec{\mathbf{0}}$.

Cindy: Wouldn't such a matrix be similar to $\mathbf{D} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$?

Similar in what sense?

Denny: What do you mean by "similar," Cindy?

Cindy: I mean for the matrix $\mathbf{D} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

the transformation $L_{\mathbf{D}} : \mathbb{R}^3 \to \mathbb{R}^3$ does all of the following:

- stretches every vector $\vec{\mathbf{x}}$ on the line $span(\vec{\mathbf{e}}_1)$ by a factor of 2,
- flips every vector $\vec{\mathbf{x}}$ on the line $span(\vec{\mathbf{e}}_2)$ to $-\vec{\mathbf{x}}$,
- and maps every vector on the line $span(\vec{\mathbf{e}}_3)$ to the origin $\vec{\mathbf{0}}$.

So Bob's matrix **A** would do the exact same thing to the lines spanned by the vectors $\vec{\mathbf{x}}_1, \vec{\mathbf{x}}_2, \vec{\mathbf{x}}_3$ as the diagonal matrix **D** does to the lines spanned by the vectors $\vec{\mathbf{e}}_1, \vec{\mathbf{e}}_2, \vec{\mathbf{e}}_3$.

Alice: Excellent observation, Cindy!

Denny: And I'm sure Theo can tell us what that "exact same thing" means, exactly.

Finding a matrix with specified eigenvectors

Theo: Will be happy to. Cindy has observed that

- $\vec{\mathbf{e}}_1$ is an eigenvector of $\mathbf{D}$ with eigenvalue $\lambda_1 = 2$,
- $\vec{\mathbf{e}}_2$ is an eigenvector of $\mathbf{D}$ with eigenvalue $\lambda_2 = -1$,
- and $\vec{\mathbf{e}}_3$ is an eigenvector of $\mathbf{D}$ with eigenvalue $\lambda_3 = 0$.

Bob's question asks us to find a matrix $\mathbf{A}$ such that

- $\vec{\mathbf{x}}_1$ is an eigenvector of $\mathbf{A}$ with eigenvalue $\lambda_1 = 2$,
- $\vec{\mathbf{x}}_2$ is an eigenvector of $\mathbf{A}$ with eigenvalue $\lambda_2 = -1$,
- and $\vec{\mathbf{x}}_3$ is an eigenvector of $\mathbf{A}$ with eigenvalue $\lambda_3 = 0$.

Bob: This wasn't even covered yet in the course.

Cindy: But perhaps we can figure out how to do it?
I mean, with Alice and Theo helping us along?

Can't we take $\mathbf{A} = \mathbf{D}$?

Frank: Can't we simply take $\mathbf{A} = \mathbf{D} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$?

Question C37.1: Would this work?

Theo: Obviously not. We are looking for eigenvectors

$$\vec{\mathbf{x}}_1 = \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix} \qquad \vec{\mathbf{x}}_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \qquad \vec{\mathbf{x}}_3 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} \quad \text{of } \mathbf{A}$$

that point in different directions than the eigenvectors

$$\vec{\mathbf{e}}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \qquad \vec{\mathbf{e}}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \qquad \vec{\mathbf{e}}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \quad \text{of } \mathbf{D}.$$

Can we translate the matrix **A** into **D**?

Frank: That would be an issue only if these vectors are written in standard coordinates. Who says they are?

Theo: This is implicitly understood unless explicitly specified otherwise.
These standard, or **Cartesian** coordinates are named after the French philosopher and mathematician René Descartes (1596–1650) who invented them.

Denny: So, in a way, the vectors are written in **French.** And Bob's matrix **A** will come out as something like **French text** where it becomes difficult to comprehend what $L_{\mathbf{A}}$ actually does.

Frank: That's exactly my point, Denny! Perhaps there is a clever way to translate this matrix **A** into **English** so that it becomes **D** and we all can see right away what it does?

Alice: Excellent idea, Frank! How would you go about the translation? What would you use as your dictionary, so to speak?

How about using alternative coordinates?

Frank: I think we should use alternative coordinates for our vectors $\vec{\mathbf{x}}_1, \vec{\mathbf{x}}_2, \vec{\mathbf{x}}_3$.

Alice: "Alternative coordinates" always means "alternative coordinates with respect to a given basis B."
What would you take here as your basis?

Frank: I would take $B = \{\vec{\mathbf{x}}_1, \vec{\mathbf{x}}_2, \vec{\mathbf{x}}_3\}$.

Question 37.2: Would this be a basis for $\mathbb{R}^3$?
How can we find out whether it is?

Bob: Since there are three vectors, we only need to verify that they are linearly independent. We can form a matrix

$$\mathbf{B} = \begin{bmatrix} 1 & 1 & 2 \\ -2 & 0 & 1 \\ 3 & 1 & 0 \end{bmatrix}$$ that has $\vec{\mathbf{x}}_1, \vec{\mathbf{x}}_2, \vec{\mathbf{x}}_3$ as its columns.

My calculations show that $\det(\mathbf{B}) = -2$, so B is a basis of $\mathbf{R}^3$.

Alternative coordinates for $\vec{x}_1, \vec{x}_2, \vec{x}_3$

Alice: Now we can express the vectors $\{\vec{x}_1, \vec{x}_2, \vec{x}_3\}$ in alternative coordinates $\{\vec{c}_1, \vec{c}_2, \vec{c}_3\}$ with respect to B.

Denny: What, again, are "alternative coordinates"?

Theo: We need to express each vector $\vec{x}_i$ as a linear combination of the vectors in the basis B. The coefficients of these linear combinations would then be the **alternative coordinates** with respect to B.

Denny: But $\vec{x}_1, \vec{x}_2, \vec{x}_3$ are already in B.
So what is there to "express as a linear combination?"

Question C37.3: What would you reply to Denny here?

Theo: $\vec{x}_1 = 1\vec{x}_1 + 0\vec{x}_2 + 0\vec{x}_3$.

Bob: $\vec{x}_2 = 0\vec{x}_1 + 1\vec{x}_2 + 0\vec{x}_3$.

Cindy: $\vec{x}_3 = 0\vec{x}_1 + 0\vec{x}_2 + 1\vec{x}_3$.

Bob's matrix in alternative coordinates

Denny: What you guys are telling me is that in **alternative coordinates** with respect to B, the vectors $\vec{\mathbf{x}}_1, \vec{\mathbf{x}}_2, \vec{\mathbf{x}}_3$ become:

$$\vec{\mathbf{c}}_1 = \vec{\mathbf{e}}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \qquad \vec{\mathbf{c}}_2 = \vec{\mathbf{e}}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \qquad \vec{\mathbf{c}}_3 = \vec{\mathbf{e}}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Frank: So I was right all along: If we do all computations in **alternative coordinates,** for Cindy's diagonal matrix **D**, the transformation $L_{\mathbf{D}}$, does exactly what Bob's $L_{\mathbf{A}}$ was supposed to do when we compute with **Cartesian coordinates.**

Question C37.4: Did Denny and Frank get this right?

Alice: Yes, Denny and Frank!

Denny: I'd call this a nice translation from **French** into **English.**

Bob: Neat. But what is the matrix **A** for **Cartesian coordinates**?

That's a little more complicated . . .

Theo: **C'est un peu plus compliqué.**

Denny: Showoff!

Alice: A little more complicated, yes. But Frank has basically already figured out how to compute $\mathbf{A}$.

Frank: Who, me? I only told you how to get away *without* $\mathbf{A}$!

Alice: OK. But could you kindly walk us through your steps of computing $L_{\mathbf{A}}(\vec{\mathbf{x}}) = \mathbf{A}\vec{\mathbf{x}}$ *without* actually using $\mathbf{A}$?

Frank: Sure. First we translate $\vec{\mathbf{x}}$ into *alternative coordinates* $\vec{\mathbf{c}}$ with respect to that basis B of desired eigenvectors.

Denny: I see now how this works when $\vec{\mathbf{x}}$ is one of the eigenvectors. But how does it work in general?

Bob: In Chapter 3 we learned how to find coefficients $\vec{\mathbf{c}}$ for expressing a given vector $\vec{\mathbf{x}}$ as a linear combination of the columns of a given matrix $\mathbf{B}$. Here is how this works:

Step 1: Express $\vec{\mathbf{x}}$ in alternative coordinates

Suppose $B = \{\vec{\mathbf{x}}_1, \vec{\mathbf{x}}_2, \dots, \vec{\mathbf{x}}_n\}$ is a basis of $\mathbb{R}^n$ and the basis vectors are written in **Cartesian coordinates.** Write these vectors in the given order as the columns of a matrix $\mathbf{B}$.

Consider a vector $\vec{\mathbf{x}}$ written in **Cartesian coordinates.**

We are looking for a vector of coefficients $\vec{\mathbf{c}}$ such that:

$$c_1\vec{\mathbf{x}}_1 + c_2\vec{\mathbf{x}}_2 + \cdots + c_n\vec{\mathbf{x}}_n = \vec{\mathbf{x}}.$$

In matrix notation, we are looking for a solution of:

$$\mathbf{B}\vec{\mathbf{c}} = \vec{\mathbf{x}}.$$

Since $r(\mathbf{B}) = n$, this system has a unique solution given by

$$\vec{\mathbf{c}} = \mathbf{B}^{-1}\vec{\mathbf{x}}.$$

This is the formula for changing **Cartesian coordinates** into **alternative coordinates** with respect to B.

Step 2: Multiply by a diagonal matrix

Alice: Thank you, Bob! So what would you do next, Frank?

Frank: Now I take the resulting vector $\vec{\mathbf{c}}$ that represents $\vec{\mathbf{x}}$ in **alternative coordinates** and multiply it with the diagonal matrix $\mathbf{D}$ that I get by writing the desired eigenvalues $\lambda_1, \dots, \lambda_n$ for the vectors $\vec{\mathbf{x}}_1, \dots, \vec{\mathbf{x}}_n$, respectively, on the diagonal. This gives the output $L_{\mathbf{A}}(\vec{\mathbf{x}})$, but written in **alternative coordinates.**

Cindy: So you form $\mathbf{D} = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}$ and then you

compute some vector $\vec{\mathbf{d}} = \mathbf{D}\vec{\mathbf{c}} = \mathbf{D}(\mathbf{B}^{-1}\vec{\mathbf{x}}) = \mathbf{D}\mathbf{B}^{-1}\vec{\mathbf{x}}$, right?

Question 37.5: Did Cindy get this right?

So how about $\mathbf{A}\vec{\mathbf{x}}$?

Frank: Now that you mentioned it, yes, that's what I do.

Denny: And how would we get the value $L_{\mathbf{A}}(\vec{\mathbf{x}})$ in **Cartesian coordinates** from your $\vec{\mathbf{d}}$?

Frank: I'd rather stick to **alternative coordinates** here, but if you insist: Translate them back to **Cartesian coordinates.**

Denny: How?

Cindy: Since in order to go from **Cartesian coordinates** to **alternative coordinates** we need to multiply by $\mathbf{B}^{-1}$ wouldn't we go back by multiplying with $\mathbf{B}$ instead?
Since $\mathbf{B}$ is the inverse matrix of $\mathbf{B}^{-1}$?

Alice: Very good observation, Cindy!

This works since in **alternative coordinates** with respect to B each column $\vec{\mathbf{x}}_i$ of $\mathbf{B}$ becomes $\vec{\mathbf{e}}_i$, as Frank had observed.

Step 3: Express $\vec{\mathbf{d}}$ in Cartesian coordinates

Suppose $B = \{\vec{\mathbf{x}}_1, \vec{\mathbf{x}}_2, \dots, \vec{\mathbf{x}}_n\}$ is a basis of $\mathbb{R}^n$ and the basis vectors are written in **Cartesian coordinates.** Write these vectors in the given order as the columns of a matrix $\mathbf{B}$.

Consider a vector $\vec{\mathbf{d}}$ written in **alternative coordinates with respect to B.**

Then $\vec{\mathbf{d}}$ is the following linear combination of basis vectors:

$$\vec{\mathbf{d}} = d_1\vec{\mathbf{e}}_1 + d_2\vec{\mathbf{e}}_2 + \cdots + d_n\vec{\mathbf{e}}_n = d_1\vec{\mathbf{x}}_1 + d_2\vec{\mathbf{x}}_2 + \cdots + d_n\vec{\mathbf{x}}_n = \vec{\mathbf{y}}.$$

The expression on the right can be written in matrix notation as:

$$\vec{\mathbf{y}} = \mathbf{B}\vec{\mathbf{d}}.$$

This is the formula for changing **alternative coordinates** with respect to B into **Cartesian coordinates** that we learned in Chapter 3.

So how about $\mathbf{A}$?

Frank: So, if I were to write the whole procedure in the terminology suggested by Cindy, it would become:

$\vec{\mathbf{y}} = L_{\mathbf{A}}(\vec{\mathbf{x}}) = \mathbf{A}\vec{\mathbf{x}} = \mathbf{B}\vec{\mathbf{d}} = \mathbf{B}\mathbf{D}\mathbf{B}^{-1}\vec{\mathbf{x}}$.

No matrix $\mathbf{A}$ needed on the right.

Bob: Granted. But what **is** the matrix $\mathbf{A}$?

Question C37.6: Indeed, what is $\mathbf{A}$?

Frank: Your $\mathbf{A}$ is simply $\mathbf{B}\mathbf{D}\mathbf{B}^{-1}$.

Denny: Cool! Does your procedure always work, Frank?

Frank: It works whenever we have a full set of eigenvectors that we can write as the columns of $\mathbf{B}$ with given eigenvalues that we can write on the diagonal of $\mathbf{D}$.

Alice: You have proved a theorem, Frank.

Frank: No way!!

A theorem

Theo: Here is your theorem:

Theorem

Let $\mathbf{A}$ *be any matrix of order* $n \times n$ *that has a full set of eigenvectors. Then there exist an invertible matrix* $\mathbf{B}$ *and a diagonal matrix* $\mathbf{D}$ *such that*

$$\mathbf{A} = \mathbf{BDB}^{-1}.$$

The matrix $\mathbf{D}$ *is called a diagonalization of* $\mathbf{A}$ *and has the eigenvalues of* $\mathbf{A}$ *on the diagonal, while the columns of* $\mathbf{B}$ *are eigenvectors of* $\mathbf{A}$*, listed in the same order as their eigenvalues on the diagonal of* $\mathbf{D}$*.*

Frank: I guess I did prove this ... with a little help from you all. Thank you!

The matrix **A** in Bob's example

The others: You are welcome. Congratulations, Frank!

Bob: Now I can calculate the matrix **A** of my original question.

Here $\mathbf{B} = \begin{bmatrix} 1 & 1 & 2 \\ -2 & 0 & 1 \\ 3 & 1 & 0 \end{bmatrix}$ and $\mathbf{D} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

We need $\mathbf{B}^{-1}$. Let's use MATLAB: $\mathbf{B}^{-1} = \begin{bmatrix} 0.5 & -1 & -0.5 \\ -1.5 & 3 & 2.5 \\ 1 & -1 & -1 \end{bmatrix}$

By Frank's theorem: $\mathbf{A} = \mathbf{BDB}^{-1} = \begin{bmatrix} 2.5 & -5 & -3.5 \\ -2 & 4 & 2 \\ 4.5 & -9 & -5.5 \end{bmatrix}$

Denny: Kinda difficult to see though what the linear transformation $L_{\mathbf{A}}$ for this matrix does.

Similar matrices

Theo: **Un peu difficile.** If we use **Cartesian coordinates.**

Frank: Which was my point all along: Use **alternative coordinates** and work with the **diagonalization D** instead, and you will see in **plain English** what is going on.

Cindy: I guess that's like what I meant when I said earlier that the matrices **A** and **D** are similar.

Theo: Excellent choice of word, Cindy! There is an important definition of this concept in the textbooks:

Definition

Let **A** and **C** be two square matrices of the same order. Then we say that **A** and **C** are *similar* if there exists an invertible matrix **B** such that

$$\mathbf{C} = \mathbf{B}^{-1}\mathbf{A}\mathbf{B}.$$

A square matrix **A** that is similar to a diagonal matrix is called *diagonalizable*.

One more theorem

Alice: Frank's theorem essentially says that if **A** has a full set of eigenvectors, then it is diagonalizable. But it works also the other way around, as the following theorem shows:

Theorem

Let **A** *be any matrix of order* $n \times n$*. Then* **A** *is diagonalizable if, and only if, it has a full set of eigenvectors.*

Frank: Sounds plausible.

Denny: But I don't see offhand the precise connection.

Bob: We will explore this connection in more detail in Module 72. Let's call it quits for today.

Take-home message

Let B be a given basis for $\mathbb{R}^n$, and let $\mathbf{B}$ be the matrix whose columns are the vectors in B. Then:

- For any vector $\vec{\mathbf{x}}$ that is written in Cartesian coordinates, $\vec{\mathbf{c}} = \mathbf{B}^{-1}\vec{\mathbf{x}}$ is the vector of alternative coordinates for $\vec{\mathbf{x}}$ with respect to B.
- $\vec{\mathbf{x}} = \mathbf{B}\vec{\mathbf{c}}$ gives the Cartesian coordinates of the vector $\vec{\mathbf{c}}$ that is written in alternative coordinates with respect to B.
- When B consists of eigenvectors of a matrix $\mathbf{A}$, then $\mathbf{A} = \mathbf{B}\mathbf{D}\mathbf{B}^{-1}$, where $\mathbf{D}$ is the *diagonalization* of $\mathbf{A}$ that lists the respective eigenvalues of the vectors in B on the main diagonal and has zero elements in all off-diagonal places.
- A square matrix $\mathbf{A}$ is *diagonalizable* if, and only if, it is *similar* to a diagonal matrix $\mathbf{D}$, which means that $\mathbf{D} = \mathbf{B}^{-1}\mathbf{A}\mathbf{B}$ for some invertible matrix $\mathbf{B}$.
- A square matrix $\mathbf{A}$ is diagonalizable if, and only if, it has a full set of eigenvectors.