Lesson Plan 4 on Bases, Coordinate Vectors and Linear Transformations under Basis Change

Frank Uhlig

Department of Mathematics and Statistics,
Auburn University, Auburn, AL 36849-5310, USA
uhligfd@auburn.edu

This Lesson Plan 4 introduces coordinate vectors with respect to general bases $\mathcal{U}$ and how to translate between coordinate vectors under basis change. We study matrix similarities, diagonalizable matrices and eigenvalues and eigenvectors of square matrices. And finish by solving linear systems of ordinary differential equations and a short look at the history of Matrix Theory.

(version August 3, 2023) iii + 17 p.; 4 - 5 class days

Concepts, Notions and Definitions in Lesson Plan 4

For [1] p. 2

Coordinate Vector

Standard $\mathcal{E} = \{e_i\}$ Basis

$\mathcal{U} = \{u_i\}$ Coordinate Vector $x_{\mathcal{U}}$

Column Vector Matrix U

Transform $\mathcal{U}$ to $\mathcal{V}$ Coordinate Vectors

Multi-augmented Matrix $(V|U)$

For [2] p. 6

Basis Change

Matrix Representation $A_{\mathcal{U}}$ for Basis $\mathcal{U}$

Matrix Similarity

(continued)

This is the transitory lesson starting from the **linear** *term in Linear Algebra progressing to nonlinear Matrix Algebra.*

Going forth from our first lesson on linear transformations and their representation as matrix $\times$ *vector products for the standard* $\mathcal{E}$ *basis of the unit vectors* $e_i \in \mathbb{R}^n$*, we now deal with arbitrary bases* $\mathcal{U} = \{u_1, ..., u_n\}$ *and how to represent vectors and linear transformations with respect to bases other than* $\mathcal{E}$.

We introduce U coordinate vectors and how to relate coordinate vectors of points in $\mathbb{R}^n$ *for different bases.*

Then we represent a linear transformation given in its standard $\mathcal{E}$ *representation* $A_\mathcal{E}$ *with respect to other bases* $\mathcal{U}$ *as* $A_\mathcal{U}$.

This introduces matrix similarities and the quest of finding bases $\mathcal{U}$ *from* $A_\mathcal{E}$ *that give* $A_\mathcal{U} = U^{-1}A_\mathcal{E}U$ *sparsity which gives us deeper insights into a linear transform's* **intrinsic qualities** *when operating on* $\mathbb{R}^n$.

[1] Coordinate Vectors

There are many, nay infinitely many different bases of $\mathbb{R}^n$ since any set of n linearly independent vectors in $\mathbb{R}^n$ is a basis.

Any basis $\mathcal{U}$ of $\mathbb{R}^n$ can be used to describe locations and actions in $\mathbb{R}^n$.

The standard basis $\mathcal{E} = \{e_1, .., e_n\}$ of $\mathbb{R}^n$ consists of the unit vectors $e_i \in \mathbb{R}^n$ with zeros in every position except for a 1 in position i, i.e.,

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}_n, \quad e_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}_n, \quad \text{or } e_{n-1} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \end{pmatrix}_n$$

for example.

Thus far we have dealt with vectors x in $\mathbb{R}^n$ in the standard basis $\mathcal{E}$. For clarity purposes we write $x = I_n \cdot x = x_{\mathcal{E}}$ with $\mathcal{E}$ subscripts for any standard vector x (for a while now) :

$$x_{\mathcal{E}} = \begin{pmatrix} \vdots & & \vdots \\ e_1 & \dots & e_n \\ \vdots & & \vdots \end{pmatrix} x_{\mathcal{E}} = I_n \cdot x_{\mathcal{E}}.$$

What should $x_{\mathcal{U}}$ describe for another basis $\mathcal{U} = \{u_1, ..., u_n\}$ with all u_i for $1 \leq i \leq n$ in standard $\mathcal{E}$ vector form $u_{i_{\mathcal{E}}}$?

Clearly the $\mathcal{U}$ coordinate vector $x_{\mathcal{U}}$ must satisfy

$$x_{\mathcal{E}} = \begin{pmatrix} \vdots & & \vdots \\ u_{1_{\mathcal{E}}} & \dots & u_{n_{\mathcal{E}}} \\ \vdots & & \vdots \end{pmatrix} x_{\mathcal{U}} = U_{n,n} \cdot x_{\mathcal{U}}$$

for the column vector matrix U of the basis vectors $\{u_{i_{\mathcal{E}}}\}$ in $\mathcal{U}$.

For any basis $\mathcal{V} = \{v_1, ..., v_n\}$ of $\mathbb{R}^n$ and its column vector matrix V comprised of the basis vectors $\{u_{i_\mathcal{E}}\}$ we likwise have

$$x_\mathcal{E} = V \cdot x_\mathcal{V}.$$

And thus $x_\mathcal{E} = U \cdot x_\mathcal{U} = V \cdot x_\mathcal{V}$ or

$$V^{-1}U \cdot x_\mathcal{U} = x_\mathcal{V} \text{ and } x_\mathcal{U} = U^{-1}V \cdot x_\mathcal{V}$$

using matrix inverses of the nonsingular matrices V and U.

Thus $V^{-1}U$ transforms U coordinate vectors to V coordinate vectors and $U^{-1}V$ transforms V coordinate vectors to U coordinate vectors.

Both coordinate transforms use the $\mathcal{E}$ coordinate vectors $x_\mathcal{E} = U \cdot x_\mathcal{U}$ and $x_\mathcal{E} = V \cdot x_\mathcal{V}$, respectively, as intermediaries.

This reminds us of trying to translate a phrase from Ukrainian to Vietnamese by using a Ukrainian to English dictionary first and then an English to Vietnamese dictionary.

How can the coordinate vector transforms $V^{-1}U$ and $U^{-1}V$ be computed quickly and accurately without much pain?

Use software that can handle vectors, matrices, and row reduction, such as Matlab, Mathematica, Python, Octave et cetera.

Start from the multi-augmented matrix

$$(V|U)_{n,2n}$$

and reduce V to I_n on the left by Gaussian elimination that is performed across each complete row of the n by $2n$ matrix $(V|U)_{n,2n}$:

$$\begin{array}{c|c} V & U \\ \hline \vdots & \vdots \\ \text{row} & \vdots \\ \vdots & \text{reduce} \\ \downarrow & \downarrow \\ \hline I_n & V^{-1} \cdot U \end{array} \qquad (*)$$

How could students use the previous scheme $(*)$ *to compute the inverse* A^{-1} *of a square matrix* A *or check if a matrix* $A_{n,n}$ *is invertible?*

[2] Coordinate Vectors under Basis Change

Our first lesson studied linear transformations

$$f : x \in \mathbb{R}^n \to f(x) \in \mathbb{R}^n$$

by using the associated matrix × vector product of the standard matrix representation $A_{\mathcal{E}}$ of f and the standard vector representation $x_{\mathcal{E}}$:

$$A_{\mathcal{E}} = \begin{pmatrix} \vdots & & \vdots \\ f(e_1) & \dots & f(e_n) \\ \vdots & & \vdots \end{pmatrix} \cdot x_{\mathcal{E}}.$$

How can another basis $\mathcal{U}$ represent the same linear transformation f as $A_{\mathcal{U}}$ that maps $\mathcal{U}$ vectors to $\mathcal{U}$ vectors directly?

This task is solved in two steps:

(a) We know that $x_{\mathcal{E}} = Ux_{\mathcal{U}}$ is an $\mathcal{E}$ vector and therefore $A_{\mathcal{E}} \cdot (Ux_{\mathcal{U}})_{\mathcal{E}}$ remains an $\mathcal{E}$ vector.

(b) Since $y_{\mathcal{U}} = U^{-1} y_{\mathcal{E}}$ for all $\mathcal{E}$ vectors y, (a) tells us that $U^{-1}(A_{\mathcal{E}} Ux_{\mathcal{U}})_{\mathcal{E}}$ is a $\mathcal{U}$ vector. I.e., $U^{-1}A_{\mathcal{E}}U$ maps all $\mathcal{U}$ vectors $x_{\mathcal{U}}$ to $\mathcal{U}$ vectors and $A_{\mathcal{U}} = U^{-1}A_{\mathcal{E}}U$ is as desired.

The last 5 pages have contained very complicated algebraic manipulations of subscripted vectors and matrices. Ideally these equations and their development should be covered several times, first by the instructor - slowly and carefully; then repeated in student presentations in parts and again in later lessons when problems arise.

Double subscripts turn everybody off, sorry. But for clarity and understanding of the what and the which, they are essential here and necessary in nonlinear Modern Matrix Theory.

Once these notions are mastered, the subscripts can be dropped and every basis change operation can be implemented as simple basis vector matrix multiplications.
Good luck and hope for the best.

[3] Matrix Representations under Special Basis Changes

Most, (almost all) square matrices $A_{n,n}$ allow a sparse diagonal matrix representation for a matrix associated basis $\mathcal{U}_A$.

[Exceptions to this 'almost rule' will be dealt with later.]

An exercise problem (use software, please) :

For $U = \begin{pmatrix} -3 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$ *and* $A_{\mathcal{E}} = \begin{pmatrix} 2 & 0 & 3 \\ 0 & -1 & 0 \\ 1 & 0 & 4 \end{pmatrix}$ *find the matrix representation* $A_{\mathcal{U}} = U^{-1} \cdot A_{\mathcal{E}} \cdot U$ *of the underlying linear transformation for the basis* $\mathcal{U}$ *of* $\mathbb{R}^3$.

What does $A_{\mathcal{U}}$ *look like? What is* $A_{\mathcal{E}}u_1$*,* $A_{\mathcal{E}}u_2$*, and* $A_{\mathcal{E}}u_3$*?*

Why and how are the basis vectors in $\mathcal{U}$ *special for* A*?*

Real or complex nonzero vectors in n-space that are replicated in their direction when multiplied by a matrix A such as u_3 in $A_{\mathcal{E}}u_3 = 5\ u_3$ above are called eigenvectors of A in English and the scaling factor 5 above is called an eigenvalue of A.

The eigenvalue/eigenvector equation comes in two forms for diagonalizable matrices A :

$$U^{-1} \cdot A \cdot U = D \qquad \text{or} \qquad A \cdot U = U \cdot D.$$

In the second form of the eigen-equation the eigenvector basis matrix U appears to be dancing from the right side to the left. This is due to the non-commuting nature of matrix products.

Soon we will understand that all square matrices have eigenvalues and eigenvectors - and that very few matrices commute.

At this time students should be able to answer these questions:

Does the matrix $C = \begin{pmatrix} 4 & 8 \\ 1 & 2 \end{pmatrix}$ *have the eigenvalue 1, 6 or 0?*

and

Are $v = \begin{pmatrix} 4 \\ 2 \end{pmatrix}$, $u = \begin{pmatrix} 8 \\ 2 \end{pmatrix}$, $r = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ *or* $w = \begin{pmatrix} 4 \\ 1 \end{pmatrix}$ *eigenvectors of* C*?*

[4] Eigenpairs under Basis Change and Systems of Linear Differential Equations

If $A = A_{\mathcal{E}}$ is the standard matrix representation of a linear transformation $f : \mathbb{R}^n \to \mathbb{R}^n$ and A has the eigenpair λ and $w \in \mathbb{R}^n$, i.e., if $Aw = \lambda w$, what happens to the eigenpair λ and w of A when we represent f with respect to another basis $\mathcal{U}$ as $A_{\mathcal{U}} = U^{-1} A_{\mathcal{E}} U$?

Multiplying $A_{\mathcal{U}} = U^{-1}A_{\mathcal{E}}U$ on the right by $\mathcal{U}^{-1}$ we obtain

$$A_{\mathcal{U}}U^{-1} = U^{-1}A_{\mathcal{E}}$$

and thus

$$A_{\mathcal{U}}U^{-1}w = U^{-1}A_{\mathcal{E}}w = U^{-1}Aw = U^{-1}\lambda w = \lambda U^{-1}w \ .$$

Reading the above equation as $A_{\mathcal{U}}(U^{-1}w) = \lambda(U^{-1}w)$ makes $U^{-1}w$ an eigenvector of $A_{\mathcal{U}}$ for the same eigenvalue λ as the standard matrix representation $A_{\mathcal{E}}$ of the linear transformation f and it modifies the eigenvector w of A.

Since $x_{\mathcal{E}} = Ux_{\mathcal{U}}$ holds for basis changes from $\mathcal{E}$ to $\mathcal{U}$ and $w = w_{\mathcal{E}}$, we see that $U^{-1}w = U^{-1}w_{\mathcal{E}} = w_{\mathcal{U}}$ is the $\mathcal{U}$ basis vector $w_{\mathcal{U}}$ of the $\mathcal{E}$ eigenvector $w_{\mathcal{E}}$ of $A = A_{\mathcal{E}}$.

Hence eigenvalues remain the same under any basis change $\mathcal{V} \to \mathcal{U}$, while the corresponding eigenvectors morph from $\mathcal{V}$ vectors to $\mathcal{U}$ vectors and describe the same point in space.

With a little help from Calculus and Matrix Algebra we can now solve linear differential equations

$$\dot{x}(t) = A \cdot x(t)$$

for diagonalizable system matrices $A_{n,n}$.

Given $U^{-1}AU = D$ for an eigenvector basis matrix U of A and and D a diagonal matrix, we multiply the linear differential equation by U^{-1} from left and obtain

$$\begin{aligned} U^{-1}\dot{x}(t) &= \frac{d(U^{-1}x(t))}{dt} = U^{-1}Ax(t) = U^{-1}A(UU^{-1})x(t) \\ &= (U^{-1}AU)U^{-1}x(t) = D \cdot U^{-1}x(t) \; . \end{aligned}$$

This is a differential equation in $y(t) = U^{-1}x(t)$.

If $n = 2$ and $A_{2,2}$ is diagonalizable for $D = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$, then

$$\dot{y}(t) = \begin{pmatrix} \dot{y}_1(t) \\ \dot{y}_2(t) \end{pmatrix} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} y_1(t) \\ y_2(t) \end{pmatrix} = \begin{pmatrix} \lambda_1 \, y_1(t) \\ \lambda_2 \, y_2(t) \end{pmatrix}.$$

[From Calculus we know that $f(x) = e^{\alpha x}$ has the derivative $f'(x) = \alpha \, e^{\alpha x}$ and thus $y_i(t) = e^{\lambda_i t} + c_i$ for $i = 1, 2$.]

Finally using $x(t) = Uy(t)$, the general solution $x(t)$ of the linear differential equation $\dot{x}(t) = Ax(t)$ is

$$x(t) = \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} = U \begin{pmatrix} e^{\lambda_1 t} \\ e^{\lambda_2 t} \end{pmatrix} + \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$$

for arbitrary constants c_i.

[5] Namings, Names and History

The real or complex scalars λ in the eigenvalue-eigenvector equation $Ax = \lambda x$ were named around 1820 by Cauchy as "valeurs propres de matrices" in French. Later they were called Matrizeneigenwerte in Germany where this subject was then developed.

Now they go by the names of "eigenvalue" and "eigenvector" in English, having been anglicized from both French and German, using Greek and Latin letters, respectively - in quite an international endeavor.

Even real matrices can have complex eigenvalues and eigenvectors due to Gauss' Ph. D. thesis of 1799 that established the 'Fundamental Theorem of Algebra' for roots of polynomials.

The characteristic polynomial $f(x) = \det(A - xI_n)$ of matrices $A_{n,n}$ and polynomial root finding idea of Cauchy ($\sim$1820) has dominated Linear Algebra for 100 ++ years without ever finding ways or means to extract matrix eigenvalues reliably.

In the 1930s there were attempts to let matrices take care of matrix problems themselves through Krylov vector iterations.

But the full benefit of matrix eigenvalues came for us only with Francis' and Kublanowskaya's QR based eigenvalue algorithms of 1961.

Francis implicit QR method was not fully understood for 20 years until Watkins established its foundation in subspace iteration and his insight was not implemented in computable form until another 20 years later in 2002 in the multishift implicit QR version by Braman, Byers and Mathias.

This multishift matrix factorization based method reached Matlab only in the early 2000s and lets us now solve matrix eigen problems efficiently until dimensions around $n = 11,000$.

The short bits of math matrix history reveal the sad fact that 90+ % of our elementary Linear Algebra text books only contain, and that 95 % of our sophomore students are only taught the dead-end path of determinants, characteristic polynomials, and subsequent polynomial root finding methods for matrix eigenvalues.

This unprecedented and unjustifiable situation justifies us to use the term 'modern' in our efforts to bring Modern Matrix Theory into the college classroom now.

Future Lesson Plans will study the Krylov subspace method that computes matrix eigenvalues of square matrices iteratively through vector iteration.

Thereafter we study orthogonal matrix factorizations and special matrices that help with QR factorizations and let us find orthonormal basis for matrix subspaces accurately.

This will help us to comprehend the basics of modern eigenvalue evaluations via Francis' QR algorithm and Krylov vector iteration and lead to further applications.