CX1104: **Linear Algebra for Computing**

$$\underbrace{\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{bmatrix}_{m\times n}}_{A} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix}_{n\times 1}}_{x} = \underbrace{\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}_{m\times 1}}_{b}$$

Chap. No : **8.1.5B**

Lecture : **Linear Transformation and EigenVectors**

Concept : **Linear Transformation and Change of Basis using eigen Basis**

Note: Revise 8.1.5A first if unfamiliar with coordinate system and change of basis.

Instructor: **A/P Chng Eng Siong**

TAs: **Zhang Su, Vishal Choudhari**

Revised: 24 Oct 2021

Rev: 3rd July 2020

1) Overview: $Ax = y$, Least Squares vs Linear Transformation

1) **In Gaussian Elimination for system of equations and Least Squares**, the equation $Ax = y$ is presented.
   In this case, A and y represent the given equation values and target respectively
   - The problem there is to find x that minimize the error to approximate y.

2) **In Linear Transformation**, the problem also has the same equation $Ax = y$,
   in this case, the interpretation is different to (1). Here,
   - x is the input vector to be transformed to desired vector y
     through the linear transformation $\mathrm{y} = \mathrm{T}(x)$, evaluated as $\mathrm{y} = \mathrm{A}x$

   - In some machine learning problems, x represents the input, y the target
     and A the model. The machine learning algorithm strive to find the best A.
   - Here, the ML community do not use linear models A, but nonlinear models.
   - and the number of examples (x_i, y_i) typically number into the millions.

1.1) Linear Transformation $y = T(x) = Ax$

Given input vector $x \in R^n$, with coordinates in the standard basis and linear transformation $T: R^n \rightarrow R^m$ represented by matrix A, then $y = T(x) = Ax \in R^m$,
The columns of A are:

$$A = \left[T(\mathbf{e}_1) \quad T(\mathbf{e}_2) \quad \cdots \quad T(\mathbf{e}_n) \right]$$

and $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ *denote the standard basis vectors for* $\mathbb{R}^n$.
This A *is called the* **matrix of** T.

$$\begin{array}{lccc} \text{abstract} & x & \xrightarrow{\quad T \quad} & y = T(x) \\ & \downarrow & & \downarrow \\ \text{Std basis} & x \in R^n & \xrightarrow{\quad A \quad} & y = Ax \in R^m \end{array}$$

$$A \in R^{mxn}$$

Fig: representation of a linear transformation

Ref:
1) https://math.libretexts.org/Bookshelves/Linear_Algebra/Book%3A_A_First_Course_in_Linear_Algebra_(Kuttler)/05%3A_Linear_Transformations (Ch5.1, 5.2)

1.2) Proof that $T(x)$ can be represented as a matrix product Ax

Theorem. Consider a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$. The transformation T can be represented as a matrix product $\mathbf{x} \mapsto A\mathbf{x}$, for some matrix $A \in \mathbb{R}^{n\times n}$.

Proof. Consider a matrix $\mathbf{x} \in \mathbb{R}^n$ given by

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

We will construct a matrix $A \in \mathbb{R}^{n\times n}$ such that $T(\mathbf{x}) = A\mathbf{x}$.

The vector $\mathbf{x}$ can also be written as

$$\begin{aligned}\mathbf{x} &= x_1 \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} + \cdots + x_n \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \\ &= x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \cdots + x_n\mathbf{e}_n \\ &= \sum_{i=1}^n x_i\mathbf{e}_i,\end{aligned}$$

where $\mathbf{e}_i$ are the standard basis vectors in $\mathbb{R}^n$.

Consider the transformation $T(\mathbf{x})$. Rewriting $\mathbf{x}$ as above, we have

$$\begin{aligned}T(\mathbf{x}) &= T\left(\sum_{i=1}^n x_i\mathbf{e}_i\right) \\ &= \sum_{i=1}^n T(x_i\mathbf{e}_i) \\ T(\mathbf{x}) &= \sum_{i=1}^n x_i T(\mathbf{e}_i). \qquad (1)\end{aligned}$$

Let the matrix $A \in \mathbb{R}^{n\times n}$ be defined by

$$\begin{aligned}A &= [\, T(\mathbf{e}_1) \quad T(\mathbf{e}_2) \quad \cdots \quad T(\mathbf{e}_n) \quad] \\ &= \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix},\end{aligned}$$

where each $T(\mathbf{e}_i)$ is a column of A, and each $a_{ij} = T(\mathbf{e}_i) \cdot \mathbf{e}_j$ is the jth component of $T(\mathbf{e}_i)$. Then, by the definition of matrix-vector multiplication, we have

$$\begin{aligned}A\mathbf{x} &= \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \\ &= \begin{bmatrix} x_1 a_{11} + \cdots + x_n a_{1n} \\ \vdots \\ x_1 a_{n1} + \cdots + x_n a_{nn} \end{bmatrix} \\ &= x_1 \begin{bmatrix} a_{11} \\ \vdots \\ a_{n1} \end{bmatrix} + \cdots + x_n \begin{bmatrix} a_{n1} \\ \vdots \\ a_{nn} \end{bmatrix} \\ &= x_1 T(\mathbf{e}_1) + \cdots + x_n T(\mathbf{e}_n) \\ A\mathbf{x} &= \sum_{i=1}^n x_i T(\mathbf{e}_i). \qquad (2)\end{aligned}$$

Therefore, by eqs. (1) and (2), we have that

$$T(\mathbf{x}) = \sum_{i=1}^n x_i T(\mathbf{e}_i) \qquad A\mathbf{x} = \sum_{i=1}^n x_i T(\mathbf{e}_i),$$

and we reach $T(\mathbf{x}) = A\mathbf{x}$, as was to be shown.

The crux of the proof is that using the standard basis and by linearity, $\mathbf{x} = \sum \mathbf{e}_i x_i \implies T(\mathbf{x}) = \sum T(\mathbf{e}_i) x_i = \sum \mathbf{a}_i x_i$ where the $\mathbf{a}_i$ can be arranged as the columns of the matrix. – user65203

Ref:

1) https://math.stackexchange.com/questions/916192/proving-any-linear-transformation-can-be-represented-as-a-matrix

2.1) If A is a dense matrix, computing Ax can sometimes be simplified by converting $\mathrm{A}\ and\ x$ into another basis.

$$
\begin{tikzcd}
\text{abstract} & x \arrow[r, "T(x)"] \arrow[d] & y = T(x) \arrow[d] \\
\text{Std basis} & x \in R^n \arrow[r, "A"] \arrow[d, "P^{-1}"] & y = Ax = P[y]_B = PDP^{-1}x \\
\text{B-basis} & [x]_B = P^{-1}x \arrow[r, "D"] & {[y]_B = D[x]_B} \arrow[u, "P"]
\end{tikzcd}
$$

Motivation: why change basis?
We wish to perform change of basis on x to another basis $[x]_B$.
In basis B, the transformation A will be a very sparse diagonal matrix D, hence calculating $\mathrm{A}x$ is cheap.

P is a matrix with columns containing a basis,

$$P = \{b_1, b_2, \dots, b_n\}$$

How is standard basis related to B-basis:

$$x = P[x]_B$$
$$P^{-1}x = [x]_B$$

$$y = \mathrm{T(x)}$$
$$y = Ax$$
$$= P(D\ \underbrace{(P^{-1}x)}_{\text{Changing from Std to B-basis, this is } [x]_B})$$

Changing from B-basis to Std basis

Transformation T in B-basis

2.2) Linear Transformation from V into V

THEOREM 8

Diagonal Matrix Representation

Suppose $A = PDP^{-1}$, where D is a diagonal $n \times n$ matrix. If $\mathcal{B}$ is the basis for $\mathbb{R}^n$ formed from the columns of P, then D is the $\mathcal{B}$-matrix for the transformation $\mathbf{x} \mapsto A\mathbf{x}$.

Lay,4th,pg291,ch4.4

Note: P contains the set of eigenvectors of A,
and for P^{-1} to exist, it means that P contains n number of independent eigenvectors, and
D is a diagonal matrix containing the eigenvalues.
Not all matrix have this characteristics. When matrix have this characteristic, we say these matrixes are diagonalizable.

See Tut8, Q7A youTube: https://youtu.be/zaSqGGmNokw

2.3) D is the transformation matrix in B basis.

See pg 7 of slide 8.1.5A

$$\boxed{\mathbf{x} = P_{\mathcal{B}}[\,\mathbf{x}\,]_{\mathcal{B}}} \qquad \boxed{P_{\mathcal{B}}^{-1}\mathbf{x} = [\,\mathbf{x}\,]_{\mathcal{B}}} \qquad (4)$$

We call $P_{\mathcal{B}}$ the **change-of-coordinates matrix** from $\mathcal{B}$ to the standard basis in $\mathbb{R}^n$.

$$y = Ax$$

if A can be decomposed into $A = PDP^{-1}$, therefore

$$y = PDP^{-1}x,$$

We can view $P^{-1}x$ as changing coordinate to B, i.e $[x]_B$

$$\begin{aligned} y &= P\,D(P^{-1}x) \\ &= P\,D[x]_B \\ &= P\,[y]_B \end{aligned}$$

We can view $D[x]_B$ as transformation matrix *D* applied on $[x]_B$, $(Theorem\ 8)$
i.e D is the transformation matrix in B basis,
and $[y]_B = D[x]_B$ (y in the B-basis). Lastly

$$y = P\,[y]_B$$

can be viewed as converting y in B-basis back to standard basis.

2.4) Linear Transformation wrt B-basis pictorially

$$
\begin{tikzcd}
x \arrow[r, "A = PDP^{-1}"] \arrow[d, shift right, "P^{-1}"'] & y = Ax = PDP^{-1}x \arrow[d, shift right, "P^{-1}"'] \\
{[x]_B} \arrow[u, shift right, "P"'] \arrow[r, "{[T]_B = D}"] & {[y]_B = [Ax]_B = [T]_B\,[x]_B} \arrow[u, shift right, "P"']
\end{tikzcd}
$$

$$
\begin{aligned}
x &= P[x]_B \\
P^{-1}x &= [x]_B
\end{aligned}
\qquad
\begin{aligned}
y &= P[y]_B \\
P^{-1}y &= [y]_B
\end{aligned}
$$

Lay,4th,pg291,ch4.4

THEOREM 8

Diagonal Matrix Representation

Suppose $A = PDP^{-1}$, where D is a diagonal $n \times n$ matrix. If $\mathcal{B}$ is the basis for $\mathbb{R}^n$ formed from the columns of P, then D is the $\mathcal{B}$-matrix for the transformation $\mathbf{x} \mapsto A\mathbf{x}$.

Lets start from y:

$$y = P\ \ [y]_B$$
$$y = P\ \ D\ \ [x]_B$$
$$y = P\ \ D\ \ P^{-1}x$$

Therefore

$$A = PDP^{-1}$$
$$P^{-1}AP = D$$

PROOF Denote the columns of P by $\mathbf{b}_1, \dots, \mathbf{b}_n$, so that $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ and $P = [\,\mathbf{b}_1 \ \cdots \ \mathbf{b}_n\,]$. In this case, P is the change-of-coordinates matrix $P_{\mathcal{B}}$ discussed in Section 4.4, where

$$P[\,\mathbf{x}\,]_{\mathcal{B}} = \mathbf{x} \quad \text{and} \quad [\,\mathbf{x}\,]_{\mathcal{B}} = P^{-1}\mathbf{x}$$

If $T(\mathbf{x}) = A\mathbf{x}$ for $\mathbf{x}$ in $\mathbb{R}^n$, then

$$
\begin{aligned}
[\,T\,]_{\mathcal{B}} &= [\,[\,T(\mathbf{b}_1)\,]_{\mathcal{B}} \ \cdots \ [\,T(\mathbf{b}_n)\,]_{\mathcal{B}}\,] && \text{Definition of } [\,T\,]_{\mathcal{B}} \\
&= [\,[\,A\mathbf{b}_1\,]_{\mathcal{B}} \ \cdots \ [\,A\mathbf{b}_n\,]_{\mathcal{B}}\,] && \text{Since } T(\mathbf{x}) = A\mathbf{x} \\
&= [\,P^{-1}A\mathbf{b}_1 \ \cdots \ P^{-1}A\mathbf{b}_n\,] && \text{Change of coordinates} \\
&= P^{-1}A[\,\mathbf{b}_1 \ \cdots \ \mathbf{b}_n\,] && \text{Matrix multiplication} \\
&= P^{-1}AP && (6)
\end{aligned}
$$

Since $A = PDP^{-1}$, we have $[\,T\,]_{\mathcal{B}} = P^{-1}AP = D$. ■

2.5) Important point: if we are asked to find x[k] , for k very large $x[n+1] = Ax[n]$ and A is diagonalizable, it is computationally efficient if we convert $x \to [x]_B$ to work on the problem!

Given A and $x[0]$, and $x[n+1] = Ax[n]$,
we convert problem to B-basis (eigenvectors of A if exists) since its computationally cheaper.

Proof:

$x[1] = Ax[0]$

$x[2] = Ax[1] = A(Ax[0]) = A^2x[0]$

Therefore, $x[k] = A^kx[0]$

if $A = PDP^{-1}$,

then $A^2 = (PDP^{-1})(PDP^{-1}) \quad = PD^2P^{-1}$ and

$\quad A^3 \;= (PDP^{-1})(PD^2P^{-1}) = PD^3P^{-1}$

Therefore $A^k = PD^kP^{-1}$

$\Rightarrow x[k] = A^kx[0] = PD^kP^{-1}x[0]$

$\Rightarrow \; x[k] = P\,D^k\;\; x[0]_B$

Note:
Computing A^k is expensive if A is a dense matrix.
Computing D^k is cheap since it is a diagonal matrix.

$$\begin{tikzcd} x \arrow[d, shift right, "P^{-1}"'] \\ {[x]_B} \arrow[u, shift right, "P"'] \end{tikzcd}$$

Change of Basis

$$\begin{aligned} x \quad &= P[x]_B \\ P^{-1}x &= \quad [x]_B \end{aligned}$$

3) Lay's notes: General Equation of Matrix of Linear Transformation T: $R^n \rightarrow R^m$

The Matrix of a Linear Transformation

Let V be an n-dimensional vector space, let W be an m-dimensional vector space, and let T be any linear transformation from V to W. To associate a matrix with T, choose (ordered) bases $\mathcal{B}$ and $\mathcal{C}$ for V and W, respectively.

Given any $\mathbf{x}$ in V, the coordinate vector $[\,\mathbf{x}\,]_{\mathcal{B}}$ is in $\mathbb{R}^n$ and the coordinate vector of its image, $[\,T(\mathbf{x})\,]_{\mathcal{C}}$, is in $\mathbb{R}^m$, as shown in Figure 1.

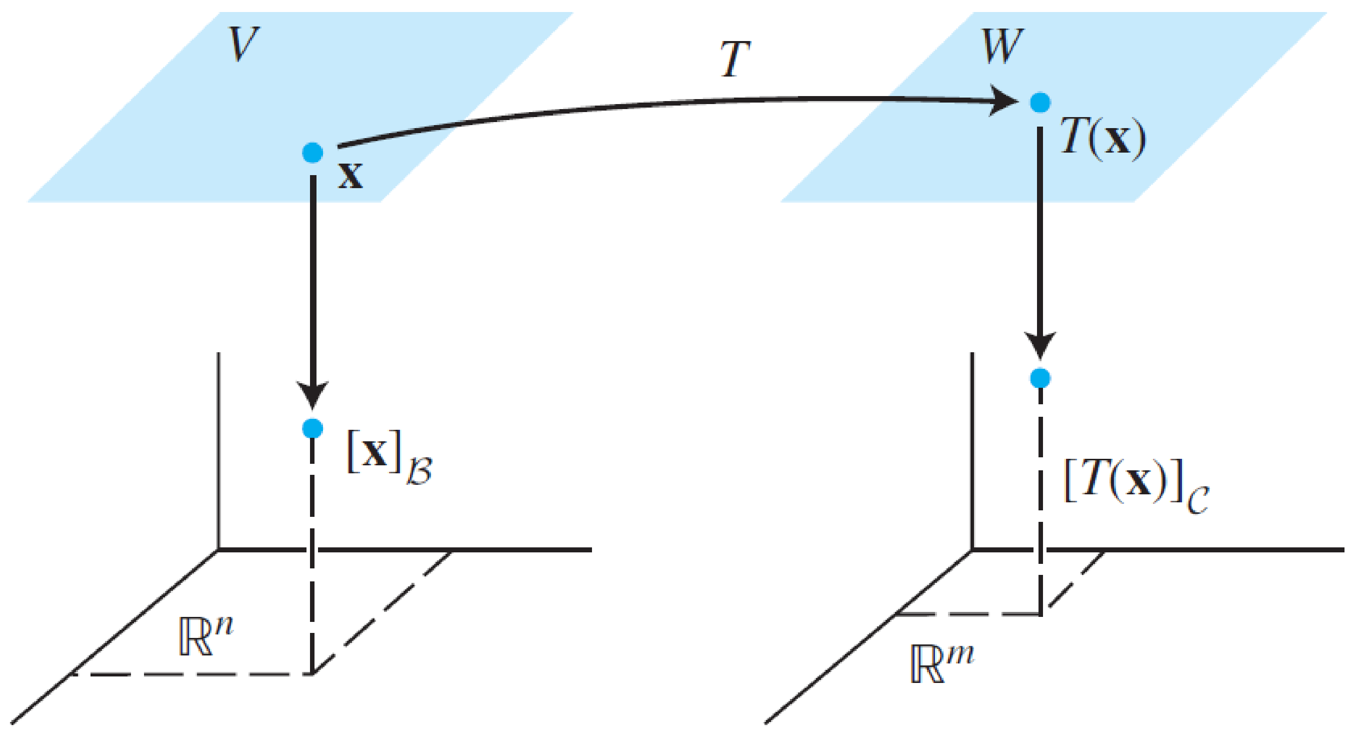


FIGURE 1 A linear transformation from V to W.

$$\begin{tikzcd}
\mathbf{x} \arrow[r, "T"] \arrow[d] & T(\mathbf{x}) \arrow[d] \\
{[\mathbf{x}]_{\mathcal{B}}} \arrow[r, "\text{Multiplication}", "\text{by } M"'] & {[T(\mathbf{x})]_{\mathcal{C}}}
\end{tikzcd}$$

FIGURE 2

The connection between $[\,\mathbf{x}\,]_{\mathcal{B}}$ and $[\,T(\mathbf{x})\,]_{\mathcal{C}}$ is easy to find. Let $\{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ be the basis $\mathcal{B}$ for V. If $\mathbf{x} = r_1\mathbf{b}_1 + \cdots + r_n\mathbf{b}_n$, then

$$[\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} r_1 \\ \vdots \\ r_n \end{bmatrix}$$

and

$$T(\mathbf{x}) = T(r_1\mathbf{b}_1 + \cdots + r_n\mathbf{b}_n) = r_1T(\mathbf{b}_1) + \cdots + r_nT(\mathbf{b}_n) \qquad (1)$$

because T is linear. Now, since the coordinate mapping from W to $\mathbb{R}^m$ is linear (Theorem 8 in Section 4.4), equation (1) leads to

$$[\,T(\mathbf{x})\,]_{\mathcal{C}} = r_1[\,T(\mathbf{b}_1)\,]_{\mathcal{C}} + \cdots + r_n[\,T(\mathbf{b}_n)\,]_{\mathcal{C}} \qquad (2)$$

Since $\mathcal{C}$-coordinate vectors are in $\mathbb{R}^m$, the vector equation (2) can be written as a matrix equation, namely,

$$[\,T(\mathbf{x})\,]_{\mathcal{C}} = M[\,\mathbf{x}\,]_{\mathcal{B}} \qquad (3)$$

where

$$M = \big[\,[\,T(\mathbf{b}_1)\,]_{\mathcal{C}} \quad [\,T(\mathbf{b}_2)\,]_{\mathcal{C}} \quad \cdots \quad [\,T(\mathbf{b}_n)\,]_{\mathcal{C}}\,\big] \qquad (4)$$

The matrix M is a matrix representation of T, called the **matrix for T relative to the bases $\mathcal{B}$ and $\mathcal{C}$**. See Figure 2.

Equation (3) says that, so far as coordinate vectors are concerned, the action of T on $\mathbf{x}$ may be viewed as left-multiplication by M.

Lay 5e, pg 290, Ch5.4

3.1) Example: Linear Transformation

EXAMPLE 1 Suppose $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2\}$ is a basis for V and $\mathcal{C} = \{\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3\}$ is a basis for W. Let $T : V \to W$ be a linear transformation with the property that

$$T(\mathbf{b}_1) = 3\mathbf{c}_1 - 2\mathbf{c}_2 + 5\mathbf{c}_3 \quad \text{and} \quad T(\mathbf{b}_2) = 4\mathbf{c}_1 + 7\mathbf{c}_2 - \mathbf{c}_3$$

Find the matrix M for T relative to $\mathcal{B}$ and $\mathcal{C}$.

SOLUTION The $\mathcal{C}$-coordinate vectors of the *images* of $\mathbf{b}_1$ and $\mathbf{b}_2$ are

$$[T(\mathbf{b}_1)]_{\mathcal{C}} = \begin{bmatrix} 3 \\ -2 \\ 5 \end{bmatrix} \quad \text{and} \quad [T(\mathbf{b}_2)]_{\mathcal{C}} = \begin{bmatrix} 4 \\ 7 \\ -1 \end{bmatrix}$$

Hence

$$M = \begin{bmatrix} 3 & 4 \\ -2 & 7 \\ 5 & -1 \end{bmatrix}$$

If $\mathcal{B}$ and $\mathcal{C}$ are bases for the same space V and if T is the identity transformation $T(\mathbf{x}) = \mathbf{x}$ for $\mathbf{x}$ in V, then matrix M in (4) is just a change-of-coordinates matrix (see Section 4.7).

Lay 5e, pg 291, Ch5.4

Revision:

Let $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ and $\mathcal{C} = \{\mathbf{c}_1, \dots, \mathbf{c}_n\}$ be bases of a vector space V. Then there is a unique $n \times n$ matrix $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$ such that

$$[\mathbf{x}]_{\mathcal{C}} = \underset{\mathcal{C}\leftarrow\mathcal{B}}{P}[\mathbf{x}]_{\mathcal{B}} \tag{4}$$

The columns of $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$ are the $\mathcal{C}$-coordinate vectors of the vectors in the basis $\mathcal{B}$. That is,

$$\underset{\mathcal{C}\leftarrow\mathcal{B}}{P} = \begin{bmatrix} [\mathbf{b}_1]_{\mathcal{C}} & [\mathbf{b}_2]_{\mathcal{C}} & \cdots & [\mathbf{b}_n]_{\mathcal{C}} \end{bmatrix} \tag{5}$$

The matrix $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$ in Theorem 15 is called the **change-of-coordinates matrix from $\mathcal{B}$ to $\mathcal{C}$**. Multiplication by $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$ converts $\mathcal{B}$-coordinates into $\mathcal{C}$-coordinates.[2] Figure 2 illustrates the change-of-coordinates equation (4).

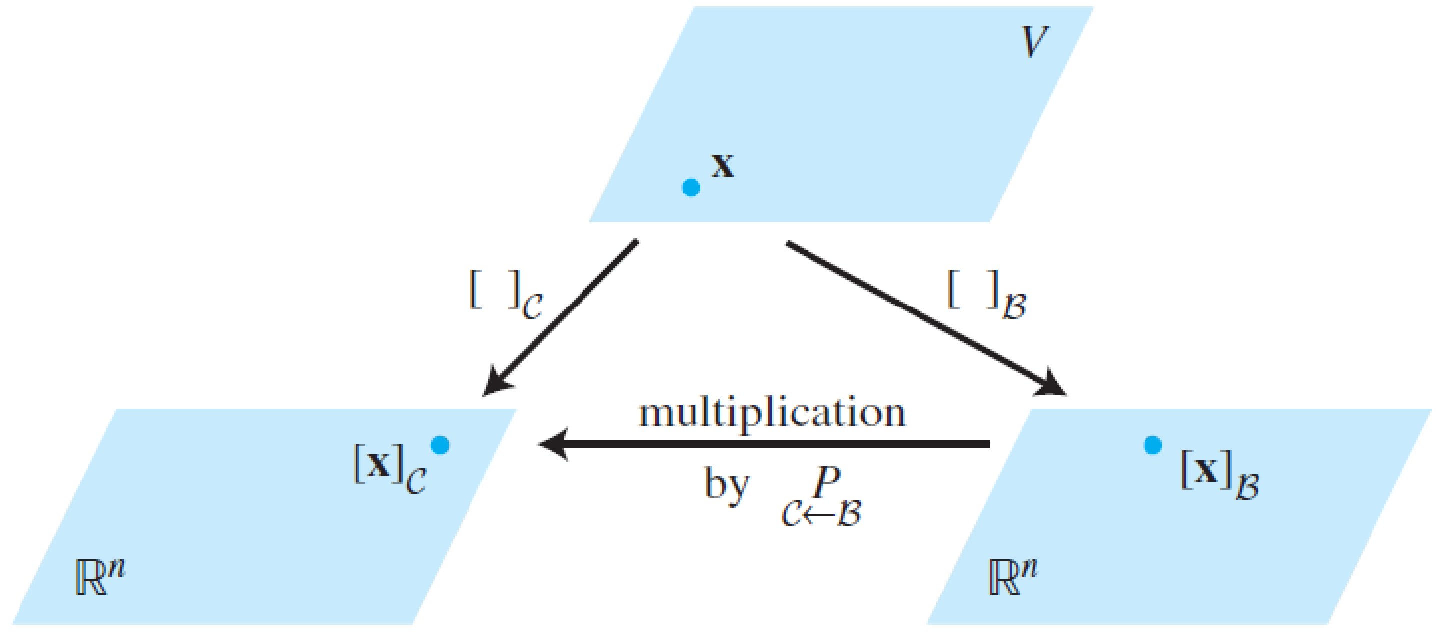


FIGURE 2 Two coordinate systems for V.

Lay 5e, pg 242, Ch 4.7

Appendix: Some useful information

References:

1)Youtube:

a)3Blue1Brown: Linear Transformation (Ch3) https://www.youtube.com/watch?v=kYB8IZa5AuE

b) Strang Lect 30, 18.06 (2005) https://www.youtube.com/watch?v=Ts3o2I8_Mxc

c) Adams Panagos: "Finding A" https://www.youtube.com/watch?v=61knWwBM3gQ

d) Technion: L54 Matrix Representation of Linear Map : https://www.youtube.com/watch?v=tRbXrnoVJI8

2) Problems in Yutsumura.com:

a) https://yutsumura.com/find-matrix-representation-of-linear-transformation-from-r2-to-r2/

b) https://yutsumura.com/linear-transformation-tr2-to-r2-given-in-figure/

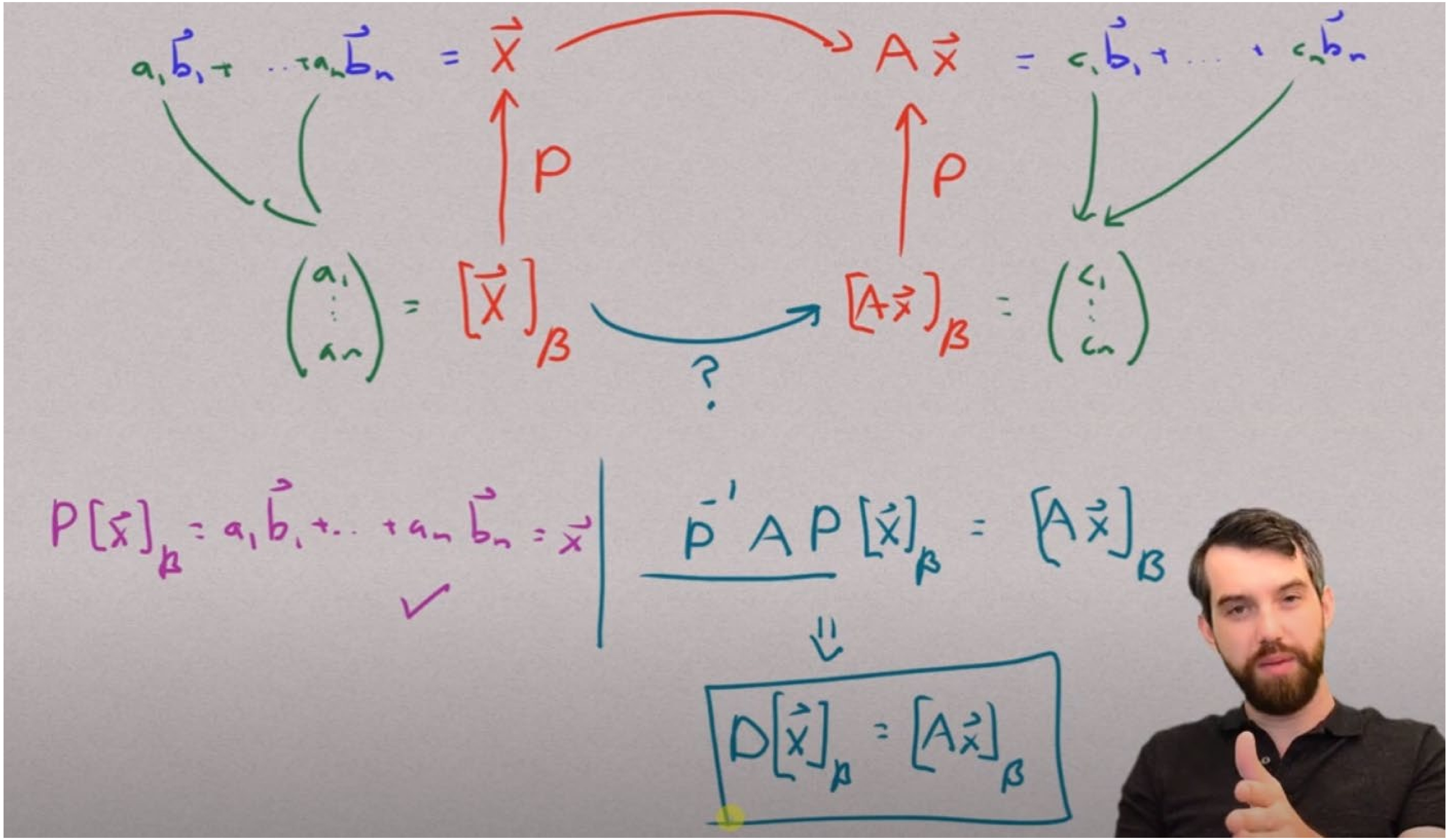
b) https://yutsumura.com/find-a-general-formula-of-a-linear-transformation-from-r2-to-r3/#more-2526

3) Reference from others:

a) Upenn: https://www2.math.upenn.edu/~moose/240S2013/slides7-23.pdf

b) Abbasi notes on deriving the transformation matrix: https://www.12000.org/my_notes/similarity_transformation_and_SVD/ind

Trefor Bazett: The Similarity Relationship Represents a Change of Basis
Youtube: https://www.youtube.com/watch?v=s4c5LQ5a4ek



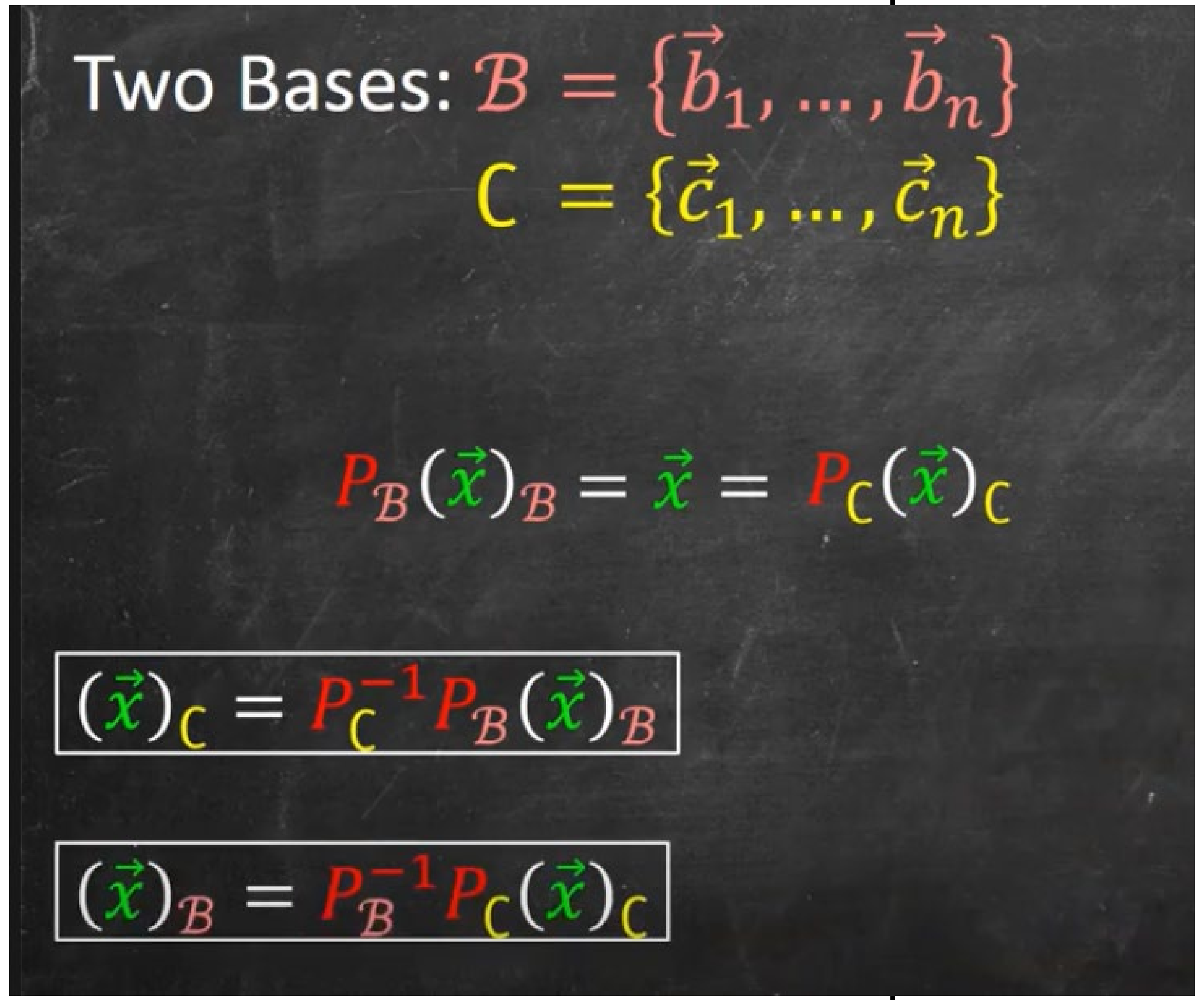
Change of Basis from Trefor Bazett

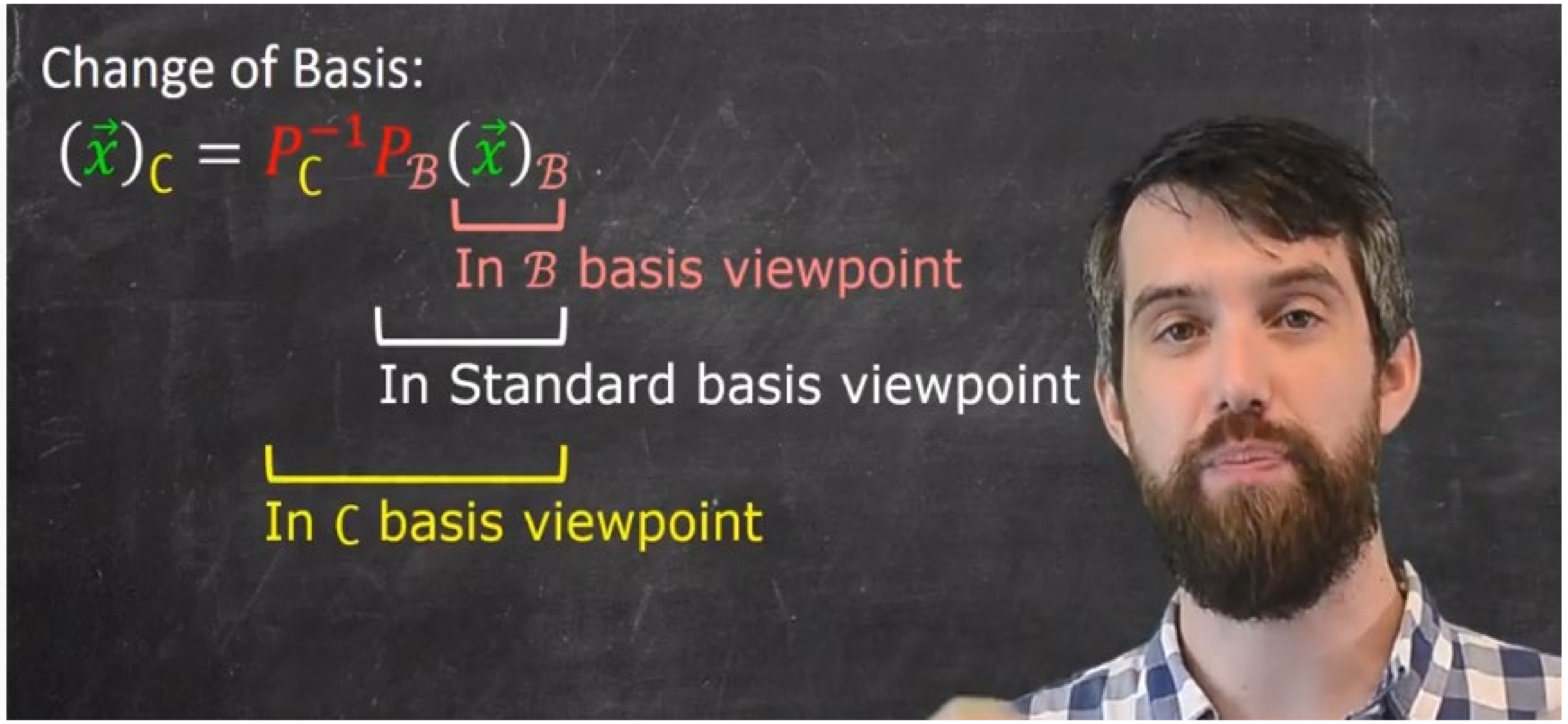
$$\begin{tikzcd}
\mathbf{x} \arrow[r, "T"] \arrow[d] & T(\mathbf{x}) \arrow[d] \\
{[\mathbf{x}]_{\mathcal{B}}} \arrow[r, "\text{Multiplication}", "\text{by } M"'] & {[T(\mathbf{x})]_{\mathcal{C}}}
\end{tikzcd}$$

FIGURE 2

Trefor Bazett: Changing between 2 bases
https://www.youtube.com/watch?v=KjlTOLhaI9s

Remember: $x = P_B[x]_B = P_C\ [x]_C$
The vector x represented in basis B and C





Lay4th, sec5.4,pg289