Review of similarity transformation and Singular Value Decomposition

Nasser M. Abbasi

Applied Mathematics Department, California State University, Fullerton. July 8 2007

Compiled on January 31, 2024 at 3:17am

Contents

1

1 Similarity transformation

A similarity transformation is

$$B = M^{-1}AM$$

Where B, A, M are square matrices. The goal of similarity transformation is to find a B matrix which has a simpler form than A so that we can use B in place of A to ease some computational work. Lets set our goal in having B be a diagonal matrix (a general diagonal form is called block diagonal or Jordan form, but here we are just looking at the case of B being a diagonal matrix).

The question becomes: Given A, can we find M such that $M^{-1}AM$ is diagonal?

The standard method to show the above is via an algebraic method, which we show first. Next we show a geometric method that explains similarity transformation geometrically.

1.1 Derivation of similarity transformation based on algebraic method

Starting with

$$B = M^{-1}AM$$

Our goal is to find a real matrix B such that it is diagonal. From the above, by pre multiplying each side by M we obtain

$$AM = MB$$

Now, since our goal is to make B diagonal, let us select the eigenvalues of A to be the diagonal of B. Now we write the above in expanded form as follows

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} m_{11} & \cdots & m_{1n} \\ \vdots & \ddots & \vdots \\ m_{n1} & \cdots & m_{nn} \end{bmatrix} = \begin{bmatrix} m_{11} & \cdots & m_{1n} \\ \vdots & \ddots & \vdots \\ m_{n1} & \cdots & m_{nn} \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_n \end{bmatrix}$$

The above can be written as n separate equations by looking at the column view of matrix multiplication

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} m_{11} \\ \vdots \\ m_{n1} \end{bmatrix} = \lambda_1 \begin{bmatrix} m_{11} \\ \vdots \\ m_{n1} \end{bmatrix}$$

And

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} m_{12} \\ \vdots \\ m_{n2} \end{bmatrix} = \lambda_2 \begin{bmatrix} m_{12} \\ \vdots \\ m_{n2} \end{bmatrix}$$

All the way to the nth column of M

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} m_{1n} \\ \vdots \\ m_{nn} \end{bmatrix} = \lambda_n \begin{bmatrix} m_{1n} \\ \vdots \\ m_{nn} \end{bmatrix}$$

Each of the above n equations is in the form

$$A\mathbf{x} = \lambda\mathbf{x}$$

and this is the key idea.

The above shows that if we take the diagonal elements of B to be the eigenvalues of A then the M matrix is the (right) eigenvectors of A.

Hence, if we want the B matrix to be diagonal and real, this means the A matrix itself must have some conditions put on it. Specifically, A eigenvalues must be all distinct and real. This happens when A is real and symmetric. Therefore, we have the general result that given a matrix A which has distinct and real eigenvalues, only then can we find a similar matrix to it which is real and diagonal, and these diagonal values are the eigenvalues of A.

1.2 Derivation of similarity transformation based on geometric method

It turns out that this derivation is more complicated than the algebraic approach. It involves a change of basis matrix (which will be our M matrix) and the representation of linear transformation T as a matrix under some basis (which will be our A matrix). Let us start from the beginning.

Given the vector $\mathbf{x}$ in $\mathbb{R}^n$, it can have many different representations (or coordinates) depending on which basis are used. There are an infinite number of basis that span $\mathbb{R}^n$, hence there are infinite representations for the same vector. Consider for example $\mathbb{R}^2$. The standard basis vectors are $\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$, but another basis are $\left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$, and yet another are $\left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \end{bmatrix} \right\}$, and so on.
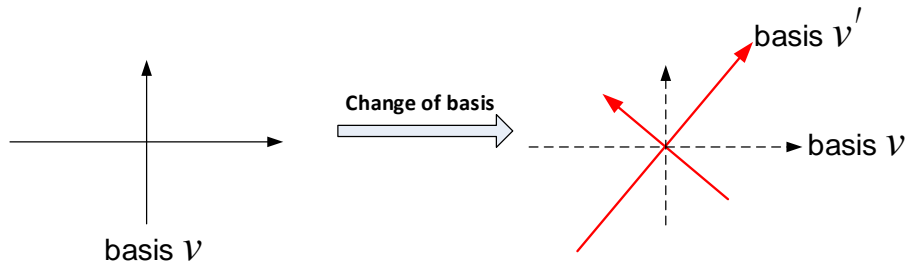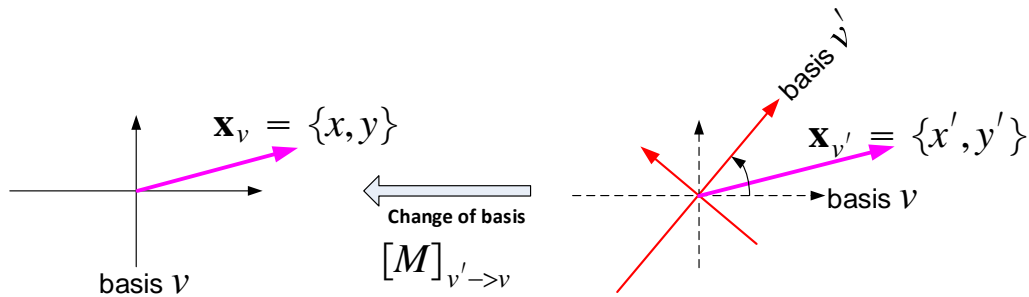
Figure 1: Change of basis

In $\mathbb{R}^n$ any n linearly independent vectors can be used as basis. Let $\mathbf{x}_v$ be the vector representation in w.r.t. basis v, and let $\mathbf{x}_{v'}$ be the vector representation w.r.t. basis v'.

Same vector X has different representation depending on which basis are used

Figure 2: Same vector under different basis

An important problem in linear algebra is to obtain $\mathbf{x}_v$ given $\mathbf{x}_{v'}$ and the change of basis matrix $[M]_{v' \to v}$.

This requires finding a matrix representation for the change of basis. Once the $[M]_{v' \to v}$ matrix is found, we can write

$$\mathbf{x}_v = [M]_{v' \to v} \, \mathbf{x}_{v'} \tag{1}$$

Where $[M]_{v' \to v}$ is the change of basis matrix which when applied to $\mathbf{x}_{v'}$ returns the coordinates of the vector w.r.t. basis v.

From (1) we see that given $\mathbf{x}_v$, then to obtain $\mathbf{x}_{v'}$ we write

$$\mathbf{x}_{v'} = [M]_{v' \to v}^{-1} \mathbf{x}_v \tag{1A}$$

Another important problem is that of linear transformation T, where now we apply some transformation on the whole space and we want to find what happens to the space coordinates (all the position vectors) due to this transformation.

Consider for example T which is a rotation in $\mathbb{R}^2$ by some angle θ. Here, every position vector in the space is affected by this rotation *but the basis remain fixed*, and we want to find the new coordinates of each point in space.

Let $\mathbf{x}_v$ be the position vector before applying the linear transformation, and let $\mathbf{y}_v$ be the new position vector. Notice that both vectors are written with respect to the same basis v. Similar to change basis, we want a way to find $\mathbf{y}_v$ from $\mathbf{x}_v$, hence we need a way to represent this linear transformation by a matrix, say $[T]_v$ with respect to the basis v and so we could write the following

$$\mathbf{y}_v = [T]_v \, \mathbf{x}_v \tag{2}$$

Same basis v but T causes change in vector representation
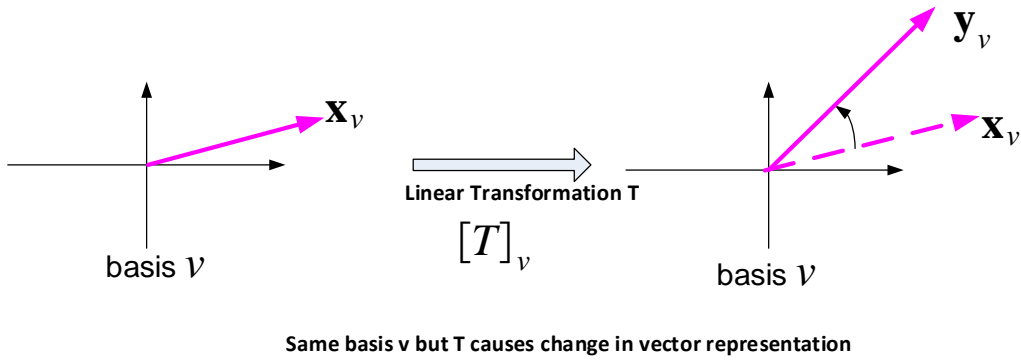
Figure 3: Linear transformation effect

Assume the basis is changed to v'. Let the representation of the same linear transformation T w.r.t. basis v' be called $[T]_{v'}$, so we write

$$\mathbf{y}_{v'} = [T]_{v'} \, \mathbf{x}_{v'} \tag{2A}$$

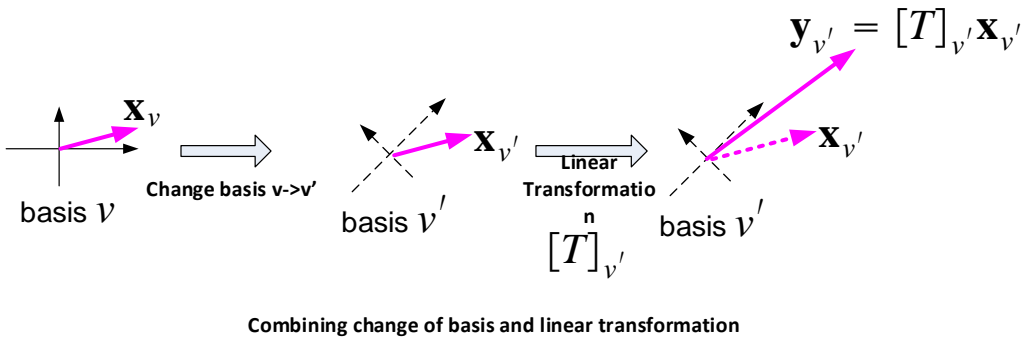**Combining change of basis and linear transformation**

Figure 4: Combining change of basis and linear transformation

Hence when the basis changes from v to v', T representation changes from $[T]_v$ to $[T]_{v'}$.

Now assume we are given some linear transformation T and its representation $[T]_v$ w.r.t. basis v, how could we find T representation $[T]_{v'}$ w.r.t. to new basis v'?

From (2) we have

$$\mathbf{y}_v = [T]_v \, \mathbf{x}_v$$

But from (1) $\mathbf{x}_v = [M]_{v' \to v} \, \mathbf{x}_{v'}$, hence the above becomes

$$\mathbf{y}_v = [T]_v \, \left([M]_{v' \to v} \, \mathbf{x}_{v'}\right)$$

Pre-multiplying from left the above equation by $[M]_{v' \to v}^{-1}$ we obtain

$$[M]_{v' \to v}^{-1} \mathbf{y}_v = [M]_{v' \to v}^{-1} [T]_v [M]_{v' \to v} \mathbf{x}_{v'}$$

But since $[M]_{v' \to v}^{-1} \mathbf{y}_v = \mathbf{y}_{v'}$, then above reduces to

$$\mathbf{y}_{v'} = [M]_{v' \to v}^{-1} [T]_v [M]_{v' \to v} \mathbf{x}_{v'}$$

But from (2A) $\mathbf{y}_{v'} = [T]_{v'} \, \mathbf{x}_{v'}$, hence the above becomes

$$[T]_{v'} \mathbf{x}_{v'} = [M]_{v' \to v}^{-1} [T]_v [M]_{v' \to v} \mathbf{x}_{v'}$$

Therefore

$$[T]_{v'} = [M]_{v' \to v}^{-1} [T]_v [M]_{v' \to v} \tag{3}$$

Notice the difference between change of basis and linear transformation. In change of basis, the vector $\mathbf{x}$ remained fixed in space, but the basis changed, and we want to find the coordinates of the vector w.r.t the new basis. With linear transformation, the vector itself is changed from $\mathbf{x}_v$ to $\mathbf{y}_v$, but both vectors are expressed w.r.t. the same basis.

Equation (3) allows us to obtain a new matrix representation of the linear transformation T by expressing the same T w.r.t. to different basis. Hence if we are given a representation of T which is not the most optimal representation, we can, by change of basis, obtain a different representation of the same T by using (3). The most optimal matrix representation for linear transformation is a diagonal matrix.

Equation (3) is called a similarity transformation. To make it easier to compare (3) above with what we wrote in the previous section when we derived the above using an algebraic approach, we let $[T]_{v'} = B, [T]_v = A$, hence (3) is

$$B = M^{-1} A M$$

The matrix $[T]_{v'}$ is similar to $[T]_v$. (i.e. B is similar to A). Both matrices represent the same linear transformation applied on the space. We will show below some examples of how to find $[T]_{v'}$ given $[T]_v$ and $[M]$. But first we show how to obtain matrix representation of T.

1.2.1 Finding matrix representation of linear transformation T

Given a vector $\mathbf{x} \in \mathbb{R}^n$ and some linear transformation (or a linear operator) T that acts on the vector $\mathbf{x}$ transforming this vector into another vector $\mathbf{y} \in \mathbb{R}^n$ according to some prescribed manner $T : \mathbf{x} \to \mathbf{y}$. Examples of such linear transformation are rotation, elongation, compression, and reflection, and any combination of these operations, but it can not include the translation of the vector, since translation is not linear.

The question to answer here is how to write down a representation of this linear transformation? It turns out that T can be represented by a matrix of size $n \times n$, and the actual numbers that go into the matrix, or the shape of the matrix, will depend on which basis in $\mathbb{R}^n$ we choose to represent $\mathbf{x}$.

We would like to pick some basis so that the final shape of the matrix is the most simple shape.

Let us pick a set of basis $v = \{\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_n\}$ that span $\mathbb{R}^n$ hence

$$\mathbf{x}_v = a_1 \mathbf{e}_1 + a_2 \mathbf{e}_2 + \cdots + a_n \mathbf{e}_n$$

Now

$$T\mathbf{x}_v = T(a_1 \mathbf{e}_1 + a_2 \mathbf{e}_2 + \cdots + a_n \mathbf{e}_n)$$

And since T is linear we can rewrite the above as

$$T\mathbf{x}_v = a_1 T\mathbf{e}_1 + a_2 T\mathbf{e}_2 + \cdots + a_n T\mathbf{e}_n \tag{1}$$

We see from above that the new transformed vector $T\mathbf{x}_v$ has the same coordinates as $\mathbf{x}_v$ if we view $T\mathbf{e}_i$ as the new basis.

Now $T\mathbf{e}_i$ itself is an application of the linear transformation but now it is being done on each basis vector $\mathbf{e}_i$. Hence it will cause each specific basis vector to transform in some manner to a new vector. Whatever this new vector is, we must be able to represent it in terms of the same basis vectors $\{\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_n\}$, therefore, we write

$$T\mathbf{e}_1 = \xi_{11}\mathbf{e}_1 + \xi_{21}\mathbf{e}_2 + \cdots + \xi_{n1}\mathbf{e}_n$$

Where ξ_{ij} is the ith coordinate of the vector $T\mathbf{e}_j$. And we do the same for $T\mathbf{e}_2$

$$T\mathbf{e}_2 = \xi_{12}\mathbf{e}_1 + \xi_{22}\mathbf{e}_2 + \cdots + \xi_{n2}\mathbf{e}_n$$

And so on until

$$T\mathbf{e}_n = \xi_{1n}\mathbf{e}_1 + \xi_{2n}\mathbf{e}_2 + \cdots + \xi_{nn}\mathbf{e}_n$$

Or

$$T\mathbf{e}_j = \xi_{ij}\mathbf{e}_i$$

Now plug the above into (1) and obtain

$$
\begin{aligned}
T\mathbf{x}_v = {} & a_1(\xi_{11}\mathbf{e}_1 + \xi_{21}\mathbf{e}_2 + \cdots + \xi_{n1}\mathbf{e}_n) \\
& + a_2(\xi_{12}\mathbf{e}_1 + \xi_{22}\mathbf{e}_2 + \cdots + \xi_{n2}\mathbf{e}_n) \\
& + \cdots \\
& + a_n(\xi_{1n}\mathbf{e}_1 + \xi_{2n}\mathbf{e}_2 + \cdots + \xi_{nn}\mathbf{e}_n)
\end{aligned}
$$

Hence, if we take the $\mathbf{e}_i$ as common factors, we have

$$
\begin{aligned}
T\mathbf{x} = {} & \mathbf{e}_1(a_1\xi_{11} + a_2\xi_{12} + \cdots + a_n\xi_{1n}) \\
& + \mathbf{e}_2(a_1\xi_{21} + a_2\xi_{22} + \cdots + a_n\xi_{2n}) \\
& + \cdots \\
& + \mathbf{e}_n(a_1\xi_{n1} + a_2\xi_{n2} + \cdots + a_n\xi_{nn})
\end{aligned}
$$

Hence, since $T\mathbf{x}_v = a_1 T\mathbf{e}_1 + a_2 T\mathbf{e}_2 + \cdots + a_n T\mathbf{e}_n$ from (1), then by comparing coefficients of each basis vector $\mathbf{e}_i$ we obtain

$$
\begin{aligned}
a_1 T &= a_1\xi_{11} + a_2\xi_{12} + \cdots + a_n\xi_{1n} \\
a_2 T &= a_1\xi_{21} + a_2\xi_{22} + \cdots + a_n\xi_{2n} \\
& \quad \cdots \\
a_n T &= a_1\xi_{n1} + a_2\xi_{n2} + \cdots + a_n\xi_{nn}
\end{aligned}
$$

Or in matrix form

$$
T\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} \xi_{11} & \xi_{12} & \cdots & \xi_{1n} \\ \xi_{21} & \xi_{22} & \cdots & \xi_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ \xi_{n1} & \xi_{n2} & \cdots & \xi_{nn} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}
$$

Hence we finally obtain that

$$
T \Rightarrow [T]_v = \begin{pmatrix} \xi_{11} & \xi_{12} & \cdots & \xi_{1n} \\ \xi_{21} & \xi_{22} & \cdots & \xi_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ \xi_{n1} & \xi_{n2} & \cdots & \xi_{nn} \end{pmatrix}
$$

Let us call the matrix that represents T under basis v as $[T]_v$.

We see from the above that the jth column of $[T]_v$ contain the coordinates of the vector $T\mathbf{e}_j$. This gives us a quick way to construct $[T]_v$: Apply T to each basis vector $\mathbf{e}_j$, and take the resulting vector and place it in the jth column of $[T]_v$.

We now see that $[T]_v$ will have a different numerical values if the basis used to span $\mathbb{R}^n$ are different from the ones used to obtain the above $[T]_v$.

Let use pick some new basis, say $v' = \{\mathbf{e}_1', \mathbf{e}_2', \cdots, \mathbf{e}_n'\}$. Let the new representation of T now be the matrix $[T]_{v'}$, then the jth column of $[T]_{v'}$ is found from applying T on the new basis $\mathbf{e}_j'$

$$T\mathbf{e}_j' = \zeta_{ij}\mathbf{e}_i'$$

Where now ζ_{ij} is the ith coordinates of the the vector $T\mathbf{e}_j'$ which will be different from ξ_{ij} since in one case we used the basis set $v' = \{\mathbf{e}_i'\}$ and in the other case we used the basis set $v = \{\mathbf{e}_i\}$. Hence we see that $[T]_{v'}$ will numerically be different depending on the basis used to span the space even though the linear transformation T itself did not change.

1.2.2 Finding matrix representation of change of basis

Now we show how to determine $[M]_{v \to v'}$, the matrix which when applied to a vector $\mathbf{x}_v$ will result in the vector $\mathbf{x}_{v'}$.

Given a vector $\mathbf{x}$, it is represented w.r.t. basis $v = \{\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_n\}$ as

$$\mathbf{x}_v = a_1\mathbf{e}_1 + a_2\mathbf{e}_2 + \cdots + a_n\mathbf{e}_n$$

and w.r.t. basis $v' = \{\mathbf{e}_1', \mathbf{e}_2', \cdots, \mathbf{e}_n'\}$ as

$$\mathbf{x}_{v'} = b_1\mathbf{e}_1' + b_2\mathbf{e}_2' + \cdots + b_n\mathbf{e}_n'$$
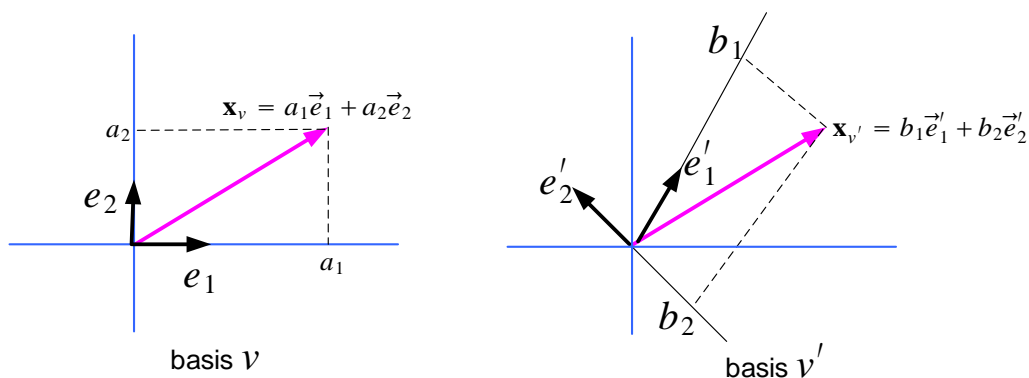
The same vector having different representation depending on basis used

Figure 5: Vector representation and basis

But the vector itself is invariant under any change of basis, hence $\mathbf{x}_v = \mathbf{x}_{v'}$

$$a_1\mathbf{e}_1 + a_2\mathbf{e}_2 + \cdots + a_n\mathbf{e}_n = b_1\mathbf{e}_1' + b_2\mathbf{e}_2' + \cdots + b_n\mathbf{e}_n' \tag{1}$$

Now write each basis $\mathbf{e}'_i$ in terms of the basis $\mathbf{e}_j$, hence we have

$$\mathbf{e}'_1 = c_{11}\mathbf{e}_1 + c_{12}\mathbf{e}_2 + \cdots + c_{1n}\mathbf{e}_n$$
$$\mathbf{e}'_2 = c_{21}\mathbf{e}_1 + c_{22}\mathbf{e}_2 + \cdots + c_{2n}\mathbf{e}_n$$
$$\vdots$$
$$\mathbf{e}'_n = c_{n1}\mathbf{e}_1 + c_{n2}\mathbf{e}_2 + \cdots + c_{nn}\mathbf{e}_n \qquad (2)$$

Substitute (2) into RHS of (1) we obtain

$$a_1\mathbf{e}_1 + a_2\mathbf{e}_2 + \cdots + a_n\mathbf{e}_n = b_1(c_{11}\mathbf{e}_1 + c_{12}\mathbf{e}_2 + \cdots + c_{1n}\mathbf{e}_n) +$$
$$b_2(c_{21}\mathbf{e}_1 + c_{22}\mathbf{e}_2 + \cdots + c_{2n}\mathbf{e}_n) +$$
$$\cdots$$
$$+ b_n(c_{n1}\mathbf{e}_1 + c_{n2}\mathbf{e}_2 + \cdots + c_{nn}\mathbf{e}_n)$$

Factor out the basis $\mathbf{e}_i$ from the RHS, we obtain

$$a_1\mathbf{e}_1 + a_2\mathbf{e}_2 + \cdots + a_n\mathbf{e}_n = (b_1c_{11}\mathbf{e}_1 + b_1c_{12}\mathbf{e}_2 + \cdots + b_1c_{1n}\mathbf{e}_n) +$$
$$(b_2c_{21}\mathbf{e}_1 + b_2c_{22}\mathbf{e}_2 + \cdots + b_2c_{2n}\mathbf{e}_n) +$$
$$\cdots$$
$$+ (b_nc_{n1}\mathbf{e}_1 + b_nc_{n2}\mathbf{e}_2 + \cdots + b_nc_{nn}\mathbf{e}_n)$$

Hence

$$a_1\mathbf{e}_1 + a_2\mathbf{e}_2 + \cdots + a_n\mathbf{e}_n = \mathbf{e}_1(b_1c_{11} + b_2c_{21} + \cdots + b_nc_{n1}) +$$
$$\mathbf{e}_2(b_1c_{12} + b_2c_{22} + \cdots + b_nc_{n2}) +$$
$$\cdots$$
$$+ \mathbf{e}_n(b_1c_{1n} + b_2c_{2n} + \cdots + b_nc_{nn})$$

Now comparing coefficients of each of the basis $\mathbf{e}_i$ we obtain the following result

$$a_1 = b_1c_{11} + b_2c_{21} + \cdots + b_nc_{n1}$$
$$a_2 = b_1c_{12} + b_2c_{22} + \cdots + b_nc_{n2}$$
$$\cdots$$
$$a_n = b_1c_{1n} + b_2c_{2n} + \cdots + b_nc_{nn}$$

Or in Matrix form, we write

$$\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} c_{11} & c_{21} & c_{31} & \cdots & c_{n1} \\ c_{12} & c_{22} & c_{32} & \cdots & c_{n2} \\ \vdots & \cdots & \cdots & \ddots & \vdots \\ c_{1n} & c_{2n} & c_{3n} & \cdots & c_{nn} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \qquad (3)$$

The above gives a relation between the coordinates of the vector $\mathbf{x}$ w.r.t. basis v (these are the a_i coordinates) to the coordinates of the same vector w.r.t. to basis v' (these are the coordinates b_j). The mapping between these coordinates is the matrix shown above which we call the $[M]$ matrix. Since the above matrix returns the coordinates of the vector w.r.t. v when it is multiplied by the coordinates of the vector w.r.t. basis v', we write it as $[M]_{v'\to v}$ to make it clear from which basis to which basis is the conversion taking place.

Looking at (2) and (3) above, we see that column j in $[M]_{v'\to v}$ is the coordinates of the $\mathbf{e}'_j$ w.r.t. basis v.

Hence the above gives a method to generate the change of basis matrix $[M]_{v'\to v}$. Simply represent each basis in v' w.r.t. to basis v. Take the result of each such representation and write it as a column in $[M]_{v'\to v}$. We now show few examples to illustrate this.

Example showing how to find change of basis matrix

Given the $\mathbb{R}^2$ basis $v = \{\mathbf{e}_1, \mathbf{e}_2\} = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$, and new basis $v' = \{\mathbf{e}'_1, \mathbf{e}'_2\} = \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\}$, find the change of basis matrix $[M]_{v\to v'}$

Column 1 of $[M]_{v\to v'}$ is the coordinates of $\mathbf{e}_1$ w.r.t. basis v'. i.e.

$$\mathbf{e}_1 = c_{11}\mathbf{e}'_1 + c_{21}\mathbf{e}'_2 \tag{4}$$

and column 2 of $[M]_{v\to v'}$ is the coordinates of $\mathbf{e}_2$ w.r.t. basis v'. i.e.

$$\mathbf{e}_2 = c_{12}\mathbf{e}'_1 + c_{22}\mathbf{e}'_2 \tag{5}$$

But (4) is

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} = c_{11} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + c_{21} \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

Hence $1 = c_{11} - c_{21}$ and $0 = c_{11} + c_{21}$, solving, we obtain $c_{11} = \frac{1}{2}, c_{21} = \frac{-1}{2}$ and (5) is

$$\begin{bmatrix} 0 \\ 1 \end{bmatrix} = c_{12} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + c_{22} \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

Hence $0 = c_{12} - c_{22}$ and $1 = c_{12} + c_{22}$, solving we obtain $c_{12} = \frac{1}{2}$, $c_{22} = \frac{1}{2}$, hence

$$[M]_{v\to v'} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{-1}{2} & \frac{1}{2} \end{bmatrix}$$

Now we can use the above change of basis matrix to find coordinates of the vector under v' given its coordinates under v. For example, given $\mathbf{x}_v = \begin{bmatrix} 10 \\ 5 \end{bmatrix}$, then

$$\mathbf{x}_{v'} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{-1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 10 \\ 5 \end{bmatrix} = \begin{bmatrix} 7.5 \\ -2.5 \end{bmatrix}$$

1.2.3   Examples of similarity transformation

Example 1   This example from Strang book (Linear Algebra and its applications, 4th edition), but shown here with little more details.

Consider $\mathbb{R}^2$, and let T be the projection onto a line at angle $\theta = 135^0$. Let the first basis $v = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$ and let the second basis $v' = \left\{ \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$.

Figure 6: Change of basis illustration

The first step is to find $[M]_{v' \to v}$. The first column of $[M]_{v' \to v}$ is found by writing e_1' w.r.t. basis v, and the second column of $[M]_{v' \to v}$ is found by writing e_2' w.r.t. basis v, Hence

$$e_1' = c_{11} e_1 + c_{21} e_2$$

$$\begin{bmatrix} 1 \\ -1 \end{bmatrix} = c_{11} \begin{bmatrix} 1 \\ 0 \end{bmatrix} + c_{21} \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

12

Or $1 = c_{11}$ and $c_{21} = -1$. And

$$e_2' = c_{12}e_1 + c_{22}e_2$$

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix} = c_{12} \begin{bmatrix} 1 \\ 0 \end{bmatrix} + c_{22} \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Hence $c_{12} = 1$ and $c_{22} = 1$. Hence

$$[M]_{v' \to v} = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$$

Now we need to find $[T]_v$ the representation of T w.r.t. basis v. The first column of $[T]_v$ is the new coordinates of the basis v_1 after applying T onto it. but

$$Te_1 = T \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} 0.5 \\ -0.5 \end{bmatrix}$$

and the second column of $[T]_v$ is the new coordinates of the basis v_2 after applying T onto it. but

$$Te_2 = T \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} -0.5 \\ 0.5 \end{bmatrix}$$

Hence

$$[T]_v = \begin{bmatrix} 0.5 & -0.5 \\ -0.5 & 0.5 \end{bmatrix}$$

Therefore

$$[T]_{v'} = [M]_{v' \to v}^{-1}[T]_v[M]_{v' \to v}$$

$$= \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 0.5 & -0.5 \\ -0.5 & 0.5 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$$

$$[T]_{v'} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

Hence we see that the linear transformation T is represented as $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ w.r.t. basis v', while it is represented as $\begin{bmatrix} 0.5 & -0.5 \\ -0.5 & 0.5 \end{bmatrix}$ w.r.t. basis v'. Therefore, it will be better to perform all calculations involving this linear transformation under basis v' instead of basis v

13

Figure 7: Change of basis

Example 2 Consider $\mathbb{R}^2$, and let T be a rotation of space by $\theta = \frac{\pi}{4}$. Let the first basis $v = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$ and let the second basis be $v' = \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\}$.
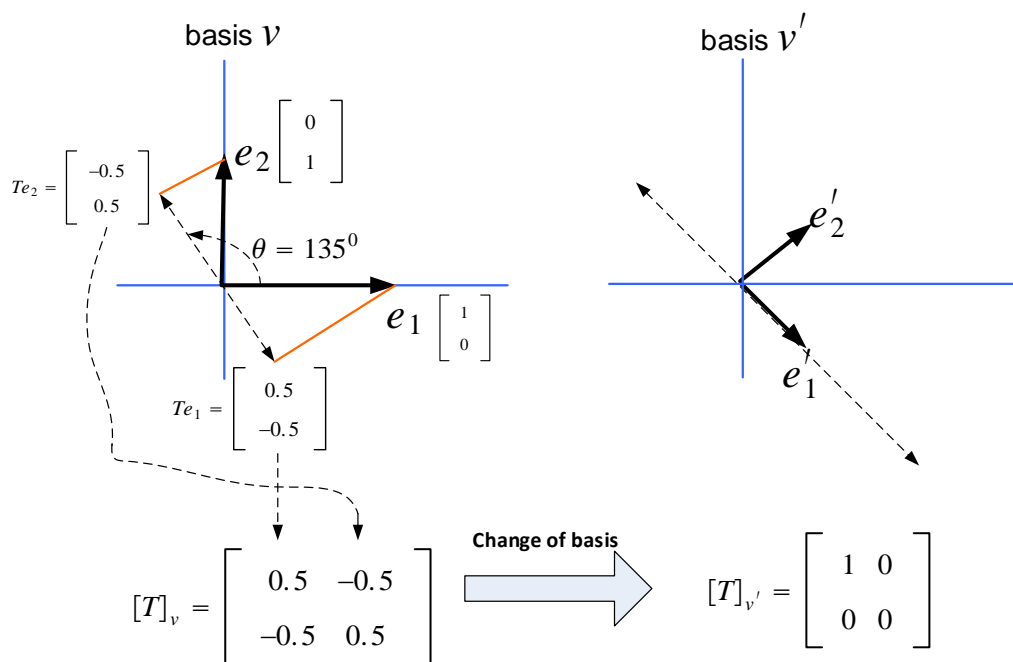
The first step is to find $[M]_{v' \to v}$. The first column of $[M]_{v' \to v}$ is found by writing e_1' w.r.t. basis v, and the second column of $[M]_{v' \to v}$ is found by writing e_2' w.r.t. basis v, Hence

$$e_1' = c_{11}e_1 + c_{21}e_2$$

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix} = c_{11} \begin{bmatrix} 1 \\ 0 \end{bmatrix} + c_{21} \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Or $1 = c_{11}$ and $c_{21} = 1$ and

$$e_2' = c_{12}e_1 + c_{22}e_2$$

$$\begin{bmatrix} -1 \\ 1 \end{bmatrix} = c_{12} \begin{bmatrix} 1 \\ 0 \end{bmatrix} + c_{22} \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Hence $c_{12} = -1$ and $c_{22} = 1$. Hence

$$[M]_{v' \to v} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$$

14

Now we need to find $[T]_v$ the representation of T w.r.t. basis v. The first column of $[T]_v$ is the new coordinates of the basis v_1 after applying T onto it. but

$$Te_1 = T\begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}$$

and the second column of $[T]_v$ is the new coordinates of the basis v_2 after applying T onto it. but

$$Te_2 = T\begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$Te_2 = \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix}$$

Hence

$$[T]_v = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

$$= \begin{bmatrix} \cos\frac{\pi}{4} & -\sin\frac{\pi}{4} \\ \sin\frac{\pi}{4} & \cos\frac{\pi}{4} \end{bmatrix}$$

$$[T]_v = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

Therefore

$$[T]_{v'} = [M]_{v'\to v}^{-1}[T]_v[M]_{v'\to v}$$

$$= \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 0 & -\sqrt{2} \\ \sqrt{2} & 0 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

Hence we see that the linear transformation T is represented as $\begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$ w.r.t. basis

v', while it is represented as $\begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$ w.r.t. basis v'. Therefore the change of basis selected above did not result in making the representation of T any simpler (in fact there was no change in the representation). This means we need to find a more direct method of finding the basis under which T has the simplest representation. Clearly we can't just keep trying different basis to find if T has simpler representation under the new basis.

1.2.4 How to find the best basis to simplify the representation of T?

Our goal of simpler representation of T is that of a diagonal matrix or as close as possible to being diagonal matrix (i.e. Jordan form). Given $[T]_v$ and an infinite number of basis v' that we could select to represent T under in the hope we can find a new representation $[T]_{v'}$ such that it is simpler than $[T]_v$, we now ask, how does one go about finding such basis?

It turns out the if we select the eigenvectors of $[T]_v$ as the columns of $[M]_{v' \to v}$, this will result in $[T]_{v'}$ being diagonal or as close to being diagonal as possible (block diagonal or Jordan form).

Let us apply this to the second example in the previous section. In that example, we had

$$[T]_v = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

The eigenvectors of $[T]_v$ are $\begin{bmatrix} -i \\ 1 \end{bmatrix}$ and $\begin{bmatrix} i \\ 1 \end{bmatrix}$, $[M]_{v' \to v} = \begin{bmatrix} -i & i \\ 1 & 1 \end{bmatrix}$. Therefore

$$[T]_{v'} = [M]_{v' \to v}^{-1}[T]_v[M]_{v' \to v}$$

$$= \begin{bmatrix} -i & i \\ 1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} -i & i \\ 1 & 1 \end{bmatrix}$$

$$[T]_{v'} = \begin{bmatrix} \left(\frac{1}{2} - \frac{1}{2}i\right)\sqrt{2} & 0 \\ 0 & \left(\frac{1}{2} + \frac{1}{2}i\right)\sqrt{2} \end{bmatrix}$$

Which is diagonal representation. Hence we see that the linear transformation T is represented as $\begin{bmatrix} \left(\frac{1}{2} - \frac{1}{2}i\right)\sqrt{2} & 0 \\ 0 & \left(\frac{1}{2} + \frac{1}{2}i\right)\sqrt{2} \end{bmatrix}$ w.r.t. basis v'

1.3 Summary of similarity transformation

To obtain $B = M^{-1}AM$ such that B is real and diagonal requires that A be real and symmetric. The eigenvalues of A goes into the diagonal of B and the eigenvectors of A go into the columns of M. This is an algebraic view.

Geometrically, A is viewed as the matrix representation under some basis v of a linear transformation T. And B is the matrix representation of the same linear transformation T but now under a new basis v' and M is the matrix that represents the change of basis from v' to v.

The question then immediately arise: If A must be real and symmetric (for B to be real and diagonal), what does this mean in terms of the linear transformation T and change of basis matrix M? This clearly mean that, under some basis v, not every linear transformation represented by A can have a similar matrix B which is real and diagonal. Only those linear transformations which result in real and symmetric representation can have a similar matrix which is real and diagonal. This is shown in the previous examples, where we saw that T defined as rotation by 45^0 under the standard basis

$v = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$ resulted in $A = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$ and since this is not symmetric, hence

we will not be able to factor this matrix using similarity transformation to a diagonal matrix, no matter which change of basis we try to represent T under. The question then, under a given basis v what is the class of linear transformation which leads to a matrix representation that is symmetric? One such example was shown above which is the projection into the line at 135^0. This question needs more time to look into.

We now write Λ to represent a diagonal matrix, hence the similarity transformation above can be written as

$$\Lambda = M^{-1}AM$$

Or

$$A = M\Lambda M^{-1}$$

Eigenvalues of A go into the diagonal

A is real and symmetric

$$A = M\,\Lambda\,M^{-1}$$

Eigenvectors of A go into the columns
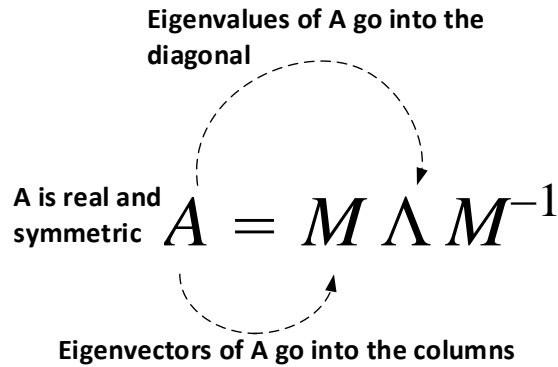
Figure 8: Illustration of similarity transformation

2 Singular value decomposition (SVD)

Using similarity transformation, we found that for a real and symmetric matrix A we are able to decompose it as $A = M\Lambda M^{-1}$ where Λ is diagonal and contains the eigenvalues of A on the diagonal, and M contains the right eigenvectors of A in its columns.

2.1 What is right and left eigenvectors?

Right eigenvectors are the standard eigenvectors we have been talking about all the time. When λ is an eigenvalues of A then $\mathbf{x}$ is a right eigenvector when we write

$$A\mathbf{x} = \lambda\mathbf{x}$$

However, $\mathbf{x}$ is a left eigenvector of A when we have

$$\mathbf{x}^H A = \lambda\mathbf{x}^H$$

where $\mathbf{x}^H$ is the Hermitian of $\mathbf{x}$

2.2 SVD

Given any arbitrary matrix $A_{m\times n}$ it can be factored into 3 matrices as follows $A_{m\times n} = P_{m\times m}D_{m\times n}Q_{n\times n}$ where P is a unitary matrix ($P^H = P^{-1}$ or $P^H P = I$), and Q is also unitary matrix.

These are the steps to do SVD

1. Find the rank of A, say r

2. Let $B = A_{n \times m}^H A_{m \times n}$, hence $B_{n \times n}$ is a square matrix, and it is semi positive definite, hence its eigenvalues will all be ≥ 0. Find the eigenvalues of B, call these σ_i^2. There will be n such eigenvalues since B is of order $n \times n$. But only r of these will be positive, and $n - r$ will be zero. Arrange these eigenvalues such that the first r non-zero eigenvalues come first, followed by the zero eigenvalues:

$$\overbrace{\sigma_1^2, \sigma_2^2, \cdots, \sigma_r^2, 0, 0, \cdots, 0}^{n\ eigenvalues}$$

3. Initialize matrix $D_{m \times n}$ to be all zeros. Take the the first r eigenvalues from above (non-zero ones), and take the square root of each, hence we get

$$\overbrace{\sigma_1, \sigma_2, \cdots, \sigma_r}^{r\ \text{singular values}}$$

and write these down the diagonal of D starting at $D(1,1)$, i.e. $D(1,1) = \sigma_1, D(2,2) = \sigma_2, \cdots, D(r,r) = \sigma_r$. Notice that the matrix D need not square matrix. Hence we can obtain an arrangement such as the following for $r = 2$

$$D = \begin{pmatrix} \sigma_1 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

where the matrix A was 3×4 for example.

4. Now find each eigenvalue of $A^H A$. For each eigenvalue, find the corresponding eigenvector. Call these eigenvectors $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n$.

5. Normalize the eigenvectors found in the above step. Now $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n$ eigenvector will be an orthonormal set of vectors. Take the Hermitian of each eigenvector $\mathbf{u}_1^H, \mathbf{u}_2^H, \cdots, \mathbf{u}_n^H$ and make one of these vectors (now in row format instead of column format) go into a row in the matrix Q.i.e. the first row of Q will be $\mathbf{u}_1^H$, the second row of Q will be $\mathbf{u}_2^H$, etc... $\left(A^H A\right)_{n \times n} \xrightarrow{\text{find eigenvectors and normalize}}$

$$\left\{ \mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_r, \quad \overbrace{\mathbf{u}_{r+1}, \cdots, \mathbf{u}_n}^{\text{ortho basis (n-r) for N(A)}} \right\} \to Q_{n \times n} = \begin{bmatrix} \mathbf{u}_1^T \\ \mathbf{u}_2^T \\ \vdots \\ \mathbf{u}_r^T \\ \mathbf{u}_{r+1}^T \\ \vdots \\ \mathbf{u}_n^T \end{bmatrix}$$

19

6. Now we need to find a set of m orthonormal vectors, these will be the columns of the matrix P: find a set of r orthonormal eigenvector $\mathbf{v}_i = \frac{1}{\sigma_i} A\mathbf{u}_i$, for $i = 1 \cdots r$. Notice that here we only use the first r vectors found in step 5. Take each one of these $\mathbf{v}_i$ vectors and make them into columns of P. But since we need m columns in P not just r, we need to come up with $m - r$ more basis vectors such that all the m vectors form an orthonormal set of basis vectors for the row space of A, i.e. $\mathbb{C}^m$. If doing this by hand, it is easy to find the this $m - r$ by inspection. In a program, we could use the same process we used with Gram-Schmidt, where we learned how find a new vector which is orthonormal to a an existing set of other vectors. Another way to find the matrix P is to construct the matrix AA^H and find its eigenvectors, those go into the columns of the matrix P as follows

$$\left(AA^H\right)_{m \times m} \xrightarrow{\text{find eigenvectors and normalize}} \left\{ \overbrace{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_r}^{\text{orthonormal basis for range A}}, \mathbf{v}_{r+1}, \cdots, \mathbf{v}_m \right\} \rightarrow P_{m \times m} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{bmatrix}$$

7. This completes the factorization, now we have $A = P\ D\ Q$

The following diagram help illustrate the SVD process described above.

SVD factorization

A(MxN) matrix, Rank=r

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

$$\boxed{A = PDQ}$$

$$A_{m \times n} = P_{m \times m} D_{m \times n} Q_{n \times n}$$

Row space(A)

Dimension=r

R^n

Dimension= n-r

Null space(A)

Column space(A) (or range of A)

Dimension=r

R^m

Dimension= m-r

Left Null space(A)

Orthonormal Basis for Null Space of A

$(A^H A)_{n \times n} \xrightarrow[\text{And normalize}]{\text{find eigenvectors}} \left\{ \vec{u}_1, \vec{u}_2, \cdots, \vec{u}_r, \overbrace{\vec{u}_{r+1}, \cdots, \vec{u}_n}^{\text{ortho basis for N(A)}} \right\} \to Q_{n \times n} = \begin{bmatrix} \vec{u}_1^T \\ \vec{u}_2^T \\ \vdots \\ \vec{u}_r^T \\ \vec{u}_{r+1}^T \\ \vdots \\ \vec{u}_n^T \end{bmatrix}$

n eigenvalues

$\overbrace{\sigma_1^2, \sigma_2^2, \cdots, \sigma_r^2, 0, 0, \cdots, 0}$

For these $n - r$ vectors

$A\vec{u}_i = 0$

Hence basis for $N(A)$

From these r vectors we obtain

$\vec{v}_i = \frac{A\vec{u}_i}{\sigma_i}$

Hence r basis for *Range A*

$(AA^H)_{m \times m} \xrightarrow[\text{And normalize}]{\text{find eigenvectors}} \left\{ \overbrace{\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_r}^{\text{r ortho basis for range A}}, \vec{v}_{r+1}, \cdots, \vec{v}_m \right\} \to P_{m \times m} = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_r & \vec{v}_{r+1} & \cdots & \vec{v}_m \end{bmatrix}$

D1.vsd
By Nasser M. Abbasi
July 3,2007

Figure 9: SVD process

3 Conclusion

As a conclusion, the following diagram shows the difference between similarity transformation factorization of a matrix and SVD.

Eigenvalues of A go into the diagonal

Diagonal matrix

A is real, sqaure and symmetric

$$A = M \Lambda M^{-1}$$

$n \times n$ $n \times n$ $n \times n$ $n \times n$

Eigenvectors of A go into the columns

$$\boxed{\textbf{Similarity Transformation}}$$

$$A^H A$$

Normalized Eigenvectors of A'A go into the rows of Q

A is arbitrary mxn matrix

$$A = P D Q$$

$m \times n$ $m \times m$ $m \times n$ $n \times n$

The square root of the nonzero eigenvalues of AA' (or A'A) go into the diagonal of matrix D. These are called the singular values of A

$$AA^H$$

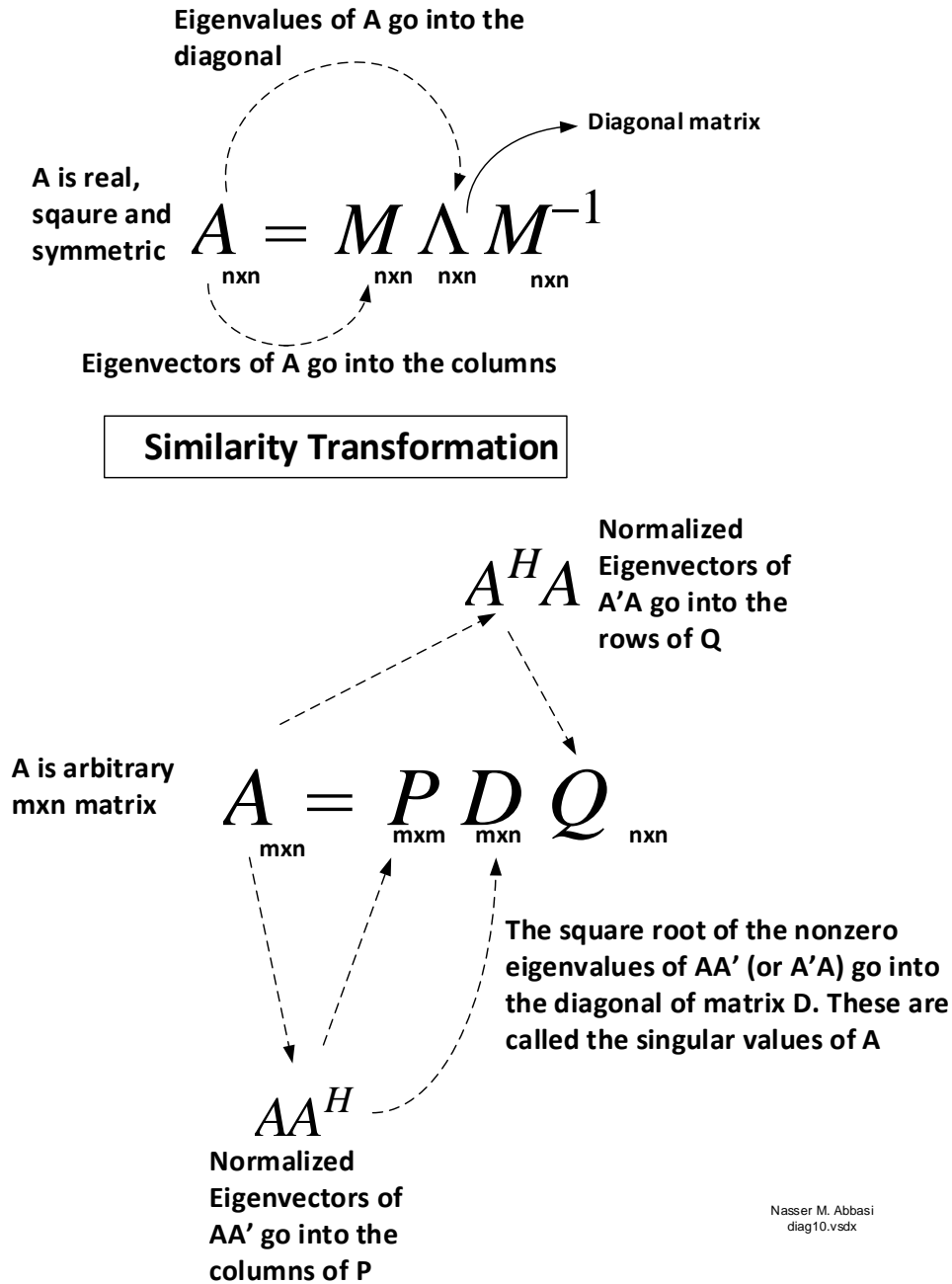**Normalized Eigenvectors of AA' go into the columns of P**

Figure 10: similarity transformation factorization of a matrix and SVD

Trying to understand SVD from geometrical point view requires more time, and this is left for future research into this subject.

4   References

1. Linear Algebra and its applications 4th edition by Gilbert Strang

2. Course notes, Linear Algebra, Math 307, CSUF mathematics department, spring 2007 by Dr Angel Pineda.

3. Principles and techniques of applied Mathematics by Bernard Friedman, Spectral theory of operators chapter.

4. Matrix computation, 3rd edition by Gene Golub and Charles Van Loan.

5. Numerical analysis: Mathematics of scientific computing, by David Kincaid and Ward Cheney.