On the Decomposition of Square Matrices

Nicholi Caron
Honors Project

Abstract

This paper provides a brief overview of some of the requirements, computational procedures, benefits and applications of the decomposition (i.e. factorization) of square matrices into the products of several simpler matrices. Some of the factorizations discussed are LU decomposition, symmetric eigenvalue decomposition, Jordan decomposition, and the singular value decomposition. An important takeaway is that these decompositions are related by their form with progressively demanding requirements and constructions.

> ”In the language of Computer Science, the expression of [a matrix] A as a product amounts to a pre-processing of the data in A, organizing that data into two or more parts whose structures are more useful in some way, perhaps more accessible for computation.”
> - David C. Lay

Acknowledgements

First and foremost, I want to say thanks to my Linear Algebra professor Dr. Paul J. Kapitza for taking the time out of his schedule to supervise this extracurricular honors project, and for humoring my tangential inquiries. Dr. Kapitza does an amazing job of elucidating the beauties which underpin Linear Algebra. He is incredibly thorough in his lectures to the point that I suspect he may rehearse them in advance.

I tend to take a top-down approach when learning new materials. I look at a subject from a macro-level to understand the context of its pieces before delving into the minutia. Professor Kapitza has shown me that inverting my approach, and spending more time on the fundamentals can lead to substantial revelations.

Next, I have to give thanks to my grandfather for emphasizing the importance of mathematics to me at an early age. I have no clue where I would be if it were not for his mentorship and foresight.

Lastly, I am forever grateful for the open-source movement and the democratization of knowledge. The wealth of information freely available at my fingertips never ceases to amaze me. It is awe-inspiring to be able to explore my curiosities as deeply as I wish to, whenever I choose to. The selflessness of the many contributors is a beautiful thing. I hope to repay them by one day contributing something myself.

> ”What drove me? I think most creative people want to express appreciation for being able to take advantage of the work that’s been done by others before us. I didn’t invent the language or mathematics I use, I make little of my own food, none of my own clothes. Everything I do depends on other members of our species and the shoulders that we stand on.”
> - Steve Jobs

Contents

PART I

Decompositions

CHAPTER 1

Symmetric Eigenvalue Decomposition (SED)

1.1 Eigenvalues and Eigenvectors

A matrix A can be thought of as a function which takes a vector $\vec{x}$ as input, and outputs a transformed vector A$\vec{x}$.
Eigenvectors are vectors for which A$\vec{x}$ is parallel to $\vec{x}$

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

where the factor by which $\vec{x}$ is stretched (i.e. the constant λ) is the eigenvalue corresponding to that particular eigenvector.

Determining Eigenvalues

Eigenvalues of a matrix A can be found by solving for λ in the following equation:

$$det(A - \lambda I) = 0$$

The polynomial this reduces to is referred to as the characteristic polynomial of A.
If there are any repeated eigenvalues, the numbers of times they repeat are referred to as their multiplicities.

Determining Eigenvectors

To find an eigenvector $\vec{e}$ associated with a particular eigenvalue λ, we must substitute λ into the equation det(A - λI) = 0 and bring the resulting matrix to row reduced echelon form, solving for all x_n. This will be shown in more detail in the examples section

1.2 Orthogonal Matrices

Orthogonal matrices consist of orthogonal (i.e. perpendicular) column vectors. The dot product of two orthogonal vectors is equal to zero.
Conveniently, orthogonal matrices have the property

$$Q^T Q = I$$

When dealing with orthogonal square matrices, they have the additional property

$$Q^T = Q^{-1}$$

1.3 Symmetric Matrices

A symmetric matrix is one for which:

$$A = A^T$$

If a symmetric matrix has some special properties, its eigenvalues and eigenvectors likely have special properties as well

- If A consists of real number entries, it's eigenvalues are real and there exists a complete set of eigenvectors that are orthogonal and of unit length (orthonormal).

Spectral Theorem

Theorem 1.3.1. *We can decompose any Hermitian (i.e. real-symmetric) matrix A with the symmetric eigenvalue decomposition (SED)*

$$A = \sum_{i=1}^{n} \lambda_i p_i p_i^T = P\Lambda P^T = P\Lambda P^{-1},$$

$$\Lambda = diag(\lambda_1 \cdots \lambda_n)$$

$$P \text{ is orthonormal}$$

The following corollaries are used in the examples

1.3.1.1. *For any real-symmetric matrix, there are exactly n (not necessarily distinct) real eigenvalues*

1.3.1.2. *The associated eigenvectors can be chosen to form an orthonormal basis*

1.3.1.3. *A factorization of the form $P\Lambda P^T$ can be crafted given the previous information*

1.4 A Concrete Example

Find the eigenvalues, eigenvectors, and the symmetric eigenvalue decomposition of the matrix

$$A = \begin{bmatrix} -2 & 6 \\ 6 & -2 \end{bmatrix}$$

- First, let's find A's eigenvalues

$$det\left(\begin{bmatrix} -2-\lambda & 6 \\ 6 & -2-\lambda \end{bmatrix}\right) = 0$$

$$\Rightarrow \lambda^2 + 4\lambda - 32 = 0$$
$$\Rightarrow (\lambda + 8)(\lambda - 4) = 0$$
$$\lambda_1 = -8$$
$$\lambda_2 = 4$$

- Next, we'll substitute our eigenvalues into (A - λ I) to find their associated eigenvectors

  1. λ_1 = -8:
  $$\begin{bmatrix} -2 & 6 \\ 6 & -2 \end{bmatrix} - \begin{bmatrix} -8 & 0 \\ 0 & -8 \end{bmatrix} = \begin{bmatrix} 6 & 6 \\ 6 & 6 \end{bmatrix}$$
  To find the eigenvector $\vec{e_1}$ corresponding with this eigenvalue, we must bring this matrix to row reduced echelon form.
  $$\begin{bmatrix} 6 & 6 \\ 6 & 6 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$$
  $$\begin{cases} \mathrm{x}_1 & = -x_2 \\ \mathrm{x}_2 & \text{is free} \end{cases} \tag{1.1}$$
  $$\Rightarrow \vec{x} = \begin{bmatrix} -x_2 \\ x_2 \end{bmatrix} = x_2 \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$
  $$\vec{e_1} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

  2. λ_2 = 4:
  $$\begin{bmatrix} -2 & 6 \\ 6 & -2 \end{bmatrix} - \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix} = \begin{bmatrix} -6 & 6 \\ 6 & -6 \end{bmatrix}$$
  To find the eigenvector $\vec{e_2}$ corresponding with λ_2, we must bring this matrix to row reduced echelon form.
  $$\begin{bmatrix} -6 & 6 \\ 6 & -6 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix}$$
  $$\begin{cases} \mathrm{x}_1 & = x_2 \\ \mathrm{x}_2 & \text{is free} \end{cases} \tag{1.2}$$
  $$\Rightarrow \vec{x} = \begin{bmatrix} x_2 \\ x_2 \end{bmatrix} = x_2 \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$
  $$\vec{e_2} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

- Therefore the eigenvectors of A are:
  $$\vec{e_1}, \vec{e_2} = \left\{ \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$$
  These vectors are already orthogonal, as $\vec{e_1} \cdot \vec{e_2} = 0$. Let's normalize them to form our orthonormal matrix P

$$\vec{p_1} = \frac{\vec{e_1}}{||\vec{e_1}||} = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

$$\vec{p_2} = \frac{\vec{e_2}}{||\vec{e_2}||} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

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

- Applying Theorem 1.3.1:

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

  - Because P happens to by symmetric, P = P^T
  - Also, since P is orthogonal and square, $P^T = P^{-1}$
  - Therefore P = $P^T = P^{-1}$

$$\begin{bmatrix} -2 & 6 \\ 6 & -2 \end{bmatrix} = \begin{bmatrix} \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} -8 & 0 \\ 0 & 4 \end{bmatrix} \begin{bmatrix} \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

$$\surd$$

CHAPTER 2

Jordan Decomposition

2.1 Background

Positive Definiteness

A symmetric, square matrix A is positive definite if:

- $\vec{x}^T$ A $\vec{x}$ is positive except when $\vec{x} = \vec{0}$.

For a symmetric matrix A;

1. All of A's eigenvalues are greater than 0,
2. The determinant of A is greater than 0,
3. All pivots of A are greater than 0.

Similarity

Two matrices A and D are similar if

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

for some matrix P. This allows us to group together matrices that perform similar transformations.

The Factorization

The Jordan Decomposition is the factorization of a square matrix A into the product of three matrices:

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

where:

1. A and D are similar matrices,
2. D is a diagonal matrix,
3. P^{-1} is the matrix inverse of P.

2.2 Powers of a Matrix

Let's say we want to square a matrix A; supposing A has a Jordan decomposition

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

$$A^2 = PDP^{-1}PDP^{-1}$$
$$\Rightarrow A^2 = PD^2P^{-1}$$

Clearly, this pattern can be extrapolated to higher powers, giving us the general form:

$$A^n = PD^nP^{-1}$$

Since D is diagonal, this works out quite nicely:

$$\Rightarrow A^n = P \begin{bmatrix} d_1^n & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & d_m^n \end{bmatrix} P^{-1}$$

This makes for a great reduction in complexity. We can avoid n matrix multiplications, which can be very expensive computations – namely O(m^3) scalar multiplications each, using the classical algorithm. And instead only perform three matrix multiplications and raise m scalar values to the n^{th} power, which is relatively trivial.

A Concrete Example

> "The heart of mathematics consists of concrete examples and concrete problems."
> - P. R. Halmos

Given the matrices

$$A = \begin{bmatrix} 1 & -2 & 8 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix},$$

$$P = \begin{bmatrix} 1 & -4 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

Confirm that P diagonalizes A, then compute A^{1000}

1. First, we have to find P^{-1}:

$$P^{-1} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & -1 & 4 \end{bmatrix}.$$

2. By rearranging the previous equation, the product of P^{-1}AP should give us a diagonal matrix D

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

$$D = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & -1 & 4 \end{bmatrix} \begin{bmatrix} 1 & -2 & 8 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & -4 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

$$\Rightarrow D = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

P diagonalizes A. $\surd$

3. Next we have to solve for A:

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

Since A and D are similar:

$$A^n = PD^nP^{-1}$$

Thus

$$A^{1000} = PD^{1000}P^{-1}$$

$$A^{1000} = P\begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}^{1000} P^{-1}$$

$$\Rightarrow A^{1000} = P\begin{bmatrix} -1^{1000} & 0 & 0 \\ 0 & -1^{1000} & 0 \\ 0 & 0 & 1^{1000} \end{bmatrix} P^{-1}$$

$$\Rightarrow A^{1000} = PI_3P^{-1}$$

$$\Rightarrow A^{1000} = PP^{-1}$$

$$A^{1000} = I_3$$

I_3 = 3X3 Identity Matrix

2.3 The Fibonacci Sequence

The state of a difference equation can be represented by a vector $\vec{u_k}$ where the second entry is the preceding state, and the first entry is the current state. Difference equations can be represented by the product of some growth matrix A and a seed vector $\vec{u_0}$:

$$\vec{u_0} = A^k u_0 = \sum_{i=1}^{n} c_n \lambda_n^k \vec{x_n}$$

The Fibonacci Sequence is an example of an additive recurrence with initial values of 0 and 1. The next digit of the Fibonacci Sequence is the sum of its two preceding values and can be represented by the form

$$F_{k+2} = F_{k+1} + F_k.$$

Although this is a second order scalar equation, it can be converted to a first order linear system by letting

$$\vec{u}_k = \begin{bmatrix} F_{k+1} \\ F_k \end{bmatrix}$$

1. $F_{k+2} = F_{k+1} + F_k$,
2. $F_{k+1} = F_1$.

This can be rewritten as the linear system

$$\vec{u}_{k+1} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \vec{u}_k.$$

Thus, our transformation matrix

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}.$$

Finding a Formula for the k^{th} Fibonacci Number

- First, we must calculate the eigenvalues of A

$$det\left(\begin{bmatrix} 1-\lambda & 1 \\ 1 & -\lambda \end{bmatrix}\right) = 0$$

$$\Rightarrow \lambda^2 - \lambda - 1 = 0$$

$$\lambda = \frac{1 \pm \sqrt{2}}{2}.$$

- Next, let's find our eigenvectors:

$$(A - \lambda I)\vec{x} = \begin{bmatrix} 1-\lambda & 1 \\ 1 & -\lambda \end{bmatrix} \vec{x} = \vec{0}$$

when:

$$\vec{x} = \begin{bmatrix} \lambda \\ 1 \end{bmatrix}$$

$$\Rightarrow \vec{x_1} = \begin{bmatrix} \lambda_1 \\ 1 \end{bmatrix}; \vec{x_2} = \begin{bmatrix} \lambda_2 \\ 1 \end{bmatrix}.$$

- We can use our initial conditions to solve for our constants:

$$\vec{u_0} = \begin{bmatrix} F_1 \\ F_0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} = c_1 x_1 + c_2 x_2$$

$$\Rightarrow c_1 = -c_2 = \frac{1}{\sqrt{5}}.$$

- Finally, since $\begin{bmatrix} F_{k_1} \\ F_k \end{bmatrix} = c_1\lambda_1 x_1 + c_2\lambda_2 x_2$, we can plug in our now-known quantities to reveal a closed form expression:

$$F_k = \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^k - \frac{1}{\sqrt{5}}\left(\frac{1-\sqrt{5}}{2}\right)^k.$$

The eigenvalues of a first order system can determine how the system performs as the inputs increase. Since $\lambda_1 = \frac{1+\sqrt{2}}{2}$ is the only eigenvalue with an absolute value greater than 1, the growth of the Fibonacci Sequence in the limit is determined by λ_1, and can thus be approximated by $F_k \approx \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^k$.

CHAPTER 3

LU Decomposition

3.1 Background

The LU decomposition of a matrix factorizes a square matrix A into two triangular matrices L and U:

$$A = LU$$

- L is a lower triangular matrix, i.e., all entries above the diagonal are equal to zero.
- U is an upper triangular matrix, i.e., all entries below the diagonal are equal to zero.

Finding L and U

1. To compute a LU decomposition, we perform Gaussian elimination on our original matrix A, until it is of upper triangular form (not necessarily reduced echelon form). This newfound upper triangular matrix will serve as U.
2. Along the way, we must keep track of which row operations we have used, as well as the order in which they were applied. The elementary matrices which represent the row operations will be multiplied successively on the left of A.
3. Lastly, we have to isolate A. To do so, we must multiply on the left of both sides by the inverse product of our elementary matrices to reveal our lower triangular matrix L.

3.2 A Concrete 2X2 Example

> "Practice yourself, for heavens sake, in little things; and thence proceed to greater"
> - Epictetus

Find an LU Decomposition of the previously established Fibonacci matrix:

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}.$$

- To turn A into an upper triangular matrix, we must subtract Row 1 from Row 2:
  - Representing this row operation as an elementary matrix, gives us

$$E_1 = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}.$$

- Multiplying our original matrix A on the left by this elementary matrix is equivalent to performing the row operation, which gives us the equation

$$E_1 A = U$$

$$\begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & -1 \end{bmatrix}.$$

- All that's left is to Multiply on the left of both sides by E_1^{-1}

$$E_1^{-1} E_1 A = E_1^{-1} U$$

$$\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & -1 \end{bmatrix}.$$

- This simplifies down to

$$I_2 A = E^{-1} U$$

Since E^{-1} is of lower triangular form,

$$I_2 A = LU$$

$$\Rightarrow A = LU$$

$$\Rightarrow \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & -1 \end{bmatrix}.$$

$$\checkmark$$

3.3 A 3X3 Example with Variables

Find an LU Decomposition of the matrix A

$$A = \begin{bmatrix} 1 & 0 & 1 \\ a & a & a \\ b & b & a \end{bmatrix},$$

for which real numbers a and b does it exist?

1. Our primary objective is to bring A to an upper triangular matrix by way of Gaussian elimination.

a) First we'd like to clear the row 2, column 1 entry.

To do so, we will subtract a times row 1 from row 2, which is equivalent to multiplying A on the left by the matrix E_1

$$E_1 = \begin{bmatrix} 1 & 0 & 0 \\ -a & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$E_1 A = U^*$$

$$\begin{bmatrix} 1 & 0 & 0 \\ -a & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ a & a & a \\ b & b & a \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & a & 0 \\ b & b & a \end{bmatrix} .$$

b) Next, we want to eliminate the $U^*_{3,1}$ entry.

To complete this, we will subtract b times row 1 from row 3, or effectively multiply E_1A on the left by E_2

$$E_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -b & 0 & 1 \end{bmatrix}$$

$$E_2 E_1 A = U^*$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -b & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ -a & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ a & a & a \\ b & b & a \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & a & 0 \\ 0 & b & a-b \end{bmatrix} .$$

c) Our last objective is to clear the $U^*_{3,2}$ entry.

We will multiply on the left by the matrix E_3 which corresponds to a subtraction of b/a times row 2 from row 3;

$$E_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -b/a & 1 \end{bmatrix}$$

$$E_3 E_2 E_1 A = U$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -b/a & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -b & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ -a & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ a & a & a \\ b & b & a \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & a & 0 \\ 0 & 0 & a-b \end{bmatrix} .$$

2. Subsequently, we have to isolate A by multiplying both sides of the equation on the left by the inverse product of our elimination matrices:

$$(E_3 E_2 E_1)^{-1}(E_3 E_2 E_1)A = (E_3 E_2 E_1)^{-1} U$$
$$\Rightarrow I_3 A = E_1^{-1} E_2^{-1} E_3^{-1} U .$$

The product of inverse elimination matrices will give us our lower triangular matrix L:

$$L = E_1^{-1} E_2^{-1} E_3^{-1}$$

$$\Rightarrow L = \begin{bmatrix} 1 & 0 & 0 \\ a & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ b & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & b/a & 1 \end{bmatrix}$$

$$\Rightarrow L = \begin{bmatrix} 1 & 0 & 0 \\ a & 1 & 0 \\ b & b/a & 1 \end{bmatrix}.$$

Therefore we have that $A = LU$, where

$$\Rightarrow \begin{bmatrix} 1 & 0 & 1 \\ a & a & a \\ b & b & a \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ a & 1 & 0 \\ b & b/a & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ 0 & a & 0 \\ 0 & 0 & a-b \end{bmatrix}.$$

$$\checkmark$$

Conclusion: The matrix A has an LU Decomposition for all real values of a and b; given a $\neq 0$.

CHAPTER 4

The Singular Value Decomposition (SVD)

4.1 Background

The singular value decomposition is widely regarded as the one of the most valuable matrix factorizations. SVD is efficiently computed for large matrices, and can be used to approximate a matrix to arbitrary precision using the Eckart-Young Low Rank Approximation Theorem which will be discussed briefly in the applications section.

The singular value decomposition takes the form

$$A = U\Sigma V^T$$

where

- U is an orthogonal matrix,
- Σ is a diagonal matrix and consists of singular values, i.e. σ_n (square roots of the eigenvalues of A), and
- V is an orthogonal matrix.

A singular value decomposition exists for any matrix, square or not.

If a matrix A is symmetric positive definite – $A = A^T$, its eigenvectors are orthogonal and consist of positive, real values – we are allowed to use a special case of the SVD where U = V = P, allowing for the familiar form

$$A = P\Lambda P^T$$

In section 1.3, we established that, given a symmetric nxn matrix, we can form a set of n eigenvectors.
It is known that the matrix A^TA is symmetric for a square matrix A and shares many important properties with A. These properties allow an SVD of A to be performed on the symmetric matrix A^TA, even if our original matrix A is not symmetric.

Calculation

The key part of finding the singular value decomposition is finding an orthonormal basis $\vec{v}$ for the row space of A where

$$A\begin{bmatrix} v_1 & v_2 & \dots & v_n \end{bmatrix} = \begin{bmatrix} \sigma_1 u_1 & \sigma_2 u_2 & \dots & \sigma_n u_n \end{bmatrix}$$

where $\vec{u}$ is an orthonormal basis for the column space of A, and σ_n are termed singular values.
After incorporating the nullspaces, and making $\vec{v}$ and $\vec{u}$ to be orthonormal bases for the entire space $\mathbb{R}^n$ the equation becomes

$$AV = U\Sigma$$

Because V is orthogonal, we can multiply both sides on the right by $V^{-1} = V^T$ to isolate A:

$$A = U\Sigma V^T$$

It wouldn't be wise to attempt to solve for U, V, and Σ simultaneously, so instead we will multiply both sides on the left by their transpose $A^T = V\Sigma^T U^T$ where Σ can be substituted for Σ^T since Σ is diagonal and U^{-1} can be substituted for U^T since U is orthogonal, to get the result

$$A^T A = V\Sigma U^{-1} U\Sigma V^T$$

$$\Rightarrow A^T A = V\Sigma^2 V^T$$

$$A^T A = V \begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \sigma_n^2 \end{bmatrix} V^T$$

We can find V by diagonalizing the symmetric positive definite matrix $A^T A$. The eigenvectors of $A^T A$ form the columns of V. Similarly, the eigenvalues of $A^T A$ form the diagonal elements of Σ^2 (σ_i^2), where σ_i are the corresponding positive square roots of λ_i.

To find U, we can either repeat this process with AA^T, or use the now known matrices A, V, and Σ to solve for U

$$AV = U\Sigma$$

$$\Rightarrow AV\Sigma^{-1} = U\Sigma\Sigma^{-1}$$

$$\Rightarrow AV\Sigma^{-1} = U.$$

4.2 A Concrete 2X2 Example

" ... a concrete life preserver thrown to students sinking in a sea of abstraction"
- W. Gottschalk

Determine the singular value decomposition of the matrix

$$A = \begin{bmatrix} 5 & 5 \\ -1 & 7 \end{bmatrix}.$$

- First, we must find $A^T A$:

$$A^T A = \begin{bmatrix} 5 & -1 \\ 5 & 7 \end{bmatrix} \begin{bmatrix} 5 & 5 \\ -1 & 7 \end{bmatrix}$$

$$\Rightarrow A^T A = \begin{bmatrix} 26 & 18 \\ 18 & 74 \end{bmatrix}.$$

- Next, let's find the eigenvalues of $A^T A$:

$$det\left(A^T A - \lambda I\right) = 0$$

$$det\left(\begin{bmatrix} 26-\lambda & 18 \\ 18 & 74-\lambda \end{bmatrix}\right) = 0$$

$$\Rightarrow (26-\lambda)(74-\lambda) - 18^2 = 0$$

$$\Rightarrow \lambda^2 - 100\lambda + 1600 = 0$$

$$\Rightarrow (\lambda - 20)(\lambda - 80) = 0$$

$$\lambda_1 = 20,$$

$$\lambda_2 = 80.$$

- After finding our eigenvalues of $A^T A$, we must find their associated eigenvectors to form V

1. $\lambda_1 = 20$

$$A^T A - 20I = \begin{bmatrix} 6 & 18 \\ 18 & 54 \end{bmatrix}$$

$$v_1 = \begin{bmatrix} -3 \\ 1 \end{bmatrix}.$$

Time to normalize v_1

$$v_1 = \begin{bmatrix} \frac{-3}{\sqrt{10}} \\ \frac{1}{\sqrt{10}} \end{bmatrix}.$$

2. $\lambda_2 = 80$

$$A^T A - 80I = \begin{bmatrix} -54 & 18 \\ 18 & -6 \end{bmatrix}$$

$$v_2 = \begin{bmatrix} 1 \\ 3 \end{bmatrix}.$$

Time to normalize v_2

$$v_2 = \begin{bmatrix} \frac{1}{\sqrt{10}} \\ \frac{3}{\sqrt{10}} \end{bmatrix}.$$

3. Thus,

$$V = \begin{bmatrix} \frac{-3}{\sqrt{10}} & \frac{1}{\sqrt{10}} \\ \frac{1}{\sqrt{10}} & \frac{3}{\sqrt{10}} \end{bmatrix}.$$

4. We'll need V^T later, so let's go ahead and find it:

$$\text{Since V is symmetric,}$$
$$V^T = V$$
$$V^T = \begin{bmatrix} \frac{-3}{\sqrt{10}} & \frac{1}{\sqrt{10}} \\ \frac{1}{\sqrt{10}} & \frac{3}{\sqrt{10}} \end{bmatrix}.$$

- Now let's use the eigenvalues of $A^T A$ to form Σ

$$\Sigma^2 = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} = \begin{bmatrix} 20 & 0 \\ 0 & 80 \end{bmatrix}$$
$$\Rightarrow \Sigma = \begin{bmatrix} \sqrt{20} & 0 \\ 0 & \sqrt{80} \end{bmatrix}$$
$$\Sigma = \begin{bmatrix} 2\sqrt{5} & 0 \\ 0 & 4\sqrt{5} \end{bmatrix}.$$

- Next we find U:

$$U = AV\Sigma^-1$$
$$U = \begin{bmatrix} 5 & 5 \\ -1 & 7 \end{bmatrix} \begin{bmatrix} \frac{-3}{\sqrt{10}} & \frac{1}{\sqrt{10}} \\ \frac{1}{\sqrt{10}} & \frac{3}{\sqrt{10}} \end{bmatrix} \begin{bmatrix} \frac{1}{2\sqrt{5}} & 0 \\ 0 & \frac{1}{4\sqrt{5}} \end{bmatrix}$$
$$U = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}.$$

- Finally, we can conclude our factorization of A:

$$A = U\Sigma V^T$$
$$A = \begin{bmatrix} 5 & 5 \\ -1 & 7 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 2\sqrt{5} & 0 \\ 0 & 4\sqrt{5} \end{bmatrix} \begin{bmatrix} \frac{-3}{\sqrt{10}} & \frac{1}{\sqrt{10}} \\ \frac{1}{\sqrt{10}} & \frac{3}{\sqrt{10}} \end{bmatrix}.$$
$$\surd$$

4.3 Applications

'The signal is the truth. The noise is what distracts us from the truth.'
- Nate Silver

Eckart-Young Theorem

Theorem 4.3.1. *If a matrix B has rank k then*

$$||A - A_k|| \leq ||A - B||.$$

This can be interpreted as saying that truncating A to rank k gives a better approximation to A than any other matrix B, given B is rank k or lower.

Consequences of the Eckart-Young Theorem

The Eckart-Young theorem is an important extension of the singular value decomposition in which a matrix may be approximated to arbitrary precision by truncating our matrices U, Σ, and V^T to a common rank r, of our choice. The truncated matrix A is not only a good approximation, it is the best rank r approximation to A possible.

In order to accurately approximate our original matrix A, our matrices must be arranged in a particular fashion. The diagonal entries of Σ, σ_i correspond to the column vectors $\vec{u_i}$, and the row vectors $\vec{v_i}^T$. The values of σ_i indicate the relative importance of their corresponding $\vec{u_i}$ and $\vec{v_i}^T$ in communicating the signal of A. Therefore, the more important a set of column vectors and row vectors are, the greater their analogous σ_i value will be. Once our matrices are in order, we can begin extracting signal from our original matrix.

1. U will be truncated to only include its first r columns,
2. Σ will be truncated to include its first r columns and r rows, becoming an rxr matrix,
3. V^T will be truncated to include its first r rows;

$$A \approx \begin{bmatrix} | & & | \\ \vec{u_1} & \dots & \vec{u_r} \\ | & & | \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_r \end{bmatrix} \begin{bmatrix} | & & | \\ \vec{v_1} & \dots & \vec{v_r} \\ | & & | \end{bmatrix}^T .$$

Image Compression

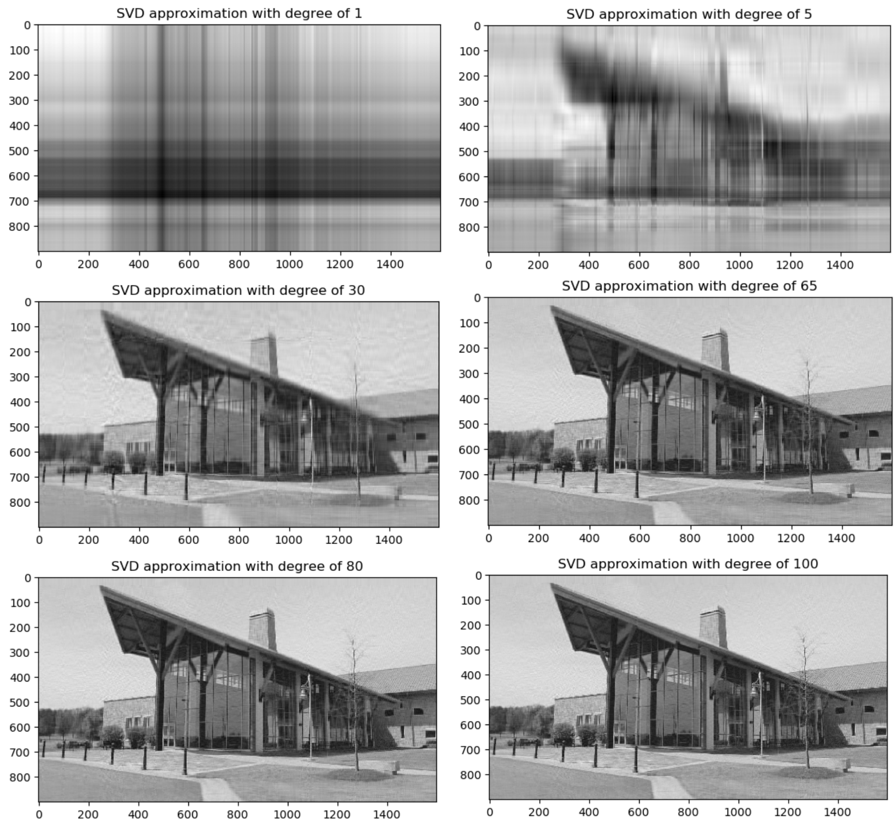
Every screen we see is composed of a grid of discrete components called pixels. Nearly every color we see on said screen is not the true color we believe it to be, but rather an approximate assemblage of the primary colors red, green, and blue. Each primary color in each pixel is given a saturation value typically ranging from 0 to 255 to determine its intensity. These combinations can be represented in hexadecimal by a six-digit hex color code.

A matrix can be thought of as a structure to organize and store numerical data. Because screens typically form a rectangular pattern of discrete values, matrices are a trivial abstraction for representing them. To simplify our picture model, let's take a greyscale image. The shades of grey are represented by the set of hex color codes in which the red, green, and blue values are all equal to one another. Thus, to represent a greyscale image, we can easily return to the comfortable world of integers by storing only the greyscale intensity.

The average desktop monitor resolution is 1980 pixels wide by 1080 pixels high. To represent a greyscale image of this value, we would need a matrix with 2,138,400 entries (note – if we wanted to represent a colored image, we would need three matrices of this size).

With the rise of social media sites such as Instagram, Facebook, and Tumblr, the number of images transmitted on a daily basis has skyrocketed. The amount of images on the internet as a whole is well into the hundreds of billions, if not trillions by now. By compressing the largest of these images to a standardized size, not only do these sites become easier to manage, but there are sizeable reductions in server loads, and hence carbon emissions.

Image Compression Example



Georgia Highlands Cartersville Campus Entrance

Collaborative Filtering

Collaborative filtering is a technique most commonly used by recommender systems to extract patterns and make predictions from the inputs of multiple agents. In companies such as Netflix or Amazon, collaborative filtering is used to try and determine which shows/products one user may like to view/purchase next based on their past preferences as well as the preferences of other users with similar taste. While this is not accomplished solely using SVD, the factorization seems to be a common denominator among many approaches to the problem.

Latent Semantic Indexing

Another technology with its roots in the singular value decomposition is Latent Semantic Indexing (LSI). LSI is a way for computers to process natural language and better understand the meaning behind the text it reads, rather than only the literal lexical arrangements. Latent Semantic Indexing uses the truncated singular value decomposition in addition to a few other methods to attempt to determine the word usage structure across documents despite variability in word choice. Because of this, LSI is able to see past synonymy (i.e. different words with the same meaning) and polysemy (i.e. one word with multiple meanings).

CHAPTER 5

References

[1] Anton, H., Rorres, C. (2019). Elementary Linear Algebra, 12th Ed. John Wiley Sons, Inc.

[2] Chowning, S. (2020). The Singular Value Decomposition. https://www.dam.brown.edu/drp/proposals/SamChowning.pdf

[3] Cline, A., Dhillion, I. (n.d.). Computation of the Singular Value Decomposition. https://www.cs.utexas.edu/users/inderjit/public_papers/HLA_SVD.pdf

[4] Eckart, C., Young, G.M. (1936). The approximation of one matrix by another of lower rank. Psychometrika, 1, 211-218.

[5] Gower, S. (2014). Netflix Prize and SVD. http://buzzard.pugetsound.edu/courses/2014spring/420projects/math420-UPS-spring-2014-gower-netflix-SVD.pdf

[6] Lay, D. C., Lay, S. R., Mcdonald, J. (1999). Linear Algebra and its Applications, 2nd Ed. Addison Wesley Publishing Company.

[7] Rosario, B. (n.d.). Latent Semantic Indexing: An Overview Latent Semantic Indexing https://www.cse.msu.edu/~cse960/Papers/LSI/LSI.pdf

[8] Strang, G. (2010). Linear Algebra. Massachussetts Institute of Technology: MIT OpenCourseWare, https://ocw.mit.edu. License: Creative Commons BY-NC-SA.