

Linear Algebra Primer

David Doria
daviddoria@gmail.com

Wednesday 3rd December, 2008

Contents

1	Why is it called “Linear Algebra”?	4
2	What is a Matrix?	4
2.1	Input and Output	4
2.1.1	Example	4
3	Definitions	4
3.1	Linear Independence	4
3.2	Vector Space	4
3.3	Basis	5
4	Matrix Terminology	5
5	Gaussian Elimination	5
5.1	Idea	5
5.2	Definitions	6
5.3	Elementary Row Operations	6
5.4	Procedure	6
5.5	Example	6
5.6	When Elimination Breaks	7
5.6.1	Case 1	7
5.6.2	Case 2	7
6	Finding the Inverse (Gauss Jordan Elimination)	7
7	Matrix Factorizations	8
7.1	LU Factorization	8
7.1.1	Example	8
8	QR Factorization	8
8.1	Example	9
9	RQ Factorization	9

10 Special Matrices	9
10.1 Permutation Matrices	9
10.1.1 Properties	9
10.1.2 Examples	9
10.2 Projection Matrices	9
10.2.1 Example	9
10.3 Symmetric Matrices	10
10.4 Positive Definite (PD) Matrices	10
10.5 Similar Matrices	10
10.5.1 Properties	10
10.6 Orthogonal Matrices	10
10.6.1 Properties	10
11 The Four Subspaces	11
11.1 Column Space $C(A)$	11
11.1.1 Example	11
11.2 Nullspace $N(A)$	11
11.2.1 Procedure	11
11.2.2 Example	12
11.3 Complete Solution to $Ax=b$	12
11.3.1 Example	12
11.4 Row Space	12
11.4.1 Connection to the Nullspace	13
11.5 Left Null Space	13
11.6 Summary	13
12 Least Squares	13
12.1 Example	14
13 Gram Schmidt Orthogonalization	14
13.1 Example	15
14 Determinants	15
14.1 Idea	15
14.2 Procedure	15
14.2.1 2×2	15
14.2.2 3×3	15
14.3 Rules	16
15 Eigenvalues and Eigenvectors	16
15.1 Procedure	16
15.1.1 Finding Eigenvalues	16
15.1.2 Finding Corresponding Eigenvectors	16
15.2 Properties	17
15.3 Example	17
15.4 Repeated Eigenvalues/Generalized Eigenvectors	17
15.4.1 Example	17
15.5 Properties	17
15.6 Diagonalization	17

15.6.1 Exponential of a Matrix	18
16 Singular Value Decomposition (SVD)	18
16.1 U	18
16.2 Sigma	18
16.3 V	18
16.4 Geometric Interpretation	19
16.5 Mini-Proof	19
16.6 Example	19
16.7 Solving Linear Equations	19
17 Summary	19
17.1 Finding the Rank of a Matrix	19
18 Practical Matrix Operations	20

1 Why is it called “Linear Algebra”?

Linear algebra is the study of linear functions/equations. A linear equation is one in which each term is either a constant or the product of a constant and the first power of a single variable. A linear function is one which obeys the two properties $f(x) + f(y) = f(x + y)$ and $f(\alpha x) = \alpha f(x)$. These equations can always be written in the form

$$ax_3 + bx_2 + cx_1 + x_0 = 0$$

If an equation cannot be written in this form, it is not linear.

The name “linear” comes from the two variable case $x_2 + x_1 + c = 0$, because the set of all solutions to this equation forms a line. In general, the solution to an n th order linear equation is a hyperplane of dimension $n - 1$.

2 What is a Matrix?

A matrix is simply a function. We are used to the notation $y = f(x)$, where the input x is operated on by the function f to produce a value y . A matrix is just one particular type of function. The notation used is $Fx = y$, where F is the matrix, but generally captial letters at the beginning of the alphabet are used to represent matrices, so you will often see $Ax = b$ instead.

2.1 Input and Output

A matrix is a function that takes a vector as input and produces a new vector as output. The input and output vectors are constrained to be a particular size, given by the number of columns and number of rows of the matrix, respectively. The size of a matrix is stated as $m \times n$ (read “ m by n ”), where m is the number of rows, and n is the number of columns.

2.1.1 Example

A 4×3 matrix takes vectors of length 3 as input and produces vectors of length 4.

3 Definitions

3.1 Linear Independence

The only solution to $Ax = 0$ is $x = 0$. Cannot combine any of the vectors with any weights to obtain 0, ie. vectors are not multiples of each other and can not be formed by any combination of the other vectors.

3.2 Vector Space

A set of vectors which has defined scalar multiplication and vector addition. A vector space must contain the zero element so that it is “closed” under addition. This means that if you add two vectors in the space, the result must also be in the space. This is impossible if it does not contain 0 because if you add a and $-a$, you get 0.

3.3 Basis

n independent vectors that span a space exactly n dimensional. If there are $n + 1$ or more vectors in an n dimensional space, then there are “too many”, which means they cannot be linearly independent. If there are “not enough” (less than n) vectors, then they cannot span the space.

4 Matrix Terminology

- Minor: The determinant of a submatrix.
- Principal minor: If a minor came from a submatrix where the list of indices of the rows used is the same as the list of indices of the columns used.
- Leading principal minor: a minor that came from the top left corner of a matrix (a principal minor where the index list that is shared by rows and columns is $1,2,3,\dots, k$).
- Positive definite: All leading principal minors are positive. This means for any vector x , $x^T Ax > 0$.
- Positive semi-definite: All leading principal minors are positive. This means for any vector x , $x^T Ax \geq 0$.
- Negative definite: All k th order leading principal minors are negative if k is odd and positive if k is even. This means for any vector x , $x^T Ax < 0$.
- Negative semi-definite: All k th order leading principal minors are negative if k is odd and positive if k is even. This means for any vector x , $x^T Ax \leq 0$.

5 Gaussian Elimination

5.1 Idea

Convert a matrix into row echelon form using elementary row operations. What this does is find a different set of equations with the same solution.

This is used to:

- Solve a system of linear equations
- Find the rank of a matrix
- Calculate the inverse of a matrix

A system of equations is very easy to solve after it has been put in row echelon form because the last equation is immediately a solution (there is only one variable), and each row above it has only one extra variable. Solving by this process is called back substitution.

5.2 Definitions

- Row Echelon Form:
 - A matrix where the leading coefficient of a row is always strictly to the right of the leading coefficient of the row above it.
 - All nonzero rows are above any rows of all zeros.
- Reduced Row Echelon Form: A row echelon matrix where the leading coefficients of each row is the only nonzero entry in its column.

5.3 Elementary Row Operations

These actions can be performed on the matrix in any order

- Subtract a multiple of one row from another row.
- Exchange two rows.
- Multiply a row by a scalar.

5.4 Procedure

- Use elementary row operations to produce an echelon matrix (all entries below the diagonal are 0).
- The first nonzero entry in each row is called a pivot. The number of pivot columns is called the rank of A .

5.5 Example

??? May not be correct Consider the system of equations:

$$\begin{aligned}2x + 4y - 2z &= 2 \\4x + 9y - 3z &= 9 \\-2x - 3y + 7z &= 10\end{aligned}$$

Step 1: $r_2 = r_2 - 2r_1$ ($l_{2,1} = 2$)

$$\begin{aligned}2x + 4y - 2z &= 2 \\0x + 1y + 1z &= 4 \\-2x - 3y + 7z &= 10\end{aligned}$$

Step 2: $r_3 = r_3 - (-r_1)$ ($l_{3,1} = -1$)

$$\begin{aligned}2x + 4y - 2z &= 2 \\0x + 1y + 1z &= 4 \\0x + y + 5z &= 12\end{aligned}$$

Step 3: $r_3 = r_3 - (r_2)$ ($l_{3,2} = 1$)

$$2x + 4y - 2z = 2$$

$$0x + 1y + 1z = 4$$

$$0x + 0y + 4z = 8$$

On the i th step, both sides of the equation are multiplied by an elimination matrix: $E_iAx = E_ib$

$$\begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 2 & 4 & -2 \\ 4 & 9 & -3 \\ -2 & -3 & 7 \end{pmatrix} x = \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ 8 \\ 10 \end{pmatrix}$$

So the complete elimination procedure is: $E_3E_2E_1Ax = E_3E_2E_1b$

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 2 & 4 & -2 \\ 4 & 9 & -3 \\ -2 & -3 & 7 \end{pmatrix} x = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ 8 \\ 10 \end{pmatrix}$$

5.6 When Elimination Breaks

5.6.1 Case 1

$$x - 2y = 1$$

$$3x - 6y = 11$$

After elimination, we obtain:

$$x - 2y = 1$$

$$0x + 0y = 8$$

This can never be true, so there is no solution!

5.6.2 Case 2

$$x - 2y = 1$$

$$3x - 6y = 3$$

After elimination, we obtain:

$$x - 2y = 1$$

$$0x + 0y = 0$$

This is always true no matter the choice of x and y , so there are infinite solutions!!

6 Finding the Inverse (Gauss Jordan Elimination)

Augment the matrix by the an identity matrix. Perform Gaussian elimination, and then continue to produce zeros above the diagonal as well. This is called reduced row echelon form. Then divide each row by its pivot. The result is

$$[A|I] \rightarrow [I|A^{-1}]$$

7 Matrix Factorizations

7.1 LU Factorization

$$A = LU$$

- L is a Lower triangular matrix
- U is an Upper triangular matrix

During elimination, the factors $l_{i,j}$ are exactly the i, j th entries of the matrix L .

With elimination we obtain $EA = U$. From this we can see that $A = E^{-1}U$. This is exactly the LU factorization.

$$U = EA$$
$$L = E^{-1}$$

Of course if there is more than one elimination step the inverses of the elimination matrices must be multiplied in reverse order to get back A. That is:

$$E_2E_1A = U$$
$$L = E_1E_2$$

7.1.1 Example

??? May not be correct From the Gaussian elimination example:

$$A = \begin{pmatrix} 2 & 4 & -2 \\ 4 & 9 & -3 \\ -2 & -3 & 7 \end{pmatrix}$$
$$E = E_3E_2E_1 = \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ -1 & -1 & 1 \end{pmatrix}$$
$$U = EA =$$

8 QR Factorization

$$A = QR$$

- Q is an orthogonal matrix.
- R is a Right (upper) triangular matrix.

Let A be defined by column vectors:

$$A = [\vec{a} \ \vec{b} \ \vec{c}]$$

$$A = QR = [\vec{q}_1 \ \vec{q}_2 \ \vec{q}_3] \begin{pmatrix} q_1^T a & q_1^T b & q_1^T c \\ 0 & q_2^T b & q_2^T c \\ 0 & 0 & q_3^T c \end{pmatrix}$$

Q can be found by using the Gram-Schmidt process on the columns of A. Then $R = Q^T A$. R is upper triangular because the dot product with a vector with any of the vectors before it is 0 (they are orthogonal). The diagonal entries of R are the lengths of a, b, and c, respectively.

8.1 Example

$$A = \begin{pmatrix} 1 & 2 & 3 \\ -1 & 0 & -3 \\ 0 & -2 & 3 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ 0 & -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{pmatrix} \begin{pmatrix} \sqrt{2} & \sqrt{2} & \sqrt{18} \\ 0 & \sqrt{6} & -\sqrt{6} \\ 0 & 0 & \sqrt{3} \end{pmatrix}$$

9 RQ Factorization

$$A = RQ$$

The RQ factorization is similar to QR factorization, except the Gram-Schmidt is applied to the rows of A instead of the columns of A .

10 Special Matrices

10.1 Permutation Matrices

Has the rows of I in any order (ie. each row has exactly one 1, and each column has exactly 1 one). Left multiplying by a permutation matrix re-orders the rows of A , while right multiplying by a permutation matrix re-orders the columns of A .

10.1.1 Properties

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

10.1.2 Examples

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} P = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} PA = AP =$$

10.2 Projection Matrices

For any subspace, a matrix can be constructed such that when it multiplies any vector, the vector is taken into the subspace. For a one dimensional subspace (a vector),

$$P = \frac{aa^T}{a^T a}$$

For a higher dimensional subspace,

$$P = A(A^T A)^{-1} A^T$$

10.2.1 Example

To project any vector on to the vector $a = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}$, left multiply by

$$P = \frac{aa^T}{a^T a} = \frac{1}{9} \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} (1 \ 2 \ 2) = \frac{1}{9} \begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 4 \\ 2 & 4 & 4 \end{pmatrix}$$

10.3 Symmetric Matrices

A symmetric matrix A has the following properties.

- $A = A^T$
- The eigenvalues of A are real.
- The eigenvectors of A are orthonormal. This means that A can be diagonalized with an orthogonal matrix ($A = Q\Lambda Q^T$)

10.4 Positive Definite (PD) Matrices

If a matrix A has one of these properties, it has them all. It is called Positive Definite because $x^T Ax > 0$ when $x \neq 0$.

- A is symmetric.
- All of the eigenvalues are > 0 .
- All leading principal minors are > 0 .
- All pivots are positive.

$$\det(A) > 0$$

10.5 Similar Matrices

B is similar to A if it can be written

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

10.5.1 Properties

B and A have the same eigenvalues.

10.6 Orthogonal Matrices

An orthogonal matrix is generally denoted Q .

10.6.1 Properties

- For any shape Q : $Q^T Q = I$
- For square Q : $Q^T Q = Q Q^T = I$. Only when square is it called an orthogonal matrix (vectors must also be normalized).
- Multiplication by an orthogonal matrix does not change the length of a vector: $\|Qx\| = \|x\|$
- $\det(Q) = \pm 1$
- Dot product of any two columns is 0. This is another way of saying that the columns of Q are orthogonal.
- Each column is length 1.

11 The Four Subspaces

For the following discussion, let A be an m by n matrix.

11.1 Column Space $C(A)$

- The column space of A is all linear combinations of the columns of A . This is all possible vectors Ax . The column space can be written

$$C(A) = Ax = x_1A_1 + x_2A_2 + \dots + x_nA_n$$

where A_n is the n th column of A .

- $Ax = b$ is solvable only if b is in $C(A)$.
- Pivot columns are a basis for $C(A)$.
- The column space is a subspace of R^m ($C(A) \subseteq R^m$).
- After row reduction, the column space is not the same! ($C(R) \neq C(A)$).
- The dimension of the column space is called the rank.
- The column space is also called the range. With a function $y = f(x)$, the range is all possible values that y can obtain. The column space is exactly analogous. The range $C(A)$ is the set of vectors y that $y = Ax$ can obtain.

11.1.1 Example

$$A = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$$

The columns are not linearly independent, so $C(A)$ is 1 dimensional. It is spanned by $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$.

11.2 Nullspace $N(A)$

- All solutions to $Ax = 0$.
- The nullspace is a subspace of R^n ($N(A) \subseteq R^n$).
- The dimension of the nullspace is called the nullity.
- The nullspace of the row reduced matrix is the same as the nullspace of A ($N(R) = N(A)$).

11.2.1 Procedure

To find the special solutions(x_s): Set one free variable (x_f) at a time to 1 (set the rest to 0) and solve the system $Ax = 0$. Each solution is a vector in $N(A)$. We do this to obtain linearly independent vectors for a basis of the nullspace.

11.2.2 Example

$$A = \begin{pmatrix} 1 & 1 & 2 & 3 \\ 2 & 2 & 8 & 10 \\ 3 & 3 & 10 & 13 \end{pmatrix}$$

Perform elimination to obtain:

$$A = \begin{pmatrix} 1 & 1 & 2 & 3 \\ 0 & 0 & 4 & 4 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Pivot variables are x_1 and x_3 since columns 1 and 3 contain pivots.

Free variables are x_2 and x_4 since columns 2 and 4 do not contain pivots.

Find Special solutions:

Set $x_2 = 1, x_4 = 0$ and solve the system ($x_1 + x_2 + 2x_3 + 3x_4 = 0, 4x_3 + 4x_4 = 0 \rightarrow x_1 + 1 + 2x_3 + 0 = 0, 4x_3 + 0 = 0$) The result is $x_3 = 0, x_2 = 1$. Since x_2 is the free variable that was set to 1, this special vector times any value of x_2 is a special solution.

Set $x_4 = 1, x_2 = 0$ and solve the system again. This time the result is $x_3 = -1, x_1 = -1$. Any value of x_4 times this vector is a special solution.

The special solution is : $x_2 \begin{pmatrix} -1 \\ 1 \\ 0 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} -1 \\ 0 \\ -1 \\ 1 \end{pmatrix}$

The nullspace matrix is has the special solutions as its columns: $N = \begin{pmatrix} -1 & -1 \\ 1 & 0 \\ 0 & -1 \\ 0 & 1 \end{pmatrix}$

These vectors are a basis for the nullspace.

11.3 Complete Solution to $Ax=b$

The complete solution only involves the column space and the nullspace. Geometrically, the solution to $Ax = b$ is simply a translation of the solution to $Ax = 0$.

Particular solutions(x_p):

Set all the free variables to 0 and solve $Ax = b$.

The complete solution is $x = x_p + x_{f_1}x_{s_1} + \dots + x_{f_n}x_{s_n}$

11.3.1 Example

Taking the same matrix from the nullspace example, we set $x_2 = x_4 = 0$ and solve the system.

11.4 Row Space

- The row space is the span of the rows of A . It is identically the column space of A^T , so it is denoted $C(A^T)$. The pivot rows are a basis for the row space.
- The row space is not changed during row reduction ($C(R^T) = C(A^T)$). This is because we are simply producing linear combinations of the rows. Since the row space is spanned by the initial rows, it is also spanned by linear combinations of them.

11.4.1 Connection to the Nullspace

Ax can also be written $\begin{pmatrix} r_1 \cdot x \\ r_2 \cdot x \\ \vdots \\ r_n \cdot x \end{pmatrix}$ where r_n is the n th row of A . From this we see another interpretation of the nullspace. $Ax = 0$ only if the dot product of each row with x is 0. That means each row is orthogonal to x .

11.5 Left Null Space

- $N(A^T) \subseteq R^m$
- $A^T y = 0$
- Contains the error of the least squares solution.
- Orthogonal to $C(A)$.

11.6 Summary

The number of solutions can be summarized with the following table:

Shape	Shape Description	Number of Solutions
$r = m = n$	square	1
$r = m, r < n$	short and fat	∞
$r < m, r = n$	tall and skinny	0 or 1
$r < m, r < n$	degenerate	0 or ∞

The relationship of the four subspaces can be summarized with the following table:

Space	\perp Complement	Subspace of	Dimension
$C(A^T)$	$N(A)$	R^n	r
$C(A)$	$N(A^T)$	R^m	r
$N(A)$	$C(A^T)$	R^n	$n - r$
$N(A^T)$	$C(A)$	R^m	$m - r$

12 Least Squares

When $Ax = b$ is not solvable, use $A^T Ax = A^T b$, which is $x = (A^T A)^{-1} A^T b$. This minimizes $\|Ax - b\|^2$.

This comes from knowing that the minimum error is achieved when the error is orthogonal to the estimate. We can write

$$A^T(b - A\hat{x}) = 0$$

by expanding this, we obtain the equation $A^T b = A^T A\hat{x}$. The solution to this is the least squares solution.

12.1 Example

!!!Test

$$x + y + z = -2x$$

$$x + y + z = 6$$

$$x + \quad + z = 1$$

$$x + y + z = -2$$

$$x + y + z = 6$$

$$x + \quad + z = 1$$

Consider the equations:

$$x_1 + 4x_2 = -2$$

$$x_1 + 2x_2 = 6$$

$$2x_1 + 3x_2 = 1$$

They are written in matrix form as:

$$A = \begin{pmatrix} 1 & 4 \\ 1 & 2 \\ 2 & 3 \end{pmatrix} \quad b = \begin{pmatrix} -2 \\ 6 \\ 1 \end{pmatrix}$$

We find $A^T A = \begin{pmatrix} 6 & 12 \\ 12 & 29 \end{pmatrix}$ and $A^T b = \begin{pmatrix} 6 \\ 7 \end{pmatrix}$.

Therefore the least squares solution is $x^* = \begin{pmatrix} 3 \\ -1 \end{pmatrix}$.

13 Gram Schmidt Orthogonalization

Consider three vectors: a , b , and c . The goal is to produce orthonormal vectors which span the same space as the original vectors. The first orthogonal vector is chosen to be exactly vector a (this is an arbitrary choice, the procedure can start with any of them). A vector orthogonal to a can be produced by projecting b onto a and subtracting the result from b . The next orthogonal vector can be found by subtracting the projection of c onto a and the projection of c onto b from c . These three vectors are orthogonal, so we simply normalize them to obtain our orthonormal basis.

$$\tilde{q}_1 = a$$

$$\tilde{q}_2 = b - \text{proj}_{q_1} b = b - \frac{q_1 q_1^T}{q_1^T q_1} b$$

$$\tilde{q}_3 = c - \text{proj}_{q_1} c - \text{proj}_{q_2} c = c - \frac{q_1 q_1^T}{q_1^T q_1} c - \frac{q_2 q_2^T}{q_2^T q_2} c$$

$$q_1 = \frac{\tilde{q}_1}{\|\tilde{q}_1\|}$$

$$q_2 = \frac{\tilde{q}_2}{\|\tilde{q}_2\|}$$

$$q_3 = \frac{\tilde{q}_3}{\|\tilde{q}_3\|}$$

During these operations, it is convenient to notice that if we compute the actual projection matrix for each vector, it results in a big matrix multiplication, ie. $q_2 = b - \text{proj}_{q_1} b = b - \frac{q_1 q_1^T}{q_1^T q_1} b$; the denominator is a scalar. The numerator becomes a 3x3 matrix if we perform $(q_1 q_1^T) b$. However, if we notice $q_1^T b$ is a scalar, this can be performed first, resulting in a vector times a scalar $q_1 (q_1^T b)$. This is clearly much easier.

13.1 Example

Consider the three vectors

$$a = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \quad b = \begin{pmatrix} 2 \\ 0 \\ -2 \end{pmatrix} \quad c = \begin{pmatrix} 3 \\ -3 \\ 3 \end{pmatrix}$$

The solution is:

$$\tilde{q}_1 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \quad \tilde{q}_2 = \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix} \quad \tilde{q}_3 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

Where $q_1 = \frac{\tilde{q}_1}{\sqrt{2}}$, $q_2 = \frac{\tilde{q}_2}{\sqrt{6}}$, and $q_3 = \frac{\tilde{q}_3}{\sqrt{3}}$.

14 Determinants

14.1 Idea

- Determinants are a way to determine if a system of linear equations has a unique solution. If the determinant is 0, the matrix is singular.
- The geometric interpretation is to take a unit cube in R^n and multiply each of its corners by A . The area of the resulting volume (a parallelepiped) is equal to the determinant of A . In R^2 , $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ brings a square with corners $(0,0)$, $(1, 0)$, $(0, 1)$, $(1, 1)$ to a parallelogram with corners $(0, 0)$, (a, c) , $(a+b, c+d)$, (b, d) .
- The determinant function associates a scalar value with a matrix.

14.2 Procedure

14.2.1 2x2

$$\det \left(\begin{bmatrix} a & b \\ c & d \end{bmatrix} \right) = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

$$A^{-1} = \frac{1}{\det(A)} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

14.2.2 3x3

There are two methods:

Method 1: Augment

$$\left(\begin{array}{ccc|cc} a & b & c & a & b \\ d & e & f & e & e \\ g & h & i & g & h \end{array} \right)$$

Calculate the sum of the products of the right diagonals and then subtract the sum of the products of the left diagonals.

Method 2: Cofactors

$$a \det \begin{pmatrix} e & f \\ h & i \end{pmatrix} - b \det \begin{pmatrix} d & f \\ g & i \end{pmatrix} + c \det \begin{pmatrix} d & e \\ g & h \end{pmatrix}$$

14.3 Rules

1. $\det(I) = 1$
2. Sign of determinant changes when rows are exchanged.
3. Subtracting a multiple of one row from another leaves det unchanged
4. If A is triangular, the product of the diagonal is the determinant. ($\det(A) = \prod_i a_{ii}$)
5. If A is singular, $\det(A) = 0$
6. If A is invertible, $\det(A) \neq 0$
7. $\det(AB) = \det(A) \det(B)$
8. $\det(A^T) = \det(A)$

15 Eigenvalues and Eigenvectors

Eigenvectors are vectors which when multiplied by a matrix do not rotate, but simply are scaled.

$$Ax = \lambda x$$

The vector which is not rotated is x and the value λ is the amount that x is stretched when multiplied by A .

15.1 Procedure

15.1.1 Finding Eigenvalues

Solve $Ax = \lambda x$,

$$Ax - \lambda x = (A - \lambda I)x = 0$$

This is called the characteristic equation. For there to be a non-trivial solution, $(A - \lambda I)$ must not be invertible. We assume this is true by saying $\det(A - \lambda I) = 0$ and then solving for λ .

15.1.2 Finding Corresponding Eigenvectors

To find the eigenvector corresponding to the i th eigenvalue, solve $(A - \lambda_i I)x = 0$

15.2 Properties

- Geometric Multiplicity (GM) = The nullspace of $A - \lambda_n I$ is called the eigenspace of λ_n . The dimension of the eigenspace is called the geometric multiplicity.
- Algebraic Multiplicity (AM) = The number of times an eigenvalue is repeated as a root of the characteristic equation.
- If $GM < AM$, A is not diagonalizable.
- When a matrix is raised to a power n , this is the same as applying the matrix n times. Therefore, when it acts on an eigenvector v , each application of A stretches v by the associated eigenvalue. This means the eigenvectors of A^n are λ^n .

15.3 Example

15.4 Repeated Eigenvalues/Generalized Eigenvectors

If the dimension of the nullspace of $A - \lambda_n I$ is less than the multiplicity of the eigenvalue, we must find generalized eigenvectors and the best we can do is $A = SJS^{-1}$ (we can't diagonalize the matrix, we can only write it as a Jordan form).

We find these vectors by solving:

$$(A - \lambda I)x_k = x_{k-1}$$

where $x_0 = 0$.

This is equivalent to saying

$$(A - \lambda I)^k x_k = 0$$

15.4.1 Example

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

The eigenvalues of A are $\lambda_1 = 1$ and $\lambda_2 = 1$. We find one eigenvector by solving $(A - \lambda_1 I)x_1 = (A - I)x_1 = 0$. We find the eigenvector $x_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. We then solve $(A - \lambda_2 I)x_2 = (A - I)x_2 = x_1$.

We obtain $x_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. This is a generalized eigenvector.

15.5 Properties

- $\prod_i \lambda_i = \det(A)$
- $\sum_i \lambda_i = \text{trace}(A)$ (The trace is $\text{tr}(A) = \sum_i A_{ii}$)

15.6 Diagonalization

Let S be the matrix with the eigenvectors of A as its columns. Λ is the matrix with the eigenvalues of A on the diagonal.

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

This is very helpful in raising matrices to powers. If A is in the form $A = S\Lambda S^{-1}$ then $A^2 = S\Lambda S^{-1}S\Lambda S^{-1} = S\Lambda^2 S^{-1}$ and in general $A^n = S\Lambda^n S^{-1}$.

If the matrix cannot be diagonalized, then the SVD must be used.

15.6.1 Exponential of a Matrix

We wish to find

$$e^{At}$$

Look at the Taylor series expansion of e^x around 0.

$$e^x = 1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3 + \dots$$

Substitute $x = At$ to obtain

$$e^{At} = I + At + \frac{1}{2}(At)^2 + \frac{1}{6}(At)^3 + \dots$$

Substitute the diagonalized version of A ($S\Lambda S^{-1}$)

$$e^{At} = I + S\Lambda S^{-1}t + \frac{1}{2}(S\Lambda S^{-1}t)^2 + \dots = I + tS\Lambda S^{-1} + \frac{1}{2}(t\Lambda)^2 S S^{-1} S S^{-1} + \dots$$

The same series appears again in the middle of the expression.

$$e^{At} = S e^{\Lambda t} S^{-1}$$

It's eigenvalues are $e^{\lambda t}$.

16 Singular Value Decomposition (SVD)

We want to find an orthogonal matrix that will diagonalize A . This is impossible, so we must find an orthogonal basis for the column space and an orthogonal basis for the row space. We also want $Av_1 = \sigma_1 u_1$ and $Av_2 = \sigma_2 u_2$, or $AV = U\Sigma$.

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

Let A be an m by n matrix.

16.1 U

U is a m by m orthonormal matrix. The columns of U are the eigenvectors of AA^T . These are called the "left singular vectors". The left singular vectors corresponding to the non-zero singular values of A (there are r of them) span $C(A)$. The columns of U which correspond to zero singular values (there are $m - r$ of them) span the left null space of A .

16.2 Sigma

Σ is a m by n diagonal matrix. Σ is the matrix with σ_i on the diagonal. The eigenvalues of AA^T and $A^T A$ are σ_i^2 . The rank is equal to the number of non-zero singular values.

16.3 V

V is a n by n orthonormal matrix. The columns of V are eigenvectors of $A^T A$. These are called the "right singular vectors". The columns of V which correspond to non-zero singular values (there are r of them) span the row space of A . The right singular values corresponding to the zero singular values (there are $n - r$ of them) span the $N(A)$.

16.4 Geometric Interpretation

Let x be a vector. When x is multiplied by V^T , this is simply a change of basis, so x is now represented in a new coordinate system, where the basis vectors are the columns of V . We will call this resulting vector x_v .

When x_v is multiplied by Σ , it is scaled in each dimension (now the columns of V) by the corresponding singular values of A . We will call this resulting vector x_s .

When x_s is multiplied by U , it is simply rotated back to the original coordinate system.

??? A takes a vector x and computes the coefficients of x along the input directions v_1, \dots, v_n . The input vector space is R^n because the matrix must be multiplied by an $n \times n$ Something matrix. Then it scales the vector by the singular values. Then it writes the vector as a combination of u_1, \dots, u_n .

Uses a different basis for row and column space.

16.5 Mini-Proof

To show that the eigenvalues of AA^T and $A^T A$ are the singular values of A , start with $Av = \sigma u$ and $A^T u = \sigma v$. Solve for u in the first equation and substitute into the second to obtain $A^T Av = \sigma^2 v$. You can do the reverse (solve for v in the second equation and substitute into the first) to obtain $AA^T u = \sigma^2 u$.

Using this, we can now show that the SVD diagonalizes A . Start with $A^T Av_i = \sigma_i^2 v_i$ (from above). Multiply by v_i^T : $v_i^T A^T Av_i = \sigma_i^2 v_i^T v_i \rightarrow \|Av_i\|^2 = \sigma_i^2$, so $\|Av_i\| = \sigma_i$. This was from seeing $(v_i^T A^T)(Av_i)$ is a vector times its transpose. Then again start with $A^T Av_i = \sigma_i^2 v_i$, but now multiply by A : $AA^T Av_i = \sigma_i^2 Av_i \rightarrow u_i = Av_i / \sigma_i$. This shows that Av_i is an eigenvector of AA^T .

16.6 Example

$$\begin{aligned} A &= \begin{pmatrix} .96 & 1.72 \\ 2.28 & .96 \end{pmatrix} \\ &= U\Sigma V^T = \begin{pmatrix} .6 & -.8 \\ .8 & .6 \end{pmatrix} \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} .8 & .6 \\ .6 & -.8 \end{pmatrix}^T \end{aligned}$$

We can see that the columns of U and V are unit length and mutually orthogonal. ex. $.6^2 + .8^2 = 1$ $u_1 \cdot u_2 = 0$

16.7 Solving Linear Equations

17 Summary

Now that we have an introduction to the topic, below is a useful summary that can be used to quickly answer common question.

17.1 Finding the Rank of a Matrix

- Gaussian elimination.
- If the matrix is square, find the number of non zero eigenvalues.
- If the matrix is not square, find the number of non zero singular values.

18 Practical Matrix Operations

- Inverse: $(AB)^{-1} = B^{-1}A^{-1}$
- Transpose: $(AB)^T = B^T A^T$
- Derivative: $\frac{d}{dx} Ax = A$
- Derivative of quadratic form: $\frac{d}{dx} x^T Ax = 2Ax$