Matrix Methods Notes

William Farmer

December 17, 2013

Contents

List of Figures

List of Equations

1 Matrices and Systems of Linear Equations

Up until this point we've been solving systems of linear equations through fiddling with them (solving for different variables, etc.) until we get an answer. Using matrices we can solve them a lot more effectively. Not only that, but any process we use will turn the matrix into an equivalent system of equations, i.e., one that has the same solutions.

We can have systems of linear equations represented in matrices, and if all equations are equal to zero, the system is homogeneous. The solution is defined as the point in $\mathbb{R}^n$ whose coordinates solve the system of equations.

We have a couple of methods to solve systems of linear equations when they are in matrix form, but first we need to define a couple different terms and operations.

1.1 Augmented Matrix

An augmented matrix is where two different matrices are combined to form a new matrix.

$$[\mathbf{A}|\mathbf{b}] = \left[\begin{array}{cccc|c} A_{11} & A_{12} & \cdots & A_{1m} & b_1 \\ A_{21} & A_{22} & \cdots & A_{2m} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ A_{n1} & A_{n2} & \cdots & A_{nm} & b_n \end{array}\right] \tag{1}$$

This is usually used to show the coefficients of the variables in a system of equations as well as the constants they are equal to.

1.2 Elementary Row Operations

We have a couple of different options to manipulate augmented matrices, which are as follows.

- Interchange row i and i

$$R_i^* = R_j, R_j^* = R_i$$

- Multiply row i by a constant.

$$R_i^* = cR_i$$

- Leaving j untouched, add to i a constant times j.

$$R_i^* = R_i + cR_j$$

These are handy when dealing with matrices and trying to obtain Reduced Row Echelon Form (1.6).

1.3 Gaussian Elimination

Our first method for solving matrices is to use Gaussian Elimination.

Our end goal with this strategy is to get to an triangular matrix or triangular form, which is easy to solve through back substitution.

$$\left(\begin{array}{ccc|c} 1 & 2 & 1 & 2 \\ 0 & 2 & -1 & 3 \\ 0 & 0 & \frac{5}{2} & 3 \end{array}\right)$$

The corresponding linear system has vector form

$$U\mathbf{x} = \mathbf{c}$$

The coefficient matrix U is upper triangular.

The method of using solely Elementary Row Operation 1 is called regular Gaussian Elimination. A square matrix A is called regular if the algorithm successfully reduces it to upper triangular form U with all non-zero pivots.

1.4 Pivoting and Permutations

Besides the above Elementary Row Operations, we also have pivoting at our disposal (which if you'll notice, is the same as Elementary Row Operation 1).

1.4.1 Pivoting

Definition 1. A square matrix is called nonsingular if it can be reduced to upper triangular form with all non-zero elements on the diagonal, the pivots, by elementary row operations.

A singular square matrix cannot be reduced by such operations.

Theorem 1. *A linear system $A\mathbf{x} = \mathbf{b}$ has a unique solution for every choice of right hand side $\mathbf{b}$ if and only if its coefficient matrix A is square and nonsingular.*

1.4.2 Permutations

Definition 2. A Permutation Matrix is a matrix obtained from the identity matrix by any combination of row interchanges.

Essentially just a method to change matrices.

There are six different 3×3 permutation matrices:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$
$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \tag{2}$$

1.4.3 Permuted LU Factorization

Because we are also allowed pivoting in Gaussian Elimination, we can get the permuted LU Factorization formula:

$$PA = LU$$

Theorem 2. *Let A be an $n \times n$ matrix. Then the following conditions are equivalent:*
1. A is non-singular.
2. A has n non-zero pivots.
3. A admits a permuted LU factorization: $PA = LU$.

1.4.4 Factorization of Symmetric Matrices

Definition 3. A square matrix is called symmetric if it equals its own transpose. $A = A^T$. Any symmetric matrix A is regular if and only if it can be factored as

$$A = LDL^T$$

1.4.5 Pivoting Strategies

There are a couple strategies we can use to ensure that both our solutions are good, and that our relative error is minimal.

Partial Pivoting says that at each stage we should use the largest (in absolute value) element as the pivot, even if the diagonal element is nonzero. This helps suppress round-off errors.

Full Pivoting lets us also exchange columns so that the greatest values are closes to the upper left.

1.5 Elementary Matrices

Definition 4. The elementary matrix E associated with an elementary row operation for m rowed matrices is the matrix obtained by applying the row operation to the $m \times m$ identity matrix I_m.

In other words, if we were to (for example), take our Identity Matrix I, add two times the first row to the second, and then multiply it by our original matrix, it's the same as the elementary row operation by itself.

These are very important for LU decomposition.

1.5.1 LU Decomposition

Theorem 3. *A Matrix A is regular if and only if it can be factored*

$$A = LU$$

Where L is a special lower triangular, having all ones on the diagonal, and U is an upper triangular matrix with nonzero diagonal entries.

In general to find the LU decomposition, apply the regular Gaussian Elimination to reduce A to its upper triangular form, and fill in the identity matrix with values used (elementary matrix). These two matrices are the Upper and Lower matrices.

1.5.2 Forward and Back Substitution

Once we have LU decomposition, we can solve the system.

1. Solve the Lower system:
$$L\mathbf{c} = \mathbf{b}$$
with Forward Substitution.
2. Solve the resulting Upper system:
$$U\mathbf{x} = \mathbf{c}$$
with Back Substitution.

1.6 Reduced Row Echelon Form

When dealing with systems of linear equations in augmented matrix form we need to get it to a solution, which can be found with Reduced Row Echelon Form (RREF). This form looks similar to the following.

$$[\mathbf{A}|\mathbf{b}] = \left[\begin{array}{ccc|c} 1 & 0 & 0 & b_1 \\ 0 & 1 & 0 & b_2 \\ 0 & 0 & 1 & b_3 \end{array}\right] \tag{3}$$

This can be characterized by the following:

- 0 rows are at the bottom.
- Leftmost non-zero entry is 1, also called the pivot (or leading 1).
- Each pivot is further to the right than the one above.
- Each pivot is the only non-zero entry in its column.

A less complete process gives us row echelon form, which allows for nonzero entries are allowed above the pivot.

1.7 Gauss Jordan Reduction

This procedure will let us solve any given matrix/linear system. The steps are as follows.

1. Given a system $A\mathbf{x} = \mathbf{b}$
2. Form augmented matrix $[A|b]$

3. Transform to RREF (1.6) using elementary row operations.
4. The linear matrix formed by this process has the same solutions as the initial system, however it is much easier to solve.

1.7.1 LDV Factorization

This sophisticated version of Gauss-Jordan elimination leads us to a more detailed version of the LU factorization. Let D be the diagonal matrix having the same diagonal entries as U. Let V be the special upper triangular matrix obtained from U by dividing each row by its pivot.

Theorem 4. *A matrix A is regular if and only if it admits a factorization*

$$A = LDV$$

Theorem 5. *A matrix A is nonsingular if and only if there is a permutation matrix P such that*

$$PA = LDV$$

where the matrices $L, D,$ and V are the same as defined above.

1.8 Existence and Uniqueness

If the RREF has a row that looks like:

$$[0, 0, 0, \cdots, 0|k]$$

where k is a non-zero constant, then the system has no solutions. We call this inconsistent.

If the system has one or more solutions, we call it consistent.

In order to be unique, the system needs to be consistent.

- If every column is a pivot, the there is only one solution (unique solution).
- Else If most columns are pivots, there are multiple solutions (possibly infinite).
- Else the system is inconsistent.

1.9 Superposition, Nonhomogeneous Principle, and RREF

For any nonhomogeneous linear system $\mathbf{Ax} = \mathbf{b}$, we can write the solutions as:

$$\mathbf{x} = \mathbf{x}_h + \mathbf{x}_p$$

Where $\mathbf{x}_h$ represents vectors in the set of homogeneous solutions, and $\mathbf{x}_p$ is a particular solution to the original equation.

We can use RREF to find $\mathbf{x}_p$, and then, using the same RREF with $\mathbf{b}$ replaced by $\mathbf{0}$, find $\mathbf{x}_h$.

The rank of a matrix r equals the number of pivot columns in the RREF. If r equals the number of variables, there is a unique solution. Otherwise if there is less, then it is not unique.

Definition 5. A square matrix of size $n \times n$ is nonsingular if and only if its rank is equal to n.

Theorem 6. *A homogeneous linear system $A\mathbf{x} = \mathbf{0}$ of m equations in n unknowns has a non-trivial solution $\mathbf{x} \neq \mathbf{0}$ if and only if the rank of A is $r < n$. If $m < n$ the system always has a nontrivial solution. If $m = n$ the system has a nontrivial solution if and only if A is singular.*

1.10 Inverse of a Matrix

When given a system of equations like:

$$\begin{cases} x + y = 1 \\ 4x + 5y = 6 \end{cases}$$

we can rewrite it in the form:

$$\begin{bmatrix} 1 & 1 \\ 4 & 5 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 6 \end{bmatrix}$$

For this sort of matrix, we can find the inverse which is defined as the matrix that, when multiplied with the original, equals an Identity Matrix. In other words:

$$A^{-1}A = AA^{-1} = I$$

1.10.1 Properties

- $\left(A^{-1}\right)^{-1} = A$
- A and B are invertible matrices of the same order if $(AB) = A^{-1}B^{-1}$
- If A is invertible, then so is A^T and $\left(A^{-1}\right)^T = \left(A^T\right)^{-1}$

1.10.2 Inverse Matrix by RREF

For an $n \times n$ matrix A, the following procedure either produces A^{-1}, or proves that it's impossible.

1. Form the $n \times 2n$ matrix $M = [A|I]$
2. Transform M into its RREF, R.
3. If the first n columns produce an Identity Matrix, then the last n are its inverse. Otherwise A is not invertible.

1.11 Invertibility and Solutions

The matrix vector equation $A\mathbf{x} = b$ where A is an $n \times n$ matrix has:

- A unique solution $x = A^{-1}b$ if and only if A is invertible.
- Either no solutions or infinitely many solutions if A is not invertible.

For the homogeneous equation $A\mathbf{x} = 0$, there is always one solution, $x = 0$ called the trivial solution.

Let $\mathbf{A}$ be an $n \times n$ matrix. The following statements apply.

- $\mathbf{A}$ is an invertible matrix.
- $\mathbf{A}^T$ is an invertible matrix.
- $\mathbf{A}$ is row equivalent to I_n.
- $\mathbf{A}$ has n pivot columns.
- The equation $\mathbf{Ax} = \mathbf{0}$ has only the trivial solution, $\mathbf{x} = \mathbf{0}$.
- The equation $\mathbf{Ax} = \mathbf{0}$ has a unique solution for every $\mathbf{b}$ in $\mathbb{R}^n$.

1.12 Determinants and Cramer's Rule

The determinant of a square matrix is a scalar number associated with that matrix. These are very important.

1.12.1 2×2 Matrix

To find the determinant of a 2×2 matrix, the determinant is the diagonal products subtracted. This process is demonstrated below.

$$\begin{aligned} A &= \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \\ |A| &= a_{22} \cdot a_{11} - a_{12} \cdot a_{21} \end{aligned} \tag{4}$$

1.12.2 Definitions

Every element of a $n \times n$ matrix has an associated minor and cofactor.

- Minor $\rightarrow$ A $(n-1) \times (n-1)$ matrix obtained by deleting the ith row and jth column of A.
- Cofactor $\rightarrow$ The scalar $C_{ij} = (C-1)^{i+j} \, |M_{ij}|$

1.12.3 Recursive Method of an $n \times n$ matrix A

We can now determine a recursive method for any $n \times n$ matrix.

Using the definitions declared above, we use the recursive method that follows.

$$|A| = \sum_{j=1}^{n} a_{ij} C_{ij} \tag{5}$$

Find j and then finish with the rules for the 2×2 matrix defined above in (1.12.1).

1.12.4 Row Operations and Determinants

Let A be square.

- If two rows of A are exchanged to get B, then $|B| = -|A|$.
- If one row of A is multiplied by a constant c, and then added to another row to get B, then $|A| = |B|$.
- If one row of A is multiplied by a constant c, then $|B| = c|A|$.
- If $|A| = 0$, A is called singular.

For an $n \times n$ A and B, the determinant $|AB|$ is given by $|A||B|$.

1.12.5 Properties of Determinants

- If two rows of $\mathbf{A}$ are interchanged to equal $\mathbf{B}$, then
$$|\mathbf{B}| = -|\mathbf{A}|$$
- If one row of $\mathbf{A}$ is multiplied by a constant k, and then added to another row to produce matrix $\mathbf{B}$, then
$$|\mathbf{B}| = |\mathbf{A}|$$
- If one row of $\mathbf{A}$ is multiplied by k to produce matrix $\mathbf{B}$, then
$$|\mathbf{B}| = k|\mathbf{A}|$$
- If $|AB| = 0$, then either $|A|$ or $|B|$ must be zero.
- $|A^T| = A$
- If $|\mathbf{A}| \neq 0$, then $|\mathbf{A}^{-1}| = \frac{1}{|\mathbf{A}|}$.

- If A is an upper or lower triangle matrix[1], then the determinant is the product of the diagonals.
- If one row or column consists of only zeros, then $|A| = 0$.
- If two rows or columns are equal, then $|A| = 0$.
- A is invertible.
- A^T is also invertible.
- A has n pivot columns.
- $|A| \neq 0$
- If $|A| = 0$ it is called singular, otherwise it is nonsingular.

1.12.6 Cramer's Rule

For the $n \times n$ matrix A with $|A| \neq 0$, denote by A_i the matrix obtained from A by replacing its ith column with the column vector $\mathbf{b}$. Then the ith component of the solution of the system is given by:

$$x_i = \frac{|A_i|}{|A|} \tag{6}$$

2 Vector Spaces and Subspaces

A vector space $\mathcal{V}$ is a non-empty collection of elements that we call vectors, for which we can define the operation of vector addition and scalar multiplication:

1. Addition: $\mathbf{x} + \mathbf{y}$
2. Scalars: $c\mathbf{x}$ where c is a constant.

that satisfy the following properties:

1. $\mathbf{x} + \mathbf{y} \in \mathcal{V}$
2. $c\mathbf{x} \in \mathcal{V}$

   which can be condensed into a single equation:

$$c\mathbf{x} + d\mathbf{y} \in \mathcal{V}$$

which is called closure under linear combinations.

[1] A triangle matrix is one where either the lower or upper half is zero, e.g. $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}$.

2.1 Properties

We have the properties from before, as well as new ones.

1. $\mathbf{x} + \mathbf{y} \in \mathcal{V} \leftarrow$ Addition
2. $c\mathbf{x} \in \mathcal{V} \leftarrow$ Scalar Multiplication
3. $\mathbf{x} + \mathbf{0} = \mathbf{x} \leftarrow$ Zero Element
4. $\mathbf{x} + (-\mathbf{x}) = (-\mathbf{x}) + \mathbf{x} = \mathbf{0} \leftarrow$ Additive Inverse
5. $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z}) \leftarrow$ Associative Property
6. $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x} \leftarrow$ Commutativity
7. $1 \cdot \mathbf{x} = \mathbf{x} \leftarrow$ Identity
8. $c(\mathbf{x} + \mathbf{y}) = c\mathbf{x} + c\mathbf{y} \leftarrow$ Distributive Property
9. $(c + d)\mathbf{x} = c\mathbf{x} + d\mathbf{x} \leftarrow$ Distributive Property
10. $c(d\mathbf{x}) = (cd)\mathbf{x} \leftarrow$ Associativity

2.2 Vector Function Space

A vector function space is just a unique vector space where the elements of the space are functions.

Note, the solutions to linear and homogeneous differential equations form vector spaces.

2.2.1 Closure under Linear Combination

$$c\mathbf{x} + d\mathbf{y} \in \mathbb{V} \text{ whenever } \mathbf{x}, \mathbf{y} \in \mathbb{V} \text{ and } c, d \in \mathbb{R} \tag{7}$$

2.2.2 Prominent Vector Function Spaces

- $\mathbb{R}^2 \rightarrow$ The space of all ordered pairs.
- $\mathbb{R}^3 \rightarrow$ The space of all ordered triples.
- $\mathbb{R}^n \rightarrow$ The space of all ordered n-tuples.
- $\mathbb{P} \rightarrow$ The space of all polynomials.
- $\mathbb{P}_n \rightarrow$ The space of all polynomials with degree $\leq n$.
- $\mathbb{M}_{mn} \rightarrow$ The space of all $m \times n$ matrices.
- $\mathbb{C}(I) \rightarrow$ The space of all continuous functions on the interval I (open, closed, finite, and infinite).
- $\mathbb{C}^n(I) \rightarrow$ Same as above, except with n continuous derivatives.
- $\mathbb{C}^n \rightarrow$ The space of all ordered n-tuples of complex numbers.

2.3 Vector Subspaces

Theorem: A non-empty subset $\mathbb{W}$ of a vector space $\mathbb{V}$ is a subspace of $\mathbb{V}$ if it is closed under addition and scalar multiplication:

- If $\mathbf{u}, \mathbf{v} \in \mathbb{W}$, than $\mathbf{u} + \mathbf{V} \in \mathbb{W}$.
- If $\mathbf{u} \in \mathbb{W}$ and $c \in \mathbb{R}$, than $c\mathbf{u} \in \mathbb{W}$.

We can rewrite this to be more efficient:

$$\text{If } \mathbf{u}, \mathbf{v} \in \mathbb{W} \text{ and } a, b \in \mathbb{R}, \text{ than } a\mathbf{u} + b\mathbf{v} \in \mathbb{W}. \tag{8}$$

Note, vector space does not imply subspace. All subspaces are vector spaces, but not all vector spaces are subspaces.

To determine if it is a subspace, we check for closure with the above theorem.

There are only a couple subspaces for $\mathbb{R}^2$:

- The zero subspace $\{(0, 0)\}$.
- Lines passing through the origin.
- $\mathbb{R}^2$ itself.

We can call the zero and the set $\mathbb{V}$ themselves trivial subspaces, calling the subspace of lines passing through the origin the only non-trivial subspace in $\mathbb{R}^2$.

We can classify $\mathbb{R}^3$ similarly:

- Trivial:
  - Zero subspace
  - $\mathbb{R}^3$

  Non-Trivial
  - Lines that contain the origin.
  - Places that contain the origin.

2.3.1 Examples

- The set of all even functions.
- The set of all solutions to $y''' - y''t + y = 0$.
- $\{P \in \mathbb{P}; P(2) = P(3)\}$

3 Span, Basis and Dimension

Given one or more vectors in a vector space, we can create more vectors through addition and scalar multiplication. These vectors that result from this process are called linear combinations.

3.1 Span

The span of a set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ of vectors in a vector space $\mathbb{V}$, denoted by $\text{Span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is the set of all linear combinations of the vectors.

3.1.1 Example

$$\text{For example, If } \mathbf{u} = \begin{bmatrix} 3 \\ 2 \\ 0 \end{bmatrix} \text{ and } \mathbf{v} = \begin{bmatrix} 0 \\ 2 \\ 2 \end{bmatrix}$$

Then we can write their span as

$$a\mathbf{u} + b\mathbf{v} = a\begin{bmatrix} 3 \\ 2 \\ 0 \end{bmatrix} + b\begin{bmatrix} 0 \\ 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 3a \\ 2a + 2b \\ 2b \end{bmatrix}$$

3.2 Spanning Sets in $\mathbb{R}^n$

A vector $\mathbf{b}$ in $\mathbb{R}^n$ is in $\text{Span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ where $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ are vectors in $\mathbb{R}^n$, provided that there is at least one solution of the matrix-vector equation $A\mathbf{x} = \mathbf{b}$, where A is the matrix whose column vectors are $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$.

3.3 Span Theorem

For a set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ in vector space $\mathbb{V}$, $\text{Span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is subspace of $\mathbb{V}$.

3.4 Column Space

For any $m \times n$ matrix A, the column space, denoted Col A, is the span of the column vectors of A, and is a subspace of $\mathbb{R}^n$.

3.5 Linear Independence

A set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ of vectors in vector space $\mathbb{V}$ is linearly independent if no vector of the set can be written as a linear combination of the others. Otherwise it is linearly dependent.

This notion of linear independence also carries over to function spaces. A set of vector functions $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ in a vector space $\mathbb{V}$ is linearly independent on an interval I if for all t in I the only solution of

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n = \mathbf{0}$$

for $(c_1, c_2, \ldots, c_n \in \mathbb{R})$ is $c_i = 0$ for all i.

If for any value t_0 of t there is any solution with $c_i \neq \mathbf{0}$, the vector functions are linearly dependent.

Example 1. The vectors

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} 0 \\ 3 \\ 1 \end{pmatrix}, \mathbf{v}_3 = \begin{pmatrix} -1 \\ 4 \\ 3 \end{pmatrix}$$

are linearly dependent because

$$\mathbf{v}_1 - 2\mathbf{v}_2 + \mathbf{v}_3 = \mathbf{0}$$

however the first two vectors are linearly independent because the only solution to

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 = \mathbf{0}$$

is $c_1 = c_2 = 0$.

3.5.1 Testing for Linear Independence

1. (a) Put the system in matrix-vector form:

$$\begin{bmatrix} \uparrow & \uparrow & \cdots & \uparrow \\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \\ \downarrow & \downarrow & \cdots & \downarrow \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} = \mathbf{0}$$

   (b) Analyze Matrix:
   The column vectors of A are linearly independent if and only if the solution $\mathbf{x} = \mathbf{0}$ is unique, which means $c_i = 0$ for all i.
   Any of the following also satisfy this condition for a unique solution:

   - A is invertible.
   - A has n pivot columns.
   - $|A| \neq 0$

2. Suppose we have a set of vectors $\mathbf{v}$.

$$\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\} \in \mathbb{R}^n, \dim(\mathbf{v}) = m$$

   Then the set $\mathbf{v}$ is linearly dependent if $n > m$ where n is the number of elements in $\mathbf{v}$. *Note, this cannot prove the opposite. It only goes one way.*

$$\left\{ \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ -3 \\ 7 \end{pmatrix} \right\} \text{Is dependent}$$

3. Columns of A are linearly independent if and only if $\mathbf{Ax} = \mathbf{0}$ has only the trivial solutions of n.

3.5.2 Linear Independence of Functions

One way to check a set of functions is to consider them as a one dimensional vector.

$$\mathbf{v}_i(t) = f_n(t)$$

Another method is the Wronskian:

$$\text{To find the Wronskian of functions } f_1, f_2, \ldots, f_n \text{ on } I:$$
$$W[f_1, f_2, \ldots, f_n] = \begin{bmatrix} f_1 & f_2 & \cdots & r_n \\ f_1^{'} & f_2^{'} & \cdots & r_n^{'} \\ \vdots & \vdots & \ddots & \vdots \\ f_1^{n-1} & f_2^{n-1} & \cdots & r_n^{n-1} \end{bmatrix} \quad (9)$$

If $W \neq 0$ for all t on the interval I, where $f_1, f_2, \ldots, f_n$ are defined, then the function space is a linearly independent set of functions on I.

3.6 Basis of a Vector Space

The set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is a basis for vector space $\mathbb{V}$ provided that

- $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is linearly independent.
- $\text{Span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\} = \mathbb{V}$

Theorem 7. *Every basis of $\mathbb{R}^n$ consists of exactly n vectors. Furthermore, a set of n vectors $\{\mathbf{v}_1, \cdots, \mathbf{v}_n \in \mathbb{R}^n\}$ is a basis iff the $n \times n$ matrix $A = (\mathbf{v}_1 \cdots \mathbf{v}_n$ is nonsingular: rank $A = n$*

Suppose the vector space V has a basis $\mathbf{v}_1, \cdots, \mathbf{v}_n$. Then every other basis of V has the same number of elements in it. This number is called the dimension of V, and is written $\dim V = n$.

Suppose V is an n-dimensional vector space, then

3. *Every set of more than n elements of V is linearly dependent.*
4. *No set of less than n elements spans V.*
5. *A set of n elements forms a basis iff it spans V.*
6. *A set of n elements forms a basis iff it is linearly independent.*

3.6.1 Standard Basis for $\mathbb{R}^n$

$$\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$$
$$\text{where}$$
$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \ldots, \mathbf{e}_n = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \quad (10)$$

are the column vectors of the identity matrix I_n.

3.6.2 Example

A vector space can have different bases.

The standard basis for $\mathbb{R}^n$ is:

$$\{\mathbf{e}_1, \mathbf{e}_2\} \text{ for } \mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \text{ and } \mathbf{e}_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} \text{ giving } \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$$

But another basis for $\mathbb{R}^2$ is given by:

$$\left\{ \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right\}$$

3.7 Dimension of the Column Space of a Matrix

Essentially, the number of vectors in a basis.

3.7.1 Properties

- The pivot columns of a matrix A form a basis for Column A.
- The dimension of the column space, called the rank of A, is the number of pivot columns in A.
$$\text{rank } A = \dim(\text{Col}(A))$$

3.7.2 Invertible Matrix Characterizations

Let A be an $n \times n$ matrix. The following are true.

- A is invertible.
- The column vector of A is linearly independent.
- Every column of A is a pivot column.
- The column vectors of A form a basis for $\text{Col}(A)$.
- Rank $A = n$

3.8 The Fundamental Matrix Subspaces

3.8.1 Kernel and Range

The range of an $m \times n$ matrix A is the subspace $rng \subset \mathbb{R}^m$ spanned by its columns. The kernel of A is the subspace $kerA \subset \mathbb{R}^n$ consisting of all vectors which are annihilated by A, so

$$kerA = \{\mathbf{z} \in \mathbb{R}^n | A\mathbf{z} = \mathbf{0}\} \subset \mathbb{R}^n$$

The range is also known as the column space or image of the matrix, while the kernel is also called the null space.

At its core, the null space, or kernel is the set of solutions $\mathbf{z}$ to the homogeneous system $A\mathbf{z} = \mathbf{0}$.

If $\mathbf{z}_1, \cdots, \mathbf{z}_k$ are individual solutions to the same homogeneous linear system, then so is any linear combination of $c_1\mathbf{z}_1 + \cdots + c_k\mathbf{z}_k$.

As we've seen before, for inhomogeneous systems, once we know the homogeneous solution, we can generalize with the inhomogeneous solution.

$$\mathbf{x} = \mathbf{x}^* + \mathbf{z}$$

This gives us a couple different properties that are equivalent.

Theorem 8. *If A is an $m \times n$ matrix, then the following conditions are equivalent:*

1. *$ker A = \{\mathbf{0}\}$, i.e. the homogeneous system $A\mathbf{x} = \mathbf{0}$ has the unique solution $\mathbf{x} = \mathbf{0}$.*
2. *$rank A = n$*
3. *The linear system $A\mathbf{x} = \mathbf{b}$ has no free variables.*
4. *The system $A\mathbf{x} = \mathbf{b}$ has a unique solution for each $\mathbf{b} \in rng A$.*

If A is a square, $n \times n$ matrix, then the following conditions are equivalent:

1. *A is nonsingular*
2. *$rank A = n$*
3. *$ker A = \{\mathbf{0}\}$*
4. *$rng A = \mathbb{R}^n$*

3.8.2 The Superposition Principle

The Superposition Principle is the key to linearity. When we have homogeneous solutions, we can generate new solutions by combining new solutions. For inhomogeneous systems, the superposition principle allows us to combine solutions corresponding to different inhomogeneities.

If we know the particular solutions of two inhomogeneous linear systems

$$A\mathbf{x} = \mathbf{b}_1, A\mathbf{x} = \mathbf{b}_2$$

that have the same coefficient matrix A, then we can combine the two systems

$$A\mathbf{x} = c_1\mathbf{b}_1 + c_2\mathbf{b}_2$$

which gives us the particular solution

$$\mathbf{x}^* = c_1\mathbf{x}_1^* + c_2\mathbf{x}_2^*$$

3.8.3 Adjoint Systems, Cokernel, and Corange

The adjoint to a linear system $A\mathbf{x} = \mathbf{b}$ of m equations in n unknowns is the linear system

$$A^T\mathbf{y} = \mathbf{f}$$

consisting of n equations in m unkonwns $\mathbf{y} \in \mathbb{R}^m$ with $\mathbf{f} \in \mathbb{R}^n$.

The corange of an $m \times n$ matrix A is the range of its transpose.

$$corngA = rngA^T = \{A^T\mathbf{y}|\mathbf{y} \in \mathbb{R}^m\} \subset \mathbb{R}^n$$

The cokernel, or left null space of A is the kernel of its transpose.

$$cokerA = kerA^T = \{\mathbf{w} \in \mathbb{R}^m|A^T\mathbf{w} = \mathbf{0}\} \subset \mathbb{R}^m$$

3.8.4 The Fundamental Theorem of Linear Algebra

$$\begin{aligned} dimcorngA = dimrngA = rankA = rankA^T &= r \\ dimkerA &= n - r \\ dimcokerA &= m - r \end{aligned}$$

4 Inner Products and Norms

The most basic example of an inner product is the familiar dot product

$$\langle \mathbf{v}, \mathbf{w} \rangle = \mathbf{v} \cdot \mathbf{w} = v_1 w_1 + v_2 w_2 + \cdots + v_n w_n$$

It's important to note here that this dot product is equal to the matrix product of $\mathbf{v}^T$ and $\mathbf{w}$.

Any vector, when dotted with itself yields the sum of the squares of its entries, which leads us to the Euclidean Norm, or the length of the vector, which is the square root.

$$\|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}}$$

Theorem 9. *An inner product on the real vector spave V is a pairing that takes two vectors $\mathbf{v}, \mathbf{w} \in V$ and produces a real number $\langle \mathbf{v}, \mathbf{w} \rangle \in \mathbb{R}$. The inner product is required to satisfy the following three axioms for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and scalars $c, d \in \mathbb{R}$*

- ***Bilinearity***

$$\langle c\mathbf{u} + d\mathbf{v}, \mathbf{w} \rangle = c\langle \mathbf{u}, \mathbf{w} \rangle + d\langle \mathbf{v}, \mathbf{w} \rangle$$
$$\langle \mathbf{u}, c\mathbf{v} + d\mathbf{w} \rangle = c\langle \mathbf{u}, \mathbf{v} \rangle + d\langle \mathbf{u}, \mathbf{w} \rangle$$

- ***Symmetry***

$$\langle \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{w}, \mathbf{v} \rangle$$

- ***Positivity***

$$\langle \mathbf{v}, \mathbf{v} \rangle > 0 \textit{ whenever } \mathbf{v} \neq \mathbf{0} \textit{ while } \langle \mathbf{0}, \mathbf{0} \rangle = 0$$

Given an inner product, the associated norm of a vector $\mathbf{v} \in V$ *is defined as the positive square root of the inner product of the vector with itself.*

4.1 Inequalities

4.1.1 The Cauchy-Schwarz Inequality

Any Euclidean dot product between two vectors can be expressed geometrically as

$$\mathbf{v} \cdot \mathbf{w} = \|\mathbf{v}\|\|\mathbf{w}\| \cos\theta$$

Therefore, the absolute value of the dot product is bounded by the product of the lengths of the vectors

$$|\mathbf{v} \cdot \mathbf{w}| \leq \|\mathbf{v}\|\|\mathbf{w}\|$$

Every inner product satisfies the Cauchy-Schwarz inequality.

Two elements $\mathbf{v}, \mathbf{w} \in v$ of an inner product space V are called orthogonal if their inner product vanishes: $\langle \mathbf{v}, \mathbf{w} \rangle = 0$

4.1.2 The Triangle Inequality

The norm associated with an inner product satisfies the triangle inequality

$$\|\mathbf{v} + \mathbf{w}\| \leq \|\mathbf{v}\| + \|\mathbf{w}\| \text{ for all } \mathbf{v}, \mathbf{w} \in V$$

Equality holds iff $\mathbf{v}, \mathbf{w}$ are parallel.

4.2 Norms

Theorem 10. *A norm on the vector space* V *assigns a real number* $\|\mathbf{v}\|$ *to each vector* $\mathbf{v} \in V$ *subject to the following axioms for every* $\mathbf{v}, \mathbf{w} \in V$ *and* $c \in \mathbb{R}$.

- ***Positivity***
$$\|\mathbf{v}\| \geq 0, (\|\mathbf{v}\| = 0 \Leftrightarrow \mathbf{v} = \mathbf{0})$$
- ***Homogeneity***
$$\|c\mathbf{v}\| = |c|\|\mathbf{v}\|$$
- ***Triangle Inequality***
$$\|\mathbf{v} + \mathbf{w}\| \leq \|\mathbf{v}\| + \|\mathbf{w}\|$$

4.2.1 Unit Vectors

In any vector space V, the elements $\mathbf{u} \in V$ where $\|\mathbf{u}\| = 1$ are very important and are referred to as unit vectors.

If $\mathbf{v} \neq \mathbf{0}$ is any nonzero vector, then the vector $\mathbf{u} = \mathbf{v}/\|\mathbf{v}\|$ obtained by dividing $\mathbf{v}$ by its norm is a unit vector parallel to $\mathbf{v}$.

4.2.2 Equivalence of Norms

Even though there are many different types of norms, in a finite dimensional vector space they are all more or less equivalent.

Let $\|\cdot\|_1$ and $\|\cdot\|_2$ be any two norms on $\mathbb{R}^n$. Then there exist positive constants $c^*, C^* > 0$ such that

$$c^*\|\mathbf{v}\|_1 \leq \|\mathbf{v}\|_2 \leq C^*\|\mathbf{v}\|_1, \forall \mathbf{v} \in \mathbb{R}^n$$

4.3 Positive Definite Matrices

An $n \times n$ matrix K is called positive definite if it is symmetric, $K^T = K$ and satisfies the positivity condition

$$\mathbf{x}^T K \mathbf{x} > 0 \text{ for all } \mathbf{0} \neq \mathbf{x} \in \mathbb{R}^n$$

This is sometimes denoted as $K > 0$. Any positive definite matrix is nonsingular.

Every inner product on $\mathbb{R}^n$ is given by

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T K \mathbf{y} \text{ for } \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$$

Where K is a positive definite matrix as is defined above.

Given any symmetric matrix K, the homogeneous quadratic polynomial

$$q(\mathbf{x}) = \mathbf{x}^T K \mathbf{x} = \sum_{i,j=1}^{n} k_{ij} x_i x_j$$

is known as a quadratic form on $\mathbb{R}^n$. The quadratic form is called the positive definite if

$$q(\mathbf{x}) > 0 \text{ for all } 0 \neq \mathbf{x} \in \mathbb{R}^n$$

thus a quadratic form is positive definite iff its coefficient matrix is.

4.3.1 Gram Matrices

Let V be an inner product space, and let $\mathbf{v}_1, \cdots, \mathbf{v}_n \in V$. The associated Gram matrix

$$K = \begin{pmatrix} \langle \mathbf{v}_1, \mathbf{v}_1 \rangle & \langle \mathbf{v}_1, \mathbf{v}_2 \rangle & \cdots & \langle \mathbf{v}_1, \mathbf{v}_n \rangle \\ \langle \mathbf{v}_2, \mathbf{v}_1 \rangle & \langle \mathbf{v}_2, \mathbf{v}_2 \rangle & \cdots & \langle \mathbf{v}_2, \mathbf{v}_n \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \mathbf{v}_n, \mathbf{v}_1 \rangle & \langle \mathbf{v}_n, \mathbf{v}_2 \rangle & \cdots & \langle \mathbf{v}_n, \mathbf{v}_n \rangle \end{pmatrix}$$

is the $n \times n$ matrix whose entries are the inner products between the selected vector space elements.

Symmetry of the inner product implies symmetry of the Gram matrix:

$$k_i j = \langle \mathbf{v}_i, \mathbf{v}_j \rangle = \langle \mathbf{v}_j, \mathbf{v}_i \rangle = k_j i$$

and hence $K^T = K$.

All Gram matrices are positive semi-definite. The Gram matrix is positive definite iff $\mathbf{v}_1, \cdots, \mathbf{v}_n$ are linearly independent. This leads us to more details about a given $m \times n$ matrix A.

1. The $n \times n$ Gram matrix $K = A^T A$ is positive definite.
2. A has linearly independent columns.
3. $rank A = n \leq m$
4. $ker A = \{0\}$

Suppose A is an $m \times n$ matrix with linearly independent columns. Suppose C is any positive definite $m \times m$ matrix. Then the matrix $K = A^T C A$ is a positive definite $n \times n$ matrix.

4.3.2 Completing the Square

Gram matrices give us a virtually unlimited supply of positive definite matrices, however we still need to determine how to figure out how to determine whether or not a given matrix is positive definite.

A symmetric matrix is positive definite iff it is regular and has all positive pivots.

In other words, a square matrix K is positive definite iff it can be factored $K = LDL^T$ where L is a special lower triangular and D is diagonal with all positive definite entries.

4.3.3 The Cholesky Factorization

We know how to write any regular quadratic form as a linear combination of squares. We can push this slightly further and deduce the Cholesky Factorization

$$K = LDL^T = LSS^T L^T = MM^T | M = LS$$

4.4 Complex Vector Spaces

Remember that complex numbers are expressed in the form $z = x + iy$ where $x, y \in \mathbb{R}$ and $i^2 = -1$.

For this we also need the complex conjugate. The complex conjugate of $z = x + iy$ is $\bar{z} = x - iy$.

We can also define complex vector spaces and inner products, the only difference is that the scalar entries are now complex scalars.

An Inner Product on the complex vector space V is a pairing that takes two vectors, $\mathbf{v}, \mathbf{w} \in V$ and produces a complex number $\langle \mathbf{v}, \mathbf{w} \rangle \in \mathbb{C}$ subject to the following requirements for $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and $c, d \in \mathbb{C}$:

1. Sesquilinearity:

$$\langle c\mathbf{u} + d\mathbf{v}, \mathbf{w} \rangle = c \langle \mathbf{u}, \mathbf{w} \rangle + d \langle \mathbf{v}, \mathbf{w} \rangle$$
$$\langle \mathbf{u}, c\mathbf{v} + d\mathbf{w} \rangle = c \langle \mathbf{u}, \mathbf{v} \rangle + d \langle \mathbf{u}, \mathbf{w} \rangle$$

2. Conjugate Symmetry:
$$\langle \mathbf{v}, \mathbf{w} \rangle = \overline{\langle \mathbf{w}, \mathbf{v} \rangle}$$

3. Positivity:
$$||\mathbf{v}||^2 = \langle \mathbf{v}, \mathbf{v} \rangle \geq 0 \wedge \langle \mathbf{v}, \mathbf{v} \rangle = 0 \Leftrightarrow \mathbf{v} = \mathbf{0}$$

The Cauchy-Schwarz inequality is also valid on any complex inner product space.

5 Orthogonality

5.1 Orthogonal Bases

Remember two elements are orthogonal if their inner product vanishes. In the case of Euclidean space, this means that the two vectors are at right angles.

Theorem 11. *A basis $\mathbf{u}_1, \cdots, \mathbf{u}_n$ of V is called orthogonal if $\langle \mathbf{u_i}, \mathbf{u_j} \rangle = 0$ for all $i \neq j$. The basis is called orthonormal if, in addition, each vector has unit length: $||\mathbf{u_i}|| = 1$ for all $i = 1, \cdots, n$.*

Also, if $\mathbf{u}_1, \cdots, \mathbf{u}_n$ is an orthogonal basis of a vector space V, then the normalized vectors $\mathbf{u}_i = \mathbf{v}_i / ||\mathbf{v_i}||, i = 1, \cdots, n$, form an orthonormal basis.

Associated with this theorem, if $\mathbf{v}_1, \cdots, \mathbf{v}_k \in V$ are nonzero, mutually orthogonal elements, so $\mathbf{v}_i \neq \mathbf{0}$ and $\langle \mathbf{v_i}, \mathbf{v_j} \rangle = 0$ for all $i \neq j$, then they are linearly independent.

Theorem 12. *Suppose $\mathbf{v}_1, \cdots, \mathbf{v}_n \in V$ are nonzero, mutually orthogonal elements of an inner product space V. Then $\mathbf{v}_1, \cdots, \mathbf{v}_n$ form an orthogonal basis for their span $W = span\{\mathbf{v}_1, \cdots, \mathbf{v}_n\} \subset V$, which is therefore a subspace of dimension $n = dimW$. In particular, if $dimV = n$, then $\mathbf{v}_1, \cdots, \mathbf{v}_n$ form a orthogonal basis for V.*

5.1.1 Computations in Orthogonal Bases

The advantage of an orthogonal or orthonormal base is that we can express other elements as linear combinations of the base elements, in other words, find their coordinates.

Theorem 13. *Let $\mathbf{u}_1, \cdots, \mathbf{u}_n$ be an orthonormal basis for an inner product space V. Then one can write any element $\mathbf{v} \in V$ as a linear combination in which its coordinates*
$$c_i = \langle \mathbf{v}, \mathbf{u_i} \rangle, i = 1, \cdots, n$$
are explicitly given as inner products. Moreover, its norm is the square root of the sum of the squares of its orthonormal basis coordinates.

We also can say that if $\mathbf{v}_1, \cdots, \mathbf{v}_n$ form an orthogonal basis, then the corresponding coordinates of a vector are given by
$$a_i = \frac{\langle \mathbf{v}, \mathbf{v_i} \rangle}{||\mathbf{v_i}||^2}$$

5.2 The Gram-Schmidt Process

Now we know that orthogonal and orthonormal bases are useful, we need to determine how to construct them.

Let V be a finite-dimensional inner product space. We will construct the basis elements one by one, and since there are no conditions on the first element we can choose the first element of V, $\mathbf{v}_1 = \mathbf{w}_1$.

The second basis vector must be orthogonal to the first, which we attempt to ensure by setting $\mathbf{v}_2 = \mathbf{w}_2 - c\mathbf{v}_1$ where c is a scalar to be determined. We can expand $\langle \mathbf{v_1}, \mathbf{v_2} \rangle$ and determine that $c = \langle \mathbf{w_2}, \mathbf{v_1} \rangle / ||\mathbf{v_1}||^2$.

We can extrapolate this process to all vectors in the space, giving us the general Gram-Schmidt formula

$$\mathbf{v}_k = \mathbf{w}_k - \sum_{j=1}^{k-1} \frac{\langle \mathbf{w_k}, \mathbf{v_j} \rangle}{||\mathbf{v_j}||^2} \mathbf{v}_j, k = 1, \cdots, n$$

We also can say that every non-zero finite-dimensional inner product space has an orthonormal basis. In fact, if the dimension is greater that 1, there are infinitely many.

5.2.1 Modifications of the Gram-Schmidt Process

We can modify the Gram-Schmidt process a little to gain additional benefit.

First step is to replace each orthogonal basis vector with its normalized version: $\mathbf{u}_j = \mathbf{v}_j / ||\mathbf{v_j}||$. This allows us to compute

$$r_{ij} = \langle \mathbf{w_j}, \mathbf{u_i} \rangle, i = 1, \cdots, j-1$$

we obtain the next orthonormal basis vector with

$$r_{ij} = \sqrt{||\mathbf{w_j}||^2 - r_{1j}^2 - \cdots - r_{j-1,j}^2}$$
$$\mathbf{u}_j = \frac{\mathbf{w}_j - r_{1j}\mathbf{u}_1 - \cdots - r_{j-1,j}\mathbf{u}_{j-1}}{r_{jj}}$$

5.3 Orthogonal Matrices

A square matrix Q is called an orthogonal matrix if it satisfies

$$Q^T Q = 1$$

This also implies that

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

A matrix is orthogonal iff its columns form an orthonormal basis with respect to the Euclidean dot product on $\mathbb{R}^n$.

An orthogonal matrix has determinant $\det Q = \pm 1$ and the product of two orthogonal matrices is also orthogonal.

5.3.1 The QR Factorization

Any nonsingular matrix A can be factored, $A = QR$, into the product of an orthogonal matrix Q and an upper triangular matrix R. The factorization is unique if all the diagonal entries of R are assumed to be positive.

This strategy can be employed as an alternative to traditional Gaussian elimination

$$A\mathbf{x} = \mathbf{b} \equiv QR\mathbf{x} = \mathbf{b} \equiv R\mathbf{x} = Q^T\mathbf{b}$$

We also can say that if we let $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ with $||\mathbf{v}|| = ||\mathbf{w}||$. Set $\mathbf{u} = (\mathbf{v} - \mathbf{w})/\|\mathbf{v} - \mathbf{w}\|$ and let $H = 1 - 2\mathbf{u}\mathbf{u}^T$ be the corresponding elementary reflection matrix. Then $H\mathbf{v} = \mathbf{w}$ and $H\mathbf{w} = \mathbf{v}$.

In other words, what we're doing is applying the Gram-Schmidt process to each column vector of the original matrix, and then creating the upper triangle matrix as an upper triangular Gram matrix.

$$A = [\mathbf{a}_1, \cdots, \mathbf{a}_n] Q = [\mathbf{e}_1, \cdots, \mathbf{e}_n] R = \begin{pmatrix} \langle \mathbf{e}_1, \mathbf{a}_1 \rangle & \langle \mathbf{e}_1, \mathbf{a}_2 \rangle & \cdots & \langle \mathbf{e}_1, \mathbf{a}_n \rangle \\ 0 & \langle \mathbf{e}_2, \mathbf{a}_2 \rangle & \cdots & \langle \mathbf{e}_2, \mathbf{a}_n \rangle \\ 0 & 0 & \cdots & \langle \mathbf{e}_3, \mathbf{a}_n \rangle \\ \vdots & \vdots & \ddots & \vdots \end{pmatrix}$$

5.4 Orthogonal Polynomials

Orthogonal Polynomials can be very useful in functions spaces. We'll start by discussing the Legendre Polynomials.

To construct the Legendre Polynomials, we start by constructing an orthonormal basis for vector space$P^{(n)}$ of polynomials of degree $\leq n$. This construction will be based on the L^2 inner product

$$\langle \mathbf{p}, \mathbf{q} \rangle = \int_{-1}^{1} p(t)q(t)\, dt$$

We then apply the Gram-Schmidt orthogonalization process to the elementary, but non-orthogonal monomial basis $1, t, t^2, \cdots, t^n$, and compute the next orthogonal polynomials through recursive application of the Gram-Schmidt Process.

The resulting polynomials are known as the monic[2] Legendre Polynomials. However there is also a way to explicitly solve for the classical Legendre Polynomials.[3]

[2]Leading coefficient is equal to 1

[3]Classical Legendre Polynomials are those that are certain scalar multiples, namely

$$P_k(t) = \frac{(2k)!}{2^k(k!)^2} q_k(t), k = 0, 1, 2, \cdots$$

and so also define a system of orthogonal polynomials. The multiple is fixed by the requirement that

$$P_k(1) = 1$$

The Rodrigues formula for the classical Legendre Polynomials is

$$P_k(t) = \frac{1}{2^k k!}\frac{d^k}{dt^k}(t^2-1)^k, ||\mathbf{P_k}|| = \sqrt{\frac{2}{2k+1}}, k = 0, 1, 2, \cdots$$

If $j \leq k$, then the polynomial $R_{j,k}(t)$ is orthogonal to all polynomials of degree $\leq j-1$.

The transformed Legendre Polynomials

$$\widetilde{P}_k(t) = P_k\left(\frac{2t-b-a}{b-a}\right), ||\widetilde{\mathbf{P}}_\mathbf{k}|| = \sqrt{\frac{b-a}{2k+1}}, k = 0, 1, 2, \cdots$$

form an orthogonal system of polynomials with respect to the L^2 inner product on the interval $[a, b]$.

5.5 Orthogonal Projections and Least Squares

An Orthogonal Projection of a point onto a subspace is finding the nearest distance between that point and the subspace.

A vector $\mathbf{z} \in V$ is said to be orthogonal to the subspace $W \subset V$ if it is orthogonal to every vector in W, so $\langle \mathbf{z}, \mathbf{w} \rangle = 0$ for all $\mathbf{w} in W$.

The orthogonal projection of $\mathbf{v}$ onto the subspace W is the element $\mathbf{w} \in W$ that makes the difference $\mathbf{z} = \mathbf{v} - \mathbf{w}$ orthogonal to W.

Let $\mathbf{u}_1, \cdots, \mathbf{u}_n$ be an orthonormal basis for the subspace $W \subset V$. Then the orthogonal projection of a vector $\mathbf{v} \in V$ onto W is

$$\mathbf{w} = c_1\mathbf{u}_1 + \cdots + c_n\mathbf{u}_n \text{ where } c_i = \frac{\langle \mathbf{v}, \mathbf{u_i} \rangle}{||\mathbf{u_i}||^2}, i = 1, \cdots, n$$

5.5.1 Orthogonal Least Squares

The orthogonal projection of a vector onto a subspace is also the least squares vector, the closest point in the subspace.

Let $W \subset V$ be a finite-dimensional subspace of an inner product space. Given a vector $\mathbf{v} \in W$, the closest point or least squares minimizer $\mathbf{w} \in W$ is the same as the orthogonal projection of $\mathbf{v}$ onto W.

5.5.2 Orthogonal Polynomials and Least Squares

The orthogonality of Legendre polynomials and their relatives helps us construct least squares approximates.

We can write the least squares approximate as a linear combination of Legendre Polynomials

$$p(t) = a_0P_0(t) + a_1P_1(t) + \cdots + a_nP_n(t) = a_0 + a_1t + a_2(\frac{3}{2}t^2 - \frac{1}{2}) + \cdots$$

The least squares coefficients can also be computed by the inner product formula, giving us the Rodrigues formula:

$$a_k = \frac{\langle \mathbf{e^t}, \mathbf{P_k} \rangle}{||\mathbf{P_k}||^2} = \frac{2k+1}{2} \int_{-1}^{1} e^t P_k(t)\, dt$$

5.6 Orthogonal Subspaces

We can extend the notion of orthogonailty from elements to subspaces.

Two subspaces $W, Z \subset V$ are called orthogonal if every vector in W is orthogonal to every vector in Z.

If $\mathbf{w}_1, \cdots, \mathbf{w}_k$ span W and $\mathbf{z}_1, \cdots, \mathbf{z}_l$ span Z, then W and Z are orthogonal subspaces if and only if $\langle \mathbf{w_i}, \mathbf{z_j} \rangle = 0$ for all $i = 1, \cdots, k$ and $j = 1, \cdots, l$.

The orthogonal complement to a subspace $W \subset V$, denoted $W^{\perp} = \{\mathbf{v} \in V | \langle \mathbf{v}, \mathbf{w} \rangle = 0 \text{ for all } \mathbf{w} \in W\}$.

Suppose that $W \subset V$ is a finite-dimensional subspace of an inner product space. Then every vector $\mathbf{v} \in V$ can be uniquely decomposed into $\mathbf{v} = \mathbf{w} + \mathbf{z}$ where $\mathbf{w} \in W$ and $\mathbf{z} \in W^{\perp}$.

If dim $W = m$ and dim $V = n$, then dim $W^{\perp} = n - m$.

If W is a finite-dimensional subspace of an inner product space, then $(W^{\perp})^{\perp} = W$.

5.6.1 Orthogonality of the Fundamental Matrix Subspaces and the Fredholm Alternative

Let A be a real $m \times n$ matrix. Then its kernel and corange are orthogonal complements as subspaces of $\mathbb{R}^n$ under the dot product, while its cokernel and range are orthogonal complements in $\mathbb{R}^m$, also under the dot product:

$$ker A = (corng A)^{\perp} \subset \mathbb{R}^n, coker A = (rng A)^{\perp} \subset \mathbb{R}^m$$

The linear system $A\mathbf{x} = \mathbf{b}$ has a solution if and only if $\mathbf{b}$ is orthogonal to the cokernel of A.

In other words, $A\mathbf{x} = \mathbf{b}$ has a solution ($\mathbf{b} \in Im(A)$) if and only if for any $\mathbf{y}$ such that $A^T\mathbf{y} = \mathbf{0}, \mathbf{y}^T\mathbf{b} = \mathbf{0}(\mathbf{b} \in ker(A^T)^{\perp}$.[4]

Multiplication by an $m \times n$ matrix A of rank r defines a one-to-one correspondence between the r-dimensional subspaces corng $A \subset \mathbb{R}^n$ and rng $A \subset \mathbb{R}^m$. Moreover, if $\mathbf{v}_1, \cdots, \mathbf{v}_r$ forms a basis of corng A then their images $A\mathbf{v}_1, \cdots A\mathbf{v}_r$ form a basis for rng A.

A compatible linear system $A\mathbf{x} = \mathbf{b}$ with $\mathbf{b} \in rng A = (coker A)^{\perp}$ has a unique solution $\mathbf{w} \in corng A$ satisfying $A\mathbf{w} = \mathbf{b}$. The general solution is $\mathbf{x} = \mathbf{w} + \mathbf{z}$ where $\mathbf{z} \in ker A$. The particular solution $\mathbf{w}$ is distinguished by the fact that it has the smallest Euclidean norm of all possible solutions: $||\mathbf{w}|| \leq ||\mathbf{x}||$ whenever $A\mathbf{x} = \mathbf{b}$.

[4]This is the specific Fredholm Alternative

6 Least Squares

Using Least Squares we can find the element in a subspace that is closest to a given point.

When trying to solve this problem, it's important to note that the goal is to minimize the squared distance

$$\|\mathbf{v}-\mathbf{b}\|^2 = \langle \mathbf{v}-\mathbf{b}, \mathbf{v}-\mathbf{b}\rangle = \|\mathbf{v}\|^2 - 2\langle \mathbf{v}, \mathbf{b}\rangle + ||\mathbf{b}||^2$$

over all possible $\mathbf{v} \in V$.

In other words, Let $\mathbf{v}_1, \cdots, \mathbf{v}_n$ form a basis for the subspace $V \subset \mathbb{R}^m$. Given $\mathbf{b} \in \mathbb{R}^m$, the closest point $\mathbf{v}* = x*_1 \mathbf{v}_1 + \cdots + x*_n \mathbf{v}_n \in V$ is prescribed by the solution $\mathbf{x}* = K^{-1}\mathbf{f}$ to the linear system $K\mathbf{x} = \mathbf{f}$, where K and $\mathbf{f}$ are given by

1. K is a symmetric $n \times n$ Gram matrix formed by

$$k_{ij} = \langle \mathbf{v}_i, \mathbf{v}_j \rangle$$

2. $\mathbf{f} \in \mathbb{R}^n$ formed by

$$f_i = \langle \mathbf{v}_i, \mathbf{b}\rangle$$

The distance between the point and the subspace is

$$d* = \|\mathbf{v}* - \mathbf{b}\| = \sqrt{||\mathbf{b}||^2 - \mathbf{f}^T\mathbf{x}*}$$

6.1 Least Squares

A least squares solution to a linear system of equations $A\mathbf{x} = \mathbf{b}$ is a vector $\mathbf{x}* \in \mathbb{R}^n$ that minimizes the Euclidean norm $\|A\mathbf{x} - \mathbf{b}\|$.

If the system has a solution, then it is automatically the least squares solution, therefore the concept of a least squares solution is only new when the system doesn't have a solution.

The least squares solution is unique if $kerA = \{\mathbf{0}\}$, or if the columns of A are linearly independent ($rankA = n$).

Assume $kerA = \{\mathbf{0}\}$. Set $K = A^TA$ and $\mathbf{f} = A^T\mathbf{b}$. Then the least squares solution to the linear system $A\mathbf{x} = \mathbf{b}$ is the unique solution $\mathbf{x}*$ to the so called normal equations

$$K\mathbf{x} = \mathbf{f} \equiv (A^TA)\mathbf{x} = A^T\mathbf{b}$$

namely

$$\mathbf{x}* = (A^TA)^{-1}A^T\mathbf{b}$$

and the least squares error is

$$\|A\mathbf{x}* - \mathbf{b}\|^2 = \|\mathbf{b}\|^2 - \mathbf{f}^T\mathbf{x}* = \|\mathbf{b}\|^2 - \mathbf{b}^TA(A^TA)^{-1}A^T\mathbf{b}$$

7 Linear Functions

Let V and W be real vector spaces. A function $L : V \to W$ is called linear if it obeys two basic rules:

1. $L[\mathbf{v} + \mathbf{w}] = L[\mathbf{v}] + L[\mathbf{w}]$
2. $L[c\mathbf{v}] = cL[\mathbf{v}]$

for all $\mathbf{v}, \mathbf{w} \in V$ and all scalars c.

Every linear function $L : \mathbb{R}^n \to \mathbb{R}^m$ is given by matrix multiplication $L[\mathbf{v}] = A\mathbf{v}$ where A is an $m \times n$ matrix.

7.1 The Space of Linear Functions

Given the vector spaces V, W, we use $\mathcal{L}(V, W)$ to denote the set of all linear functions: $L : V \to W$.

The dual space to a vector space V is defined as the vector space $V* = \mathcal{L}(V, \mathbb{R})$ consisting of all real-valued linear functions $L : V \to \mathbb{R}$.

Let V be a finite dimensional real inner product space. Then every linear function $L : V \to \mathbb{R}$ is given by an inner product with a fixed vector $\mathbf{a} \in V$:

$$L[\mathbf{v}] = \langle \mathbf{a}, \mathbf{v} \rangle$$

7.2 Composition

Besides adding and multiplying by scalars, one can also compose linear functions.

Let V, W, Z be vector spaces. If $L : V \to W$ and $M : W \to Z$ are linear functions, then the composite function $M \circ L : V \to Z$, defined by $(M \circ L)[\mathbf{v}] = M[L[\mathbf{v}]]$ is linear.

7.3 Inverses

Let $L : V \to W$ be a linear function. If $M : W \to V$ is a function such that both compositions

$$L \circ M = I_W, M \circ L = I_V$$

are equal to the identity function, then we call M the inverse of L and write $M = L^{-1}$.

If it exists, the inverse of a linear function is also a linear function.

8 Linear Transformations

If we consider a linear function that maps n dimensional space to itself, we can also consider that the function maps a point $\mathbf{x} \in \mathbb{R}^n$ to its image point $L[\mathbf{x}] = A\mathbf{x}$, where A is its $n \times n$ representative. This can be referred to as a linear transformation.

Most of the important classes of linear transformations already appear in the two dimensional case. Every linear function $L : \mathbb{R}^2 \to \mathbb{R}^2$ has the form

$$L\begin{pmatrix}x\\y\end{pmatrix}=\begin{pmatrix}ax+by\\cx+dy\end{pmatrix},\text{ where } A=\begin{pmatrix}a & b\\c & d\end{pmatrix}$$

is an arbitrary 2×2 matrix.

8.1 Change of Basis

Let $L : V \to W$ be a linear function. Suppose V has basis $\mathbf{v}_1, \cdots, \mathbf{v}_n$ and W has basis $\mathbf{w}_1, \cdots, \mathbf{w}_m$. We can write

$$\mathbf{v} = x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_n \in V, \mathbf{w} = y_1\mathbf{w}_1 + \cdots + y_m\mathbf{w}_m \in W$$

9 Eigenvalues and Eigenvectors

Let A be an $n \times n$ matrix. A scalar, λ is called an eigenvalue of A if there is a non-zero vector $\mathbf{v} \neq \mathbf{0}$, called an eigenvector, such that

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

In other words, the matrix A stretches the vector $\mathbf{v}$ by a certain value, λ.

To find these values and vectors, we construct the equation:

$$(A - \lambda\mathbf{I}) = \mathbf{0}$$

Note, the scalar λ is an eigenvalue of the matrix A iff $A - \lambda\mathbf{I}$ is singular $(rank < n)$. The corresponding eigenvectors are the nonzero solutions to the eigenvalue question.

Also, a scalar λ is an eigenvalue of the matrix A iff λ is a solution to the characteristic equation

$$\det(A - \lambda\mathbf{I}) = 0$$

If A is a real matrix with a complex eigenvalue and corresponding complex eigenvector, then the complex conjugate is also an eigenvalue.

9.1 Basic Properties of Eigenvalues

If A is an $n \times n$ matrix, then its characteristic polynomial is

$$p_A(\lambda) = \det(A - \lambda\mathbf{I}) = c_n\lambda^n + c_{n-1}\lambda^{n-1} + \cdots + c_1\lambda + c_0$$

An $n\times n$ matrix A has at least one, and at most n distinct complex eigenvalues.

The sum of the eigenvalues of a matrix equals its trace, while the product equals its determinant.

9.2 Eigenvector Bases

If $\lambda_1, \cdots, \lambda_k$ are distinct eigenvalues of the same matrix A, then the corresponding eigenvectors $\mathbf{v}_1, \cdots, \mathbf{v}_k$ are linearly independent.

And if the $n \times n$ real matrix A has n distinct real eigenvalues $\lambda_1, \cdots, \lambda_n$, then the corresponding real eigenvectors form a basis of $\mathbb{R}^n$. If A (which may now be either a real or a complex matrix) has n distinct complex eigenvalues, then the corresponding eigenvectors form a basis of $\mathbb{C}^n$.

An eigenvalue λ of a matrix A is called complete if the corresponding eigenspace $V_\lambda = ker(A - \lambda\mathbf{I})$ has the same dimension as its multiplicity. The matrix A is complete if all its eigenvalues are.

An $n \times n$ real or complex matrix A is complete iff its eigenvectors span $\mathbb{C}^n$. In particular, any $n \times n$ matrix that has n distinct eigenvalues is complete.

9.3 Diagonalization

A square matrix is called diagonalizable if there exists a nonsingular matrix S and a diagonal matrix $\Lambda = diag(\lambda_1, \cdots, \lambda_n)$ such that

$$S^{-1}AS = \Lambda \text{ or } A = S\Lambda S^{-1}$$

A matrix is complex diagonalizable iff it is complete. A matrix is real diagonalizable iff it is complete and has all real eigenvalues.

9.4 Eigenvalues of Symmetric Matrices

Let $A = A^T$ be a real symmetric $n \times n$ matrix. Then

1. All the eigenvalues of A are real.
2. Eigenvectors corresponding to distinct eigenvalues are orthogonal.
3. There is an orthonomral basis of $\mathbb{R}^n$ consisting on n eigenvectors of A.

In particular, all symmetric matrices are complete.

A symmetric matrix $K = K^T$ is positive definite iff all of its eigenvalues are strictly positive.

Let $A = A^T$ be an $n \times n$ symmetric matrix. Let $\mathbf{v}_1, \cdots, \mathbf{v}_n$ be an orthogonal eigenvector basis such that $\mathbf{v}_1, \cdots, \mathbf{v}_r$ correspond to nonzero eigenvalues, while $\mathbf{v}_{r+1}, \cdots, \mathbf{v}_n$ are null eigenvectors corresponding to the zero eigenvalue (if any). Then $r = rank(A)$, the non-null eigenvectors form an orthogonal basis for $rng(A) = corng(A)$, while the null eigenvectors form an orthogonal basis for $ker(A) = coker(A)$.

9.5 The Spectral Theorem

Let A be a real, symmetric matrix. Then there exists an orthogonal matrix Q such that

$$A = Q\Lambda Q^{-1} = Q\Lambda Q^T$$

10 Singular Values

The singular values of an $m \times n$ matrix A are the positive square roots $\sigma_i = \sqrt{\lambda_i} > 0$, of the nonzero eigenvalues of the associated Gram matrix $K = A^T$. The corresponding eigenvectors of K are known as the singular vectors of A.

Any nonzero, real $m \times n$ matrix A of rank $r > 0$ can be factored $A = P\Sigma Q^T$ into the product of an $m \times r$ matrix P with orthonormal columns, so $P^T P = I$, the $r \times r$ diagonal matrix $\Sigma = diag(\sigma_1, \cdots, \sigma_r)$ that has the singular values of A as its diagonal entries, and an $r \times n$ matrix Q^T with orthonormal rows, so $Q^T Q = I$.

Given the singular value decomposition $A = P\Sigma Q^T$, the columns of Q form an orthonormal basis for for $corng(A)$, while the columns of P form an orthonormal basis for $rng(A)$.

The condition number of a matrix is the ratio between its largest and smallest singular values: $K(A) = \sigma_1/\sigma_2$.

The pseudoinverse of a nonzero $m \times n$ matrix with singular value decomposition $A = P\Sigma Q^T$ is the $n \times m$ matrix $A^+ = Q\Sigma^{-1}P^T$.

Let A be an $m \times n$ matrix of rank n. Then $A^+ = \left(A^T A\right)^{-1} A^T$

Consider the linear system $A\mathbf{x} = \mathbf{b}$. Let $\mathbf{x}^* = A^+\mathbf{b}$, where A^+ is the pseudoinverse of A. If $\ker A = \{\mathbf{0}\}$, then $\mathbf{x}^*$ is the Euclidean least squares solution to the linear system. If, more generally, $\ker A \neq \{\mathbf{0}\}$ then $\mathbf{x}^* \in corng(A)$ is the least squares solution of minimal Euclidean norm among all vectors that minimize least squares error $\|A\mathbf{x} - \mathbf{b}\|$.

11 Incomplete Matrices

A Complex, square matrix U is called unitary if is satisfies $U^\dagger U = I$ where $U^\dagger = \overline{U^T}$ denotes the Hermitian transpose where one first transposes and then takes the complex conjugate of all the entries.

If two matrices are unitary, then so is their product.

A Attachments

LaTeX Source Code