Linear Algebra Review

1. The Field

Definition of a Field

The notion of a *field* is a collection of values with a *plus* operation and a *times* operation. Three examples are the fields of *real numbers*, *complex numbers*, and *GF(2)*(the field consisting just of zero and one).

Complex Number Defitions and Properties

- $\boldsymbol{i}^2 = -1$ or alternatively, $\boldsymbol{i} = \sqrt{-1}$
- The sum of a real number and an imaginary number is called a *complex number.* A complex number has a *real part* and an *imaginary part.*
- The *conjugate* of a complex number z, written $\overline{z}$, is defined as $z.real - z.imag$.
- An alternative definition for the absolute value or the distance from the origin to the point in the complex plane: $|z|^2 = z \cdot \overline{z}$
- A translation is $f(z) = z_0 + z$ where z_0 is a complex number.

Definition of Galois Field GF(2)

It consists of two elements, 0 and 1.

- The arithmetic operation is $a + b \mod 2$. It can be thought of as as an *exclusive-or.*
- The multiplication should be treated as standard multiplication.

Abstracting over Fields

Linear Algebra concepts, theorems, and procedures work because of the definition of a *field* and that these fields satisfy basic laws such commutativity and distributivity.

2. The Vector

Definition 2.1.1

A vector with four entries, each of which is a real number, is called a 4-*vector over* $\mathbb{R}$. The entries of a vector must all be drawn from a single field.

Definition 2.1.2

For a field $\mathbb{F}$ and a positive integer n, a vector with n entries, each belonging to $\mathbb{F}$, is called an n-vector over $\mathbb{F}$. The set of n-vectors over $\mathbb{F}$ is denoted $\mathbb{F}^n$.

Definition 2.2.2

For a finite set D and a field $\mathbb{F}$, a D-vector over $\mathbb{F}$ is a function from D to $\mathbb{F}$.

Sparsity

- spare representation is omission of key-value pairs whose values are zero
- a vector most of whose values are zero is called a *sparse* vector
- an k-sparse vector is defined as a vector that contains no more than k non-zero entries

Definition 2.4.1

Addition of n-vectors is defined in terms of addition of corresponding entries:

$$[u_1, u_2, ..., u_n] + [v_1, v_2, ..., v_n] = [u_1 + v_1, u_2 + v_2, ..., u_n + v_n]$$

Proposition 2.4.5 Associativity and Commutativity of Vector Addition

For any vectors **u**, **v**, **w**,

$$(\boldsymbol{u} + \boldsymbol{v}) + \boldsymbol{w} = \boldsymbol{u} + (\boldsymbol{v} + \boldsymbol{w})$$

and

$$\boldsymbol{u} + \boldsymbol{v} = \boldsymbol{v} + \boldsymbol{u}$$

Zero Element

Every field $\mathbb{F}$ has a zero element, so the set $\mathbb{F}^D$ of D-vectors over $\mathbb{F}$ necessarily has a *zero vector*, a vector all of whose entries have value zero. This is denoted by $\mathbf{0}_D$, or merely by $\mathbf{0}$ if it is not necessary to specify D.

Proposition 2.5.5 Associativity of scalar-vector multiplication

$$\alpha(\beta\boldsymbol{v}) = (\alpha\beta)\boldsymbol{v}$$

Proposition 2.6.5 Scalar-vector multiplication distributes over scalar and vector addition

$$(\alpha + \beta)\boldsymbol{u} = \alpha\boldsymbol{u} + \beta\boldsymbol{u}$$
$$\alpha(\boldsymbol{u} + \boldsymbol{v}) = \alpha\boldsymbol{u} + \alpha\boldsymbol{v}$$

Convex Combination

An expression of the form $\alpha\boldsymbol{v} + \beta\boldsymbol{v}$ where $\alpha, \beta >= 0$ and $\alpha + \beta = 1$ is called a *convex combination* of $\boldsymbol{u}$ and $\boldsymbol{v}$. This is true for any pair $\boldsymbol{u}, \boldsymbol{v}$ of distinct n-vectors over $\mathbb{R}$. The $\boldsymbol{u}$-to-$\boldsymbol{v}$ line segments consists of the set of convex combinations of $\boldsymbol{u}$ and $\boldsymbol{v}$.

Affine Combination

An expression of the form $\alpha\boldsymbol{u} + \beta\boldsymbol{v}$ where $\alpha + \beta = 1$ is called an *affine combination* of $\boldsymbol{u}$ and $\boldsymbol{v}$. The line through $\boldsymbol{u}$ and $\boldsymbol{v}$ consists of the set of affine combinations of $\boldsymbol{u}$ and $\boldsymbol{v}$.

Invertible Functions

A function is invertible if it is *one-to-one* and *onto*. Every function has at most *one functional inverse.*

Identity Function

Function g is the functional inverse of f if $f \circ g$ and $g \circ f$ are the identity functions on their domains.

Vector Subtraction

Vector subtraction is defined in terms of vector addition and negative: **u-v** is defined as $\boldsymbol{u} + (-\boldsymbol{u})$. Vector subtraction is the functional inverse of vector addition, and this fact can be verified with a composition.

2.9 Dot-product / Scalar product

For two D-vectors $\boldsymbol{u}$ and $\boldsymbol{v}$, the dot-product is the sum of the product of the corresponding entries:

$$\boldsymbol{u} \cdot \boldsymbol{v} = \sum_{k \in D} \boldsymbol{u}[k]\boldsymbol{v}[k]$$

Dot-product can be used to measure similarity between vectors over $\mathbb{R}$.

2.9.8 Algebraic properties of the dot-product

These properties hold regardless of the choice of the field.

Dot-product Commutativity When you take a dot-product of two vectors, the order of the two does not matter:

$$\boldsymbol{u} \cdot \boldsymbol{v} = \boldsymbol{v} \cdot \boldsymbol{u}$$

Dot-product Homogeneity

$$(\alpha\boldsymbol{u}) \cdot \boldsymbol{v} = \alpha(\boldsymbol{u} \cdot \boldsymbol{v})$$

Dot-product distributes over vector addition

$$(\boldsymbol{u} + \boldsymbol{v}) \cdot \boldsymbol{w} = \boldsymbol{u} \cdot \boldsymbol{w} + \boldsymbol{v} \cdot \boldsymbol{w}$$

Definition 2.9.6 Linear Equation

A *linear equation* is an equation of the form $\boldsymbol{\alpha} \cdot \boldsymbol{x} = \beta$, where $\boldsymbol{\alpha}$ is a vector, β is a scalar, and $\boldsymbol{x}$ is a vector variable. The scalar β is called the right-hand size of the linear equation because it is conventionally written on the right of the equals sign.

Definition 2.9.10 Linear System

In general, a system of *linear equations*(often abbreviated linear system) is a collection of equations:

$$\boldsymbol{a}_1 \cdot \boldsymbol{x} = \beta_1$$
$$\boldsymbol{a}_2 \cdot \boldsymbol{x} = \beta_2$$
$$\vdots$$
$$\boldsymbol{a}_m \cdot \boldsymbol{x} = \beta_m$$

where $\boldsymbol{x}$ is a vector variable. A solution is a vector $\hat{\boldsymbol{x}}$ that satifies all the equations.

Linear Filter

Searching for a short audio segment in a longer audio segment by padding the short audio segment with zeroes is called a *linear filter.* The short segment plays the role of a *kernel.*

Triangular System of Linear Equations

This representation of a "square" linear system whose upper-right triangle consists of possibly non-zero values but the lower-left must consist of zero values.

Backward Substitution A triangular system can be solved using *backward substitution* by starting by solving the last row first and solving each previous row one by one.

Proposition 2.11.5 For a triangular system specified by a length-n list rowlist of n-vectors and an n-vector $\boldsymbol{b}$, if rowlist$[i][i] \neq 0$ for $i = 0, 1, \ldots, n-1$ then the solution found by triangularsolven(rowlist, b) is the only solution to the system.

Proposition 2.11.6 For a length-n list rowlist of n-vector, if rowlist$[i][i] = 0$ for integer i then there is a vector $\boldsymbol{b}$ for which the triangular system has no solution.

3. Vector Space

3.1.1 Definition of Linear Combination

Suppose $\boldsymbol{v}_1, ..., \boldsymbol{v}_n$ are vectors. We define a *linear combination* of $\boldsymbol{v}_1, ..., \boldsymbol{v}_n$ to be a sum $\alpha_1\boldsymbol{v}_1 + ... + \alpha_n\boldsymbol{v}_n$ where $\alpha_1, ..., \alpha_n$ are scalars.

Trivial Linear Combination A linear combination is considered a trivial linear combination if all of its coefficients are zero.

Forward Problem

Given an element of the domain, find the image under the function.

Backward Problem

Given an element of the co-domain, find the pre-image, if one exists.

3.2.1 Span

The set of all linear combinations of vectors $\boldsymbol{v}_1, ..., \boldsymbol{v}_n$ is called the span of these vectors, and is written Span $\{\boldsymbol{v}_1, ..., \boldsymbol{v}_n\}$.

- The span of an empty set is the *zero vector*, which can be expressed as Span$\{\} = \{0\}$. The zero vector is also the origin.
- A geometric object such as a point, a line, or a plane is called a *flat.* It can also be considered an affine space that is a subset of $\mathbb{R}^n$ for some n.

Mathematically expressed as

$$Span\{\boldsymbol{v}\} = \{\alpha\ \boldsymbol{v} : \alpha \in \mathbb{R}\}$$

Flat Dimensionality The span of k vectors over $\mathbb{R}$ forms a k-dimensional flat containing the origin or a flat of a lower dimension containing the origin. A flat geometric object such as a point, a line, or a plane is called a *flat.* Flats exist in higher dimensions as well.

Definition 3.2.9 Generators

Let V be a set of vectors. If $\boldsymbol{v}_1, ..., \boldsymbol{v}_n$ such that $V = Span\{\boldsymbol{v}_1, ..., \boldsymbol{v}_n\}$ then $\{\boldsymbol{v}_1, ..., \boldsymbol{v}_n\}$ is a *generating set* for V, and we refer to the vectors $\boldsymbol{v}_1, ..., \boldsymbol{v}_n$ as *generators* for V.

3.2.9 Generators

Let $\mathcal{V}$ be a set of vectors. If $\boldsymbol{v}_1, ..., \boldsymbol{v}_n$ are vectors such that $\mathcal{V}$ = Span $\{\boldsymbol{v}_1, ..., \boldsymbol{v}_n\}$ then we say that $\{\boldsymbol{v}_1, ..., \boldsymbol{v}_n\}$ is the *generating set* for $\mathcal{V}$, and we refer to the vectors $\boldsymbol{v}_1, ..., \boldsymbol{v}_n$ as the *generators* for $\mathcal{V}$.

Standard Generators

$$[x, y, z] = x[1, 0, 0] + y[0, 1, 0] + z[0, 0, 1]$$

The coordinate representation of [x,y,z] in terms of these generators is [x,y,z]. These are the *standard* generators for $\mathbb{R}^3$. We denote them by $\boldsymbol{e}_0$, $\boldsymbol{e}_1$, and $\boldsymbol{e}_2$. For any finite domain D and field $\mathbb{F}$, there are standard generators for $\mathbb{F}^D$. For each $k \in D$, $\boldsymbol{e}_k$ is the function $k : 1$. It maps k to 1 and the maps all the other domain elements to zero.

3.3.8 Definition Homogeneous linear equation

A linear equation with a right-hand size of zero is called a *homogeneous* linear equation.

3.3.11 Definition Homogeneous linear system

A linear system (collection of linear equations) with all right-hand sides zero is called a *homogeneous* linear system.

Note: A *flat* containing the origin is the solution set of a homogeneous linear system.

Vector Space Properties The following three properties apply regardless of whether this is a solution set of a linear system or a subset $\mathcal{V}$ of $\mathbb{F}^D$, where $\mathcal{V}$ is the span of some D-vectors over $\mathbb{F}$

1. $\mathcal{V}$ contains the zero vector
2. For every vector $\boldsymbol{v}$, if $\mathcal{V}$ contains $\boldsymbol{v}$ then it contains $\alpha\boldsymbol{v}$ for every scalar α, is closed scalar-vector multiplication.
3. For every pair $\boldsymbol{u}$ and $\boldsymbol{v}$ of vectors, if $\mathcal{V}$ contains $\boldsymbol{u}$ and $\boldsymbol{v}$, then it contains $\boldsymbol{u}+\boldsymbol{v}$. In other words, it's closed under vector addition.

Definition 3.4.1 Vector Space Definition A set $\mathcal{V}$ of vectors is called a *vector space* if it satisfies the three vector space properties.

- $\mathbb{F}$ is a vector space because it contains the zero vector and is closed under scalar-vector multiplication and vector addition.
- the span of some vectors is a vector space
- the solution set of a homogeneous linear system is a vector space
- flats, such as a line or a plane, that contain the origin can be written as the span of some vectors or as the solution set of a homogeneous linear system, and therefore such a flat is a vector space

Proposition 3.4.6

For any field $\mathbb{F}$ and any finite domain D, the singleton set consisting of the zero vector $\mathbf{0}_D$ is a vector space. A flat containing the origin is the solution set of a homogeneous linear system.

Definition 3.4.7

A vector space consisting only of the zero vector is a *trivial* vector space. Furthermore, the minimum number of vectors whose span is $\{\mathbf{0}_D\}$ is zero, which is computed from the empty set of D-vectors.

Definition 3.4.9

$\mathcal{V}$ and $\mathcal{W}$ are vector spaces and $\mathcal{V}$ is a subspace of $\mathcal{W}$, we say $\mathcal{V}$ is a *subspace* of $\mathcal{W}$.

Affine Spaces

- $\{\boldsymbol{a}+\boldsymbol{v} : \boldsymbol{v} \in \mathcal{V}\}$ can be abbrevated as $\boldsymbol{a} + \mathcal{V}$, where $\mathcal{V}$ is a vector space

Definition 3.5.2 Affine Combination

A linear combination of $\alpha_1\boldsymbol{v}_1 + ... + \alpha_n\boldsymbol{v}_n$ is said to be an *affine combination*, if the coefficients $\alpha_1 + ... + \alpha_n$ sum to one.

Affine Hull

The set of all affine combinations of a collection of vectors is called an *affine hull* of that collection.

- the affine hull of a one-vector collection is a single point (the one vector in the collection), i.e. a 0-dimensional object
- the affine hull of a two-vector collection is a line(the line through the two vectors), i.e. a 1-dimensional object
- the affine hull of a three-vector collection is a plane (the plane through the three vectors), i.e. a 2-dimensional object

Definition 3.5.8 Affine Space

An *affine space* is the result of translating a vector space. That is, a set $\mathcal{A}$ is an affine space if there is a vector $\boldsymbol{a}$ and a vector space $\mathcal{V}$ such that

$$\mathcal{A} = \{\boldsymbol{a}+\boldsymbol{v} : \boldsymbol{v} \in \mathcal{V}\}$$

i.e. $\mathcal{A} = \boldsymbol{a} + \mathcal{V}$. A *flat* is just an affine space that is a subset of $\mathbb{R}^n$ for some n.

Lemma 3.5.10

For any vectors $\boldsymbol{u}_1, ..., \boldsymbol{u}_n$,

$$\left\{\alpha_1\boldsymbol{u}_1 + ... + \alpha_n\boldsymbol{u}_n : \sum_{i=1}^{n}\alpha_i = 1\right\} = \{\boldsymbol{u}_1 + \boldsymbol{v} : \boldsymbol{v} \in Span\,\{\boldsymbol{u}_2 - \boldsymbol{u}_1, ..., \boldsymbol{u}_n - \boldsymbol{u}_1\}\}.$$

In words, the affine hull of $\boldsymbol{u}_1, ..., \boldsymbol{u}_n$ equals the set obtained by adding $\boldsymbol{u}_1$ to each vector in the span of $\boldsymbol{u}_2 - \boldsymbol{u}_1, \ldots, \boldsymbol{u}_n - \boldsymbol{u}_1$. This shows that the affine hull of vectors is an affine space.

- The solution set of a homogeneous linear system is a vector space. The solution set of an arbitrary linear system is an affine space except in the case where the solution set is empty.

Two Representations of an Affine Space

- as $\boldsymbol{a} + \mathcal{V}$ where $\mathcal{V}$ is the span of some vectors, and
- as the affine hull of some vectors.

Theorem 3.6.2 Solution is empty or an Affine Space

For any linear system, the set of solutions either is empty or is an affine space. Every vector space is the solution space of a homogeneous system. Every affine space is a solution set of a linear system.

Corollary 3.6.4 Solution Uniqueness

Suppose a linear system has a solution. The solution is unique if and only if the only solution to the corresponding homogeneous linear system is the zero vector.

4. Matrix

As we have defined a D over of $\mathbb{F}$ to be a function from a set D to $\mathbb{F}$, so we define a $R \times C$ matrix over $\mathbb{F}$ to be a function from the Cartesian product $R \times C$. We refer to the elements of R as row labels and we refer to the elements of C as column labels.

Definition 4.1.6 Identity Matrix

For a finite set D, the $D \times D$ matrix is the matrix whose row-labels and column-labels both belong to D. There is a 1 for every $(d, d), d \in D$ entry, there is a 1 in its place. All the remaining entries are 0. This is denoted $\mathbb{1}_D$ or simply $\mathbb{1}$.

Definition 4.2.1 Column Space and Row Space

For a matrix M,

- *column space* of M, written Col M, is the *vector space* spanned by the columns of M, and
- *row space* of M, written Row M, is the *vector space* spanned by the rows of M.

For example, if c1, c2, and c3 are the columns of a matrix M, then the $ColM = Span\{c1, c2, c3\}$.

Definition 4.4.1 Transpose

Transposing a matrix means swapping its rows and columns. The transpose of an $P \times Q$ matrix, written M^T, is a $Q \times P$ matrix such that $(M^T)_{j,i} = M_{i,j}$ for every $i \in P, j \in Q$. We say a matrix M is *symmetric matrix* if $M^T = M$.

Matrix-vector and Vector-matrix Multiplication Outline

- Linear-combinations of Matrix-Vector multiplication: multiplying *columns*
- Dot-product of Matrix-Vector multiplication: multiplying *rows*
- Linear-combinations of Vector-Matrix multiplication: multiplying *rows*
- Dot-product of Vector-Matrix multiplication: multiplying *columns*

Definition 4.5.1 Linear-combinations definition of matrix-vector multiplication

Let M be an $R \times C$ matrix over $\mathbb{F}$. Let $\boldsymbol{v}$ be a C-vector over $\mathbb{F}$. Then $M * \boldsymbol{v}$ is the linear combination

$$\sum_{c \in C} \boldsymbol{v}[c] - \text{column c of M}$$

If M is an $R \times C$ matrix but v is not a C-vector then the product $M * v$ is illegal. In the traditional-matrix sense, if M is an $m \times n$ matrix over $\mathbb{F}$ then $M * v$ is legal only if v is an n-vector over $\mathbb{F}$. That is, the number of columns of the matrix must match the number of entries of the vector.

Definition 4.5.6 Linear-combinations definition of vector-matrix multiplication

Let M be an $R \times C$ matrix over $\mathbb{F}$. Let $\boldsymbol{w}$ be a R-vector. Then $\boldsymbol{w} * M$ is the linear combination

$$\sum_{r \in R} \boldsymbol{w}[r] - \text{row r of M}$$

If M is an $R \times C$ matrix but $\boldsymbol{w}$ is not an R-vector then the product $\boldsymbol{w} * M$ is illegal. This is a good moment to point out that matrix-vector multiplication is different from vector-matrix multiplication; in fact, often $M * \boldsymbol{v}$ is a legal product but $\boldsymbol{v} * M$ is not or vice versa. Because we are use to assuming commutativity when we multiply numbers, the noncommutativity of multiplication between matrices and vectors can take some getting used to.

Computational Residual

The computed value of x in the linear equation $M * x = \boldsymbol{b}$ might not yield a solution. Furthemore, the computed vector might not be the exact solution. It is necessary to compute the residual, $(\boldsymbol{b} - M * x)^2$, to determine if it is indeed a solution. If it's an exact solution, the residual will be 0. Depending on the domain, this computation might be not be necessary if the domain is exact, e.g., $GF(2)$.

Definition 4.6.1 (Dot-Product Definition of Matrix-Vector Multiplication)

If M is an $R \times C$ matrix and $\boldsymbol{u}$ is a C-vector then $M * \boldsymbol{u}$ is the R-vector $\boldsymbol{v}$ such that $\boldsymbol{v}[r]$ is the dot-product of row r of M with $\boldsymbol{v}$.

Definition 4.6.3 (Dot-Product Definition of Vector-Matrix Multiplication)

If M is an $R \times C$ matrix and $\boldsymbol{u}$ is a R-vector then $\boldsymbol{u} * M$ is the C-vector $\boldsymbol{v}$ such that $\boldsymbol{v}[c]$ is the dot-product of $\boldsymbol{v}$ with column c of M.

Upper Triangular Matrix

A $n \times n$ upper-triangle matrix is a matrix whose $A_{i,j} = 0$ if $i > j$. Visually, this means the lower left portion of the matrix is 0. The entries forming the triangle can be zero or non-zero. The definition applies to traditional matrices. To generalize our matrices with arbitrary row- and column-label sets, we specify the orderings of the label-sets.

Definition 4.10.20 Diagonal Matrix

For a domain $D \times D$, a matrix M is a diagonal matrix if for all $M_{i,j} = 0$ where $i \neq j$. Visually, only the diagonal is allowed to be non-zero. The identity matrix is a special case of a diagonal matrix, where $x = 1$ instead of a variable coefficient.

Proposition 4.6.13 Algebraic properties of matrix-vector multiplication

Let M be an $R \times C$ matrix. * For any C-vector v and any scalar α,

$$M * (\alpha \boldsymbol{v}) = \alpha(M * \boldsymbol{v})$$

- For any C-vectors v and u,

$$M * (\boldsymbol{v} + \boldsymbol{u}) = M * \boldsymbol{v} + M * \boldsymbol{u}$$

Definition 4.7.1 Null Space

The *Null Space* of a matrix is the set $\{\boldsymbol{v} : A * \boldsymbol{v} = \boldsymbol{0}\}$. It is written $NullA$. This is the equivalent of a homogeneous linear sytem formulated as a matrix-vector equation. Since the Null Space is a homogeneous solution, it inherits its properties, such as being a vector space.

Lemma 4.7.4

For any $R \times C$ matrix A and C-vector $\boldsymbol{v}$, a vector z is in the null space of A if and only if $A * (\boldsymbol{v} + z) = A * \boldsymbol{v}$.

Corollary 4.7.5

We know that if two solutions to a system of linear equations differ by a vector that is a solution that solves a corresponding homogeneous linear system. The same principle applies to the matrix-vector representation.

Suppose $\boldsymbol{u}_1$ is a solution fo the matrix equation $A * \boldsymbol{x} = \boldsymbol{b}$. Then $\boldsymbol{u}_2$ is also a solution of and only if $\boldsymbol{u}_1 - \boldsymbol{u}_2$ belongs to the null space of A.

Corollary 4.7.6

Suppose a matrix-vector equation $M * \boldsymbol{x} = \boldsymbol{b}$ has a solution. The solution is unique if and only if the null space of A consists solely of the zero vector, the trivial solution.

Definition 4.10.1 Linear Function

Let U and V be vector spaces over $\mathbb{F}$. Let $f : U \longrightarrow V$ is called a *linear function* if it satisfies the following two properties:

- For any vector $\boldsymbol{u}$ in the domain of f and a scalar α, $f(\alpha \boldsymbol{u}) = \alpha f(\boldsymbol{u})$
- For any two vectors $\boldsymbol{v}$ and $\boldsymbol{u}$ in the domain of f, $f(\boldsymbol{u} + \boldsymbol{v}) = f(\boldsymbol{u}) + f(\boldsymbol{v})$

A synonym for a *linear function* is a *linear transformation.* Let M be an $R \times C$ matrix over a field $\mathbb{F}$, and define

$$f : \mathbb{F}^C \longrightarrow \mathbb{F}^R$$

by $f(\boldsymbol{x}) = M * \boldsymbol{x}$. The domain and co-domain are the vectors spaces. By the algebraic properties of matrix-vector multiplication(proposition 4.6.13), f is also a linear function. For any matrix M, the function $\boldsymbol{x} \mapsto M * \boldsymbol{x}$ is a linear function.

Bilinearity of dot-product

The dot product function $f(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{x} \cdot \boldsymbol{y}$ is linear in both its first and second arguments. We say that the dot-product function is bilinear to mean that is linear in each of its arguments.

Linear Functions: Zero Vector

If $f : U \longrightarrow V$ is a linear function, then the zero vector of U maps to the zero vector of V.

Definition 4.10.11 Kernels of Linear Functions

Analogous to the definition of the null space of a matrix, the kernel of a linear function is $\{\boldsymbol{v} : f(\boldsymbol{v}) = 0\}$. It is written as the Ker f.

Lemma 4.10.15 One-to-One Lemma

A linear function is one-to-one if and only if kernel is a trivial vector space.

Lemma 4.10.19

If $f : \mathbb{F}^C \longrightarrow \mathbb{F}^R$ is a linear function, then there is an $R \times C$ over $\mathbb{F}$ $f(\boldsymbol{x}) = M * \boldsymbol{x}$ where $\boldsymbol{x}$ is a vector $\boldsymbol{x} \in \mathbb{F}^C$.

Matrix-Multiplication Lemma of Functional Composition

The matrices A and B define linear functions via matrix-vector multiplication: $f_A(\boldsymbol{y}) = A{*}\boldsymbol{y}$ and $f_B(\boldsymbol{x}) = B{*}\boldsymbol{x}$. Naturally, the matrix AB resulting from multiplying the two functions is $f_{AB}(\boldsymbol{x}) = (AB) * \boldsymbol{x}$. This implies function composition.

$$f_{AB} = f_A \circ f_B$$

Matrix Multiplication is Associative

$$(AB)C = A(BC)$$

Proposition 4.11.14 Transpose of Matrix-Matrix Product

For matrices A and B, $(AB)^T = B^T A^T$

Column Vector

A column vector is a matrix M with one column that acts like vector multiplied from its left.

Row Vector

A row vector is a matrix M with one row that acts like a vector multiplied from the right.

Inner Product

Let $\boldsymbol{u}$ and $\boldsymbol{v}$ be two D-vectors. Consider the matrix-matrix product of $\boldsymbol{u}^T\boldsymbol{v}$, the result of which is a dot product that yields a single value.

Outer Product

Let $\boldsymbol{u}$ and $\boldsymbol{v}$ be two vectors not necessarily of the same domain. Consider the matrix-matrix product of $\boldsymbol{u}\boldsymbol{v}^T$.

Inverse of a Linear Function is linear

If f is a linear function and g is its inverse, then g is also a linear function.

Matrix Inverse

Let A be an $R \times C$ matrix over $\mathbb{F}$, and B be a $C \times R$ matrix over $\mathbb{F}$. Define function $f : \mathbb{F}^C \longrightarrow \mathbb{F}^R$ by $f(\boldsymbol{x}) = A * \boldsymbol{x}$ and $g : \mathbb{F}^R \longrightarrow \mathbb{F}^C$ by $g(\boldsymbol{y}) = B * \boldsymbol{y}$. If f and g are functional inverses, then matrices A and B are said to be inverses of each other. If A has an inverse, then we say A is an *invertible matrix*. It can be shown using the uniqueness of a functiona inverse(Lemma 0.3.19) that a matrix has at most one inverse. Matrix inverses are denoted by A^{-1}. A matrix that is not invertible is called a *singular matrix*.

Uses of Matrix Inverse

If $R \times C$ has matrix A has an inverse A^{-1}, then AA^{-1} is the $R \times R$ identity matrix.

Lemma 4.13.13 Diagonal Matrix Invertability

Suppose A is an upper-triangle matrix. Then A is invertible if and only if none of its diagonal elements is zero.

Proposition 4.13.14

If A and B are invertible matrices and matrix product AB is defined, the AB is an invertible matrix and $(AB)^{-1} = B^{-1}A^{-1}$.
Note: A and B do not necessarily have to be inverses of one another! These could be any random A and B matrices that have the aforementioned properties.

Corollary 4.13.19

Matrices A and B are inverses of each other if and only if both AB and BA are identity matrices.

5. Vector Space

Coordinate System

The coordinate system for a vector space V is specified by its generators $\boldsymbol{a_1}, ..., \boldsymbol{a_n}$ of V. Every vector $\boldsymbol{v}$ can be written as a linear combination whose coefficients, or coordinates, are α and whose generators are a : $\boldsymbol{v} = \alpha_1\boldsymbol{a}_1 + ... + \alpha_n\boldsymbol{a}_n$. The vector $[\alpha_1, ..., \alpha_n]$ is called the *coordinate representation* of $\boldsymbol{v}$ in terms of $\boldsymbol{a_1}, ..., \boldsymbol{a_n}$.

Greedy algorithms for finding a set of generators

Given a vector space V, find the minimum number of vectors whose span is equal to V.

Grow algorithm Start with an empty set B whose span will ultimately contain the generators for V. For each vector v in V, insert it into B if it's not in the *Span* B. This algorith is not resctrive because it doesn't stop until it iterates through all the vectors in V or finds a vector v that isn't in the span of *Span* B. This is a greedy algorithm makes a choice without giving thought to the future.

Shrink algorithm Start with a non-empty set B whose span will ultimately contain the generators for V. For each vector v in B, remove it only if $Span(B - v) = V$. Continue repeating this while possible. The algorithm stops when there is no vector whose removal would leave a spanning set. This greedy algorithm makes a choice without giving thought to the future.

Dominating set

A *dominating* set is a set of nodes in a graph such that every node in the graph is in the set or is a neighbor(reachable by one edge from a node in the dominating set). The goal of the *minimum-dominating-set problem* is to find a dominating set of minimum size.

Definition 5.4.1 x-to-k path

For a sequence of edges in graph G,

$$[\{x_1, x_2\}, ..., \{x_{k-1}, x_k\}]$$

is called an x_1-to-x_k path.

Definition 5.4.2 Spanning edges

A set S of edges is *spanning*, if for every edge $\{x, y\}$ in G, there is a path *x-to-y* (sequence of edges from node x to node y) of edges in S.

Definition 5.4.3 Forest

A forest is a set of edges containing no cycles(loops consisting of several edges).

Definition 5.5.1 Superfluous-Vector Lemma

For a vector $\boldsymbol{v}$ in S, if $\boldsymbol{v}$ can be written as a linear combination of the other vectors in S, then $Span(S - \{\boldsymbol{v}\}) = Span(S)$. The vector is then superfluous.

Linear Dependence: Vectors

Vectors $\boldsymbol{v}_1, ..., \boldsymbol{v}_n$ are linearly *dependent* if there is a nontrivial linear combination of the vectors that equals the *zero vector.*

$$\mathbf{0} = \alpha_1\boldsymbol{v}_1, +... + \alpha_1\boldsymbol{v}_n$$

We refer to the linear combination as a *linear dependency* in $\boldsymbol{v}_1, +... + \boldsymbol{v}_n$.

Definition 5.5.2 Linear Independence: Vectors

Vectors $\boldsymbol{v}_1, ..., \boldsymbol{v}_n$ are linearly *independent* if the only solution is the trivial solution. Stated alternatively, if the only linear combination is the trivial linear combination, then it said these vectors are linearly independent.

$$\mathbf{0} = \alpha_1\boldsymbol{v}_1, +... + \alpha_1\boldsymbol{v}_n$$

This is the same as asking if the null space of a matrix contains a non-trivial vector as a solution and if the solution of a homogeneous linear system is trivial.

Definition 5.5.8 Subset Linearity

The subset of a linearly independent set of vectors is also linearly independent.

Definition 5.5.9 Span Lemma

Let $\boldsymbol{v}_1, ..., \boldsymbol{v}_n$ be vectors. A vector $\boldsymbol{v_i}$ is in the span of the other vectors if and only if the zero vector vector can be a written as a linear combination of all the vecotrs with a non-zero coeficient for $\boldsymbol{v_i}$.

Corollary 5.5.10 Grow-Algorithm Linearity

The vectors obtained by the Grow algorithm are linearly independent.

Corollary 5.5.11 Shrink-Algorithm Linearity

The vectors obtained by the Shrink algorithm are linearly independent.

Definition 5.6.1 Basis

Let $\boldsymbol{V}$ be a vector space. A *basis* of this vector space is linearly independent set of generators for $\boldsymbol{V}$. A set B of vectors that is a basis for $\boldsymbol{V}$ satisfies two properties:
1. the generators are linearly independent
2. $Span\ B = \boldsymbol{V}$

Lemma 5.6.9 Standard Basis Vectors

The standard generators for $\boldsymbol{F}^D$ for a basis.

Lemma 5.6.11 Subset-Basis Lemma

Any set T of vectors contains a subset B that forms a basis for $Span\ T$.

Lemma 5.7.1 Unique-Representation Lemma

Let $\boldsymbol{a}_1, ..., \boldsymbol{a}_2$ be a basis for a vector space $\boldsymbol{V}$. For any vector $\boldsymbol{v} \in \boldsymbol{V}$, there is exactly one representation of $\boldsymbol{v}$ in terms of the basis vectors. We often represent each vector in a vector space by its coordinate representation $[\alpha_1, ..., \alpha_n]$. This implies that the funcion that maps the set of basis vectors is both one-to-one and onto. It's both one-to-one, where it is presented uniquely in both the forwards and backwards direction, and it is onto because it spans the entire original set by the definition of bases(otherwise it wouldn't be a basis for this vector space).

Change of Basis

Suppose you have two non-equal sets of bases for $\boldsymbol{V}$. Given the properties of a basis, we're able to deduce that transformation(function that maps coordinate to vector) that transforms the first basis to the second, and vice-versa, is invertible. This is intuitive because the basis functions are also one-to-one and onto.

Two Natural Ways of Specifying a Vector Space $\boldsymbol{V}$

1. You can specify the generators for $\boldsymbol{V}$. This is equivalent to specifying a matrix A such that $\boldsymbol{V} = ColA$.
2. Specifying a homogeneous linear system whose solution set is $\boldsymbol{V}$. This is equivalent to specifying a matrix A such that $\boldsymbol{V} = NullA$.

Lemma 5.11.1 Exchange Lemma

Suppose S is a set of vectors and A is a subset of S. Suppose z is a vector in Span S but not in A such that $A \cup \{z\}$ is linearly independent. Then there is a vector $w \in S - A$ such that Span S = Span $\{z\} \cup S - \{w\}$.

The proof involves setting z as a linear of combination of vectors from S and A. In the process, you'll realize that there must be at least one non-zero coefficient of the combinations of S; otherwise, A and $\{z\}$ would be linearly dependent. Now you express the linear combination on one side, including z, but move the non-zero vector($w \in S$) to the other side. With this new formulation, you can remove this superfluous vector by the Superfluous-Lemma Vector Lemma and introduce z. You've now *exchanged* a vector.

6. Dimension

Lemma 6.1.1 Morphing Lemma

Let V be a vector space. Suppose S be any generating set for V and B, a linearly independent set, belonging to V. Then $|S| \geq |B|$.

Theorem 6.1.2 Basis Theorem

Let V be a vector space. All bases for V have the same size.

This can be proved by selecting any two bases of V and then expressing each as S in the Morphing Lemma. Since in both cases the inequality is such that the size of the generators is greater than or equal to the size of either basis, it must be that the bases are the same size.

Theorem 6.1.3 Smallest Set for Basis

Let V be a vector space. Then a set of generators for V is a *smallest* set for V if and only if the set is a basis for V.

The proof involves identifying T as a basis for V and S as the smallest generating set for V. By the Morphing Lemma, $|T| \leq |S|$, so T is also the smallest set of generators. If T is not a basis for V, then there is a set of generators smaller than T.

Definition 6.2.2 Dimension

We define the *dimension* of a vector space as the size of the basis for that vector space. The dimension of a vector space V is written $dimV$.

For any field $\mathbb{F}$ and any finite set D, one basis for $\mathbb{F}^D$ is the *standard basis*, which consists of $|D|$ vectors. Therefore $\mathbb{F}^D$ has dimension $|D|$.

Definition 6.2.5 Rank

We define the *rank* of a set S of vectors as the dimension of Span S.

Proposition 6.2.8

For any set S of vectors, rank $S \leq |S|$.

Definition 6.2.9

For a matrix M, the *row rank* of M is the rank of its rows, and the *column rank* is the rank of its columns. Equivalently, the row rank of M is the dimension of the rowspace of M, and the column rank of M is the dimension of column space of M.

Lemma 6.2.13 Superset-Basis Lemma

For any vector space V and any linearly independent set A of vectors, V has a basis that contains all of A.

Lemma 6.2.14 Dimension Principle

If V is a subspace of W, then

1. dim $V \leq$ dim W
2. if dim $V =$ dim W, then $V = W$

Proposition 6.2.17

Any set of D-vectors has *rank* at most $|D|$.

Lemma 6.2.18 Grow Algorithm Termination Lemma

If dim V is finite, then $GROW(V)$ terminates.

Corollary 6.2.19

For finite D, any subspace of $\mathbb{F}^D$ has a basis.

Theorem 6.2.20 Rank Theorem

For any matrix, the row rank equals the column rank.

The proof first involves identifying that the rank of a row space is at most the rank of a column space using the matrix-vector multiplication definition of basis of the column space. If you then reinterpret the same equation in terms of vector matrix-multiplication and view the rows of the basis matrix as linear combinations of the coordinates, you can see that the rank is the same or less than that of the rank of the coordinates. We then show that the row rank is less than or equal to the column rank. If you then transpose the matrix, you'll have two inequalities. You combine them using the transitive property, you see that the row rank and column rank must be equal.

Definition 6.2.21

We define the *rank* of a matrix to be its column rank, which is also equal to its row rank.

Definition 6.3.1 Direct Sum

Let U and V be two vector spaces consisting of D-vectors over a field $\mathbb{F}$. If U and V share only the zero vector, then we define the *direct sum* of U and V to be the set $\{u + v : u \in U, v \in V\}$ written $U \oplus V$. That is, $U \bigoplus V$ is the set of all sums of a vector in U and a vector in V.

Proposition 6.3.5

The direct sum $U \oplus V$ is a vector space.

Lemma 6.3.6 Generators of Direct Sum

The union of a set of generators for V and a set of generators for W is a set of generators for $U \oplus V$.

Lemma 6.3.8 Direct Sum Basis Lemma

The union of a basis of U and a basis of V is a basis of $U \oplus V$.

The proof involves expressing a linear combination of the union of the bases equal to 0 to show linear independence. By moving one of the linear combinations to the other side of the equation, you can then say that the only common vector between the vector spaces is the zero vector. Since they're already the basis for each vector space, they're already linearly independent, so the linear combination, by definition, must be trivial. The bases are also already generators, which meets the criteria for a new basis to be formed.

Corollary 6.3.9 Direct-Sum Dimension Corollary

$\dim U + \dim V = \dim U \oplus V$.

Corollary 6.3.10 Direct-Sum Unique Representation Corollary

Any vector in $U \oplus V$ has a unique representation as $u + v$ where $u \in U, v \in V$.

The proof involves expressing this as a linear combination using its basis and then using the original Unique Representation Lemma to show that its both one-to-one and onto.

Definition 6.3.11 Complementary Subspaces

If $U \oplus V = W$, we say that U and V are *complementary subspaces* of W.

Definition 6.3.15 Missing Complementary Subspace

For any vector space W and any subspace U of W, there is a subspace V of W such that $W = U \oplus V$.

6.4.2 Largest Invertible Subfunction

Let $f : V \longrightarrow W$ be a linear function that is not necessarily invertible. Let's define a function $f^* : V^* \longrightarrow W^*$ that is invertible. Let $w_1, \ldots, w_r$ be the basis for W^*, and let $v_1, \ldots, v_r$ be the pre-images of the corresponding basis in W.

- f^* is *onto* because there is one pre-image for every w_k.
- f^* is *one-to-one* because the $\text{Ker} f^* = 0$. This is proved by representing a vector in terms of its domain and then converting the domain values to the co-domain values. We use the vectors that are the pre-images of the basis in the co-domain and use these since they span V^*. We then simply use our function to convert back to the w basis vectors which are linearly independent. This show that f^* is *one-to-one.*
- $v_1, \ldots, v_r$ form a basis for V^*. This is proved by expressing 0 in terms of v_k with coefficients and then converting to corresponding basis with w_k. Because the basis of W* is linearly independent, the coefficients must be 0; therefore, $v_1, \ldots, v_r$ form a basis for V^*.

Theorem 6.4.3 Kernel Image

The construction of an invertible subfunction $f^* : V^* \longrightarrow W^*$ from a linear function f allows us to relate the domain of the subfunction to the kernel of original linear function f: $V = \text{Ker} f \oplus V^*$.

Theorem 6.4.7 Dimension of Kernel-Image Theorem

For any linear function $f : V \longrightarrow W$, $\dim \text{Ker} f + \dim \text{Im} f = \dim V$.

The proof involves writing the Kernel-Image Theorem in terms of dimensions(written as is) and then realizing that the dimension of $\dim V^* = \dim \text{Im} f$ because the size of basis of the domain and co-domain are equal.

Theorem 6.4.8 Linear-Function Invertibility Theorem

Let $f : V \longrightarrow W$ be a linear function. Then f is invertible if and only if $\dim \text{Ker} f = 0$ and $\dim V = \dim W$.

The dimension of the kernel implies that the function is one-to-one(trivial solution), and the dimension of the domain and co-domain being equal implies that the function is onto.

Nullity of a Matrix

The *nullity* of a matrix is the dimension of the null space of the matrix.

Theorem 6.4.9 Rank-Nullity Theorem

For any n-column matrix A, $\text{rank} A + \text{nullity} A = n$.

The proof involves stating that the sum of the rank of A, the dimension of the column space of A, and the nullity of A(dimension of the null space of A) is n.

Corollary 6.4.10 Matrix Invertibility

Let A be an $R \times C$ matrix. Then A is invertible if and only if $|R| = |C|$ and the columns of A are linearly independent.

The proof involves converting the Linear-Function Invertibility Theorem to its matrix counterparts.

Corollary 6.4.11 Transpose Matrix

The transpose of an invertible matrix is invertible.

The proof involves identifying that the all the columns are linearly independent, so the rank of the matrix is n. But because the matrix is square, it must mean all the rows are also linearly independent. The same applies to the transpose, so the inverse is also invertible.

Corollary 6.4.12

Suppose A and B are square matrices such that BA is the identity matrix. Then A and B are inverses of each other.

The Annihilator

There are two ways of representing a vector space:

1. as the span of a finite set vectors
2. as the solution set of a homogeneous linear system

The analogous representations of an affine space are

1. the affine hull of a finite set of vecotrs
2. the solution set of a linear system

Definition 6.5.7 Annihilator

For a subpace V of F^n, the *annihilator* of V, written V^o, is

$$V^o = \{\boldsymbol{u} \in F^n : \boldsymbol{u} \cdot \boldsymbol{v} = 0 \text{ for every vector } \boldsymbol{v} \in V\}$$

Theorem 6.5.13 Annihilator Dimension Theorem

If V and V^o are subspaces of F^n, then $\text{dim}V + \text{dim}V^o = n$.

The proof involves using the Rank-Nullity Theorem $\text{rank}A + \text{Nullity}A = n$, where it is then converted to $\text{dim}V + \text{dim}V^o = n$.

Theorem 6.5.15 Annihilator Theorem

$(V^o)^o = V$ (The annihilator of the annihilator is the original space.)

7. Gaussian Elimination

Applications of Gaussian Elimination

1. Finding a basis for the span of given vectors. This additionally gives us an algorithm for rank and therefore for testing linear dependence.
2. Finding a basis for the null space of a matrix.
3. Solving a matrix equation, which is the same as expressing a given vector as a linear combination of other given vectors, which is the same as solving a system of linear equations.

Definition 7.1.1 Echelon Form

An $m \times n$ matrix A is in echelon form if it satisfies the following condition: for any row, if that row's first non-zero entry is in position k, then the first non-zero value in the previous rows must have been in a position less than k.

The triangular matrix is a special case where the first non-zero entry in row i is in column i.

If a row of a matrix in echelon form is zero, then every subsequent row must be zero.

Lemma 7.1.2 Echelon Row Space

If a matrix is in echelon form, the non-zero rows form a basis for the row space.

It's proved by using the Grow Algorithm starting from the bottom up. As you move up, each non-zero row will have its first non-zero value in a column position that is smaller than the previous rows considered in the grow algorithm. By inducation, this means that each row is linearly independent.

Lemma 7.1.3 Row-Addition operations preserve the Row Space

For matrices A and N, Row NA $\subseteq$ Row A.

The proof involves expressing $\boldsymbol{v} = ([\boldsymbol{u}^T][N])[A]$ and then recognizing that this merely represents a linear combination of the rows of A.

Corollary 7.1.4

For matrices A and M, if M is invertible then Row MA = Row A.

Shortfalls of using Gaussian Elimination using Numerical Analysis

1. *Partial pivoting* selects rows with nonzero entries in column c, chooose row with entry having the largest absolute value.
2. *Complete pivoting* selects a column on the fly to maximize the pivot element, instead of selecting the order of the columns beforehand.

While partial pivoting is used in practice because it runs equickly and is easy to implement, it is error-prone. Complete pivoting keeps those errors under control at the expense of speed.

Proposition 7.3.1

For any matrix A, there is an invertible matrix M such that MA is in echelon form.

Theorem 7.6.1 Prime Factorization Theorem

For every positive integer N, there is a unique bag of primes whose product is N.

Composite Factoring Optimization

If N is composite, it has a nontrivial divisor that is at most $\sqrt{N}$.

8. Inner Product

Distance, length, norm, inner product

- Property N1: For any vector $\boldsymbol{v}$, $\|\boldsymbol{v}\|$ is a nonnegative real number
- Property N2: For any vector $\boldsymbol{v}$, $\|\boldsymbol{v}\|$ is zero if and only if the $\boldsymbol{v}$ is a zero vector.
- Property N3: For any vector $\boldsymbol{v}$ and any scalar α, $\|\alpha\boldsymbol{v}\| = |\alpha|\|\boldsymbol{v}\|$.
- Property N4: For any vector $\boldsymbol{v}$ and $\boldsymbol{u}$, $\|\alpha\boldsymbol{v}\| = |\alpha|\|\boldsymbol{v}\|$.

Inner Product Definition

One way to define the vector norm is to define an operation on vectors called the *inner product*: $\langle\boldsymbol{u}, \boldsymbol{v}\rangle$.

The *norm* of a vector is defined as

$$\|\boldsymbol{v}\| = \sqrt{\langle\boldsymbol{v}, \boldsymbol{v}\rangle}$$

Inner Product Properties

1. Linearity in the First Argument: $\langle\boldsymbol{u} + \boldsymbol{v}, \boldsymbol{w}\rangle = \langle\boldsymbol{u}, \boldsymbol{w}\rangle + \langle\boldsymbol{v}, \boldsymbol{w}\rangle$
2. Symmetry: $\langle\boldsymbol{u}, \boldsymbol{v}\rangle = \langle\boldsymbol{v}, \boldsymbol{u}\rangle$
3. Homogeneity:$\langle\alpha\boldsymbol{u}, \boldsymbol{v}\rangle = \alpha\langle\boldsymbol{u}, \boldsymbol{v}\rangle$

Theorem 8.3.1 Pythagorean Theorem for vectors over reals

If vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ over the reals are orthogonal, then

$$\|\boldsymbol{u} + \boldsymbol{v}\|^2 = \|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2$$

Orthogonality

Two vectors are orthogonal if their inner product is 0:

$$\langle\boldsymbol{u}, \boldsymbol{v}\rangle = 0$$

Orthongality Properties

1. Property O1: If $\boldsymbol{u}$ is orthogonal to $\boldsymbol{v}$, then $\alpha\boldsymbol{u}$ is orthogonal to $\alpha\boldsymbol{v}$ for every scalar α.
2. Property O2: If $\boldsymbol{u}$ and $\boldsymbol{v}$ are both orthogonal to $\boldsymbol{w}$, then $\boldsymbol{u} + \boldsymbol{v}$ is orthogonal to $\boldsymbol{w}$.

Lemma 8.3.3

If $\boldsymbol{u}$ is orthogonal to $\boldsymbol{v}$, then for any scalars α, β,

$$\|\alpha\boldsymbol{u} + \beta\boldsymbol{v}\|^2 = \alpha^2\|\boldsymbol{u}\|^2 + \beta^2\|\boldsymbol{v}\|^2$$

Definition 8.3.6 Decomposition of a vector into components

For any vector $\boldsymbol{b}$ and $\boldsymbol{v}$, define vector $\boldsymbol{b}^{\|\boldsymbol{v}}$ and $\boldsymbol{b}^{\perp\boldsymbol{v}}$, respectively, the *projection* of $\boldsymbol{b}$ *along* $\boldsymbol{v}$ and the *projection* of $\boldsymbol{b}$ *orthogonal* to $\boldsymbol{v}$ if

$$\boldsymbol{b} = \boldsymbol{b}^{\|\boldsymbol{v}} + \boldsymbol{b}^{\perp\boldsymbol{v}}$$

and, for some scalar $\sigma \in \mathbb{R}$,

$$\boldsymbol{b}^{\|\boldsymbol{v}} = \sigma\boldsymbol{v}$$

and

$$\boldsymbol{b}^{\perp\boldsymbol{v}} \text{ is orthogonal to } \boldsymbol{v}$$

Lemma 8.3.8 Fire Engine Lemma

Let $\boldsymbol{b}$ and $\boldsymbol{v}$ be vectors. The point in Span$\{\boldsymbol{v}\}$ closest to $\boldsymbol{b}$ is $\boldsymbol{b}^{\|\boldsymbol{v}}$, and the distance is $\|\boldsymbol{b}^{\perp\boldsymbol{v}}\|$.

Lemma 8.3.11

For any vector $\boldsymbol{b}$ and $\boldsymbol{v}$ over the reals.

1. There is a scalar σ such that $\boldsymbol{b} - \sigma\boldsymbol{v}$ is orthogonal to $\boldsymbol{v}$.
2. The point $\boldsymbol{p}$ on Span $\boldsymbol{v}$ that minimizes $\|b - p\|$ is $\sigma\boldsymbol{v}$.
3. The value of σ is $\frac{\langle \boldsymbol{b},\boldsymbol{v}\rangle}{\langle \boldsymbol{v},\boldsymbol{v}\rangle}$.

9. Orthogonalization

Definition 9.1.1

A vector $\boldsymbol{v}$ is orthogonal to a set S of vectors if $\boldsymbol{v}$ is orthogonal to every vector in S.

Lemma 9.1.3

A vector $\boldsymbol{v}$ is orthogonal to each of the vectors $\boldsymbol{a}_1, ..., \boldsymbol{a}_n$ if and only if it is orthogonal to every vector in Span $\boldsymbol{a}_1, ..., \boldsymbol{a}_n$.

The proof involves expressing w, which is in the Span of the $\boldsymbol{a}$ vectors, with its coeffiecients and then computing the inner product with v, the vector orthogonal to every vector in the Span. Because the inner products with the vectors without the coefficient are zero, the coefficients of the Span multiplied by zero is still zero.

Definition 9.1.4

For a vector $\boldsymbol{b}$ and a vector space $\boldsymbol{V}$, we define the projection of $\boldsymbol{b}$ onto $\boldsymbol{V}$ and the projection of $\boldsymbol{b}$ orthogonal to $\boldsymbol{V}$ so that $\boldsymbol{b}^{\|V}$ is in $\boldsymbol{V}$, and $\boldsymbol{b}^{\perp V}$ is orthogonal to every vector in $\boldsymbol{V}$.

$$\boldsymbol{b} = \boldsymbol{b}^{\|V} + \boldsymbol{b}^{\perp V}$$

Lemma 9.1.6 Generalized Fire Engine Lemma

Let $\boldsymbol{V}$ be a vector space, and let $\boldsymbol{b}$ be a vector. The point in $\boldsymbol{V}$ closest to $\boldsymbol{b}$ is $\boldsymbol{b}^{\|\boldsymbol{V}}$, and the distance is $\|\boldsymbol{b}^{\perp\boldsymbol{V}}\|$.

Proposition 9.5.1

Mutually orthogonal nonzero vectors are linearly independent

The proof involves using the orthogonal vectors with coefficens and setting the linear combination to be zero. By taking, for example, the inner product of v_1* with the linear combination, only v_1* is remaining and must be zero because of the other side of the equation. You proved that the first coefficient is zero. It's fully proved by iterating through all the orthogonal vectors.

Definition 9.6.1

Let W be a vector space over the reals, and let U be a subspace of W. The *orthogonal complement* of U with respect W is defined to be the set V such that

$$V = \{w \in W : wisorthogonaltoeveryvecotrinU\}$$

Furthermore, V is a subspace of W(Lemma 9.6.2).

Lemma 9.6.4

Let V be the orthogonal complement of U with respect to W. The only vector in $U \cap V$ is the zero vector.

Lemma 9.6.5

If the orthogonal complement of U with respect to W is V then

$$U \oplus V = W$$

QR Factorization

Matrix factorizations play a * mathematical role of offering insight into the nature of matrices – each factorization gives us a new way to think about a matrix * compoutational role by allowing us to compute solutions to fundamental computational problems

For example, an orthogonal marix with a coefficient matrix(Equation 9.7) be used to solve this square matrix equation and also applicable to the least-square problem. Equation 9.7 defines a matrix whose columns are mutually orthogonal and another triangular matrix that when multiplied provide the original matrix.

Definition 9.7.1

Mutually orthogonal vectors are said to be *orthonormal* if they all have a norm 1. A matrix is said to be *column-orthogonal* if its column vectors are *orthonormal.* A square column-orthogonal matrix is said to be an *orthogonal matrix.*

Lemma 9.7.2

If Q is a column-orthogonal matrix, then Q^TQ is an identity matrix.

Lemma 9.7.3 Inverse of Orthogonal Matrix

If Q is an orthogonal matrix, then its inverse is Q^T.

Definition 9.7.4 QR factorization of a matrix

The QR factorization of an $m \times n$ matrix A (where $m \geq n$) is $A = QR$ where Q is an $m \times n$ column-orthogonal matrix Q and R is a triangular matrix:

$$[A] = [Q][R]$$

Lemma 9.7.5

In the QR factorization of A, if A's columns are linearly independent then Col$Q$ = Col$A$.

Lemma 9.8.1

Suppose A is a square matrix with linearly independent columns. The vector $\hat{x}$ found by the above algorithm, namely QR Square, satisfies the equation $Ax = b$.

Lemma 9.8.3

Let Q be a column-orthogonal basis, and let $V = \text{Col}Q$. Then, for any vector $\boldsymbol{b}$ whose domain equals Q's row-label set, $Q^T\boldsymbol{b}$ is the coordinate representation of $\boldsymbol{b}^{\|\boldsymbol{V}}$ in terms of the columns of Q, and $QQ^T\boldsymbol{b}$ is $\boldsymbol{b}^{\|\boldsymbol{V}}$ itself.

Lemma 9.8.3

Let Q be a column-orthogonal basis, and let $V = \text{Col}Q$. Then, for any vector $\boldsymbol{b}$ whose domain equals Q's row-label set, $Q^T\boldsymbol{b}$ is the coordinate representation of $\boldsymbol{b}^{\|\boldsymbol{V}}$ in terms of the columns of Q, and $QQ^T\boldsymbol{b}$ is $\boldsymbol{b}^{\|\boldsymbol{V}}$ itself.

10. Special Bases

Lemma 10.2.2

Let Q be a column-orthogonal matrix. Multiplication of vectors by Q preserves inner-products. In the context of orthonormal bases, this implies they preserve norm.

$$\langle Q\boldsymbol{u}, Q\boldsymbol{v}\rangle = \langle \boldsymbol{u}, \boldsymbol{v}\rangle$$

Corollary 10.2.3

For any column-orthogonal matrix Q and vector $\boldsymbol{u}$, $||Q\boldsymbol{u}|| = ||\boldsymbol{u}||$. Let b and $\bar{b}$ be two vectors, and let x and $\overline{x}$ be the represetations of b and $\bar{b}$ with respect to an orthonormal basis. Since $Qx = b$ and $Q\overline{x} = \bar{b}$, the corrolary implies that $||b - \bar{b}|| = ||x - \overline{x}||$. This implies that finding a vector close to b is equivalent to finding a representation close to x.

11. Singular Value Decomposition

Lemma 11.1.1 Frobenius Norm

The square of the Frobenius norm of A equals the sum of the squares of the rows of A.

$$||A||_F = \sqrt{\Sigma_i \Sigma_j A[i,j]^2}$$

Definition 11.2.2

We refer to σ_1 as the first singular value of matrix A, and we refer to v_1 as the first right singular vector.

Theorem 11.2.4

The minimum sum of the squared distances is $||A||_F^2 - \sigma_1^2$ with a with a rank-one approximation.

Definition 11.2.10

The first left singular vector of matrix A is defined to be the vector u_1 such that $\sigma_1 \boldsymbol{u_1} = A\boldsymbol{v_1}$, where σ_1 and v_1 are, respectively, the first singular value and the first right singular vector.

Theorem 11.2.11

The best rank-one approximation to A is $\sigma_1 \boldsymbol{u_1} \boldsymbol{v_1}^T$ where σ_1 is the first singular value, $\boldsymbol{u}_1$ is the first left singular vector, and $\boldsymbol{v_1}$ is the first right singular vector.

Definition 11.3.2

The vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_r$ are the right singular vectors of matrix A, and the corresponding real numbers are $\sigma_1, \sigma_2, ..., \sigma_r$ are the singular values of matrix A.

Proposition 11.3.3

The right singular vectors are orthonormal. This is because we only pick vectors that are orthonormal to previously selected vectors.

Proposition 11.3.5

The singular values are nonnegative and in descending order.

Lemma 11.3.6

Every row of matrix A is in the span of the right singular vectors.

Definition 11.3.7

The vectors $u_1, u_2, ..., u_r$ such that $\sigma_j u_j = Av_j$ are the left singular vectors of matrix A.

Proposition 11.3.8

The left singular vectors are orthonormal.

Definition 11.3.9

The (reduced) form of singular value decomposition of a matrix A is a factorization of A as $A = U\Sigma V^T$ in which the matrices U, Σ, and V have three properties:

1. Σ is a diagonal matrix whose entries $\sigma_1, ..., \sigma_r$ are positive and in descending(non-negative) order.
2. V is a column-orthogonal matrix.
3. U is a column-orthogonal matrix.

Theorem 11.3.10

Every matrix A over $\mathbb{R}$ has a singular value decomposition.

Symmetry

The SVD is symmetric under transposition:

$$\begin{aligned} A^T &= (U\Sigma V^T)^T \\ &= V\Sigma^T U^T \\ &= V\Sigma U^T \end{aligned}$$

Lemma 11.3.11

Let $v_1, ..., v_k$ be an orthonormal vector basis for a vector space V. Then

$$||A||_F^2 - ||Av_1||^2 - ... - ||Av_k||^2.$$

This can be interpreted as computing the distance from a row vector of matrix A to the corresponding vector k in the right singular vectors.

Theorem 11.3.12

Let A be an $m \times n$ matrix, and let $a_1, ..., a_m$ be its rows. Let $v_1, ..., v_r$ be its right singular vectors, and let $\sigma_1, ..., \sigma_r$ be its singular values. For any positive integer $k \leq r$, Span$\{v_1, ..., v_k\}$ is the k-dimensional vector space V that minimizes the distance between the rows of A to the vector space spanned by the right singular vectors. The minimum sum of squared distances is $||A||_F^2 - \sigma_1^2 - \sigma_2^2 - ... - \sigma_k^2$.

Theorem 11.3.13

For $k \leq$ rank A, the best rank-at-most-k approximation to A is

$$\tilde{A} = \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^T + ... + \boldsymbol{u}_k \boldsymbol{v}_k^T$$

for which $||A - \tilde{A}||_F^2 = ||A||_F^2 - \sigma_1^2 - \sigma_2^2 - ... \sigma_k^2$.

Proposition 11.3.14

In the singular value decomposition $U\Sigma V^T$ of A, Col U = Col A and Row V^T = Row A.