Linear Algebra

1 Vector Spaces

1-1 Vector Spaces

A **vector space** (or linear space) V over a field F is a set on which the operations addition (+) and scalar multiplication, are defined so that for all $x, y, z \in V$ and all $a, b \in F$,

0.	$x + y$ and ax are unique elements in V.	Closure
1.	$x + y = y + x$	Commutativity of Addition
2.	$(x + y) + z = x + (y + z)$	Associativity of Addition
3.	There exists $0 \in V$ such that for every $x \in V, x + 0 = x$.	Existence of Additive Identity (Zero Vector)
4.	There exists an element $-x$ such that $x + (-x) = 0$.	Existence of Additive Inverse
5.	$1x = x$	Multiplicative Identity
6.	$(ab)x = a(bx)$	Associativity of Scalar Multiplication
7.	$a(x + y) = ax + ay$	Left Distributive Property
8.	$(a + b)x = ax + bx$	Right Distributive Property

Elements of F, V are **scalars**, **vectors**, respectively. F can be $\mathbb{R}, \mathbb{C}, \mathbb{Z}/p$, etc.

Examples:

F^n	n-tuples with entries from F
F^∞	sequences with entries from F
$M_{m \times n}(F)$ or $F^{m \times n}$	mxn matrices with entries from F
$\mathcal{F}(S, F)$	functions from set S to F
$P(F)$ or $F[x]$	polynomials with coefficients from F
$C[a, b], C^\infty$	continuous functions on $[a, b], (-\infty, \infty)$

Cancellation Law for Vector Addition: If $x, y, z \in V$ and $x + z = y + z$, then $x = y$.
Corollary: 0 and -x are unique.

For all $x \in V, a \in F$,

- $0x = 0$
- $x0 = 0$
- $(-a)x = -(ax) = a(-x)$

1-2 Subspaces

A subset W of V over F is a **subspace** of V if W is a vector space over F with the operations of addition and scalar multiplication defined on V.

$W \subseteq V$ is a subspace of V if and only if

1. $x + y \in W$ whenever $x \in W, y \in W$.
2. $cx \in W$ whenever $c \in F, x \in W$.

A subspace must contain 0.

Any intersection of subspaces of V is a subspace of V.

If S_1, S_2 are nonempty subsets of V, their sum is $S_1 + S_2 = \{x + y | x \in S_1, y \in S_2\}$.
V is the **direct sum** of W_1 and W_2 ($V = W_1 \oplus W_2$) if W_1 and W_2 are subspaces of V such that $W_1 \cap W_2 = \{0\}$ and $W_1 + W_2 = V$. Then each element in V can be written uniquely as $w_1 + w_2$ where $w_1 \in W_1, w_2 \in W_2$. W_1, W_2 are **complementary**.

$W_1 + W_2$ $(W_1 \wedge W_2)$ is the smallest subspace of V containing W_1 and W_2, i.e. any subspace containing W_1 and W_2 contains $W_1 + W_2$.

For a subspace W of V, $v + W = \{v + w | w \in W\}$ is the **coset** of W containing v.

- $v_1 + W = v_2 + W$ iff $v_1 - v_2 \in W$.
- The collection of cosets $V/W = \{v + W | v \in V\}$ is called the **quotient** (factor) **space** of V modulo W. It is a vector space with the operations
  - $(v_1 + W) + (v_2 + W) = (v_1 + v_2) + W$
  - $a(v + W) = av + W$

1-3 Linear Combinations and Dependence

A vector $v \in V$ is a **linear combination** of vectors of $S \subseteq V$ if there exist a finite number of vectors $u_1, u_2, \ldots u_n \in S$ and scalars $a_1, a_2, \ldots a_n \in F$ such that

$$v = a_1 u_1 + \cdots + a_n u_n.$$

v is a linear combination of $u_1, u_2, \ldots u_n$.

The **span** of S, span(S), is the set consisting of all linear combinations of the vectors in S. By definition, $\mathrm{span}(\phi) = \{0\}$. S **generates** (spans) V if span(S)=V.

The span of S is the smallest subspace containing S, i.e. any subspace of V containing S contains span(S).

A subset $S \subseteq V$ is **linearly (in)dependent** if there (do not) exist a finite number of distinct vectors $u_1, u_2, \ldots u_n \in S$ and scalars $a_1, a_2, \ldots a_n$, not all 0, such that

$$a_1 u_1 + \cdots + a_n u_n = 0.$$

Let S be a linearly independent subset of V. For $v \in S - V, S \cup \{v\}$ is linearly dependent iff $v \in \mathrm{span}(S)$.

1-4 Bases and Dimension

A (ordered) **basis** β for V is a (ordered) linearly independent subset of V that generates V.
Ex. $e_1 = (1,0, \ldots 0), e_2 = (0,1, \ldots 0), \ldots e_n = (0,0, \ldots 1)$ is the standard ordered basis for F^n.

A subset β of V is a basis for V iff each $v \in V$ can be uniquely expressed as a linear combination of vectors of β.

Any finite spanning set S for V can be reduced to a basis for V (i.e. some subset of S is a basis).

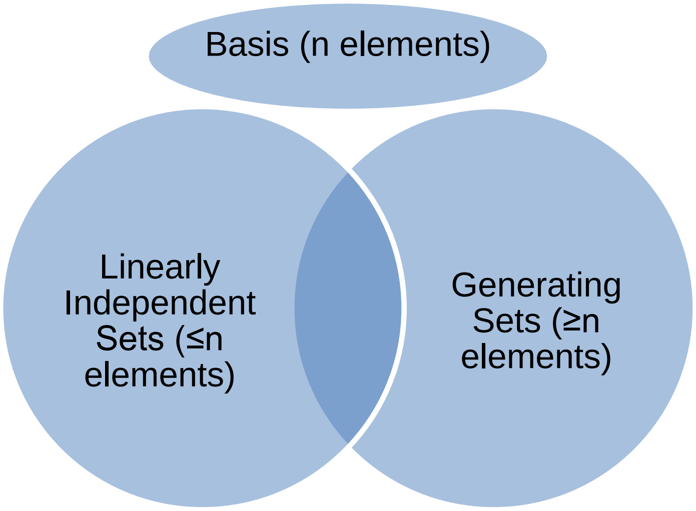
<u>Replacement Theorem</u>: (Steinitz) Suppose V is generated by a set G with n vectors, and let L be a linearly independent subset of V with m vectors. Then $m \leq n$ and there exists a

subset H of G containing $n-m$ vectors such that $L \cup H$ generates V.
Pf. Induct on m. Use induction hypothesis for $\{v_1, \dots v_m\}$; remove a u_1 and replace by v_{m+1}.

Corollaries:

- If V has a finite basis, every basis for V contains the same number of vectors. The unique number of vectors in each basis is the **dimension** of V (dim(V)).
- Suppose dim(V)=n. Any finite generating set/ linearly independent subset contains ≥n/≤n elements, can be reduced/ extended to a basis, and if the set contains n elements, it is a basis.

Subsets of V, dim(V)=n



Let W be a subspace of a finite-dimensional vector space V. Then dim(W)≤dim(V). If dim(W)=dim(V), then W=V.

$$\dim(W_1 + W_2) = \dim(W_1) + \dim(W_2) - \dim(W_1 \cap W_2)$$

$$\dim(V) = \dim(W) + \dim(V / W)$$

The dimension of V/W is called the **codimension** of V in W.

1-5 Infinite-Dimensional Vector Spaces

Let $\mathcal{F}$ be a family of sets. A member M of $\mathcal{F}$ is **maximal** with respect to set inclusion if M is contained in no member of $\mathcal{F}$ other than M. ($\mathcal{F}$ is partially ordered by ⊆.)
A collection of sets $\mathcal{C}$ is a **chain** (nest, tower) if for each A, B in $\mathcal{C}$, either $A \subseteq B$ or $B \subseteq A$. ($\mathcal{F}$ is totally ordered by ⊆.)

Maximal Principle: [equivalent to Axiom of Choice] If for each chain $\mathcal{C} \subseteq \mathcal{F}$, there exists a member of $\mathcal{F}$ containing each member of $\mathcal{C}$, then $\mathcal{F}$ contains a maximal member.

A **maximal linearly independent subset** of $S \subseteq V$ is a subset B of S satisfying
(a) B is linearly independent.
(b) The only linearly independent subset of S containing B is B.
Any basis is a maximal linearly independent subset, and a maximal linearly independent

subset of a generating set is a basis for V.

Let S be a linearly independent subset of V. There exists a maximal linearly independent subset (basis) of V that contains S. Hence, ***every vector space has a basis***.
Pf. $\mathcal{F}$ = linearly independent subsets of V. For a chain $\mathcal{C}$, take the union of sets in $\mathcal{C}$, and apply the Maximal Principle.

Every basis for a vector space has the same cardinality.

Suppose $S_1 \subseteq S_2 \subseteq V$, S_1 is linearly independent and S_2 generates V. Then there exists a basis such that $S_1 \subseteq \beta \subseteq S_2$.

Let β be a basis for V, and S a linearly independent subset of V. There exists $S_1 \subseteq \beta$ so $S \cup S_1$ is a basis for V.

1-6 Modules

A left/right R-**module** $_RM/M_R$ over the ring R is an abelian group (M,+) with addition and scalar multiplication ($R \times M \to M$ or $M \times R \to M$) defined so that for all $r, s \in R$ and $x, y \in M$,

	Left	Right
1. Distributive	$r(x+y) = rx + ry$	$(x+y)r = xr + yr$
2. Distributive	$(r+s)x = rx + sx$	$x(r+s) = xr + xs$
3. Associative	$r(sx) = (rs)x$	$(xr)s = x(rs)$
4. Identity	$1x = x$	$x1 = x$

Modules are generalizations of vector spaces. All results for vector spaces hold except ones depending on division (existence of inverse in R). Again, a basis is a linearly independent set that generates the module. Note that if elements are linearly independent, it is not necessary that one element is a linear combination of the others, and bases do not always exist.

A **free module** with n generators has a basis with n elements. V is finitely generated if it contains a finite subset spanning V. The **rank** is the size of the smallest generating set.

Every basis for V (if it exists) contains the same number of elements.

1-7 Algebras

A **linear algebra** over a field F is a vector space $\mathcal{A}$ over F with multiplication of vectors defined so that for all $x, y, z \in \mathcal{A}, c \in F$,

1. Associative	$x(yz) = (xy)z$
2. Distributive	$x(y+z) = xy + xz, (x+y)z = xz + yz$
3.	$c(xy) = (cx)y = x(cy)$

If there is an element $1 \in \mathcal{A}$ so that $1x = x1 = x$, then 1 is the identity element. $\mathcal{A}$ is commutative if $xy = yx$.
Polynomials made from vectors (with multiplication defined as above), linear transformations, and $n \times n$ matrices (see Chapters 2-3) all form linear algebras.

2 Matrices

2-1 Matrices

A $m \times n$ **matrix** has m rows and n columns arranged filled with entries from a field F (or ring R). $A_{ij} = A(i,j)$ denotes the entry in the *i*th column and *j*th row of A. Addition and scalar multiplication is defined component-wise:

$$(A+B)_{ij} = A_{ij} + B_{ij}$$
$$(cA)_{ij} = cA_{ij}$$

The $n \times n$ matrix of all zeros is denoted $\mathcal{O}_n$ or just O.

2-2 Matrix Multiplication and Inverses

Matrix product:
Let A be a $m \times n$ and B be a $n \times p$ matrix. The product AB is the $m \times p$ matrix with entries

$$(AB)_{ij} = \sum_{k=1}^{n} A_{ik} B_{kj} \text{ , } 1 \le i \le m, 1 \le j \le p$$

Interpretation of the product AB:

1. Row picture: Each row of A multiplies the whole matrix B.
2. Column picture: A is multiplied by each column of B. Each column of AB is a linear combination of the columns of A, with the coefficients of the linear combination being the entries in the column of B.
3. Row-column picture: $(AB)_{ij}$ is the dot product of row I of A and column j of B.
4. Column-row picture: Corresponding columns of A multiply corresponding rows of B and add to AB.

Block multiplication: Matrices can be divided into a rectangular grid of smaller matrices, or blocks. If the cuts between columns of A match the cuts between rows of B, then you can multiply the matrices by replacing the entries in the product formula with blocks (entry i,j is replaced with block i,j, blocks being labeled the same way as entries).

The identity matrix I_n is a nxn square matrix with ones down the diagonal, i.e.

$$(I_n)_{ij} = \delta_{ij} = \begin{cases} 1 \text{ if } i = j \\ 0 \text{ if } i \ne j \end{cases}$$

A is invertible if there exists a matrix A^{-1} such that $AA^{-1} = A^{-1}A = I$. The inverse is unique, and for square matrices, any inverse on one side is also an inverse on the other side.

Properties of Matrix Multiplication (A is mxn):

1.	$A(B+C) = AB + AC$	Left distributive
2.	$(A+B)C = AC + BC$	Right distributive
3.	$I_m A = A = AI_n$	Left/ right identity
4.	$A(BC) = (AB)C$	Associative
5.	$a(AB) = (aA)B = A(aB)$	
6.	$(AB)^{-1} = B^{-1}A^{-1}$ (A, B invertible)	

$AB \neq BA$: Not commutative
Note that any 2 polynomials of the same matrix commute.

A nxn matrix A is either a zero divisor (there exist nonzero matrices B, C such that $AB = CA = \mathcal{O}$) or it is invertible.

The **Kronecker** (tensor) **product** of pxq matrix A and rxs matrix B is $A \otimes B = \begin{bmatrix} a_{11}B & \cdots & a_{1q}B \\ \vdots & \ddots & \vdots \\ a_{p1}B & \cdots & a_{pq}B \end{bmatrix}$. If v and w are column vectors with q, s elements, $(A \otimes B)(v \otimes w) = (Av) \otimes (Bw)$. Kronecker products give nice eigenvalue relations- for example the eigenvalues are the products of those of A and B. [AMM 107-6, 6/2000]

2-3 Other Operations, Classification

The **transpose** of a mxn matrix A, A^t, is defined by $(A^T)_{ij} = A_{ji}$.
The **adjoint** or **Hermitian** of a matrix A is its conjugate transpose:

$$A^* = A^H = \bar{A}^T$$

Name	Definition	Properties
Symmetric	$A = A^T$	
Self-adjoint/ Hermitian	$A = A^*$	z^*Az is real for any complex z.
Skew-symmetric	$-A = A^T$	
Skew-self-adjoint/ Skew-Hermitian	$-A = A^*$	
Upper triangular	$A_{ij} = 0 \text{ for } i > j$	
Lower triangular	$A_{ij} = 0 \text{ for } i < j$	
Diagonal	$A_{ij} = 0 \text{ for } i \neq j$	

Properties of Transpose/ Adjoint

1. $(AB)^T = B^T A^T, (AB)^* = B^* A^*$ (For more matrices, reverse the order.)
2. $(A^{-1})^T = (A^T)^{-1}$
3. $(Ax)^T y = x^T A^T y = x^T (A^T y)$, $(Ax)^* y = x^* A^* y = x^* (A^* y)$
4. $A^T A$ is symmetric.

The **trace** of a $n \times n$ matrix A is the sum of its diagonal entries:

$$\mathrm{tr}(A) = \sum_{i=1}^{n} A_{ii}$$

The trace is a linear operator, and $\mathrm{tr}(AB) = \mathrm{tr}(A)\mathrm{tr}(B)$.

The **direct sum** $A \oplus B$ of $m \times m$ and $n \times n$ matrices A and B is the $(m+n) \times (m+n)$ matrix C given by $C = \begin{bmatrix} A & O \\ O & B \end{bmatrix}$,

$$C_{ij} = \begin{cases} A_{ij} \text{ for } 1 \leq i, j \leq n \\ B_{i-m, j-m} \text{ for } m+1 \leq i, j \leq n+m \\ 0, \text{else} \end{cases}$$

3 Linear Transformations

3-1 Linear Transformations

For vector spaces V and W over F, a function $T: V \to W$ is a **linear transformation** (homomorphism) if for all $x, y \in V$ and $c \in F$,
(a) $T(x + y) = T(x) + T(y)$
(b) $T(cx) = cT(x)$

It suffices to verify $T(cx + y) = cT(x) + T(y)$.
$T(0) = 0$ is automatic.

$$T\left(\sum_{i=1}^{n} a_i x_i\right) = \sum_{i=1}^{n} a_i T(x_i)$$

Ex. Rotation, reflection, projection, rescaling, derivative, definite integral
Identity I_v and zero transformation T_0

An **endomorphism** (or linear operator) is a linear transformation from V into itself.

T is **invertible** if it has an **inverse** T^{-1} satisfying $TT^{-1} = I_W, T^{-1}T = I_V$. If T is invertible, V and W have the same dimension (possibly infinite).
Vector spaces V and W are isomorphic if there exists a invertible linear transformation (an **isomorphism**, or automorphism if V=W) $T: V \to W$. If V and W are finite-dimensional, they are isomorphic iff dim(V)=dim(W). V is isomorphic to $F^{\dim(\mathrm{V})}$.

The space of all linear transformations $\mathcal{L}(V, W) = \mathrm{Hom}(V, W)$ from V to W is a vector space over F. The inverse of a linear transformation and the composite of two linear transformations are both linear transformations.

The **null space** or kernel is the set of all vectors x in V such that T(x)=0.
$N(T) = \{x \in V | T(x) = 0\}$
The **range** or image is the subset of W consisting of all images of vectors in V.
$R(T) = \{T(x) | x \in V\}$
Both are subspaces. **nullity**(T) and **rank**(T) denote the dimensions of N(T) and R(T), respectively.

If $\beta = \{v_1, v_2, \dots v_n\}$ is a basis for V, then $R(T) = \mathrm{span}(\{T(v_1), T(v_2), \dots T(v_n)\})$.

Dimension Theorem: If V is finite-dimensional, nullity(T)+rank(T)=dim(V).
Pf. Extend a basis for N(T) to a basis for V by adding $\{v_{k+1}, \dots, v_n\}$. Show $\{T(v_{k+1}), \dots, T(v_n)\}$ is a basis for R(T) by using linearity and linear independence.

T is one-to-one iff N(T)={*0*}.

If V and W have equal finite dimension, the following are equivalent:
(a) T is one-to-one.
(b) T is onto.
(c) rank(T)=dim(V)
(a) and (b) imply T is invertible.

A linear transformation is uniquely determined by its action on a basis, i.e., if $\beta = \{v_1, v_2, \dots v_n\}$ is a basis for V and $w_1, w_2, \dots w_n \in W$, there exists a unique linear transformation $T: V \to W$ such that $T(v_i) = w_i, i = 1,2, \dots n$.

A subspace W of V is **T-invariant** if $T(x) \in W$ for every $x \in W$. T_W denotes the restriction of T on W.

3-2

Matrix Representation of Linear Transformation

Matrix Representation:
Let $\beta = \{v_1, v_2, \dots v_n\}$ be an ordered basis for V and $\gamma = \{w_1, w_2, \dots w_n\}$ be an ordered basis for W. For $x \in V$, define $a_1, a_2, \dots a_n$ so that

$$x = \sum_{i=1}^{n} a_i u_i.$$

The coordinate vector of x relative to β is

$$\phi_\beta(x) = [x]_\beta = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$$

Note ϕ_β is an isomorphism from V to F^n. The *i*th coordinate is $f_i(x) = a_i$.
Suppose $T: V \to W$ is a linear transformation satisfying

$$T(v_j) = \sum_{i=1}^{m} a_{ij} w_i \text{ for } 1 \le j \le n$$

The matrix representation of T in β and γ is $A = [T]_\beta^\gamma = \mathcal{M}_\beta^\gamma(T)$ with entries as defined above. (i.e. load the coordinate representation of $T(v_j)$ into the *j*th column of A.)

Properties of Linear Transformations (Composition)

1.	$T(U_1 + U_2) = TU_1 + TU_2$	Left distributive
2.	$(U_1 + U_2)T = U_1T + U_2T$	Right distributive
3.	$I_V T = T = TI_W$	Left/ right identity
4.	$S(TU) = (ST)U$	Associative (holds for any functions)
5.	$a(TU) = (aT)U = T(aU)$	
6.	$(TU)^{-1} = U^{-1}T^{-1}$ (T, U invertible)	

Linear transformations [over finite-dimensional vector spaces] can be viewed as left-multiplication by matrices, so linear transformations under composition and their corresponding matrices under multiplication follow the same laws. This is a motivating factor for the definition of matrix multiplication. Facts about matrices, such as associativity of matrix multiplication, can be proved by using the fact that linear transformations are associative, or directly using matrices.

Note: From now on, definitions applying to matrices can also apply to the linear transformations they are associated with, and vice versa.

The left-multiplication transformation $L_A: F^n \to F^m$ is defined by $L_A(x) = Ax$ (A is a mxn matrix).

Relationships between linear transformations and their matrices:

1. To find the image of a vector $u \in V$ under T, multiply the matrix corresponding to T

on the left: $[T(u)]_\gamma = [T]_\beta^\gamma [u]_\beta$ i.e. $L_A \phi_\beta = \phi_\gamma T$ where $A = [T]_\beta^\gamma$.

2. Let V, W be finite-dimensional vector spaces with bases β, γ. The function $\Phi: \mathcal{L}(V, W) \rightarrow M_{m\times n}(F)$ defined by $\Phi(T) = [T]_\beta^\gamma$ is an isomorphism. So, for linear transformations $U, T: V \rightarrow W$,
   a. $[T + U]_\beta^\gamma = [T]_\beta^\gamma + [U]_\beta^\gamma$
   b. $[aT]_\beta^\gamma = a[T]_\beta^\gamma$ for all scalars a.
   c. $\mathcal{L}(V, W)$ has dimension mn.
3. For vector spaces V, W, Z with bases α, β, γ and linear transformations $T: V \rightarrow W$, $U: W \rightarrow Z$, $[UT]_\alpha^\gamma = [U]_\beta^\gamma [T]_\alpha^\beta$.
4. T is invertible iff $[T]_\beta^\gamma$ is invertible. Then $[T^{-1}]_\gamma^\beta = ([T]_\beta^\gamma)^{-1}$.

3-3 Change of Coordinates

Let β and γ be two ordered bases for finite-dimensional vector space V. The change of coordinate matrix (from β-coordinates to γ-coordinates) is $Q = [I_V]_\beta^\gamma$. Write vector j of β in terms of the vectors of γ, take the coefficients and load them in the *j*th column of Q. (This is so (0,…1,…0) gets transformed into the *j*th column.)

1. Q^{-1} changes γ-coordinates into β-coordinates.
2. $[T]_\gamma = Q[T]_\beta Q^{-1}$

Two nxn matrices are **similar** if there exists an invertible matrix Q such that $B = Q^{-1}AQ$. Similarity is an equivalence relation. Similar matrices are manifestations of the same linear transformation in different bases.

3-4 Dual Spaces

A **linear functional** is a linear transformation from V to a field of scalars F. The **dual space** is the vector space of all linear functionals on V: $V^* = \mathcal{L}(V, F)$. V** is the double dual.

If V has ordered basis $\beta = \{x_1, x_2, \dots x_n\}$, then $\beta^* = \{f_1, f_2, \dots f_n\}$ (coordinate functions—the dual basis) is an ordered basis for V*, and for any $f \in V^*$,

$$f = \sum_{i=1}^{n} f(x_i) f_i$$

To find the coordinate representations of the vectors of the dual bases in terms of the standard coordinate functions:

1. Load the coordinate representations of the vectors in β into the columns of W.
2. The desired representation are the rows of W^{-1}.
3. The two bases are biorthogonal. For an orthonormal basis (see section 5-5), the coordinate representations of the basis and dual bases are the same.

Let V, W have ordered bases β, γ. For a linear transformation $T: V \rightarrow W$, define its transpose (or dual) $T^t: W^* \rightarrow V^*$ by $T^t(\mathrm{g}) = \mathrm{g}T$. T^t is a linear transformation satisfying $[T^t]_{\gamma^*}^{\beta^*} = \left([T]_\beta^\gamma\right)^t$.

Define $\hat{x}: V^* \rightarrow F$ by $\hat{x}(\mathrm{f}) = \mathrm{f}(x)$, and $\psi: V \rightarrow V^{**}$ by $\psi(x) = \hat{x}$. (The input is a function; the output is a function evaluated at a fixed point.) If V is finite-dimensional, ψ is an

	isomorphism. Additionally, every ordered basis for V* is the dual basis for some basis for V. The **annihilator** of a subset S of V is a subspace of V^*: $S^0 = \mathrm{Ann}(S) = \{f \in V^* \mid f(x) = 0 \ \forall \ x \in S\}$

4 Systems of Linear Equations

4-1 Systems of Linear Equations

The **system of equations**

$$\begin{cases} a_{11}x_1 + \cdots + a_{n1}x_n = b_1 \\ \vdots \\ a_{m1}x_1 + \cdots a_{mn}x_n = b_m \end{cases}$$

can be written in matrix form as Ax=b, where $A = \begin{bmatrix} a_{11} & \cdots & a_{n1} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}$ and $b = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$. The augmented matrix is $[A|b]$ (the entries of b placed to the right of A).
The system is consistent if it has solution(s). It is singular if it has zero or infinitely many solutions. If b=*0*, the system is homogeneous.

1. Row picture: Each equation gives a line/ plane/ hyperplane. They meet at the solution set.
2. Column picture: The columns of A combine (with the coefficients $x_1, \ldots x_n$) to produce b.

4-2 Elimination

There are three types of **elementary row/ column operations**:
- (1) Interchanging 2 rows/ columns
- (2) Multiplying any row/ column by a nonzero scalar
- (3) Adding any multiple of a row/ column to another row/ column

An elementary matrix is the matrix obtained by performing an elementary operation on I_n.
Any two matrices related by elementary operations are (row/column-)**equivalent**.

Performing an elementary row/ column operation is the same as multiplying by the corresponding elementary matrix on the left/ right. The inverse of an elementary matrix is an elementary matrix of the same type. When an elementary row operation is performed on an augmented matrix or the equation $Ax = b$, the solution set to the corresponding system of equations does not change.

Gaussian elimination- Reduce a system of equations (line up the variables, the equations are the rows), a matrix, or an augmented matrix by using elementary row operations.
Forward pass

1. Start with the first row.
2. Excluding all rows before the current row (row j), in the leftmost nonzero column (column k), make the entry in the current row nonzero by switching rows as necessary. (Type 1 operation) The **pivot** d_i is the first nonzero in the current row, the row that does the elimination. [Optional: divide the current row by the pivot to make the entry 1. (2)]
3. Make all numbers below the pivot zero. To make the entry a_{ik} in the *i*th row 0, subtract row j times the multiplier $l_{ik} = a_{ik}/d_i$ from row i. This corresponds to multiplication by a type 3 elementary matrix M_{ik}.
4. Move on to the next row, and repeat until only zero rows remain (or rows are exhausted).

Backward pass (Back-substitution)

5. Work upward, beginning with the last nonzero row, and add multiples of each row to

the rows above to create zeros in the pivot column. When working with equations, this is essentially substituting the value of the variable into earlier equations.

6. Repeat for each preceding row except the first.

A **free variable** is any variable corresponding to a column without a pivot. Free variables can be arbitrary, leading to infinitely many solutions. Express the solution in terms of free variables.
If elimination produces a contradiction (in A|b, a row with only the last entry a nonzero, corresponding to 0=a), there is no solution.

Gaussian elimination produces the **reduced row echelon form** of the matrix: (Forward/ backward pass accomplished 1, (2), 3/ 4.)

1. Any row containing a nonzero entry precedes any zero row.
2. The first nonzero entry in each row is 1.
3. It occurs in a column to the right of the first nonzero entry in the preceding row.
4. The first nonzero entry in each row is the only nonzero entry in its column.

The reduced row echelon of a matrix is unique.

4-3

Factorization

Elimination = Factorization

Performing Gaussian elimination on a matrix A is equivalent to multiplying A by a sequence of elementary row matrices.

If no row exchanges are made, $U = (\sum E_{ij})A$, so A can be factored in the form

$$A = \left(\sum E_{ij}^{-1}\right) U = LU$$

where L is a lower triangular matrix with 1's on the diagonal and U is an upper triangular matrix (note the factors are in opposite order). Note E_{ij} and E_{ij}^{-1} differ only in the sign of entry (i,j), and the *multipliers go directly into the entries of* L. U can be factored into a diagonal matrix D containing the pivots and U' an upper triangular matrix with 1's on the diagonal:

$$A = LDU'$$

The first factorization corresponds to the forward pass, the second corresponds to completing the back substitution. If A is symmetric, $U^{'} = L^{T}$.

Using $A = LU$, $(LU)x = Ax = b$ can be split into two triangular systems:

1. Solve $Lc = b$ for c.
2. Solve $Ux = c$ for x.

A permutation matrix P has the rows of I in any order; it switches rows.
If row exchanges are required, doing row exchanges

1. in advance gives $PA = LU$.
2. after elimination gives $A = L_1 P_1 U_1$.

4-4

The Complete Solution to Ax=b, the Four Subspaces

The rank of a matrix A is the rank of the linear transformation L_A, and the number of pivots after elimination.

Properties:

1. Multiplying by invertible matrices does not change the rank of a matrix, so elementary row and column matrices are rank-preserving.
2. rank(A^t)=rank(A)
3. Ax=b is consistent iff rank(A)=rank(A|b).
4. Rank inequalities

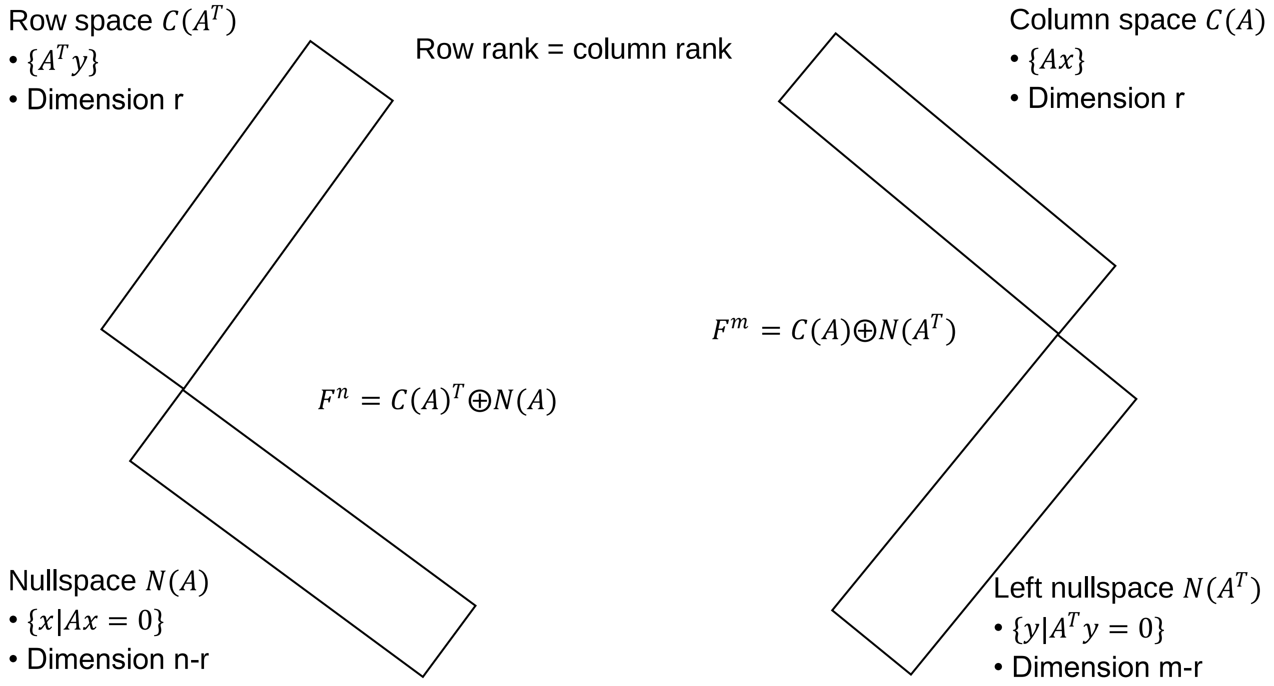
Linear transformations T, U	Matrices A, B
rank(TU) ≤ min(rank(T), rank(U))	rank(AB) ≤ min(rank(A), rank(B))

Four Fundamental Subspaces of A

1. The **row space** C(A^T) is the subspace generated by rows of A, i.e. it consists of all linear combinations of rows of A.
    a. Eliminate to find the nonzero rows. These rows are a basis for the row space.
2. The **column space** C(A) is the subspace generated by columns of A.
    a. Eliminate to find the pivot columns. These columns of A (the original matrix) are a basis for the column space. The free columns are combinations of earlier columns, with the entries of F the coefficients. (See below)
    b. This gives a technique for extending a linearly independent set to a basis: Put the vectors in the set, then the vectors in a basis down the columns of A.
3. The **nullspace** N(A) consists of all solutions to $Ax = 0$.
    a. Finding the Nullspace (after elimination)
        i. Repeat for each free variable x: Set x=1 and all other free variables to 0, and solve the resultant system. This gives a special solution for each free variable.
        ii. The special solutions found in (1) generate the nullspace.
    b. Alternatively, the nullspace matrix (containing the special solutions in its columns) is $N = \begin{bmatrix} -F \\ I \end{bmatrix}$ when the row reduced echelon form is $R = \begin{bmatrix} I & F \\ 0 & 0 \end{bmatrix}$. If columns are switched in R, corresponding rows are switched in N.
4. The **left nullspace** N(A^T) consists of all solutions to $A^T x = 0$ or $x^T A = 0$.

<u>Fundamental Theorem of Linear Algebra (Part 1)</u>:

Dimensions of the Four Subspaces: A is mxn, rank(A)=r (If the field is complex, replace A^T by A^*.)



The relationships between the dimensions can be shown using pivots or the dimension theorem.

The Complete Solution to Ax=b

1. Find the nullspace N, i.e. solve Ax=0.
2. Find any particular solution x_p to Ax=b (there may be no solution). Set free variables to 0.
3. The solution set is $N + x_p$; i.e. all solutions are in the form $x_n + x_p$, where x_n is in the nullspace and x_p is a particular solution.

4-5

Inverse Matrices

A is invertible iff it is square (nxn) and any one of the following is true:

1. A has rank n, i.e. A has n pivots.
2. $Ax = b$ has exactly 1 solution.
3. Its columns/ rows are a basis for F^n.

Gauss-Jordan Elimination: If A is an invertible nxn matrix, it is possible to transform $(A|I_n)$ into $(I_n|A^{-1})$ by elementary row operations. Follow the same steps as in Gaussian elimination, but on $(A|I_n)$. If A is not invertible, then such transformation leads to a row whose first n entries are zeros.

5 Inner Product Spaces

5-1 Inner Products

An **inner product** on a vector space V over F (ℝ or ℂ) is a function that assigns each ordered pair $(x, y) \in V$ a scalar $\langle x, y\rangle$, such that for all $x, y, z \in V$ and $c \in F$,

1. $\langle x + z, y\rangle = \langle x, y\rangle + \langle z, y\rangle$
2. $\langle cx, y\rangle = c\langle x, y\rangle$ (The inner product is linear in its first component.)
3. $\overline{\langle x, y\rangle} = \langle y, x\rangle$ (Hermitian)
4. $\langle x, x\rangle > 0$ for $x > 0$. (Positive)

V is called an inner product space, also an Euclidean/ unitary space if F is ℝ/ ℂ.
The inner product is conjugate linear in the second component:

1. $\langle x, y + z\rangle = \langle x, y\rangle + \langle x, z\rangle$
2. $\langle cx, y\rangle = \bar{c}\langle x, y\rangle$

If $\langle x, y\rangle = \langle x, z\rangle$ for all $x \in V$ then $y = z$.

The standard inner product (dot product) of $x = (a_1, \ldots, a_n)$ and $y = (b_1, \ldots, b_n)$ is

$$x \cdot y = \langle x, y\rangle = \sum_{i=1}^{n} a_i \overline{b_i}$$

The standard inner product for the space of continuous complex functions H on $[0,2\pi]$ is

$$\langle f, g\rangle = \frac{1}{2\pi}\int_0^{2\pi} f(t)\overline{g(t)}\,dt$$

A norm of a vector space is a real-valued function $\|\cdot\|$ satisfying

1. $\|cx\| = c\|x\|, c \geq 0$
2. $\|x\| \geq 0$, equality iff $x = 0$.
3. Triangle Inequality: $\|x + y\| \leq \|x\| + \|y\|$

The distance between two vectors x, y is $\|x - y\|$.

In an inner product space, the **norm** (length) of a vector is $\|x\| = \sqrt{\langle x, x\rangle}$.

Cauchy-Schwarz Inequality: $|\langle x, y\rangle| \leq \|x\|\|y\|$

5-2 Orthogonality

Two vectors are **orthogonal** (perpendicular) when their inner product is 0. A subset S is orthogonal if any two distinct vectors in S are orthogonal, **orthonormal** if additionally all vectors have length 1. Subspaces V and W are orthogonal if each $v \in V$ is orthogonal to each $w \in W$. The orthogonal complement $V^{\perp}$(V perp) of V is the subspace containing all vectors orthogonal to V. (Warning: $V^{\perp\perp} = V$ holds when V is finite-dimensional, not necessarily when V is infinite-dimensional.) ***When an orthonormal basis is chosen, every inner product on finite-dimensional V is similar to the standard inner product.*** The conditions effectively determine what the inner product has to be.

Pythagorean Theorem: If x and y are orthogonal, $\|x + y\|^2 = \|x\|^2 + \|y\|^2$.

Fundamental Theorem of Linear Algebra (Part 2):
The nullspace is the orthogonal complement of the row space.
The left nullspace is the orthogonal complement of the column space.

5-3 Projections

Take 1: Matrix and geometric viewpoint

The [orthogonal] **projection** of b onto a is

$$p = \frac{\langle b, a\rangle}{\|a\|^2} a = \frac{b \cdot a}{a \cdot a} a = \frac{a^* b}{\underbrace{a^* a}_{\hat{x}}} a$$

The last two expressions are for (row) vectors in $\mathbb{C}^n$, using the dot product. (Note: this shows that $a \cdot b = \|a\|\|b\| \cos\theta$ for 2 and 3 dimensions.)

Let S be a finite orthogonal basis. A vector y is the sum of its projections onto the vectors of S:

$$y = \sum_{v \in S} \frac{\langle y, v\rangle}{\|v\|^2} v$$

Pf. Write y as a linear combination and take the inner product of y with a vector in the basis; use orthogonality to cancel all but one term.

As a corollary, any orthogonal subset is linearly independent.

To find the projection of b onto a finite-dimensional subspace W, first find an orthonormal basis for W (see section 5-5), β. The projection is

$$p = \sum_{v \in \beta} \langle b, v\rangle\, v$$

and the error is $e = b - p$. b is perpendicular to e, and p is the vector in W so that $\|b - p\|$ is minimal. (Proof uses Pythagorean theorem)

Bessel's Inequality: (β a basis for a subspace)

$\sum_{v \in \beta} \frac{\langle y, v\rangle^2}{\|v\|^2} \le \|y\|^2$, equality iff $y = \sum_{v \in \beta} \frac{\langle y, v\rangle}{\|v\|^2} v$

If $\beta = \{v_1, \dots, v_n\}$ is an orthonormal basis, then for any linear transformation T, $\left([T]_\beta\right)_{ij} = \langle T(v_j), v_i\rangle$.

Alternatively:

Let W be a subspace of $\mathbb{C}^m$ generated by the linearly independent set $\{a_1, \dots a_n\}$. Solving $A^*(b - A\hat{x}) = 0 \Rightarrow A^* A\hat{x} = A^* b$, the projection of a onto W is

$$p = A\hat{x} = \underbrace{A(A^* A)^{-1} A^*}_{P} b$$

where P is the projection matrix. In the special case that the set is orthonormal, $Qx \approx b \Rightarrow \hat{x} = Q^T b, p = \underbrace{QQ^T}_{P} b$

A matrix P is a projection matrix iff $P^2 = P$.

Take 2: Linear transformation viewpoint

If $V = W_1 \oplus W_2$ then the **projection** on W_1 along W_2 is defined by

$$T(x) = x_1 \text{ when } x = x_1 + x_2;\ x_1 \in W_1, x_2 \in W_2$$

T is an **orthogonal projection** if $R(T)^\perp = N(T)$ and $N(T)^\perp = R(T)$. A linear operator T is an orthogonal projection iff $T^2 = T = T^*$.

5-4 Minimal Solutions and Least Squares Approximations

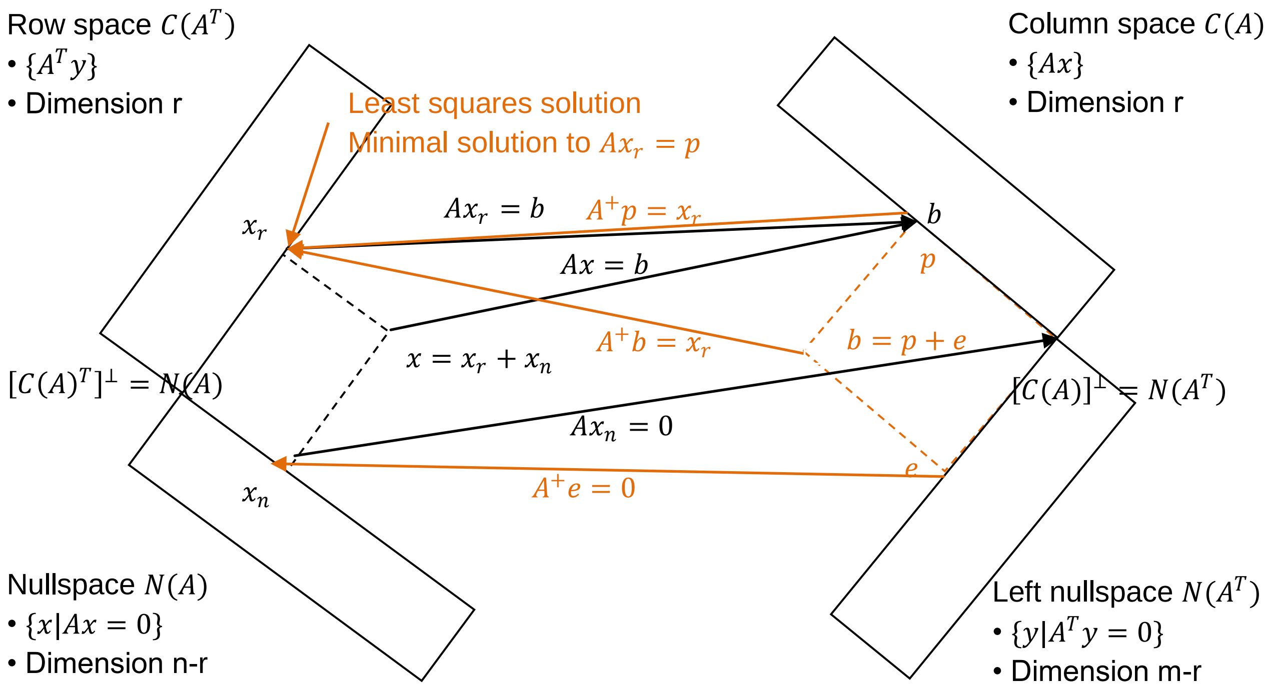
When $Ax = b$ is consistent, the **minimal solution** is the one with least absolute value.

1. There exists exactly one minimal solution s, and $s \in C(A^*)$.
2. s is the only solution to $Ax = b$ in $C(A^*)$: $(AA^*)u = b \Rightarrow s = A^*u = A^*(AA^*)^{-1}b$.

The **least squares solution** $\hat{x}$ makes $E = \|Ax - b\|^2$ as small as possible. (Generally, $Ax = b$ is inconsistent.) Project b onto the column space of A.

To find the real function in the form $y(t) = \sum_{i=1}^{m} C_i f_i(t)$ for fixed functions f_i that is closest to the points $(t_1, y_1), \ldots (t_n, y_n)$, i.e. such that the error $e = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(y_i - y(t_i)\right)^2$ is least, let A be the matrix with $A_{ij} = f_i(t_j)$, $b = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$. Then $Ax = b$ is equivalent to the system $y(t_i) = y_i$. Now find the projection of b onto the columns of A, by multiplying by A^T and solving $A^T A\hat{x} = A^T b$. Here, p is the values estimated by the best-fit curve and e gives the errors in the estimates.

Ex. Linear functions $y = C + Dt$:
$A = \begin{bmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_n \end{bmatrix}$.The equation $A^T A\hat{x} = A^T b$ becomes $\begin{bmatrix} n & \sum t_i \\ \sum t_i & \sum t_i^2 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum t_i y_i \end{bmatrix}$.
A has orthogonal columns when $\sum t_i = 0$. To produce orthogonal columns, shift the times by letting $T_i = t_i - \hat{t} = t_i - \frac{t_1 + \cdots + t_n}{n}$. Then $A^T A$ is diagonal and $C = \frac{\sum y_i}{n}, D = \frac{\sum y_i\, t_i}{\sum t_i^2}$. The least squares line is $y = C + D(t - \hat{t})$.



5-5 Orthogonal Bases

<u>Gram-Schmidt Orthogonalization Process</u>:
Let $S = \{v_1, \ldots v_n\}$ be a linearly independent subset of V. Define $S' = \{w_1, \ldots w_n\}$ by $v_1 = w_1$ and

$$v_k = w_k - \sum_{j=1}^{k-1} \frac{\langle y, v_j \rangle}{\|v_j^2\|} v_j$$

Then S’ is an orthogonal set having the same span as S. To make S’ orthonormal, divide every vector by its length. (It may be easier to subtract the projections of w_l on w_k for all $l > k$ at step k, like in elimination.)

Ex. Legendre polynomials $\frac{1}{\sqrt{2}}, \sqrt{\frac{3}{2}}x, \sqrt{\frac{5}{8}}(3x^2 - 1), \dots$ are an orthonormal basis for $\mathbb{R}[x]$ (integration from -1 to 1).

Factorization A=QR
From $a_1, \dots a_n$, Gram-Schmidt constructs orthonormal vectors $q_1, \dots q_n$. Then

$$A = QR$$

$$\begin{bmatrix} a_1 & \cdots & a_n \end{bmatrix} = \begin{bmatrix} q_1 & \cdots & q_n \end{bmatrix} \begin{bmatrix} q_1^* a_1 & q_1^* a_2 & \cdots & q_1^* a_n \\ 0 & q_2^* a_2 & \ddots & q_2^* a_n \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & q_n^* a_n \end{bmatrix}$$

Note R is upper triangular.

Suppose $S = \{v_1, \dots v_k\}$ is an orthonormal set in n-dimensional inner product space V. Then

- (a) S can be extended to an orthonormal basis $\{v_1, \dots v_n\}$ for V.
- (b) If W=span(S), $S_1 = \{v_{k+1}, \dots v_n\}$ is an orthonormal basis for $W^\perp$.
- (c) Hence, $V = W \oplus W^\perp$ and $\dim(V) = \dim(W) + \dim(W^\perp)$.

5-6 Adjoints and Orthogonal Matrices

Let V be a finite-dimensional inner product space over F, and let $g: V \to F$ be a linear transformation. The unique vector $y \in V$ such that $g(x, y) = \langle x, y \rangle$ for all $x \in V$ is given by

$$y = \sum_{i=1}^{n} \overline{g(v_i)} v_i$$

Let $T: V \to W$ be a linear transformation, and β and γ be bases for inner product spaces V, W. Define the **adjoint** of T to be the linear transformation $T^*: W \to V$ such that $[T^*]_\gamma^\beta = ([T]_\beta^\gamma)^*$. (See section 2.3) Then T^* is the unique (linear) function such that $\langle T(x), y \rangle_W = \langle x, T^*(y) \rangle_V$ for all $x \in V, y \in W$ and $c \in F$.

A linear operator T on V is an **isometry** if $\|T(x)\| = \|x\|$ for all $x \in V$. If V is finite-dimensional, T is **orthogonal** for V real and **unitary** for V complex. The corresponding matrix representations, as well as properties of T, are described below.

	Commutative property	Inverse property	Symmetry property
Real	**Normal** $AA^T = A^T A$	Orthogonal $A^T A = I$	Symmetric $A^T = A$
Complex	Normal $AA^* = A^* A$	Unitary $A^* A = I$	Self-adjoint/ Hermitian $A^* = A$
Linear Transformation	$\langle Tv, Tw \rangle = \langle T^* v, T^* w \rangle$ $\|Tv\| = \|T^* x\|$	$\langle Tv, Tw \rangle = \langle v, w \rangle$ $\|Tv\| = \|v\|$ $(Ux)^T (Uy) = x^T y$	$\langle Tv, w \rangle = \langle v, Tw \rangle$

A real matrix Q has orthonormal columns iff $Q^TQ = I$. If Q is square it is called an **orthogonal** matrix, and its inverse is its transpose.
A complex matrix U has orthonormal columns iff $U^*U = I$. If U is square it is a **unitary** matrix, and its inverse is its adjoint.
If U has orthonormal columns it leaves lengths unchanged ($\|Ux\| = \|x\|$ for every x) and preserves dot products $(Ux)^T(Uy) = x^Ty$.
A^*A is invertible iff A has linearly independent columns. More generally, A^*A has the same rank as A.

5-7

Geometry of Orthogonal Operators

A **rigid motion** is a function $f: V \to V$ satisfying $\|f(x) - f(y)\| = \|x - y\|$ for all $x, y \in V$. Each rigid motion is the composition of a translation and an orthogonal operator.

A (orthogonal) linear operator is a

1. **rotation** (around $W^\perp$) if there exists a 2-dimensional subspace $W \subseteq V$ and an orthonormal basis $\beta = \{x_1, x_2\}$ for W, and θ such that $T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$.
   and $T(y) = y$ for $y \in W^\perp$.
2. **reflection** (about $W^\perp$) if W is a one-dimensional subspace of V such that $T(x) = -x$ for all $x \in W$ and $T(y) = y$ for all $y \in W^\perp$**.**

<u>Structural Theorem for Orthogonal Operators:</u>

1. Let T be an orthogonal operator on finite-dimensional real inner product space V. There exists a collection of pairwise orthogonal T-invariant subspaces $\{W_1, \dots, W_m\}$ of V of dimension 1 or 2 such that $V = W_1 \oplus \cdots \oplus W_m$. Each T_{W_i} is a rotation or reflection; the number of reflections is even/ odd when $\det(T) = 1/ \det(T) = -1$. It is possible to choose the subspaces so there is 0 or 1 reflection.
2. If A is orthogonal there exists orthogonal Q such that
   $$QTQ^{-1} = \begin{bmatrix} I_p & & & & \\ & -I_q & & & \\ & & R_{\theta_1} & & \\ & & & \ddots & \\ & & & & R_{\theta_n} \end{bmatrix}$$ where p, q are the dimensions of N(T-I), N(T+I)
   and $R_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$.

Alternate method to factor QR:
Q is a product of reflection matrices $I - 2uu^T$ and plane rotation matrices (Givens rotation) in the form (1s on diagonal. Shown are rows/ columns i, j).

$$Q_{ij} = \begin{bmatrix} \ddots & & & & \\ & \cos(\theta) & & -\sin(\theta) & \\ & & \ddots & & \\ & \sin(\theta) & & \cos(\theta) & \\ & & & & \ddots \end{bmatrix}$$

Multiply by Q_{ij} to produce 0 in the (i,j) position, as in elimination.

$$\left(\prod Q_{ij}\right) A = R \Rightarrow A = \underbrace{\left(\prod Q_{ij}^{-1}\right)}_{Q} R$$

where the factors are reversed in the second product.

6 Determinants

6-1 Characterization

The **determinant** (denoted $|A|$ or $\det(A)$) is a function from the set of square matrices to the field F, satisfying the following conditions:

1. The determinant of the nxn identity matrix is 1, i.e. $\det(I) = 1$.
2. If two rows of A are equal, then $\det(A) = 0$, i.e. the determinant is alternating.
3. The determinant is a linear function of each row separately, i.e. it is n-linear. That is, if $a_1, \ldots a_n, u, v$ are rows with n elements,

$$\det\begin{pmatrix} a_1 \\ \vdots \\ a_{r-1} \\ u+kv \\ a_{r+1} \\ \vdots \\ a_n \end{pmatrix} = \det\begin{pmatrix} a_1 \\ \vdots \\ a_{r-1} \\ u \\ a_{r+1} \\ \vdots \\ a_n \end{pmatrix} + k\det\begin{pmatrix} a_1 \\ \vdots \\ a_{r-1} \\ v \\ a_{r+1} \\ \vdots \\ a_n \end{pmatrix}$$

These properties completely characterize the determinant.

4. The determinant changes sign when two rows are exchanged.
5. Adding a multiple of one row to another row leaves $\det(A)$ unchanged.
6. A matrix with a row of zeros has $\det(A) = 0$.
7. If A is triangular then $\det(A) = a_{11}a_{22}\cdots a_{nn}$ is the product of diagonal entries.
8. A is singular iff $\det(A) = 0$.
9. $\det(AB) = \det(A)\det(B)$
10. A^T has the same determinant as A. Therefore the preceding properties are true if "row" is replaced by "column".

6-2 Calculation

1. The Big Formula: Use n-linearity and expand everything.

$$\det(A) = \sum_{\sigma\in\mathfrak{S}_n} \operatorname{sgn}(\sigma) A_{1,\sigma(1)}A_{2,\sigma(2)}\cdots A_{n,\sigma(n)}$$

where the sum is over all $n!$ permutations of {1,…n} and $\operatorname{sgn}(\sigma) = \begin{cases} 1, \text{if } \sigma \text{ is even} \\ -1, \text{if } \sigma \text{ is odd} \end{cases}$.

2. Cofactor Expansion: Recursive, useful with many zeros, perhaps with induction.

(Row)

$$\det(A) = \sum_{j=1}^{n} a_{ij}C_{ij} = \sum_{j=1}^{n} a_{ij}(-1)^{i+j}\det(M_{ij})$$

(Column)

$$\det(A) = \sum_{i=1}^{n} a_{ij}C_{ij} = \sum_{i=1}^{n} a_{ij}(-1)^{i+j}\det(M_{ij})$$

where M_{ij} is A with the *i*th row and *j*th column removed.

3. Pivots:

If the pivots are $d_1, d_2, \ldots d_n$, and $PA = LU$, (P a permutation matrix, L is lower triangular, U is upper triangular)

$\det(A) = \det(P)\,(d_1 d_2 \cdots d_n)$ where det(P)=1/ -1 if P corresponds to an even/ odd permutation.

   a. Let A_k denote the matrix consisting of the first k rows and columns of A. If there are no row exchanges in elimination,

$$d_k = \frac{\det(A_k)}{\det(A_{k-1})}$$

4. By Blocks:
   a. $\begin{vmatrix} A & B \\ O & C \end{vmatrix} = |\mathrm{A}||C|$
   b. $\begin{vmatrix} A & B \\ C & D \end{vmatrix} = \begin{vmatrix} A & B \\ O & D - CA^{-1}B \end{vmatrix} = |A||D - CA^{-1}B|$

Tips and Tricks

Vandermonde determinant (look at when the determinant is 0, gives factors of polynomial)

$$\begin{vmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{n-1} & x_2^{n-1} & \cdots & x_n^{n-1} \end{vmatrix} = \prod_{i>j} (x_i - x_j)$$

Circulant Matrix (find eigenvectors, determinant is product of eigenvalues)

$$\begin{vmatrix} a_0 & a_1 & \cdots & a_{n-1} \\ a_{n-1} & a_0 & \cdots & a_{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ a_1 & a_2 & \cdots & a_0 \end{vmatrix} = \prod_{j=0}^{n-1} \sum_{k=0}^{n-1} \left(e^{\frac{2\pi i}{n}}\right)^{jk} a_k$$

$$\begin{vmatrix} a_1 & x & \cdots & x \\ x & a_2 & \cdots & x \\ \vdots & \vdots & \ddots & \vdots \\ x & x & \cdots & a_n \end{vmatrix} = (a_1 - x) \cdots (a_n - x) + x \sum_{i=1}^{n} \prod_{j \neq i} (a_j - x)$$

$$= x(a_1 - x) \cdots (a_n - x) \left(\frac{1}{x} + \frac{1}{a_1 - x} + \cdots + \frac{1}{a_n - x}\right)$$

For a real matrix A,

$$\det(I + A^2) = \|\det(I + iA)\|^2 \geq 0$$

If A has eigenvalues $\lambda_1, \ldots, \lambda_n$, then

$$\det(A + \lambda I) = (\lambda_1 + \lambda) \cdots (\lambda_n + \lambda)$$

In particular, if M has rank 1,

$$\det(I + M) = 1 + \mathrm{tr}(M)$$

6-3

Properties and Applications

Cramer's Rule:

If A is a nxn matrix and $\det(A) \neq 0$ then $Ax = b$ has the unique solution given by

$$x_i = \frac{\det(B_i)}{\det(A)}, 1 \leq i \leq n$$

Where B_i is A with the *i*th column replaced by b.

Inverses:

Let C be the cofactor matrix of A. Then

$$A^{-1} = \frac{C^T}{\det(A)}$$

The **cross product** of $u = (u_1, u_2, u_3)$ and $v = (v_1, v_2, v_3)$ is

$$u \times v = \begin{vmatrix} i & j & k \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix}$$

a vector perpendicular to u and v (direction determined by the right-hand rule) with length

	$\lVert u\rVert\lVert v\rVert\lvert\sin\theta\rvert$. Geometry: The area of a parallelogram with vertices sides $\langle x_1, y_1\rangle, \langle x_2, y_2\rangle$ is $\begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix}$. (Oriented areas satisfy the same properties as determinants.) The area of a parallelepiped with sides $u = (u_1, u_2, u_3)$, $v = (v_1, v_2, v_3)$, and $u = (w_1, w_2, w_3)$ is $(u \times v) \cdot w = \begin{vmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix}$ The **Jacobian** used to change coordinate systems in integrals is $\begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} & \frac{\partial y}{\partial w} \\ \frac{\partial z}{\partial u} & \frac{\partial z}{\partial v} & \frac{\partial z}{\partial w} \end{vmatrix}$.

7 Eigenvalues and Eigenvectors, Diagonalization

7-1 Eigenvalues and Eigenvectors

Let T be a linear operator (or matrix) on V. A nonzero vector $v \in V$ is an (right) **eigenvector** of T if there exists a scalar λ, called the **eigenvalue**, such that $T(v) = \lambda v$. The **eigenspace** of λ is the set of all eigenvectors corresponding to λ: $E_\lambda = \{x \in V | T(x) = \lambda x\}$.

The characteristic polynomial of a matrix A is $\det(A - \lambda I)$. The zeros of the polynomial are the eigenvalues of A. For each eigenvalue solve $Av = \lambda v$ to find linearly independent eigenvalues that span the eigenspace.

Multiplicity of an eigenvalue λ:

1. Algebraic (μ_{alg})- the multiplicity of the root λ in the characteristic polynomial of A.
2. Geometric (μ_{geom})- the dimension of the eigenspace of λ. $1 \leq \dim(E_\lambda) \leq \mu_{alg}(\lambda)$. $\dim(E_\lambda) = \dim\big(N(A - \lambda I)\big) = n - \operatorname{rank}(A - \lambda I)$.

For real matrices, complex eigenvalues come in conjugate pairs.

The product of the eigenvalues (counted by algebraic multiplicity) equals $\det(A)$.
The sum of the eigenvalues equals the trace of A.

An eigenvalue of 0 implies that A is singular.

Spectral Mapping Theorem:
Let A be a nxn matrix with eigenvalues $\lambda_1, \ldots, \lambda_n$ (not necessarily distinct, counted according to algebraic multiplicity), and P be a polynomial. Then the eigenvalues of $P(A)$ are $P(\lambda_1), \ldots, P(\lambda_n)$.

Gerschgorin's Disk Theorem:
Every eigenvalue of A is strictly in a circle in the complex plane centered at some diagonal entry A_{ii} with radius $r_i = \sum_{j \neq i} |a_{ij}|$ (because $(\lambda - A_{ii})x_i = \sum_{j \neq i} a_{ij} x_j$).

Perron-Frobenius Theorem:
Any square matrix with positive entries has a unique eigenvector with positive entries (up to multiplication by a positive factor), and the corresponding eigenvalue has multiplicity one and has strictly greater absolute value than any other eigenvalue.
Generalization: Holds for any irreducible matrix with nonnegative entries, i.e. there is no reordering of rows and columns that makes it block upper triangular.

A left eigenvalue of A satisfies $v^T A = \lambda v$ instead. Biorthogonality says that any right eigenvector of A associated with λ is orthogonal to all left eigenvectors of A associated with eigenvalues other than λ.

7-2 Invariant and T-Cyclic Subspaces

The subspace $C_x = Z(x; T) = W = \operatorname{span}(\{x. T(x), T^2(x), \ldots\})$ is the **T-cyclic subspace** generated by x. W is the smallest T-invariant subspace containing x.

1. If W is a T-invariant subspace, the characteristic polynomial of T_W divides that of T.
2. If k=dim(W) then $\beta_x = \{x, T(x), \ldots, T^{k-1}(x)\}$ is a basis for W, called the T-cyclic basis

generated by x. If $\sum_{i=0}^{k} a_i T^i(x) = 0$ with $a_k = 1$, the characteristic polynomial of T$_W$ is $(-1)^k \sum_{i=0}^{k} a_i t^i$.

3. If $V = W_1 \oplus W_2 \cdots W_k$, each W_i is a T-invariant subspace, and the characteristic polynomial of T_{W_i} is $f_i(t)$, then the characteristic polynomial of T is $\prod_{i=1}^{k} f_i(t)$.

Cayley-Hamilton Theorem:
A satisfies its own characteristic equation: if $f(t)$ is the characteristic polynomial of A, then $f(A) = \mathcal{O}$.

7-3 Triangulation

A matrix is **triangulable** if it is similar to an upper triangular matrix.
(Schur) A matrix is triangulable iff the characteristic polynomial splits over F. A real/ complex matrix A is unitarily/ orthogonally equivalent to a real/ complex upper triangular matrix. (i.e. $A = QTQ^{-1}$, Q is orthogonal/ unitary)
Pf. T=L$_A$ has an eigenvalue iff T* has. Induct on dimension n. Choose an eigenvector z of T*, and apply the induction hypothesis to the T-invariant subspace $\mathrm{span}(z)^{\perp}$.

7-4 Diagonalization

T is **diagonalizable** if there exists an ordered basis β for V such that $[T]_\beta$ is diagonal. A is diagonalizable if there exists an invertible matrix S such that $S^{-1}AS = \Lambda$ is a diagonal matrix.

Let $\lambda_1, \dots, \lambda_k$ be the eigenvalues of A. Let S_i be a linearly independent subset of E_{λ_i} for $1 \le i \le k$. Then $\bigcup S_i$ is linearly independent. (Loosely, eigenvectors corresponding to different eigenvalues are linearly independent.)

T is diagonalizable iff both of the following are true:

1. The characteristic polynomial of T splits (into linear factors).
2. For each eigenvalue, the algebraic and geometric multiplicities are equal. Hence there are n linearly independent eigenvectors

T is diagonalizable iff V is the direct sum of eigenspaces of T.

To diagonalize A, put the n linearly independent eigenvectors into the columns of A. Put the corresponding eigenvalues into the diagonal entries of Λ. Then

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

For a linear transformation, this corresponds to

$$[T]_\beta = [I]_\gamma^\beta [T]_\gamma [I]_\beta^\gamma$$

Simultaneous Triangulation and Diagonalization
Commuting matrices share eigenvectors, i.e. given that A and B can be diagonalized, there exists a matrix S that is an eigenvector matrix for both of them iff $AB = BA$. Regardless, AB and BA have the same set of eigenvalues, with the same multiplicities.
More generally, let $\mathfrak{F}$ be a commuting family of triangulable/ diagonalizable linear operators on V. There exists an ordered basis for V such that every operator in $\mathfrak{F}$ is simultaneously represented by a triangular/ diagonal matrix in that basis.

7-5 Normal Matrices

(For review see 5-6)

A nxn [real] symmetric matrix:

1. Has only real eigenvalues.
2. Has eigenvalues that can be chosen to be orthonormal. ($S = Q, Q^{-1} = Q^T$) (See below.)
3. Has n linearly independent eigenvectors so can be diagonalized.
4. The number of positive/ negative eigenvalues equals the number of positive/ negative pivots.

For real/ complex finite-dimensional inner product spaces, T is symmetric/ normal iff there exists an orthonormal basis for V consisting of eigenvectors of T.

Spectral Theorem (Linear Transformations)
Suppose T is a normal linear operator ($T^*T = TT^*$) on a finite-dimensional real/ complex inner product space V with distinct eigenvalues $\lambda_1, \ldots, \lambda_n$ (its spectrum). Let W_i be the eigenspace of T corresponding to λ_i and T_i the orthogonal projection of V on W_i.

1. T is diagonalizable and $V = W_1 \oplus \cdots \oplus W_n$.
2. W_i is orthogonal to the direct sum of W_j with $j \neq i$.
3. There is an orthonormal basis of eigenvectors.
4. Resolution of the identity operator: $I = T_1 + \cdots + T_n$
5. Spectral decomposition: $T = \lambda_1 T_1 + \cdots + \lambda_k T_n$

Pf. The triangular matrix in the proof of Schur's Theorem is actually diagonal.

1. If $Ax = \lambda x$ then $A^*x = \bar{\lambda}x$.
2. W is T-invariant iff $W^\perp$ is T^*-invariant.
3. Take a eigenvector v; let $W = \mathrm{span}(v)$. From (1) v is an eigenvector of T^*; from (2) $W^\perp$ is T-invariant.
4. Write $V = W \oplus W^\perp$. Use induction hypothesis on $W^\perp$.

(Matrices)
Let A be a normal matrix ($A^*A = AA^*$). Then A is diagonalizable with an orthonormal basis of eigenvectors:

$$A = U\Lambda U^*$$

where Λ is diagonal and U in unitary.

Type of Matrix	Condition	Factorization
Hermitian (Self-adjoint)	$A^* = A$	$A = U\Lambda U^{-1}$ U unitary, Λ real diagonal Real eigenvalues (because $\lambda v^* v = v^* A v = \bar{\lambda} v^* v$)
Unitary	$A^*A = I$	$A = U\Lambda U^{-1}$ U unitary, Λ diagonal Eigenvalues have absolute value 1
Symmetric (real)	$A^T = A$	$A = Q\Lambda Q^{-1}$ Q orthogonal, Λ real diagonal Real eigenvalues
Orthogonal (real)	$A^T A = I$	$A = Q\Lambda Q^{-1}$ Q unitary, Λ diagonal Eigenvalues have absolute value 1

7-6 Positive Definite Matrices and Operators

A real matrix A is **positive (semi)definite** if $x^*Ax > 0$ ($x^*Ax \geq 0$) for every nonzero vector x. A linear operator T on a finite-dimensional inner product space is positive (semi)definite if T is self-adjoint and $\langle T(x), x\rangle > 0$ ($\langle T(x), x\rangle \geq 0$) for all $x \neq 0$.

The following are equivalent:

1. A is positive definite.
2. All eigenvalues are positive.
3. All upper left determinants are positive.
4. All pivots are positive.

Every positive definite matrix factors into

$$A = LDU' = LDL^T$$

with positive pivots in D. The **Cholesky factorization** is

$$A = \left(L\sqrt{D}\right)\left(L\sqrt{D}\right)^T$$

7-7 Singular Value Decomposition

Every $m \times n$ matrix A has a **singular value decomposition** in the form

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

where U and V are unitary matrices and $\Sigma = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{bmatrix}$ is diagonal. The singular values $\sigma_1, \dots \sigma_r$ ($\sigma_k = 0$ for $k > r = \mathrm{rank}(A)$) are positive and are in decreasing order, with zeros at the end (not considered singular values).

If A corresponds to the linear transformation $T: V \to W$, then this says there are orthonormal bases $\beta = \{v_1, \dots, v_n\}$ and $\gamma = \{u_1, \dots, u_m\}$ such that

$$T(v_i) = \begin{cases} \sigma_i u_i \text{ if } 1 \leq i \leq r \\ 0 \text{ if } i > r \end{cases}$$

Letting β', γ' be the standard ordered bases for V, W,

$$AV = U\Sigma \Leftrightarrow [T]_{\beta'}^{\gamma'}[I]_{\beta}^{\beta'} = [I]_{\gamma}^{\gamma'}[T]_{\beta}^{\gamma}$$

Orthogonal elements in the basis are sent to orthogonal elements; the singular values give the factors the lengths are multiplied by.

To find the SVD:

1. Diagonalize A^*A, choosing orthonormal eigenvectors. The eigenvalues are the squares of the singular values and the eigenvector matrix is V.
   $$A^*A = V\Sigma^2V^* = V\begin{bmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_n^2 \end{bmatrix}V^*$$
2. Similarly,
   $$AA^* = U\Sigma^2U^*$$
   If V and the singular values have already been found, the columns of U are just the images of $v_1, \dots, v_n$ under left multiplication by A: $u_i = Av_i$, unless this gives 0.
3. If A is a mxn matrix:
   1. The first r columns of V generate the row space of A.
   2. The last n-r columns generate the nullspace of A.
   3. The first r columns of U generate the column space of A.
   4. The last m-r columns of U generate the left nullspace of A.

The **pseudoinverse** of a matrix A is the matrix A^+ such that for $y \in C(A)$, A^+y is the vector x in the row space such that $Ax = y$, and for $y \in N(A^T)$, $A^+y = 0$. For a linear transformation, replace $C(A)$ with $R(T)$ and $N(A^T)$ with $R(T)^\perp$. In other words,

1. AA^+ is the projection matrix onto the column space of A.
2. A^+A is the projection matrix onto the row space of A.

Finding the pseudoinverse:

$$A^+ = V\Sigma^+U^* = V\begin{bmatrix} \sigma_1^{-1} & & & \\ & \ddots & & \\ & & \sigma_r^{-1} & \\ & & & \end{bmatrix}U^*$$

The shortest least squares solution to $Ax = b$ is $x^+ = A^+b$.
See Section 5-4 for a picture.

The **polar decomposition** of a complex (real) matrix A is

$$A = QH$$

where Q is unitary (orthogonal) and H is semi-positive definite Hermitian (symmetric). Use the SVD:

$$A = (UV^*)(V\Sigma V^*)$$

If A is invertible, Q is positive definite and the decomposition is unique.

Summary

Type of matrix	Eigenvalues	Eigenvectors (can be chosen…)
Real symmetric	Real	Orthogonal
Orthogonal	Absolute value 1	
Skew-symmetric	(Pure) Imaginary	
Self-adjoint	Real	
Positive definite	Positive	

8 Canonical Forms

A canonical form is a standard way of presenting and grouping linear transformations or matrices. Matrices sharing the same canonical form are similar; each canonical form determines an equivalence class.
Similar matrices share…

- Eigenvalues
- Trace and determinant
- Rank
- Number of independent eigenvectors
- Jordan/ Rational canonical form

8-1 Decomposition Theorems

A **minimal polynomial** of T is the (unique) monic polynomial $p(t)$ of least positive degree such that $p(T) = T_0$. If $g(T) = T_0$ then $p(t)|g(t)$; in particular, $p(t)$ divides the characteristic polynomial of T.

Let W be an invariant subspace for T and let $x \in V$. The **T-conductor** ("T-stuffer") of x into W is the set $S_T(x;W)$ which consists of all polynomials g over F such that $(g(T))(x) \in W$. (It may also refer to the monic polynomial of least degree satisfying the condition.)
If $W = \{0\}$, T is called the **T-annihilator** of x, i.e. it is the (unique) monic polynomial $p(t)$ of least degree for which $p(T)(x) = 0$. The T-conductor/ annihilator divides any other polynomial with the same property.
The T-annihilator $p(t)$ is the minimal polynomial of T_W, where W is the T-cyclic subspace generated by x. The characteristic polynomial and minimal polynomial of T_W are equal or negatives.

Let L be a linear operator on V, and W a subspace of V. W is **T-admissible** if

1. W is invariant under T.
2. If $f(T)x \in W$, there exists $y \in W$ such that $f(T)(x) = f(T)(y)$.

Let T be a linear operator on finite-dimensional V.
Primary Decomposition Theorem (leads to Jordan form):
Suppose the minimal polynomial of T is

$$p(t) = \prod_{i=1}^{k} {p_i}^{r_i}$$

where p_i are distinct irreducible monic polynomials and r_i are positive integers. Let W_i be the null space of $p_i(T)^{r_i}$. Then

1. $V = W_1 \oplus \cdots \oplus W_k$.
2. Each W_i is invariant under T.
3. The minimal polynomial of T_{W_i} is $p_i^{r_i}$.

Pf. Let $f_i = \frac{p}{p_i^{r_i}}$. Find g_i so that $\sum_{i=1}^{n} f_i g_i = 1$. $E_i = f_i(T)g_i(T)$ is the projection onto W_i.

Cyclic Decomposition Theorem (leads to rational canonical form):
Let T be a linear operator on finite-dimensional V and W_0 (often taken to be $\{0\}$) a proper T-admissible subspace of V. There exist nonzero $x_1, \dots x_r$ with (unique) T-annihilators $p_1, \dots, p_r$, called **invariant factors** such that

1. $V = W_0 \oplus Z(x_1;T) \oplus \cdots \oplus Z(x_r;T)$
2. $p_k | p_{k-1}$ for $2 \le k \le r$.

Pf.

1. There exist nonzero vectors $\beta_1, \ldots, \beta_r$ in V such that
   a. $V = W_0 + Z(\beta_1;T) + \cdots + Z(\beta_r;T)$
   b. If $1 \le k \le r$ and $W_k = W_0 + Z(\beta_1;T) + \cdots + Z(\beta_k;T)$ then p_k has maximum degree among all T-conductors into W_{k-1}.
2. Let $f = s(\beta; W_{k-1})$. If $f(T)(\beta) = \beta_0 + \sum_{1 \le i < k} g_i(T)(\beta_i)$, $\beta_i \in W_i$ then $g_i = fh_i$ for some h_i and $f = f(T)(\gamma_0)$ for some $\gamma_0 \in W_0$. (Stronger form of condition that each W_i is T-admissible.)
3. Existence: Let $x_k = \beta_k - \gamma_0 - \sum_{1 \le i < k} h_i \beta_i$. $\beta_k - x_k \in W_{k-1}, \beta_k \in W_k$ implies $s(x_k; W_{k-1}) = s(\beta_k; W_{k-1}) = p_k$ and $W_k = W_0 + Z(x_1;T) + \cdots + Z(x_k;T)$.
4. Uniqueness: Induct. Show p_1 is unique. If p_i is unique, operate p_{i+1} on both sides of 2 decompositions of V to show that $p_{i+1} | q_{i+1}$ and vice versa.

8-2

Jordan Canonical Form

$[T]_\beta$ is a **Jordan canonical form** of T if

$$[T]_\beta = \begin{bmatrix} A_1 & O & \cdots & O \\ O & A_2 & \cdots & O \\ \vdots & \vdots & \ddots & \vdots \\ O & O & \cdots & A_n \end{bmatrix}$$

where each A_i is a Jordan block in the form

$$\begin{bmatrix} \lambda & 1 & \cdots & 0 & 0 \\ 0 & \lambda & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \lambda & 1 \\ 0 & 0 & \cdots & 0 & \lambda \end{bmatrix}$$

with λ an eigenvalue.

Nonzero $x \in V$ is a **generalized eigenvector** corresponding to λ if $(T - \lambda I)^p(x) = 0$ for some p. The **generalized eigenspace** consists of all generalized eigenvectors corresponding to λ:

$$K_\lambda = \{x \in V | (T - \lambda I)^p(x) = 0 \text{ for some positive integer } p\}$$

If p is the smallest positive integer so that $(T - \lambda I)^p(x) = 0$,

$$\{(T - \lambda I)^{p-1}(x), \ldots, (T - \lambda I)(x), x\}$$

is a cycle of generalized eigenvectors corresponding to λ. Every such cycle is linearly independent.

<u>Existence</u>

K_λ (the W_i in the Primary Decomposition Theorem) has an ordered basis consisting of a union of disjoint cycles of generalized eigenvectors corresponding to λ. Thus every linear transformation (or matrix) on a finite-dimensional vector space, whose characteristic polynomial splits, has a Jordan canonical form. V is the direct sum of the generalized eigenspaces of T.

<u>Uniqueness and Structure</u>

The Jordan canonical form is unique (when cycles are listed in order of decreasing length) up to ordering of eigenvalues.

Suppose β_i is a basis for K_{λ_i}. Let T_i be the restriction of T to K_{λ_i}. Suppose β_i is a disjoint union of cycles of generalized eigenvectors $\gamma_1, \dots, \gamma_{n_i}$ with lengths $p_1 \geq \cdots \geq p_{n_i}$. The dot diagram for T_i contains one dot for each vector in β_i, and

1. has n_i columns, one for each cycle.
2. The jth column consists of p_j dots that correspond to the vectors of γ_j, starting with the initial vector.

The dot diagram of T_i is unique: The number of dots in the first r rows equals $\mathrm{nullity}((T - \lambda iIr)$, or if rj is the number of dots in the jth row, rj=$\mathrm{rank}T-\lambda iIj-1-\mathrm{rank}(T-\lambda iIj)$. In particular, the number of cycles is the geometric multiplicity of λ_i.

The Jordan canonical form is determined by the eigenvalues and $\mathrm{nullity}((\mathrm{T} - \lambda_\mathrm{i}\mathrm{I})^\mathrm{r})$ *for every eigenvalue* λ_i*.*

So now we know…

Supposing $p(t)$ splits, let $\lambda_1, \dots, \lambda_k$ be the distinct eigenvalues of T, and let p_i be the order of the largest Jordan block corresponding to λ_i. The minimal polynomial of T is

$$p(t) = \prod_{i=1}^{k} (t - \lambda_i)^{p_i}$$

T is diagonalizable iff all exponents are 1.

8-3

Rational Canonical Form

Let T be a linear operator on finite-dimensional V with characteristic polynomial

$$f(t) = (-1)^n \prod_{i=1}^{k} \left(p_i(t)\right)^{n_i}$$

where the factors $p_i(t)$ are distinct irreducible monic polynomials and n_i are positive integers. Define

$$K_{p_i} = \{x \in V | p_i(T)^k(x) = 0 \text{ for some positive integer } k\}$$

Note this is a generalization of the generalized eigenspace.

The **companion matrix** of the monic polynomial $p(t) = a_0 + a_1 t + \cdots + a_{k-1}t^{k-1} + t^k$ is

$$C(p) = \begin{bmatrix} 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & \cdots & 0 & -a_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -a_{k-1} \end{bmatrix}$$ because the characteristic polynomial of c(p) is $(-1)^k p(t)$.

Every linear operator T on finite-dimensional V has a **rational canonical form** (Frobenius normal form) *even if the characteristic polynomial does not split.*

$$[T]_\beta = \begin{bmatrix} C_1 & O & \cdots & O \\ O & C_2 & \cdots & O \\ \vdots & \vdots & & \vdots \\ O & O & \cdots & C_r \end{bmatrix}$$

where each C_i is the companion matrix of an invariant factor p_i.

<u>Uniqueness and Structure:</u>

The rational canonical form is unique under the condition $p_{i+1} | p_i$ for each $1 \leq i < r$.

The rational canonical form is determined by the prime factorization of f(t) and $\mathrm{nullity}(\mathrm{p_i(T)^r})$ *for every positive integer r.*

Generalized Cayley-Hamilton Theorem:
Suppose the characteristic polynomial of T is

$$f(t) = \prod_{i=1}^{k} {p_i}^{r_i}$$

where p_i are distinct irreducible monic polynomials and r_i are positive integers. Then the minimal polynomial of T is

$$p(t) = \prod_{i=1}^{k} {p_i}^{d_i}$$

where $d_i = \frac{\text{nullity } (p_i(T)^{r_i})}{\deg (p_i)}$.

8-4

Calculation of Invariant Factors

For a matrix over the polynomials F[x], elementary row/ column operations include:
  (1) Interchanging 2 rows/ columns
  (2) Multiplying any row/ column by a nonzero *scalar*
  (3) Adding any *polynomial* multiple of a row/ column to another row/ column
However, note arbitrary division by polynomials is illegal in F[x].

For such a (mxn) polynomial F[x], the following are equivalent:
1. P is invertible.
2. The determinant of P is a nonzero scalar.
3. P is row-equivalent to the mxm identity matrix.
4. P is a product of elementary matrices.

A $m \times n$ matrix is in **Smith normal form** if
1. Every entry not on the diagonal is 0.
2. On the main diagonal of N, there appear polynomials $f_1, \ldots f_l$ such that $f_k | f_{k+1}, 1 \leq k < \min(m, n)$.

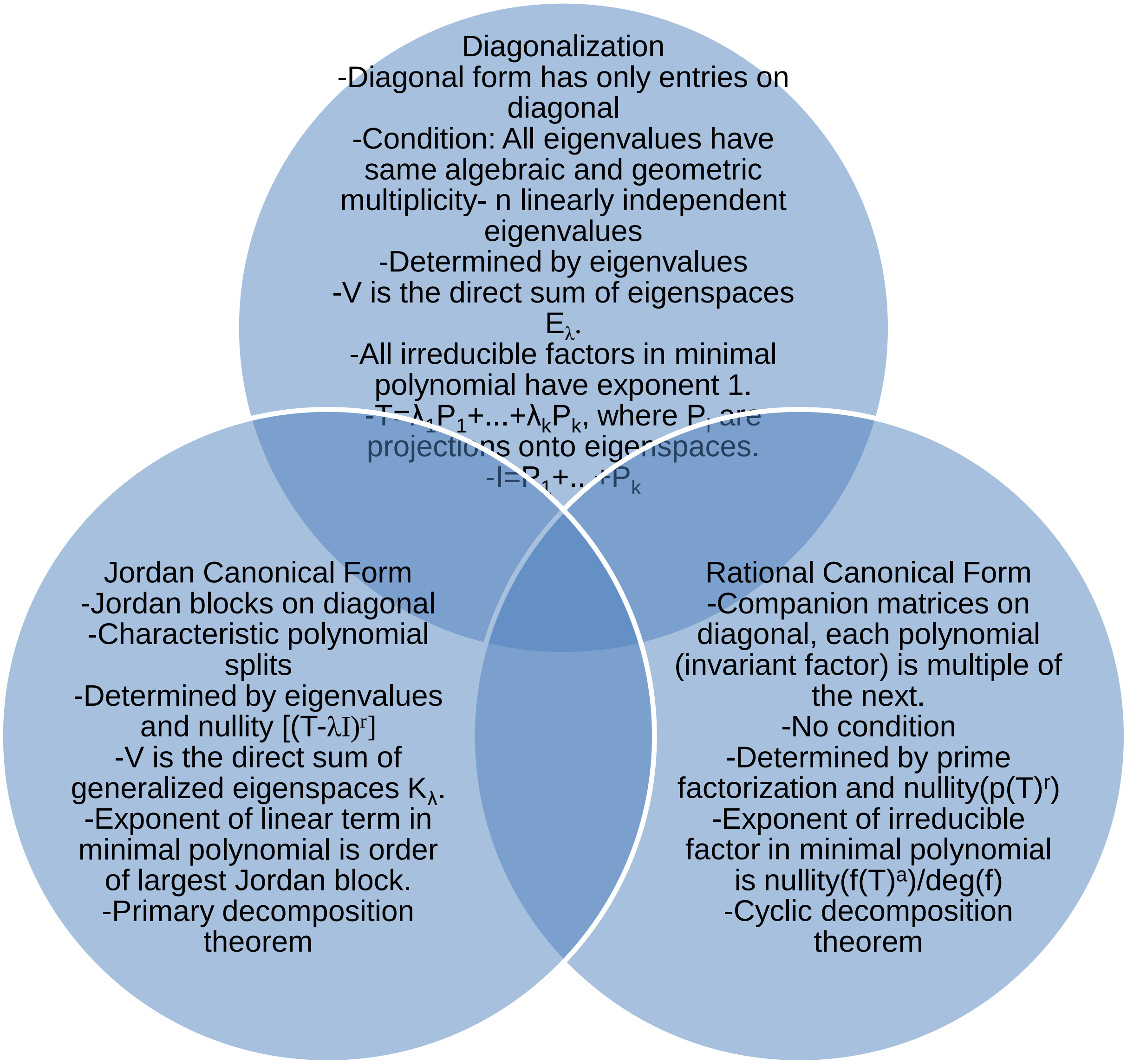
Every matrix is equivalent to a unique matrix N in normal form. For a $m \times n$ matrix A, follow this algorithm to find it:
1. Make the first column $\begin{bmatrix} p \\ 0 \\ \vdots \\ 0 \end{bmatrix}$.
    a. Choose the nonzero entry f in the first column that has the least degree.
    b. For each other nonzero entry p, use polynomial division to write $p = fq + r$, where r is the remainder upon division. Subtract q times the row with f from the row with p.
    c. Repeat a and b until there is (at most) one nonzero entry. Switch the first row with that row if necessary.
2. Put the first row in the form $[p \quad 0 \quad \cdots \quad 0]$ by following the steps above but exchanging the words “rows” and “columns”.
3. Repeat 1 and 2 until the first entry g is the only nonzero entry in its row and column. (This process terminates because the least degree decreases at each step.)
4. If g does not divide every entry of A, find the first column with an entry not divisible by g and add it to column 1, and repeat 1-4; the degree of “g” will decrease. Else, go to the next step.
5. Repeat 1-4 with the $(m-1) \times (n-1)$ matrix obtained by removing the first row and

column.

Uniqueness:
Let $\delta_k(M)$ be the gcd of the determinants of all $k \times k$ submatrices of M ($\delta_0(M) = 1$). Equivalent matrices have all these values equal. The polynomials in the normal form are $f_k = \frac{\delta_k(M)}{\delta_{k-1}(M)}$.

Let A be a $n \times n$ matrix, and $p_1, \dots, p_r$ be its invariant factors. The matrix $xI - A$ is equivalent to the $n \times n$ diagonal matrix with diagonal entries $1, \dots, 1, p_1, \dots, p_r$. Use the above algorithm.

Summary



Diagonalization
-Diagonal form has only entries on diagonal
-Condition: All eigenvalues have same algebraic and geometric multiplicity- n linearly independent eigenvalues
-Determined by eigenvalues
-V is the direct sum of eigenspaces E_λ.
-All irreducible factors in minimal polynomial have exponent 1.
-$T=\lambda_1 P_1+\dots+\lambda_k P_k$, where P_i are projections onto eigenspaces.
-$I=P_1+\dots+P_k$

Jordan Canonical Form
-Jordan blocks on diagonal
-Characteristic polynomial splits
-Determined by eigenvalues and nullity $[(T-\lambda I)^r]$
-V is the direct sum of generalized eigenspaces K_λ.
-Exponent of linear term in minimal polynomial is order of largest Jordan block.
-Primary decomposition theorem

Rational Canonical Form
-Companion matrices on diagonal, each polynomial (invariant factor) is multiple of the next.
-No condition
-Determined by prime factorization and nullity$(p(T)^r)$
-Exponent of irreducible factor in minimal polynomial is nullity$(f(T)^a)$/deg(f)
-Cyclic decomposition theorem

8-5 Semi-Simple and Nilpotent Operators

A linear operator N is **nilpotent** if there is a positive integer r such that $N^r = T_0$.
The characteristic and minimal polynomials are in the form x^n.

A linear operator is **semi-simple** if every T-invariant subspace has a complementary T-invariant subspace.

	A linear operator (on finite-dimensional V over F) is semi-simple iff the minimal polynomial has no repeated irreducible factors. If F is algebraically closed, T is semi-simple iff T is diagonalizable. Let F be a subfield of the complex numbers. Every linear operator T can be uniquely decomposed into a semi-simple operator S and a nilpotent operator N such that 1. $T = S + N$ 2. $SN = NS$ N and S are both polynomials in T. Every linear operator whose minimal (or characteristic) polynomial splits can be uniquely decomposed into a diagonalizable operator D and a nilpotent operator N such that 1. $T = D + N$ 2. $DN = ND$ N and D are both polynomials in T. If E_i are the projections in the Primary Decomposition Theorem (Section 8.1) then $D = \sum_{i=1}^{k} \lambda_i E_i\,,\ N = \sum_{i=1}^{k} (T - \lambda_i I) E_i$.

9 Applications of Diagonalization, Sequences

9-1 Powers and Exponentiation

Diagonalization helps compute matrix powers:

$$A^k = (S\Lambda S^{-1})^k = S\Lambda^k S^{-1}$$

To find $A^k x$, write x as a combination of the eigenvectors (Note S is a change of base formula that finds the coordinates $(c_1, \dots, c_n)$)

$$x = \sum_{i=1}^{n} c_i x_i$$

Then

$$A^k x = \sum_{i=1}^{n} c_i \lambda_i^k x_i$$

If diagonalization is not possible, use the Jordan form:

$$A^k = (SJS^{-1})^k = SJ^k S^{-1}$$

Use the following to take powers of a $m \times m$ Jordan block $J = \begin{bmatrix} \lambda & 1 & \cdots & 0 & 0 \\ 0 & \lambda & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \lambda & 1 \\ 0 & 0 & \cdots & 0 & \lambda \end{bmatrix}$:

$$J^r = \begin{bmatrix} \lambda^r & \binom{r}{1}\lambda^{r-1} & \cdots & \binom{r}{m-2}\lambda^{r-(m-2)} & \binom{r}{m-1}\lambda^{r-(m-1)} \\ 0 & \lambda^r & \cdots & & \binom{r}{m-2}\lambda^{r-(m-2)} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \lambda^r & \binom{r}{1}\lambda^{r-1} \\ 0 & 0 & \cdots & 0 & \lambda^r \end{bmatrix}$$

For a matrix in Jordan canonical form, use this formula for each block.

The **spectral radius** is the largest absolute value of the eigenvalues. If it is less than 1, the matrix powers converge to 0, and it determines the rate of convergence.

The matrix exponential is defined as $(A^0 = I)$

$$e^{At} = \sum_{i=0}^{\infty} \frac{(At)^n}{n!}$$

$$e^{At} = Se^{\Lambda t}S^{-1} = S\begin{bmatrix} e^{\lambda_1 t} & & \\ & \ddots & \\ & & e^{\lambda_n t} \end{bmatrix}$$

Thus the eigenvalues of e^{At} are $e^{\lambda t}$.
In general, $e^{At} = a_{n-1}A^{n-1}t^{n-1} + \cdots + a_0 I$ for some constants $a_{n-1}, \dots, a_0$. Letting $r(x) = a_{n-1}x^{n-1} + \cdots + a_0$, we have $e^{\lambda} = \frac{d^i}{d\lambda^i} r(\lambda)$ for $0 \le i < \mu_{alg}(\lambda)$ for every eigenvalue λ. Use the system of n equations to solve for the coefficients.
When A is skew-symmetric, e^{At} is orthogonal.

9-2 Markov Matrices

Let u_k be a column vector where the *i*th entry represents the probability that at the *k*th step the system is at state i. Let A be the transition matrix, that is, A_{ij} contains the probability that a system in state j at any given time will be at state i the next step. Then

$$u_k = A^k u_0$$

where u_0 contains the initial probabilities or proportions.

The **Markov matrix** A satisfies:

1. Every entry is nonnegative.
2. Every column adds to 1.

A contains an eigenvalue of 1, and all other distinct eigenvalues have smaller absolute value.
If all entries of A are positive, then the eigenvalue 1 has only multiplicity 1. The eigenvector corresponding to 1 is the steady state- approached by the probability vectors u_k and describing the probability that a long time late the system will be at each state.

9-3

Recursive Sequences

System of linear recursions:
To find the solution to the recurrence with n variables

$$\begin{cases} x_{1,k+1} = a_{11}x_{1,k} + \cdots + a_{n1}x_{n,k} \\ \vdots \\ x_{n,k+1} = a_{m1}x_{1,k} + \cdots a_{mn}x_{n.k} \end{cases}$$

let $x_k = \begin{bmatrix} x_{1,k} \\ \vdots \\ x_{n,k} \end{bmatrix}$ and use $x_k = A^k x_0$.

Pell's Equation:
If D is a positive integer that is not a perfect square, then all positive solutions to $x^2 - Dy^2 = 1$ are in the form (x, y) with

$$A^k = \begin{bmatrix} x & Dy \\ y & x \end{bmatrix}$$

where $A = \begin{bmatrix} x_1 & Dy_1 \\ y_1 & x_1 \end{bmatrix}$ and (x_1, y_1) is the fundamental solution, that is, the solution where $x_1 > 1$ is minimal.

Homographic recurrence:
A homographic function is in the form $f: \mathbb{C}\backslash\{-\frac{d}{c}\} \to C$ defined by $f(z) = \frac{az+b}{cz+d}$, $c \neq 0$. $A_f = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is the corresponding matrix. Define the sequence $\{x_n\}_{n\geq 0}$ by $x_{n+1} = f(x_n), n \geq 0$. Then $x_n = \frac{a_n x_0 + b_n}{c_n x_0 + d_n}$ where $\left(A_f\right)^n = \begin{bmatrix} a_n & b_n \\ c_n & d_n \end{bmatrix}$.

Linear recursions:
A sequence of complex numbers satisfies a linear recursion of order k if

$$x_n + a_1 x_{n-1} + \cdots + a_k x_{n-k} = 0, n \geq k$$

Solve the characteristic equation $t^k + a_1 t^{k-1} + \cdots + a_k = 0$. If the roots are $t_1, \dots, t_h$ with multiplicities $s_1, \dots s_h$, then $x_n = f_1(n)t_1^n + \cdots + f_h(n)t_h^n$ where f_i is a polynomial of degree at most s_i. Determine the polynomials from solving a system involving the first k terms of the sequence. (Note the general solution is a k-dimensional subset of $\mathbb{C}^\infty$.)

10 Linear Forms

10-1 Multilinear Forms

A function L from $V^n = \underbrace{V \times \cdots \times V}_{n}$, where V is a module over R, to R is

1. **Multilinear** (**n-linear**) if it is linear in each component separately:
   $L(x_1, \dots, cx_i + y_i, \dots, x_n) = cL(x_1, \dots, x_i, \dots, x_n) + L(x_1, \dots, y_i, \dots, x_n)$
2. **Alternating** if $L(x_1, \dots, x_n) = 0$ whenever $x_i = x_j$ with $i \neq j$.

The collection of all multilinear functions on V^n is denoted by $M^n(V)$, and the collection of all alternating multilinear functions is $\Lambda^n(V)$.

If L and M are multilinear functions on V^r, V^s, respectively, the **tensor product** of L and M is the function on V^{r+s} defined by

$$(L \otimes M)(x, y) = L(x)M(y)$$

where $x \in V^r, y \in V^s$. The tensor product is linear in each component and is associative.

For a permutation σ define $L_\sigma(x_1, \dots, x_r) = L(x_{\sigma(1)}, \dots, x_{\sigma(n)})$ and the linear transformation $\pi_r : M^r(V) \to \Lambda^r(V)$ by

$$\pi_r L = \sum_\sigma (\operatorname{sgn}(\sigma)\, L_\sigma)$$

If V is a free module of rank n, $M^r(V)$ is a free R-module of rank n^r, with basis $f_{j_1} \otimes \cdots \otimes f_{j_r}$ $(1 \le j_1, \dots, j_r \le n)$ where $\{f_1, \dots, f_n\}$ is a basis for V^*.
When $V = R^n$, and L is a r-linear form in $M^r(V)$,

$$L(x_1, \dots, x_r) = \sum_{1 \le j_1, \dots j_r \le n} A(1, j_1) \cdots A(r, j_r) L(e_{j_1}, \dots, e_{j_r})$$

where A is the rxn matrix with rows $x_1, \dots, x_r$.
$\Lambda^r(V)$ is a free R-module of rank $\binom{n}{r}$, with basis the same as before, but $j_1, \dots, j_r$ are combinations of $\{1, \dots, n\}$ $(1 \le j_1 < \cdots < j_r \le n)$.

Where the Determinant fits in:

1. $D = \sum_\sigma (\operatorname{sgn}(\sigma)\, f_{\sigma(1)} \otimes \cdots \otimes f_{\sigma(n)})$, the f_i standard coordinate functions.
2. If T is a linear operator on $V = R^n$ and $L \in \Lambda^n(V)$,
   $$L\big(T(x_1), \dots, T(x_n)\big) = \det(T)\, L(x_1, \dots, x_n)$$
   The determinant of T is the same as the determinant of any matrix representation of T.
3. The special alternating form $D_J = \pi_r(f_{j_1} \otimes \cdots \otimes f_{j_r})$ $(J = \{j_1, \dots, j_r\})$ is the determinant of the rxr matrix A defined by $A_{ik} = f_{j_k}(x_i)$, also written as $\frac{\partial(x_1, \dots x_r)}{\partial(y_{j_1}, \dots, y_{j_r})}$, where $\{f_1, \dots, f_n\}$ is the standard dual basis.

10-2 Exterior Products

Let G be the group of all permutations which permute $\{1, \dots, r\}$ and $\{r+1, \dots, s\}$ within themselves. For alternating r and s-linear forms L and M, define $\psi : \mathfrak{S}_{r+s} \to M^{r+s}(V)$ by $\psi(\sigma) = (\operatorname{sgn}(\sigma))(L \otimes M)_\sigma$. For a coset aG, define $\tilde{\psi}(aG) = \psi(a)$. The **exterior product** of L and M is

$$L \wedge M = \sum_{H \in \mathfrak{S}_{r+s}/G} \tilde{\psi}(H)$$

Then

1. $r!\,s!\,L \wedge M = \pi_{r+s}(L \otimes M)$; in particular $L \wedge M = \frac{1}{r!s!}\pi_{r+s}(L \otimes M)$ if R is a field of characteristic 0.
2. $(L \wedge M) \wedge N = L \wedge (M \wedge N)$
3. $L \wedge M = (-1)^{rs} M \wedge L$

Laplace Expansions:

Define $L(x_1, \dots x_r) = \det\left(\begin{bmatrix} A_{11} & \cdots & A_{1r} \\ \vdots & \ddots & \vdots \\ A_{r1} & \cdots & A_{rr} \end{bmatrix}\right)$ and $M(x_1, \dots x_s) = \det\left(\begin{bmatrix} A_{1,r+1} & \cdots & A_{1n} \\ \vdots & \ddots & \vdots \\ A_{s,r+1} & \cdots & A_{sn} \end{bmatrix}\right)$ where $x_i = \langle A_{i1}, \dots, A_{in} \rangle \in R^n$ and $s = n - r$. Then $L \wedge M = \det(A)$, giving

$$\det(A) = \sum_{j_1 < \cdots < j_r, k_1 < \cdots < k_s} (-1)^{j_1 + \cdots + j_r + \frac{r(r-1)}{2}} \det\left(\begin{bmatrix} A(j_1, 1) & \cdots & A(j_1, r) \\ \vdots & \ddots & \vdots \\ A(j_r, 1) & \cdots & A(j_r, r) \end{bmatrix}\right)$$
$$\det\left(\begin{bmatrix} A(k_1, r+1) & \cdots & A(k_1, n) \\ \vdots & \ddots & \vdots \\ A(k_s, r+1) & \cdots & A(k_s, n) \end{bmatrix}\right)$$

For a free R-module V of rank n, the **Grassman ring** over V^* is defined by

$$\Lambda(V) = \Lambda^0(V) \oplus \cdots \oplus \Lambda^n(V)$$

and has dimension 2^n. (The direct sum is treated like a Cartesian product.)

10-3 Bilinear Forms

A function $H: V \times V \to F$ is a **bilinear form** on V if H is linear in each variable when the other is held fixed:

1. $H(ax_1 + x_2, y) = aH(x_1, y) + H(x_2, y)$
2. $H(x, ay_1 + y_2) = aH(x, y_1) + H(x, y_2)$

The bilinear form is **symmetric** (a scalar product) if $H(x, y) = H(y, x)$ for all $x, y \in V$ and **skew-symmetric** if $H(x, y) = -H(y, x)$.
The set of all bilinear forms on V, denoted by $\mathcal{B}(V)$, is a vector space. An real inner product space is a symmetric bilinear form.

A function $K: V \to F$ is a **quadratic form** if there exists a symmetric bilinear form H such that $K(x) \equiv H(x, x)$. If F is not of characteristic 2,

$$H(x, y) = \frac{K(x + y) - K(x) - K(y)}{2}$$

Let $\beta = \{v_1, \dots, v_n\}$ be an ordered basis for V. The matrix $A = \psi_\beta(H)$ with $A_{ij} = H(v_i, v_j)$ is the matrix representation of H with respect to β.

1. ψ_β is an isomorphism.
2. Thus $\mathcal{B}(V)$ has dimension n^2.
3. If $\beta^* = \{L_1, \dots, L_n\}$ is a basis for V^* then $f_{ij}(x, y) = L_i(x) L_j(y)$ is a basis for $\mathcal{B}(V)$.
4. ψ_β is (skew-)symmetric iff H is.
5. A is the unique matrix satisfying $H(x, y) \equiv [x]_\beta^T A [y]_\beta$.

Square matrix B is **congruent** to A if there exists an invertible matrix Q such that $B = Q^T A Q$.

Congruence is a equivalence relation. For 2 bases β, γ, $\psi_\beta(H)$ and $\psi_\gamma(H)$ are congruent; conversely, congruent matrices are 2 representations of the same bilinear form.

Define $L_x(y) = \left(L_H(x)\right)(y) = H(x, y)$ and $R_y(x) = \left(R_H(y)\right)(x) = H(x, y)$.The **rank** of H is $\text{rank}(L_H) = \text{rank}(R_H)$. For n-dimensional V, the following are equivalent:

1. rank(H)=n
2. For $x \neq 0$, there exists y such that $H(x, y) \neq 0$.
3. For $y \neq 0$, there exists y such that $H(x, y) \neq 0$.

Any H satisfying 2 and 3 is nondegenerate. The **radical** of H, Rad(H), is the kernel of L_H or R_H, in other words, it is orthogonal to all other vectors.

10-4

Theorems on Bilinear Forms and Diagonalization

A bilinear form H on finite-dimensional V is **diagonalizable** if there is a basis β such that $\psi_\beta(H)$ is diagonal.
If F does not have characteristic 2, then a bilinear form is symmetric iff it is diagonalizable. If V is a real inner product space, the basis can be chosen to be orthonormal.

$$\psi_\beta(H) = A = Q^T D Q$$

where Q is the change-of-coordinate matrix changing standard β-coordinates into γ-coordinates and $\psi_\gamma(H) = D$. Diagonalize the same way as before, choosing Q to be orthonormal so $Q^T = Q^{-1}$.

A vector v is **isotropic** if $H(v, v) = 0$ (orthogonal to itself). A subspace W is isotropic if the restriction of H to W is 0. A subspace is maximally isotropic if it has greatest dimension among all isotropic subspaces. Orthogonality, projections, and adjoints for scalar products is defined the same way as orthogonality for inner products: v and w are orthogonal if $H(v, w) = 0$, and $W^\perp = \{v | H(v, w) = 0 \ \forall \ w \in W\}$.

1. If $V = \text{Rad}(H) \oplus W$ then the restriction of H to W, H_W, is nondegenerate.
2. If H is nondegenerate on V and subspace $W \subseteq V$, $W \oplus W^\perp = V$.
3. If H is nondegenerate, there exists an orthonormal basis for V.

Sylvester's Law of Inertia:
Let H be a symmetric form on finite-dimensional real V. Then the number of positive diagonal entries (the **index** p of H) and negative diagonal entries in any diagonal representation of H is the same. The **signature** is the number of positive entries and the number of negative entries. The rank, index, and signature are all invariants of the bilinear form.

1. Two real symmetric nxn matrices are congruent iff they have the same invariants.
2. A symmetric nxn matrix is congruent to
   $$J_{pr} = \begin{bmatrix} I_p & \mathcal{O} & \mathcal{O} \\ \mathcal{O} & -I_{r-p} & \mathcal{O} \\ \mathcal{O} & \mathcal{O} & \mathcal{O} \end{bmatrix}$$
3. For nondegenerate H:
   1. The maximal subspace W such that H_W is positive/ negative definite is p/ n-p.
   2. The maximal isotropic subspace W has dimension $\min\{p, n-p\}$

If f^* is the adjoint of linear transformation f, and $f^\vee$ is the dual (transpose), then $R_H f^* = f^V R_H$.

Let H be a skew-symmetric form on n-dimensional V over a subfield of $\mathbb{C}$. Then r=rank(H) is even and there exists β such that $\psi_\beta(H)$ is the direct sum of the $(n-r)\times(n-r)$ zero matrix and $\frac{r}{2}$ copies of $\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$.

10-5

Sesqui-linear Forms

A **sesqui-linear form** f on $\mathbb{R}$ or $\mathbb{C}$ is

Linear in the first component	$f(cx+y,z)=cf(x,z)+f(y,z)$
Conjugate-linear in the second component	$f(x,cy+z)=\bar{c}f(x,y)+f(x,z)$

The form is **Hermitian** if $f(x,y)=\overline{f(y,x)}$. A sesqui-linear form f is Hermitian if $f(x,x)$ is real for all x. [Note: Some books reverse x and y for sesqui-linear forms and inner products.]

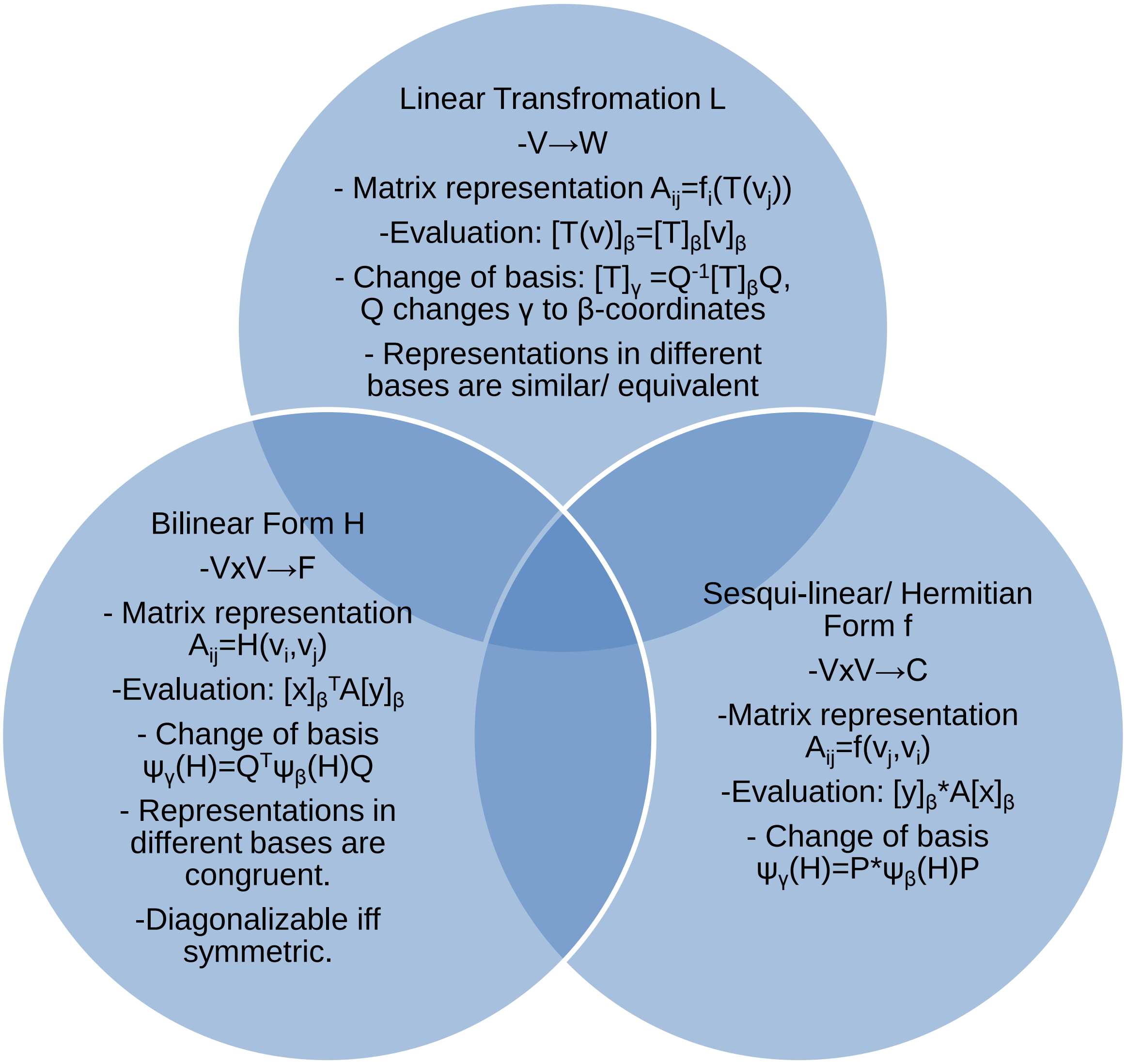
The matrix representation A of f in basis $\{v_1,\dots,v_n\}$ is given by $A_{ij} = f(x_j,x_i)$. (Note the reversal.) Then $H(x,y)\equiv[y]_\beta^* A[y]_\beta[x]_\beta$.

If V is a finite-dimensional inner product space, there exists a unique linear operator T$_f$ on V such that $f(x,y)=\langle T_f(x),y\rangle$. This map $f\to T_f$ is an isomorphism from the vector space of sesqui-linear forms onto $\mathcal{L}(V,V)$. f is self-adjoint iff T_f is self-adjoint.

f on $\mathbb{R}$ or $\mathbb{C}$ is **positive**/ nonnegative if it is Hermitian and $f(x,x)>0$ for $x\neq 0$/ $f(x,x)\geq 0$. ***A positive form is simply an inner product.*** f is positive if its matrix representation is positive definite.

Principal Axis Theorem: (from the Spectral Theorem)
For every Hermitian form f on finite-dimensional V, there exists an orthonormal basis in which f has a real diagonal matrix representation.

Summary



10-6

Application of Bilinear and Quadratic Forms: Conics, Quadrics and Extrema

An equation in 2/ 3 variables of degree 2 determines a conic/ quadric.

1. Group all the terms of degree 2 on one side, and represent them in the form $[x_1 \quad \cdots \quad x_n]A\begin{bmatrix}x_1\\ \vdots\\ x_n\end{bmatrix}$ where n=2/ 3 and A is a symmetric $n \times n$ matrix. If the coefficient of x_i^2 is c_{ii} then $A_{ii} = c_{ii}$. If the coefficient of $x_i x_j, i < j$ is c_{ij} then $A_{ij} = A_{ji} = \frac{c_{ij}}{2}$. Diagonalize $A = Q^T DQ$ and write the terms as $([x_1 \quad \cdots \quad x_n]Q^T)D\left(Q\begin{bmatrix}x_1\\ \vdots\\ x_n\end{bmatrix}\right)$. The axes the conic/ quadric are oriented along are given by the eigenvectors.
2. Write the linear terms with respect to the new coordinates, and complete the square in each variable.

Name of Quadric	Equation
Ellipsoid	$a_{11}x_1^2 + a_{22}x^2 + a_{33}x_3^2 = 1$
1-sheeted hyperboloid	$a_{11}x_1^2 + a_{22}x^2 - a_{33}x^2 = 1$
2-sheeted hyperboloid	$a_{11}x_1^2 - a_{22}x^2 - a_{33}x^2 = 1$
Elliptic paraboloid	$a_{11}x_1^2 + a_{22}x^2 = x_3$

Hyperbolic paraboloid	$a_{11}x_1^2 + a_{22}x^2 = x_3$
Elliptic cone	$a_{11}x_1^2 + a_{22}x^2 - a_{33}x_3^2 = 0$

The **Hessian** matrix $A(p)$ of $f(p)$ is defined by

$$A_{ij} = \frac{\partial^2 f(p)}{(\partial t_i)(\partial t_j)}$$

<u>Second Derivative Test</u>:

Let $f(t_1, \dots, t_n)$ be a real-valued function for which all third-order partial derivatives exist and are continuous. Let $p = (p_1, \dots, p_n)$ be a critical point (i.e. $\frac{\partial f}{\partial t_i} = 0$ for all i).

(a) If all eigenvalues of $A(p)$ are positive, f has a local minimum at p.
(b) If all eigenvalues are negative, f has a local maximum at p.
(c) If $A(p)$ has at least one positive and one negative eigenvalue, p is a saddle point.
(d) If $\operatorname{rank}(A(p)) < n$ (an eigenvalue is 0) and $A(p)$ does not have both positive and negative eigenvalues, the test fails.

11 Numerical Linear Algebra

11-1 Elimination and Factorization in Practice

Partial pivoting- For the kth pivot, choose the largest number in row k or below in that column. Exchange that row with row k. Small pivots create large roundoff error because they must be multiplied by large numbers.

A **band matrix** A with half-bandwidth w has $A_{ij} = 0$ when $|i - j| > w$.

Operation counts (A is $k \times k$ and invertible) (Multiply-subtract counted as one operation)

Process	Count (≲)	Reason
Forward elimination (A→U), A=LU factorization	$\frac{1}{3}n^3$	$\sum k^2 - k$. When there are k rows left, for all k-1 rows below, multiply-subtract k times.
Forward elimination on band matrix with half-bandwidth w	$\frac{1}{3}w^2(3n - 2w \approx w2n$ when w small	$\approx \sum w^2 - w$. There are no more than w-1 nonzeros below any pivot.
Forward elimination, right side (b)	$\frac{1}{2}n^2$	$\sum k$. When there are k rows left, multiply-subtract for all entries below the current one.
Back-substitution	$\frac{1}{2}n^2$	$\sum k$. For row k, divide by pivot and substitute into previous k-1 rows.
Factorization into QR (Gram-Schmidt)	$\frac{2}{3}n^3$	$\sum 2k^2$. When there are k columns left, divide the kth vector by its norm, find the projection of all remaining columns onto it ($\approx k^2$) then subtract ($\approx k^2$).
A^{-1} (Gauss-Jordan elimination)	n^3	$\frac{1}{3}n^3$ for A=LU, $\sum \frac{1}{2}(n-k)^2 \approx \frac{1}{6}n^3$ for right side- no work is required on the kth column on the right side until row k, $n(\frac{1}{2}n^2)$ back substitution

Note: For parallel computing, working with matrices (more concise) may be more efficient.

11-2 Norms and Condition Numbers

The **norm** of a matrix is the maximum magnification of a vector x by A:

$$\|A\| = \max_{x \neq 0} \frac{\|Ax\|}{\|x\|}$$

For a symmetric matrix, $\|A\|$ is the absolute value of the eigenvalue with largest absolute value.

Finding the norm:

$$\|A\|^2 = \max_{x \neq 0} \frac{\|Ax\|^2}{\|x\|^2} = \max_{x \neq 0} \frac{x^T A^T A x}{x^T x} = \text{Largest eigenvalue of } A^T A$$

$$\|A\| = \text{Largest singular value of } A$$

The **condition number** of A is

$$c = \text{cond}(A) = \|A\| \|A^{-1}\|$$

When A is symmetric, $c = \frac{|\lambda\ |_{\max}}{|\lambda|_{\min}}$. Anyway, $c = \sqrt{\frac{\text{Largest eigenvalue of} A^T A}{\text{Smallest eigenvalue of} A^T A}}$.

The condition number shows the sensitivity of a system $Ax = b$ to error. Problem error is inaccuracy in A or b due to measurement/ roundoff. Let Δx be the solution error and $\Delta A, \Delta b$ be the problem errors.

1. When the problem error is in b,
$$\frac{1}{c}\frac{||\Delta b||}{||b||} \le \frac{||\Delta x||}{||x||} \le c\frac{||\Delta b||}{||b||}$$
2. When the problem error is in A,
$$\frac{||\Delta x||}{||x + \Delta x||} \le c\frac{||\Delta A||}{||A||}$$

11-3 Iterative Methods

For systems:

General approach:

1. Split A into S-T. $Ax = b \Rightarrow Sx = Tx + b$
2. Compute the sequence $Sx_{k+1} = Tx_k + b$

Requirements:

1. (2) should be easy to solve for x_{k+1}, so the preconditioner S should be diagonal or triangular.
2. The error should converge to 0 quickly:
$$e_{k+1} = S^{-1}Te_k, e_k = x - x_k$$
Thus the largest eigenvalue of $S^{-1}T$ should have absolute value less than 1.

Useful for large sparse matrices, with a wide band.

Method	S	Remarks
Jacobi's method	Diagonal part of A	
Gauss-Siedel method	Lower triangular part of A	About twice as fast: Often $|\lambda|_{\max}$ is the square of the $|\lambda|_{\max}$ for Jacobi.
Successive overrelaxation	S has diagonal of original A, but below, entries are those of ωA.	Combination of Jacobi and Gauss-Siedel. Choose ω to minimize spectral radius.
Incomplete LU method	Approximate L times approximate U	Set small nonzero in L, U to 0.

Conjugate Gradients for positive definite A:

Set $x_0 = 0 \text{ (or approximate solution)}, r_0 = b, p_0 = r_0$.

	Formula	Description
1.	$\alpha_n = \frac{r_{n-1}^T r_{n-1}}{p_{n-1}^T A p_{n-1}}$	Step length x_{n-1} to x_n
2.	$x_n = x_{n-1} + \alpha_n p_{n-1}$	Approximate solution
3.	$r_n = r_{n-1} - \alpha_n A p_{n-1}$	New residual $b - Ax_n$
4.	$\beta_n = \frac{r_n^T r_n}{r_{n-1}^T r_{n-1}}$	Improvement
5.	$p_n = r_n + \beta_n p_{n-1}$	Next search direction

Computing eigenvalues

1. (Inverse) power methods: Keep multiplying a vector u by A. Typically, u approaches the direction of the eigenvector corresponding to the largest eigenvalue. Convergence is quicker when $\left|\frac{\lambda_2}{\lambda_1}\right|$ is small, where λ_1, λ_2 are eigenvalues with largest, second largest absolute values. For the smallest eigenvalue, apply the method with A^{-1} (but solve $Au_{k+1} = u_k$ rather than compute the inverse).
2. QR Method: Factor $A = QR$, reverse R and Q (eigenvalues don’t change), multiply them to get A', and repeat. Diagonal entries approach the eigenvalues. When the last diagonal entry is accurate, remove the last row and column and continue. Modifications:
   a. Factor $A_k - c_k I$ into $Q_k R_k$. $A_{k+1} = R_k Q_k + c_k I$. Choose c near an unknown eigenvalue.
   b. (Hessenberg) Obtain off-diagonal entries first by changing A to a similar matrix. Zeros in lower-left corner stay.

12 Applications

12-1 Fourier Series (Analysis)

Use the orthonormal system $\frac{1}{\sqrt{2\pi}}, \frac{\cos x}{\sqrt{\pi}}, \frac{\sin x}{\sqrt{\pi}}, \frac{\cos 2x}{\sqrt{\pi}}, \cdots$ to express a function in $[0,2\pi]$ as a Fourier series:

$$f(x) = a_0 + a_1 \cos x + b_1 \sin x + a_2 \cos 2x + b_2 \sin 2x + \cdots$$

Use projections (Section 5.3) to find the coefficients. (Multiply by the function you're trying to find the coefficient for, and integrate from 0 to 2π; orthogonality makes all but one term 0.) The orthonormal system is **closed**, meaning that f is actually equal to the Fourier series. Fourier coefficients offer a way to show the isomorphism between **Hilbert spaces** (complete, separable, infinite-dimensional Euclidean spaces). See Analysis notes for details and derivation, Differential Equations for formulas.

The exponential Fourier series uses the orthonormal system $f_n(t) = e^{int}, n \in \mathbb{Z}$ instead. This applies to functions in $[-\infty, \infty]$.

12-2 Fast Fourier Transform

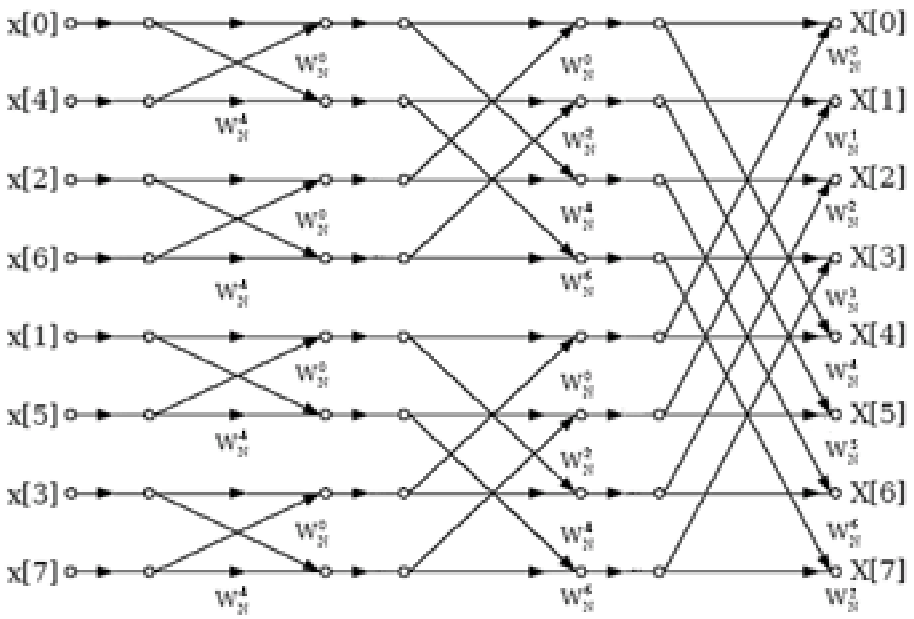
Let $\omega = e^{\frac{2\pi i}{n}}$. The Fast Fourier Transform takes as input the coefficients c_j of $\omega^j, 0 \le j < n$ and outputs the value of the function $f(x) = \sum_{j=0}^{n-1} c_j \omega^j$ at $k, 0 \le k < n$. The matrix for F satisfies $F_{jk} = \omega^{jk}$ when the rows and columns are indexed from 0. Then

$$F_n c = y, c = \begin{bmatrix} c_0 \\ \vdots \\ c_{n-1} \end{bmatrix}, y = \begin{bmatrix} y_0 \\ \vdots \\ y_{n-1} \end{bmatrix} = \begin{bmatrix} f(0) \\ \vdots \\ f(n-1) \end{bmatrix}$$

The inverse of F is $\frac{1}{n} F^* = \frac{1}{n} \bar{F}$. The inverse Fourier transform gives the coefficients from the functional values. To calculate a Fourier transform quickly when $n = 2^l$, break

$$F_n = \begin{bmatrix} I_{\frac{n}{2}} & D_{\frac{n}{2}} \\ I_{\frac{n}{2}} & -D_{\frac{n}{2}} \end{bmatrix} \begin{bmatrix} F_{\frac{n}{2}} & \\ & F_{\frac{n}{2}} \end{bmatrix} [\text{even-odd permutation}]$$

$D_{n/2}$ is the diagonal matrix with (n/2)th roots of unity. The last matrix has n/2 columns with 1's in even locations (in increasing order starting from 0) and the next n/2 rows in odd locations. Then break up the middle matrix using the same idea, but now there's two copies. Repeating to F_2, the operation count is $\frac{1}{2} nl = \frac{1}{2} n \ln(n)$. The net effect of the permutation matrices is that the numbers are ordered based on the number formed from their digits reversed.



http://cnx.org/content/m12107/latest/

Set $m = \frac{1}{2}n$. The first and last m components of $y = F_n c$ are combinations of the half-size transforms $y^{'} = F_m c'$ and $y^{''} = F_m c''$, i.e. for $0 \le j < m$,

$$\begin{cases} y_j = y_j^{'} + \omega_n^j y_j{}'' \\ y_{j+m} = y_j^{'} - \omega_n^j y_j{}'' \end{cases}$$

12-3 Differential Equations

The set of solutions to a homogeneous linear differential equation with constant coefficients

$$\sum_{i=0}^{n} a_i y^{(i)} = 0$$

is a n-dimensional subspace of C^∞. The functions $t^j e^{\lambda t}$ (λ a root of the auxiliary polynomial $\sum_{i=0}^{n} a_i x^i = 0,\ 0 \le j < m,$ where m is the multiplicity of the root) are linearly independent and satisfy the equation. Hence they form a basis for a solution space.

The general solution to the system of n linear differential equations $x^{'} = Ax$ is any sum of solutions of the form

$$e^{\lambda t}\left[f(t)(A-\lambda I)^{p-1} + f^{'}(t)(A-\lambda I)^{p-2} + \cdots + f^{(p-1)}(t)\right]x$$

where the x are the end vectors of distinct cycles that make up a Jordan canonical basis for A, λ is the eigenvalue corresponding to x, p is the order of the Jordan block, and $f(t)$ is a polynomial of degree less than p.

12-4 Combinatorics and Graph Theory

Graphs and applications to electric circuits
The **incidence matrix** A of a directed graph has a row for every edge and a column for every node. If edge i points away from/ toward node j, then $A_{ij} = -1/\ 1$, respectively. Suppose the graph is connected, and has n nodes and m edges. Each node is labeled with a number (voltage), and multiplying by A gives the vector of edge labels showing the difference between

the nodes they connect (potential differences/ flow).

1. The row space has dimension n-1. Take any n-1 rows corresponding to a spanning tree of the graph to get a basis for the row space. Rows are dependent when edges form a loop.
2. The column space has dimension n-1. The vectors in the column space are exactly the labeling of edges such that the numbers add to zero around every loop (when moving in the reverse direction as the edges, multiply by -1). This corresponds to all attainable sets of potential differences (Voltage law).
3. The nullspace has dimension 1 and contains multiples of $(1,\ldots,1)^T$. Potential differences are 0.
4. The left nullspace has dimension m-n+1. There are m-n+1 independent loops in the graph. The vectors in the left nullspace are those where the flow in equals the flow out at each node (Current law). To find a basis, find m-n+1 independent loops; for each loop choose a direction, and label the edge 1 if it goes around the loop in that direction and -1 otherwise.

Let C be the diagonal matrix assigning a conductance (inverse of resistance) to each edge. Ohm's law says $y = -CAx$. The voltages at the nodes satisfy

$$A^T CAx = f$$

where f tells the source from outside (ex. battery).

Another useful incidence matrix is where A has a row and column for each vertex, and $A_{ij} = 1$ if vertices i and j are connected by an edge, and 0 otherwise. (For directed graphs, use -1/ 1.)

Sets

The **incident matrix** A for a family of subsets $\{S_1, \ldots, S_n\}$ containing elements $\{x_1, \ldots, x_m\}$ has $A_{ij} = \begin{cases} 1 \text{ if } x_i \in S_j \\ 0 \text{ if } x_i \notin S_j \end{cases}$. Exploring AA^T and using properties of ranks, determinants, linear dependency, etc. may give conclusions about the sets. Working in the field $\mathbb{Z}_2$ on problems dealing with parity may help.

12-5

Engineering

<u>Discrete case:</u> Springs

$$K = A^T CA, Ku = f$$

Vector/ Equation	Description	Matrix
u	Movements of the n masses	
$e = Au$	Kinematic equation: Elongations of the m springs	A gives the elongations of the springs.
$y = Ce$	Constitutive law: Tensions (internal forces) in the m springs	C is a diagonal matrix that applies Hooke's Law for each spring, giving the forces.
$f = A^T y$	Static/ balance equation: External forces on n masses	Internal forces balance external forces on masses.

There are four possibilities for A:

Case	Description	Matrix A	Equations
Fixed-fixed	There are n+1 springs; each mass has 2 springs coming out of it and the top and bottom are fixed in place.	$\begin{bmatrix} 1 & & \\ -1 & \ddots & \\ & & 1 \\ & & -1 \end{bmatrix}$	$e_1 = u_1$ $e_2 = u_2 - u_1$ $\vdots$ $e_{n+1} = u_n$

Fixed-free	There are n springs; one end is fixed and the other is not. (Here we assume the top end is fixed.)	$\begin{bmatrix} 1 & & & \\ -1 & \ddots & & \\ & & 1 & \\ & & -1 & 1 \end{bmatrix}$	$e_1 = u_1$ $e_2 = u_2 - u_1$ $\vdots$ $e_n = u_n - u_{n-1}$
Free-free	No springs at either end. n-1 springs.	$\begin{bmatrix} -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{bmatrix}$	$e_1 = u_2 - u_1$ $\vdots$ $e_{n-1} = u_n - u_{n-1}$
Circular	The nth spring is connected to the first one. n springs.	$\begin{bmatrix} 1 & & & -1 \\ -1 & \ddots & & \\ & & 1 & \\ & & -1 & 1 \end{bmatrix}$	$e_1 = u_1 - u_n$ $e_2 = u_2 - u_1$ $\vdots$ $e_n = u_n - u_{n-1}$

Each spring is stretched or compressed by the difference in displacements.

Facts about K:

1. K is tridiagonal except for the circular case: only nonzero entries are on diagonal or one entry above or below.
2. K is symmetric.
3. K is positive definite for the fixed-fixed and fixed-free case.
4. K^{-1} has all positive entries for the fixed-fixed and fixed-free case.

$u = K^{-1}f$ in the fixed-fixed and fixed-free case give the movements from the forces.

For the singular case:

1. The nullspace of K is $\begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$, if the whole system moves by the same amount the forces stay the same.
2. To solve $Ku = f$, the forces must add up to 0 (equilibrium).

<u>Continuous case:</u> Elastic bar

$A^TCAu = f$ becomes the differential equation

$$-\frac{d}{dx}\left(c(x)\frac{du}{dx}\right) = f(x)$$

The discrete case can be used to approximate the continuous case. When going from the continuous to discrete case, multiply by Δx.

12-6

Physics: Special Theory of Relativity

For each event p occurring at $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$ at time t read on clock C relative to S, assign the space-time coordinates relative to C and S $\begin{bmatrix} x \\ y \\ z \\ t \end{bmatrix}$. Suppose S and S' have parallel axes and S' moves at constant velocity v relative to S in the +x direction, and they coincide when their clocks C and C' read 0. The unit of length is the light second. Define $T_v \begin{bmatrix} x \\ y \\ z \\ t \end{bmatrix} = \begin{bmatrix} x' \\ y' \\ z' \\ t' \end{bmatrix}$, where the two sets of coordinates represent the same event with respect to S and S'

Axioms:

1. The speed of light is 1 when measured in either coordinate system.
2. T$_v$ is an isomorphism.
3. $T_v \begin{bmatrix} x \\ y \\ z \\ t \end{bmatrix} = \begin{bmatrix} x' \\ y' \\ z' \\ t' \end{bmatrix}$ implies $y = y', z = z'$.
4. $T_v \begin{bmatrix} x \\ y_1 \\ z_1 \\ t \end{bmatrix} = \begin{bmatrix} x' \\ y' \\ z' \\ t' \end{bmatrix}, T_v \begin{bmatrix} x \\ y_2 \\ z_2 \\ t \end{bmatrix} = \begin{bmatrix} x'' \\ y'' \\ z'' \\ t'' \end{bmatrix}$ implies $x'' = x', t'' = t'$.
5. The origin of S moves in the negative x'-axis of S' at velocity –v as measured from S'.

These axioms complete characterize the **Lorentz transformation** T$_v$, whose representation in the standard bases is

$$[T_v]_\beta = \begin{bmatrix} \frac{1}{\sqrt{1-v^2}} & 0 & 0 & \frac{-v}{\sqrt{1-v^2}} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\frac{v}{\sqrt{1-v^2}} & 0 & 0 & \frac{1}{\sqrt{1-v^2}} \end{bmatrix}$$

1. If a light flash at time 0 at the origin is observed at $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$ is observed at time t, then $x^2 + y^2 + z^2 - t^2 = 0$.
2. Time contraction: $t' = t\sqrt{1-v^2}$
3. Length contraction: $x' = x\sqrt{1-v^2}$

12-7

Computer Graphics

3-D computer graphics use homogeneous coordinates: $\begin{bmatrix} x \\ y \\ z \\ c \end{bmatrix}$ represents the point $(\frac{x}{c}, \frac{y}{c}, \frac{z}{c})$ (the point at infinity if c=0).

The transformation…	is like multiplying (on the left side) by…
Translation by (x_0, y_0, z_0)	$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ x_0 & y_0 & z_0 & 1 \end{bmatrix}$
Scaling by a, b, c in x, y, and z directions	$\begin{bmatrix} a & 0 & 0 & 0 \\ 0 & b & 0 & 0 \\ 0 & 0 & c & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$
Rotation around z-axis (similar for others) by θ	$\begin{bmatrix} \cos\theta & -\sin\theta & 0 & 0 \\ \sin\theta & \cos\theta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$
Projection onto plane through (0,0,0) perpendicular to unit vector n	$P = \begin{bmatrix} I - nn^T & 0 \\ 0 & 1 \end{bmatrix}$
Projection onto plane passing through Q, perpendicular to unit vector n	$T_- P T_+$ where T is the translation taking Q to the origin, and P is as above

Reflection through plane through (0,0,0) perpendicular to unit vector n	$\begin{bmatrix} I-2nn^T & 0 \\ 0 & 1 \end{bmatrix}$

The matrix representation for an affine transformation is

$$\begin{bmatrix} T(1,0,0)-T(0,0,0) & 0 \\ T(0,1,0)-T(0,0,0) & 0 \\ T(0,0,1)-T(0,0,0) & 0 \\ T(0,0,0) & 1 \end{bmatrix}$$

12-8 Linear Programming

Linear programming searches for a nonnegative vector x satisfying $Ax = b$ that minimizes (or maximizes) the cost $c \cdot x$. The dual problem is to maximize $b \cdot y$ subject to $A^T y \leq c$. The extremum must occur at a corner. A corner is a vector$x$ with positive entries that satisfies the m equations $Ax = b$ with at most m positive components.

Duality Theorem:

If either problem has a best solution then so does the other. Then the minimum cost $c \ \cdot x^*$ equals the maximum income $b \cdot y^*$.

Simplex Method:

1. First find a corner. If one can't easily be found, create m new variables, start with their sum as the cost, and follow the remaining steps until they are all zero, then revert to the original problem.
2. Move to another corner that lowers the cost. Repeat for each zero component: Change it from 0 to 1, find how the nonzero components would adjust to satisfy $Ax = b$, then compute the change in the total cost $c \cdot x$. Let the entering variable be the one that causes the most negative change (per single unit). Reduce the entering variable until the first positive component hits 0.
3. When every other "adjacent" corner has higher cost, the current corner is the optimal x.

12-9 Economics

A **consumption matrix** A has the amount of product j needed to produce product i in entry (i,j). Then $u = Av$ where v/ u are the input/ output column vectors containing the amount of product i in entry i.

If the column vector y contains the demands for each product, then for the economy to meet the demands, there must exist a vector p with nonnegative entries satisfying

$$\underbrace{p}_{\text{input}} - \underbrace{Ap}_{\text{consumption}} = \underbrace{y}_{\text{demand}} \Leftrightarrow (I-A)p = y \Rightarrow p = (I-A)^{-1}y$$

if the inverse exists.

If the largest eigenvalue…	then $(I-A)^{-1}$…
is greater than 1	has negative entries
is equal to 1	fails to exist
is less than 1	has only nonnegative entries

If the spectral radius of A is less than 1, then the following expansion is valid:

$$(I-A)^{-1} = I + A + A^2 + A^3 + \cdots$$

References

Introduction to Linear Algebra (Third Edition) by Gilbert Strang
Linear Algebra (Fourth Edition) by Friedberg, Insel, and Spence
Linear Algebra (Second Edition) by Kenneth Hoffman and Ray Kunze
Putnam and Beyond by Titu Andreescu and Razvan Gelca
MIT OpenCourseWare, 18.06 and 18.700

Notes

I tried to make the notes as complete yet concise and understandable as possible by combining information from 3 books on linear algebra, as well as put in a few problem-solving tips. Strang's book offers a very intuitive view of many linear algebra concepts; for example the diagram on "Orthogonality of the Four Subspaces" is copied from the book. The other two books offer a more rigorous and theoretical development; in particular, Hoffman and Kunze's book is quite complete.

I prefer to focus on vector spaces and linear transformations as the building blocks of linear algebra, but one can start with matrices as well. These offer two different viewpoints which I try to convey: Rank, canonical forms, etc. can be described in terms of both. Big ideas are ***emphasized*** and I try to summarize the major proofs as I understand them, as well as provide nice summary diagrams.

A first (nontheoretical) course on linear algebra may only include about half of the material in the notes. Often in a section I put the theoretical and intuitive results side by side; just use the version you prefer. I organized it roughly so later chapters depend on earlier ones, but there are exceptions. The last section is applications and a miscellany of stuff that doesn't fit well in the other sections. Basic knowledge of fields and rings is required.

Since this was made in Word, some of the math formatting is not perfect. Oh well.

Feel free to share this; I hope you find it useful!

Please report all errors and suggestions by posting on my blog or emailing me at holdenlee1@yahoo.com. (I'm only a student learning this stuff myself so you can expect errors.) Thanks!

Things to add: Continuity arguments, linear algebra in a ring, proof of Sylvester's law