Orthogonality

Notes on Linear Algebra

Chia-Ping Chen

Department of Computer Science and Engineering
National Sun Yat-Sen University
Kaohsiung, Taiwan ROC

Introduction

- What is a basis?
  - By definition, it is a set of linearly independent vectors that spans a vector space.
  - Geometrically, it is a set of coordinate axes.
- A basis is implicit when we represent a vector by an array of numbers.
- The numbers represent a linear combination for the vector.

Orthonormal Basis

- A basis B is said to be orthonormal if the vectors are
  - mutually orthogonal,
  - of unit length.
- An orthonormal basis makes the task of turning geometric properties into algebraic calculations simple.

Vector Length

- With an orthonormal basis $\{e_1, \dots, e_n\}$ for a vector space S, every vector $x \in S$ can be written as

$$x = x_1 e_1 + \cdots + x_n e_n.$$

- The x_i's are called coordinates or components.
- We can write

$$x = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}^T.$$

- The length of x, denoted by $|x|$, is defined by

$$|x|^2 = x_1^2 + x_2^2 + \cdots + x_n^2.$$

Inner Product

- The inner product of vectors x, y is defined by

$$x^T y = x_1 y_1 + \cdots + x_n y_n.$$

- $x^T y$ is linear in the vectors x, y.
- $x^T y$ is only defined when x and y have the same number of components.

Orthogonality

- (Pythagoras) Vectors x and y are perpendicular if

$$|x|^2 + |y|^2 = |x - y|^2.$$

- Using the formula for vector length, two vectors x and y are perpendicular if

$$(x_1^2 + \cdots + x_n^2) + (y_1^2 + \cdots + y_n^2) = (x_1 - y_1)^2 + \cdots + (x_n - y_n)^2.$$

- It follows, after cancelling some terms, that

$$x^T y = 0.$$

Orthogonality to Independence

- (theorem) If nonzero vectors $v_1, \ldots, v_k$ are mutually orthogonal, then they are linearly independent.
- (proof) Suppose that

$$c_1 v_1 + \cdots + c_k v_k = 0.$$

Taking the inner product of both sides with v_1,

$$c_1 |v_1|^2 = 0 \ \Rightarrow c_1 = 0.$$

Similarly for other c_i's. So $v_1, \ldots, v_k$ are linearly independent.

Orthogonality from Independence

- (converse) From a linearly independent set of vectors, we can construct an orthonormal set with the same number of vectors.
- This is proved by construction by the famous Gram-Schmidt process.
- There exists an orthonormal basis for every vector space!

Orthogonal Subspaces

- Two subspaces U and V of a space S are said to be orthogonal if

$$u^T v = 0 \text{ for every } u \in U \text{ and } v \in V.$$

- This is denoted by $U \perp V$.
- Are the plane containing the floor and the plane containing the wall of the classroom orthogonal?
- Only need to check whether the basis vectors of V and U are mutually orthogonal (why?)

Orthogonality of Matrix Subspaces

- (theorem) Of the four fundamental subspaces of a matrix

$$\mathcal{C}(A^T) \perp \mathcal{N}(A),$$

$$\mathcal{C}(A) \perp \mathcal{N}(A^T).$$

- (proof) Let $x \in \mathcal{N}(A)$ and $v \in \mathcal{C}(A^T)$

$$v = A^T z \ \Rightarrow \ v^T x = z^T A x = 0.$$

Similarly, let $y \in \mathcal{N}(A^T)$ and $u \in \mathcal{C}(A)$,

$$u = Aw \ \Rightarrow \ u^T y = w^T A^T y = 0.$$

Orthogonal Complement

- A vector x is orthogonal to a set of vectors S if x is orthogonal to every vector in S.
- The orthogonal complement of a vector subspace V is the set of vectors orthogonal to V, and is denoted by $V^{\perp}$.
- Using inner product, it can be shown that $V^{\perp}$ is a subspace. (exercise)
- By this definition, every subspace orthogonal to V is a subset of $V^{\perp}$.

Complement of Matrix Subspaces

- (theorem) $\mathcal{N}(A) = (\mathcal{C}(A^T))^\perp$.
- (proof) $Ax = 0$ means x is perpendicular to every row of A, and therefore to $\mathcal{C}(A^T)$.
- (theorem) $\mathcal{C}(A^T) = (\mathcal{N}(A))^\perp$.
- (proof) $\mathcal{C}(A^T) \perp \mathcal{N}(A)$, so $\mathcal{C}(A^T) \subset (\mathcal{N}(A))^\perp$. Suppose there is a vector z orthogonal to $\mathcal{N}(A)$ but not in $\mathcal{C}(A^T)$. Then adding z in A as an extra row will not change $\mathcal{N}(A)$ but will enlarge $\mathcal{C}(A^T)$. The sum of dimensions of row space and nullspace will exceed the number of variables, which is not possible.

Fundamental Theorem Part 2

- Given an $m \times n$ matrix,
  - the row space is the orthogonal complement of the nullspace.
  - the column space is the orthogonal complement of the left nullspace.
- (corollary) The equation $Ax = b$ is solvable if and only if $b^T y = 0$ whenever $A^T y = 0$.

More on Orthogonal Complement

- Two subspaces V, W of a vector space S can be orthogonal without being complements.

$$\dim V + \dim W \leq \dim S.$$

- Orthogonal complement implies equality in the above equation

$$W = V^{\perp} \Rightarrow V = W^{\perp} \Rightarrow \dim V + \dim W = \dim S$$

Decomposition

- Given orthogonal complements V and W of a vector space S, every vector $x \in S$ can be written as a sum of two vectors
$$x = v + w,$$
with $v \in V$ and $w \in W$.
- v is called the projection (of x) onto V and w is the projection onto W.
- v, w are unique.

Using Fundamental Spaces

- A matrix multiplication Av maps v to the column space of A.
- Suppose A is $m \times n$, a vector $x \in R^n$ can be decomposed by

$$x = x_r + x_n, \;\; x_r \in \mathcal{C}(A^T), x_n \in \mathcal{N}(A).$$

- With this decomposition, we have

$$Ax = A(x_r + x_n) = Ax_r.$$

- Note that x_r is in row space, Ax_r is in column space.

$1-1$ Mapping

- The mapping from the row space to the column space by A is invertible.
- That is, if $b \in \mathcal{C}(A)$, then there is a unique $x_r \in \mathcal{C}(A^T)$ such that $Ax_r = b$.
- (proof) Suppose $Ax_r^{(1)} = b = Ax_r^{(2)}$.

$$A(x_r^{(1)} - x_r^{(2)}) = b - b = 0 \;\Rightarrow\; x_r^{(1)} - x_r^{(2)} \in \mathcal{N}(A).$$

  Combined with the fact that $x_r^{(1)} - x_r^{(2)} \in \mathcal{C}(A^T)$, we have $x_r^{(1)} - x_r^{(2)} = 0$.
- Every matrix maps its row space *onto* its column space.

Projection

- Suppose we are given a point (vector) b in space and we want to know its distance to a given line, say in the direction of a.
- We need to find a point p closest to b on that line.
- p satisfies the condition that $\overline{bp}$ is perpendicular to a.
- More generally, the projection point of b on a subspace S is the point $p \in S$ that is closest to b. $\overline{bp}$ is orthogonal to S.

On Projection

- Projection is the key to the least-squares solution to an over-determined system of linear equations $Ax = b$: we find the projection of b to the column space of A.
- Given a subspace S, we can construct a projection matrix P for S. Then the projection point of an arbitrary vector b on S is simply Pb.
- P can be constructed easily if an orthogonal basis for S is available.

Case: a Line

- The projection point p of b onto the line defined by direction a has two properties.
  - p is in the direction of a, say $p = \overline{x}a$.
  - $\overline{bp}$ is perpendicular to a.
- It follows

$$a^T(b - \overline{x}a) = 0 \;\Rightarrow\; p = \overline{x}a = \frac{a^T b}{a^T a}a = Pb.$$

- The projection matrix only depends on a,

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

A Side: Schwarz Inequality

- Since a distance is non-negative,

$$|b-p|^2 = |b - \frac{a^Tb}{a^Ta}a|^2 \geq 0.$$

- (Schwartz inequality) From above,

$$(b - \frac{a^Tb}{a^Ta}a)^T(b - \frac{a^Tb}{a^Ta}a) \geq 0 \Rightarrow |a^Tb|^2 \leq |a|^2|b|^2.$$

- The angle between vectors a and b can be defined by

$$\cos\theta = \frac{a^Tb}{|a||b|}, \quad -1 \leq \cos\theta \leq 1.$$

Properties of Projection Matrices

- Let $P = \frac{aa^T}{a^T a}$ be the matrix for a projection onto a line.
  - P is symmetric,
  - $P^2 = P$,
  - P is a matrix of rank 1.
- If $|a| = 1$, then $P = aa^T$. Can you use this result to construct the projection matrix onto the θ-direction in the xy-plane?

Over-Determined Systems

- Suppose there are more equations than variables in a system, say $Ax = b$.
- The right-hand side b is a vector in R^m and the column space is of dimension $r \leq n < m$.
- Often b is not in the column space of A and there is no exact solution.
- To illustrate, suppose that we want to solve

$$\begin{cases} a_1 x = b_1, \\ a_2 x = b_2. \end{cases}$$

A solution does not exist unless b is a multiple of a.

Least-Squares: Single Variable

- A least-squares solution is $\overline{x}$ that minimizes the sum of squares of $a_i\overline{x} - b_i$.
- We define the error function

$$E^2(x) = |ax - b|^2 = \sum_{i=1}^{m}(a_i x - b_i)^2.$$

- Setting the first derivative of E^2 with respect to x to 0, one gets the least-squares solution

$$\overline{x} = \frac{a^T b}{a^T a}.$$

- Note its relation to the projection matrix.

Least-Squares: Several Variables

- With several variables, we minimize $E^2 = |Ax - b|^2$ for $Ax = b$.
- This is indeed a projection problem from the following perspective: Ax is a vector in $\mathcal{C}(A)$, and we are looking for a point in $\mathcal{C}(A)$ that is closest to b.
- E^2 is minimized when $A\overline{x}$ is the projection point of b on $\mathcal{C}(A)$.

Normal Equation

- It follows that $b - A\overline{x}$ is perpendicular to $\mathcal{C}(A)$.

$$A^T(b - A\overline{x}) = 0 \;\Rightarrow\; A^T A\overline{x} = A^T b$$

- This is called the normal equation.
- If $A^T A$ is invertible then

$$\overline{x} = (A^T A)^{-1} A^T b \;\Rightarrow\; p = A\overline{x} = A(A^T A)^{-1} A^T b.$$

Properties of A^TA

- A^TA is symmetric.
- A^TA has the same nullspace as A.

$$Ax = 0 \;\Rightarrow\; A^TAx = 0;$$

$$A^TAx = 0 \;\Rightarrow\; x^TA^TAx = 0 \;\Rightarrow\; |Ax|^2 = 0.$$

- If A has linearly independent columns, then
  - A^TA is invertible, since the nullspace is $\{0\}$;
  - The projection point of b onto $\mathcal{C}(A)$ is

$$p = A\overline{x} = A(A^TA)^{-1}A^T b$$

An Example

- Suppose

$$A = \begin{bmatrix} 1 & 2 \\ 1 & 3 \\ 0 & 0 \end{bmatrix}, \quad b = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}.$$

- Since $A^T A$ is invertible,

$$\overline{x} = (A^T A)^{-1} A^T b = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \Rightarrow p = A\overline{x} = \begin{bmatrix} 4 \\ 5 \\ 0 \end{bmatrix}.$$

  Apparently p is in the column space of A.

Projection Matrice P

- The projection matrix P to the column space of A is

$$P = A(A^T A)^{-1} A^T \text{ since } p = A(A^T A)^{-1} A^T\, b = P\, b.$$

- Clearly, $P^2 = P$ and $P^T = P$.
- Conversely, if $P^T = P$ and $P^2 = P$, then P is a projection matrix to $\mathcal{C}(P)$, since $b - Pb$ is perpendicular to Pa for any a, b, as

$$(b - Pb)^T\, Pa = b^T Pa - b^T P^T Pa = 0.$$

Least-Squares Fitting of Data

- Suppose we are fitting data $(t_i, b_i)_{i=1}^{m}$ to a line

$$b = C + Dt.$$

- This can be written as

$$\begin{bmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}, \quad \text{or } Ax = b.$$

- Again we have a least-squares problem.

Theorem of Linear Fitting

- Define the squared error

$$E^2 = \sum (b_i - C - Dt_i)^2 = |b - Ax|^2.$$

- The line $b = \overline{C} + \overline{D}t$ which minimizes the squared error satisfies

$$A^T A \begin{bmatrix} \overline{C} \\ \overline{D} \end{bmatrix} = A^T b \Rightarrow \begin{bmatrix} m & \sum t_i \\ \sum t_i & \sum t_i^2 \end{bmatrix} \begin{bmatrix} \overline{C} \\ \overline{D} \end{bmatrix} = \begin{bmatrix} \sum b_i \\ \sum t_i b_i \end{bmatrix}.$$

Orthonormal

- Vectors $q_1, \ldots, q_k$ are said to be orthonormal if they are of unit length and perpendicular to each other, or

$$q_i^T q_j = \delta_{ij}.$$

- As an important example, the standard basis is an orthonormal one.
- Given a set of linearly independent vectors, one can construct an orthonormal set spanning the same space.
- The Gram-Schmidt process does this.

Orthogonal Matrices

- An orthogonal matrix is a square matrix whose column vectors are orthonormal.
- We use Q to denote an orthogonal matrix and $q_1, \ldots, q_n$ to denote columns in Q.
- For an orthogonal matrix Q,

$$Q^T Q = I \Rightarrow Q^T = Q^{-1} \Rightarrow QQ^T = I.$$

- If the columns of a square matrix are orthonormal, so are the rows!

Invariance under Orthogonal Matrix

- The length of a vector is preserved by an orthogonal transformation, i.e. multiplication by an orthogonal matrix,

$$|Qx|^2 = |x^T Q^T Qx| = |x^T x| = |x|^2.$$

- The inner product is also preserved,

$$(Qx)^T (Qy) = x^T Q^T Qy = x^T y.$$

Examples of Orthogonal Matrix

- rotation matrix

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

- reflection matrix

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

Orthonormal Basis

- The columns of an $n \times n$ orthogonal matrix Q is an orthonormal basis for R^n.
- We can write any vector b as a linear combination of column vectors of Q,

$$b = Qc.$$

- It is easy to get the combination coefficients: multiplying both sides by Q^T,

$$c = Q^T b.$$

Projection via Orthonormal Basis

- We now have

$$b = \sum q_i c_i = \sum q_i (q_i^T b).$$

- The component in the direction of q_i is precisely the projection of b on q_i,

$$\frac{q_i q_i^T}{q_i^T q_i} b = (q_i q_i^T) b.$$

- One can see that the projection of b onto a space, e.g. R^n, is the sum of projections on each of the vectors in an orthonormal basis of that space.

Rectangular Cases

- If Q is square, the projection matrix is the identity matrix I, which equals to QQ^T.
- What if Q is rectangular with orthonormal columns?
- Q^T is still a left inverse of Q, since $Q^TQ = I$.
- The least-squares solution to $Qx = b$ is

$$Q^TQ\overline{x} = Q^Tb \quad \Rightarrow \overline{x} = Q^Tb.$$

- The projection matrix P is

$$P = Q(Q^TQ)^{-1}Q^T = QQ^T,$$

which has the same form as the square case.

Gram-Schmidt Process

- This process constructs an orthonormal basis $\{q_1, \ldots, q_n\}$ from a basis $\{a_1, \ldots, a_n\}$.
- The idea is to subtract from a_j the components in the directions that have been settled and then normalize.
  - orthogonalization
$$a'_j = a_j - \sum_{i=1}^{j-1} (q_i^T a_j) q_i$$
  - normalization
$$q_j = \frac{a'_j}{|a'_j|}$$

Example

- Given

$$a = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \; b = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \; c = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}.$$

- By Gram-Schmidt

$$q_1 = \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{bmatrix}, \; q_2 = \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ -1/\sqrt{2} \end{bmatrix}, \; q_3 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}.$$

QR Factorization

- Let $A = [a_1 \ \ldots \ a_n]$ and $Q = [q_1 \ \ldots \ q_n]$. Then

$$A = QR, \quad \text{or } a_j = \sum R_{ij} q_i.$$

- Since q_i's are orthonormal,

$$R_{ij} = q_i^T a_j.$$

- As q_i is constructed to be orthogonal to the earlier a_j's,

$$R_{ij} = q_i^T a_j = 0, \quad i > j.$$

- R is an upper-triangular matrix.

Example

- For the previous example,

$$
\begin{aligned}
A &= \begin{bmatrix} 1 & 1 & 2 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \\
&= \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 0 & 1 \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 \end{bmatrix} \begin{bmatrix} \sqrt{2} & 1/\sqrt{2} & \sqrt{2} \\ 0 & 1/\sqrt{2} & \sqrt{2} \\ 0 & 0 & 1 \end{bmatrix} \\
&= QR
\end{aligned}
$$

Least-Squares with QR

- A least-squares problem is simplified by QR.
  - Substituting $A = QR$ in the normal equation,

$$
\begin{aligned}
A^T A\bar{x} = A^T b &\Rightarrow R^T Q^T QR\bar{x} = R^T Q^T b \\
&\Rightarrow R^T R\bar{x} = R^T Q^T b \\
&\Rightarrow R\bar{x} = Q^T b
\end{aligned}
$$

  - Since R is upper-triangular, $\bar{x}$ can be solved easily by the back substitution.

Hilbert Space

- A Hilbert space is the set of vectors with an infinite sequence of components and a finite length,

$$\{(v_1, v_2, \dots) \mid \sum v_i^2 < \infty\}.$$

- The inner product is well-defined in a Hilbert space since

$$|v^T w| \leq |v|\,|w|.$$

- Two vectors in a Hilbert space are said to be orthogonal if $v^T w = 0$.
- Hilbert space generalizes the notion of vector spaces to (countably) infinite dimension.

Function Space

- A function space is a space containing functions as vectors.
- The inner product is defined by an integral

$$(f, g) = \int f(x)g(x)dx.$$

- Length is defined by the inner product

$$|f| = (f, f).$$

- So is orthogonality

$$(f, g) = 0.$$

Fourier Series

- The Fourier series of a function $f(x)$ defined on $(0, 2\pi)$ is an expansion into sines and cosines,

$$f(x) = a_0 + \sum_{k=1}^{\infty} a_k \cos kx + b_k \sin kx.$$

- The cosines and sines consist a basis.
- Since the sines and cosines are mutually orthogonal, the coefficients are determined by projection. For example,

$$a_k = \frac{(f, \cos kx)}{(\cos kx, \cos kx)}.$$

Legendre Polynomials

- How about using the polynomials as basis vectors? Note, while $1, x, x^2, \ldots$ are independent, they are not orthogonal.
- We can construct an orthogonal basis from the powers of x via an orthogonalization process.
  - Let the interval be $[-1, 1]$. Start with $v_1(x) = 1$.
  - $v_2(x) = x$ is orthogonal to $v_1(x)$.
  - For the second-order polynomial,

$$v_3(x) = x^2 - \frac{(v_1, x^2)}{(v_1, v_1)} v_1(x) - \frac{(v_2, x^2)}{(v_2, v_2)} v_2(x) = x^2 - \frac{1}{3}.$$

  - $v_i(x)$ are the Legendre polynomials.

Approximation by Polynomial

- Suppose we want to approximate $y = x^5$ by a straight line

$$\hat{y} = c + dx$$

in the interval $(0, 1)$ in the sense of least-squares.

- That is, we want to minimize

$$|\hat{y} - y|^2.$$

- We describe a number of ways to find c, d.

Via Normal Equation

- Least-squares solution via normal equation.

$$A^T A\bar{x} = A^T b \Rightarrow \begin{bmatrix} (1,1) & (1,x) \\ (x,1) & (x,x) \end{bmatrix} \begin{bmatrix} c \\ d \end{bmatrix} = \begin{bmatrix} (1,x^5) \\ (x,x^5) \end{bmatrix}$$

- It's a matter of integration to solve for c and d.

Direct Minimization

- We can minimize $E^2 = \int_0^1 (x^5 - c - dx)^2 dx$.

Via Gram-Schmidt

- We first find an orthonormal basis based on polynomials

$$q_1 = 1, \;\; q_2 = x - \frac{1}{2}.$$

- The projection is given by

$$c + dx = \frac{(x^5, 1)}{(1,1)} 1 + \frac{(x^5, x - \frac{1}{2})}{(x - \frac{1}{2}, x - \frac{1}{2})}(x - \frac{1}{2}).$$

Discrete Fourier Transform

- While the Fourier series deals with functions defined on an interval, the discrete Fourier transform (DFT) deals with functions defined on an integer set.
- For DFT, the input is a sequence of numbers $y_0, \ldots, y_{n-1}$ and the output is another sequence $c_0, \ldots, c_{n-1}$ of the same length.
- The input-output relation can be represented by a matrix F, called the Fourier matrix, that

$$y = Fc \Leftrightarrow c = F^{-1}y.$$

Fourier Matrix I

- We require the function values y's to agree with the sequence at n discrete points x's, giving n equations.
- As an example, suppose the values of y at $x = 0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}$ are $2, 4, 6, 8.$ Then

$$c_0 + c_1 e^{ix} + c_2 e^{i2x} + c_3 e^{3ix} = \begin{cases} 2, & x = 0 \\ 4, & x = \frac{\pi}{2} \\ 6, & x = \pi \\ 8, & x = \frac{3\pi}{2}. \end{cases}$$

Fourier Matrix II

The previous equation can be written as

$$\begin{cases} c_0 + i^0 c_1 + i^0 c_2 + i^0 c_3 = 2 \\ c_0 + i^1 c_1 + i^2 c_2 + i^3 c_3 = 4 \\ c_0 + i^2 c_1 + i^4 c_2 + i^6 c_3 = 6 \\ c_0 + i^3 c_1 + i^6 c_2 + i^9 c_3 = 8 \end{cases} \Rightarrow \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & i & i^2 & i^3 \\ 1 & i^2 & i^4 & i^6 \\ 1 & i^3 & i^6 & i^9 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \\ 6 \\ 8 \end{bmatrix}$$

$$F = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & i & i^2 & i^3 \\ 1 & i^2 & i^4 & i^6 \\ 1 & i^3 & i^6 & i^9 \end{bmatrix} \Rightarrow F^{-1} = \frac{1}{4} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & (-i) & (-i)^2 & (-i)^3 \\ 1 & (-i)^2 & (-i)^4 & (-i)^6 \\ 1 & (-i)^3 & (-i)^6 & (-i)^9 \end{bmatrix}$$

Fourier Matrix III

- For n-point DFT, F and F^{-1} has entries

$$F_{jk} = w_n^{jk}, \quad F_{jk}^{-1} = \frac{1}{n} w_n^{-jk}$$

  where $w_n = e^{i\frac{2\pi}{n}}$.
- Note that the columns of F are orthogonal, since

$$1 + w_n^l + \cdots + w_n^{l(n-1)} = 0, \quad \forall\, 1 \leq l \leq n-1.$$

Fast Fourier Transform

- DFT is important in digital signal processing.
- The brute-force n-point DFT requires $O(n^2)$ floating-point multiplications.
- The cost can be reduced to $O(n \log n)$.
- We show the relation between n-point and m-point ($m = \frac{n}{2}$) DFTs. Split c into the even components c' and the odd components c''. With $y' = F_m c'$ and $y'' = F_m c''$,

$$\begin{cases} y_j = y'_j + w_n^j y''_j, & j = 0, 1, \ldots, m-1 \\ y_{j+m} = y'_j - w_n^j y''_j, & j = 0, 1, \ldots, m-1 \end{cases}$$

Time Complexity

- We first verify the previous formula.

$$y_j = \sum_{k=0}^{n-1} w_n^{jk} c_k = \sum_{k=0}^{m-1} w_n^{2jk} c_{2k} + \sum_{k=0}^{m-1} w_n^{j(2k+1)} c_{2k+1}$$

$$= \sum_{k=0}^{m-1} w_{\frac{n}{2}}^{jk} c'_k + w_n^j \sum_{k=0}^{m-1} w_{\frac{n}{2}}^{jk} c''_k.$$

- Recursively, an n-point DFT requires twice the number of multiplications of $\frac{n}{2}$-point DFT plus $\frac{n}{2}$.

$$\begin{cases} T(1) = 0 \\ T(n) = 2 * T(\frac{n}{2}) + \frac{n}{2} \end{cases} \Rightarrow T(n) = \frac{1}{2} n \log n.$$

Example

The steps for $n = 4$ is

$$\begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ c_3 \end{bmatrix} \rightarrow \begin{bmatrix} c_0 \\ c_2 \\ c_1 \\ c_3 \end{bmatrix} \rightarrow \begin{bmatrix} F_2 c' \\ \\ F_2 c'' \end{bmatrix} \rightarrow \begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

This is equivalent to

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & i & i^2 & i^3 \\ 1 & i^2 & i^4 & i^6 \\ 1 & i^3 & i^6 & i^9 \end{bmatrix} = \begin{bmatrix} 1 & & 1 & \\ & 1 & & i \\ 1 & & -1 & \\ & 1 & & -i \end{bmatrix} \begin{bmatrix} 1 & 1 & & \\ 1 & -1 & & \\ & & 1 & 1 \\ & & 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & & & \\ & & 1 & \\ & 1 & & \\ & & & 1 \end{bmatrix}$$

Summary for Linear Equations

- $Ax = b$.
- When A is invertible, the solution can be found by Gaussian elimination.
- When A is rectangular, b must be in the column space of A for a solution to exist.
- If b is not in the column space of A, there is no exact solution for $Ax = b$. In this case, we look for the least-squares solution by solving the normal equation

$$A^T A\overline{x} = A^T b.$$