Linear Algebra/Probability Theory review

MinSeok Song

- **Rank Nullity Theorem**

  Let $T : V \to W$ be a linear transformation between two vector spaces. Then $Rank(T) + Nullity(T) = dim(V)$. (this can be easily shown by using first isomorphism theorem and splitting lemma(check))

  Proof. Another method is to compare $Nullity(T)$ and $dim(V)$, and apply Steinitz exchange lemma to find the dimension of the image. □

- Steinitz exchange lemma: use induction and exchange u_k with w_i for some $i \in \{w_k \dots w_n\}$ in a induction step.

- Why is covariance matrix positive definitie? First define joint normal distribution.

  $$X_j = m_j + a_{j1}Z_1 + a_{j2}Z_2 + \cdots + a_{jm}Z_m$$

  For the case $m_j = 0$, we have $X = AZ$. AA^T is called the covariance matrix. Every symmetric semi-positive can be written as this form: use the spectral theorem.

- $Tr(A) = \sum \lambda_i$

- 1. normal iff unitarily diagonalizable
  2. Hermitian iff unitarily diagonalizable with real diagonal entries (aka Spectral theorem)

- (digression)

  X is diagonalizable iff the minimal polynomial is the product of distinct $x - \lambda_i$'s. So the degree of this polynomial is equal to the number of distinct eigenvalues.

  Characteristic polynomial of a square matrix is $\det(tI - A)$. Caley-Hamilton states that every square matrix over a commutative ring satisfies its own characteristic equation (intuition: $p(A)v_j = p(\lambda_j)v_j = 0$). Indeed, characteristic polynomial and minimal polynomial have the same roots over $\mathbb{C}$ (use $0 = \mu_A(A) \cdot v = \mu_A(\lambda) \cdot v$). In fact, it is important to note that $f(\lambda)$ is an eigenvalue of $f(A)$.

- positive eigenvalues$\not\rightarrow$positive definite

  Usually, we consider positive definite matrix AND symmetric matrix so eigenvalue is real.

- **Gerschgorin's theorem**

  Lemma 1. *Strictly diagonally dominant matrix is nonsingular.*

  Proof. There exists $x \neq 0$ such that $Ax = 0$. Use triangle inequality. □

  Theorem 2. *The number of eigenvalues in each connected component of $\cup_{i=1}^{n} G_i$ is equal to the number of Gerschgorin discs that constitute that component.*

  Proof. Suppose $z \notin \cup_{i=1}^{n} G_i$. Use the above lemma on $A - zI$. For the second statement, scale the non-diagonal elements. □

  Note that G_i is a closed disc.

  1. Schur factorization $A = QRQ^T$

     Proof. Consider E_λ and complement with Steinitz exchange lemma to form orthogonal basis of $\mathbb{C}^n$. Think in terms of change of basis. Use induction. □

  2. Singular Value Decomposition

     Proof. Apply spectral theorem on AA^T and $A^T A$. □

  3. LDU decomposition
  4. Cholesky decomposition
     A real Hermitian positive-definite matrix A can be expressed as LL^T.
  5. QR decomposition (Gram Schmidt Algorithm)
  6. LU factorization