Appendix A

Math Review

Preamble

Add a short paragraph. Cite sources. Indicate how much the reader needs to skim. Think about the audience of this paper. Guide them to the resources as we are listing the definitions accordingly. Mention the importance and rank each definition (1,2,3 stars).

A.1 Linear Algebra

A.1.1 Matrices

Definition A.1.1 (Eigenvalue). *Suppose $P \in \mathbb{F}^{n\times n}$ is a square matrix. Then, the eigenvalues of the matrix P are precisely the roots of the characteristic polynomial of P, given by $char_P(\lambda) = \det(P - \lambda I)$. The polynomial $char_P(\lambda)$ has degree n. So by the Fundamental Theorem of Algebra, P has a multiset of n eigenvalues.*

Definition A.1.2 (Inverse Matrix). *Suppose there is a matrix $P \in \mathbb{F}^{m\times n}$. Then, P^{-1} is its inverse matrix iff multiplying it by P returns the identity matrix. That is, the inverse matrix must satisfy:*

$$P^{-1}P = I$$

Definition A.1.3 (Transpose Matrix). *Suppose there is a matrix $P = (p_{ij}) \in \mathbb{F}^{m\times n}$. Then, its transpose matrix, $P^T = (q_{ij}) = (p_{ji}) \in \mathbb{F}^{n\times m}$ matrix whose columns are the rows of the original matrix.*

Definition A.1.4 (Conjugate Transpose Matrix). *Suppose there is a matrix $P = (p_{ij}) \in \mathbb{C}^{m\times n}$. Then, its conjugate transpose matrix, $P^\dagger = (q_{ij}) = (\overline{p_{ji}}) \in \mathbb{F}^{n\times m}$ is a matrix whose columns are the rows of the original matrix.*

Definition A.1.5 (Symmetric Matrix). *A matrix P is symmetric iff it is equal to its transpose:*

$$P = P^T$$

Definition A.1.6 (Hermitian Matrix). *A matrix P is Hermitian iff it is equal to its conjugate transpose:*

$$P = P^\dagger$$

Definition A.1.7 (Orthogonal Matrix). *A matrix P is called orthogonal iff its transpose is its inverse:*

$$P^T = P^{-1}.$$

Theorem A.1.1 (Orthogonal Group). *The set of* all *unitary matrices in $\mathbb{F}^{n\times n}$ is a matrix group. (It is called the orthogonal group.)*

Definition A.1.8 (Unitary Matrix). *A matrix P is called unitary iff its conjugate transpose is its inverse:*

$$P^\dagger = P^{-1}.$$

Theorem A.1.2 (Unitary Group). *The set of* all *unitary matrices in $\mathbb{C}^{n\times n}$ is a matrix group. (It is called the unitary group.)*

A.1.2 Proof: Real Symmetric Matrices have Real Eigenvectors

How does this relate to the rest of the thesis?

Tie back the relevance of this proof to the rest of this thesis.

Proof

Notation. For notational convenience, for any $N \in \mathbb{N}$, let $\widetilde{N} = \{1, \ldots, N\}$.

In this document, we prove that for any $M \times M$ real symmetric matrix, S_M, there exists for some eigenvalue λ, a corrosponding **real** eigenvector $\vec{v} \in \mathbb{R}^M$. Prior to starting the main proof, we begin with a lemma.

Lemma. Suppose we have a $M \times M$ real symmetric matrix with a some eigenvalue λ. If there we have a corrosponding eigenvector $v \in \mathbb{C}^M$, then every entry of v, say v_i is equal to a **real** linear combination of the other entries $v_j \mid j \neq i$.

So, we will show that:

$$\forall i \in \widetilde{M} : v_i = \sum_{j \neq i} c_j v_j \quad (c_j \in \mathbb{R})$$

Proof of Lemma. Begin by taking a real symmetric matrix S_M for some $M \in \mathbb{N}$. Suppose we have an eigenvalue λ. Then, if we have some eigenvector v, we know that:

$$(1) : \forall i \in \widetilde{M} : a_1 v_1 + \cdots + d_i v_i + \cdots + a_{m-1} v_m = \lambda v_i \quad (a_j \in \mathbb{R})$$

We obtain (1) by expanding the equality $Av = \lambda v$ and noticing that every row of Av is expressible as the sum of the non-diagonal entries multiplied by $v_j \mid j \neq i$ plus $d_i v_i$. Note that since our matrix is symmetric, for some rows, some of the constants a_j are not distinct but this should not raise any issues. Next, we collect the terms:

$$\forall i \in \widetilde{M} : a_1 v_1 + \cdots + a_{m-1} v_m = v_i(\lambda - d_i)$$

Since S_M is a real symmetric matrix, the a_j terms are real so we can say:

$$\forall i \in \widetilde{M} : v_i(\lambda - d_i) = \sum_{j \neq i} a_j v_j \quad (a_j \in \mathbb{R})$$

Finally, divide both sides by $(\lambda - d_i)$. Since S_M is a real symmetric matrix, we know $\lambda \in \mathbb{R}$ then also $(\lambda - d_i) \in \mathbb{R}$. On the right hand side, the coefficients of the v_j become $\frac{a_j}{(\lambda - d_i)}$. Since $a_j \in \mathbb{R}$, then also $\frac{a_j}{(\lambda - d_i)} \in \mathbb{R}$. Letting $c_j = \frac{a_j}{(\lambda - d_i)}$, we obtain:

$$\forall i \in \widetilde{M} : v_i = \sum_{j \neq i} c_j v_j \quad (\forall j : c_j \in \mathbb{R})$$

Thus, for any $M \in \mathbb{N}$, a real symmetric matrix with eigenvalue λ must have a corrosponding eigenvector v such that each of its entries is expressible as a real linear combination of the other entries. $\square$

Now, we will prove the main theorem.

Theorem (**Taqi**). Suppose we have a $M \times M$ real symmetric matrix, S_M. Then, we will show that there exists for some eigenvalue λ, a corrosponding **real** eigenvector $\vec{v} \in \mathbb{R}^M$.

Proof. For this proof we will induct on the dimension of the matrix, M. So let the inductive statement be

$$f(M) : S_M \text{ has a real eigenvector } v \text{ corrosponding to an eigenvalue } \lambda$$

Base Case. Take the base case $M = 2$. Then by **Zoom Meeting 11.12**, we know $f(2)$ is true.

Inductive Step. For our inductive step, we need to show that $f(M) \Rightarrow f(M+1)$. So, let us assume $f(M)$. This means that we can assume any real symmetric matrix S_M has a real eigenvector $v \in \mathbb{R}^M$ corrosponding to λ.

Next, we will write S_{M+1} as the matrix S_M augmented by some $u \in \mathbb{R}^M$ as follows:

$$S_{M+1} = \left[\begin{array}{c|c} S_M & u \\ \hline u^T & d_{M+1} \end{array}\right]$$

From our lemma, we use the fact that S_{M+1} is symmetric and our assumption of $f(M)$ to obtain:

$$(1) : \forall i \in \{1, \ldots, m+1\} : v_i = \sum_{j \neq i} c_j v_j \quad (c_j \in \mathbb{R})$$

$$(2) : \forall i \in \tilde{M} : v_i \in \mathbb{R}$$

In particular for (2), we know that $v_i = \left(\sum_{j \neq i} \frac{a_j}{d_i - \lambda} v_j\right)$.

From (1), we know that for row $i = m+1$: $v_{m+1} = \sum_{j \neq m+1} c_j v_j \quad (c_j \in \mathbb{R})$ By (2), this is a linear combination of real entries v_i. Since $v_{m+1} \in \mathbb{R}$, it follows that:

$$\forall i \in \{1, \ldots, m+1\} : v_i \in \mathbb{R}$$

So, we have established that $f(m) \Rightarrow f(M+1)$.

By the induction, the theorem is proved. $\square$.

A.2 Probability Theory

Please refer to **?** as a resource; the definitions were sourced from there. For the definitions and theorems covered in Section A.1, consider checking out Chapter 3: Random Variables and their distributions. For Order Statistics, consider Chapter 8, Section 6: Order Statistics.

A.2.1 Random Variables

Definition A.2.1 (Random Variable)**.** *A random variable $X : \Omega \to \mathbb{R}$ is a function from some sample space $\Omega = \{s_i\}_{i=1}^n$ to the real numbers $\mathbb{R}$. The sample space is taken to be any set of events such that the probability function corrosponding to the random variable, p_X exhausts over all the events in Ω. In other words, we expect $\int_\Omega p_X(s) = 1$.*

Definition A.2.2 (Probability Density Function)**.** *For a continuous r.v. X with CDF F, the probability density function (PDF) of X is the derivative f of the CDF, given by $f(x) = F'(x)$. The support of X, and of its distribution, is the set of all x where $f(x) > 0$.*

Theorem A.2.1 (Characterizing the PDF)**.** *A probability density function is characterized by a few properties that are necessary for it to be valid. They are as follows:*

1. *Non-negativity: the PDF must be a non-negative valued function everywhere.*

$$f(x) \geq 0$$

2. *Integrates to 1: the PDF must integrate to 1 when integrated over its entire support.*

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

Definition A.2.3 (Cumulative Distribution Function)**.** *The cumulative distribution function (CDF) of an r.v. X is the function F_X given by $F_X(x) = \boldsymbol{P}(X \leq x)$. When there is no risk of ambiguity, we sometimes drop the subscript and just write F (or some other letter) for a CDF.*

Theorem A.2.2 (Characterizing the CDF)**.** *A cumulative distribution function is characterized by a few properties that are necessary for it to be valid. They are as follows:*

1. *Monotonic function: the CDF must always be a monotonic function. That is:*

$$x_0 \leq x_1 \implies F(x_0) \leq F(x_1)$$

2. *Right-continuous: The CDF is continuous except possibly for having some jumps. Wherever there is a jump, the CDF is continuous from the right. That is, for any a, we have:*

$$F(a) = \lim_{x \to a^+} F(x)$$

3. *Converges to 0,1 in limits*

$$\lim_{x \to -\infty} F(x) = 0 \text{ and } \lim_{x \to \infty} F(x) = 1$$

Definition A.2.4 (Independence)**.** *Random variables X and Y are said to be independent if $\forall x, y \in \mathbb{R}$:*

$$\boldsymbol{P}(X \leq x, Y \leq y) = \boldsymbol{P}(X \leq x)\boldsymbol{P}(Y \leq y)$$

If the variables are discrete, then this is equivalent to the condition:

$$\boldsymbol{P}(X = x, Y = y) = \boldsymbol{P}(X = x)\boldsymbol{P}(Y = y)$$

Definition A.2.5 (i.i.d)**.** *A vector of random variables $\vec{X} = (X_i)_{i=1}^N$ is said to be i.i.d (independent and identially distributed) is each of its entries X_i are exactly so.*

A.2.2 Statistics

Definition A.2.6 (Statistic)**.** *A statistic is formally defined as a function of a vector of random variables. So, for instance f is a statistic of a vector of random variables $\bar{X}$. Its observed value is given by $f(\vec{X})$.*

Example (Mean Statistic)**.** *For instance, the mean of a random sample $\vec{X}$ is a statistic. Formally, we would define the statistic $f : \vec{X} \to \frac{\sum_i x_i}{N}$. So the mean of the random sample is defined as $\bar{x} = f(\vec{X})$.*

Order Statistics

Definition A.2.7 (Order Statistic)**.** *Suppose $\vec{X} = \{X_i\}_{i=1}^N$ is an ordered vector of random variables. Then, the i^{th} order statistic of X is given by X_i.*

Example (Order Statistic)**.** *Suppose $X = (20, 7, 2, 1)$. The smallest (fourth) order statistic is 1. The second order statistic is 7. The largest (first) order statistic is 20.*

Theoretical Results: Order Statistics

Here, talk about the results in Chapter 8 of Blitzstein with the following pairs: (Uniform, Beta) and (Exponential, Gamma). Add the simulations. Beef up this sub-subsection.

A.3 Markov Chains

How does this relate to the rest of the thesis?

Tie back the relevance of this proof to the rest of this thesis. This will come up in Appendix D. It is also important because transition matrices are representations of markov chains.

Definition A.3.1 (Markov Chain)**.** *Say a set of random variables X_i each take a value in a set, called the state space, $S_M = \{1, 2, \ldots, M\}$. Then, a sequence of such random variables $X_0, X_1, \ldots, X_n$ is called a Markov Chain if the following conditions are satisifed:*

- *$\forall X_i : X_i$ has support and range $S_M = \{1, 2, ..., M\}$.*
- *(Markov Property) The transition probability from state $i \to j$, $\boldsymbol{P}(X_{n+1} = j \mid X_n = i)$ is conditionally independent from all past events in the sequence $X_{n-1} = i', X_{n-2} = i'', \ldots, X_0 = i^{(n-1)}$, excluding the present/last event in the sequence. In other words, given the present, the past and the future are conditionally independent.*

  $$\forall i, j \in S_M : \boldsymbol{P}(X_{n+1} = j \mid X_n = i) = \boldsymbol{P}(X_{n+1} = j \mid X_n = i, X_{n-1} = i', \ldots, X_0 = i^{(n-1)})$$

Definition A.3.2 (Transition Matrixx)**.** *Let $X_0, X_1, \ldots, X_M$ be a Markov Chain with state space S_M. Letting $q_{ij} = \boldsymbol{P}(X_{n+1} = j \mid X_n = i)$ be the transition probability from $i \to j$, then the matrix $Q \in \mathcal{M}_{\mathbb{R}^+}[M \times M] : Q = (q_{ij})$ is the transition matrix of the chain. Q must satisfy the following conditions to be a valid transition matrix:*

Definition A.3.3 (Transition Matrix)**.** *Take a Markov Chain with states $1, \ldots, M$. Letting $q_{ij} = P(X_{n+1} = j \mid X_n = i)$ be the transition probability from $i \to j$, then the matrix $Q = (q_{ij})$ is the transition matrix of the chain. For this transiton matrix to be valid, its rows have to be stochastic, meaning their entries sum to 1; $\forall i \in 1, \ldots, M$: $\sum_{j \in 1,\ldots,M} q_{ij} = 1$.*

- *Q is a non-negative matrix. That is, note that $Q \in \mathcal{M}_{\mathbb{R}^+}[M \times M]$ so every $q_{ij} \in \mathbb{R}^+$. This follows because probabilities are necessarily non-negative values.*
- *The entries of every row i of Q must sum up to 1. This may be understood as applying the law of total probability to the event of transitioning from any given state $\forall i \in S_M$. In other words, the chain has to go somewhere with probability 1.*

  $$\forall i \in S_M : \sum_{j \in S_M} q_{ij} = 1$$

- *Note, it is NOT necessary that the converse holds. The columns of our transition matrix need not sum to 1 for it to be a valid transition matrix.*

Definition A.3.4 (n-Step Transition Probability)**.** *The $n-$step transition probability of $i \to j$ is the probability of being at j exactly n steps after being at i. We denote this value $q_{ij}^{(n)}$:*

$$q_{ij}^{(n)} : \boldsymbol{P}(X_n = j \mid X_0 = i)$$

Realize:

$$q_{ij}^{(2)} = \sum_{k \in S_M} q_{ik} \cdot q_{kj}$$

Because by definition, a Markov Chain is closed under a support/range of S_M so the event $i \to j$ may have taken any intermediate step $k \in S_M$. Realize by notational equivalence, $Q^2 = (q_{ij}^{(2)})$. Inducting over n, we then obtain that:

$$q_{ij}^{(n)} \textit{ is the } (i, j) \textit{ entry of } Q^n$$

Definition A.3.5 (Marginal Distribution of Xn)**.** *Let $\boldsymbol{t} = (t_1, t_2, \ldots, t_M)$ such that $\forall i \in S_M : t_i = \boldsymbol{P}(X_0 = i)$So, $\boldsymbol{t} \in \mathcal{M}_{\mathbb{R}}[1, M]$. Then, the marginal distribution of X_n is given by the product of the vector $\boldsymbol{t}Q^n \in \mathcal{M}_{\mathbb{R}}[1, M]$. That is, the j^{th} component of that vector is $P(X_n = j)$ for any $j \in S_M$. We may call $\boldsymbol{t}$ an initial state distribution.*

A.3.1 Classification of states

- *A state $i \in S_M$ is said to be **recurrent** if starting from i, the probability is 1 that the chain will eventually return to i. If the chain is not recurrent, it is **transient**, meaning that if it starts at i, there is a non-zero probability that it never returns to i.*
- *Caveat: As we let $n \to \infty$, our Markov chain will gurantee that all transient states will be left forever, no matter how small the probability is. This can be proven by letting the probability be some ε, then realizing that by the support of Geom(ε) is always some finite value, then the equivalence between the Markov property and independent Geometric trials gurantees the existence of some finite value such that there is a success of never returning to i.*

Definition A.3.6 (Reducibility)**.** *A Markov chain is said to be **irreducible** if for any $i, j \in S_M$, it is possible to go from $i \to j$ in a finite number of steps with positive probability. In other words:*

$$\forall i, j \in S_M : \exists n \in \mathbb{N} : q_{ij}^{(n)} > 0$$

- *From our quantifier formulation of irreducible Markov chains, note that we can equivalently say that a chain is irreducible if there is integer $n \in \mathbb{N}$ such that the (i, j) entry of Q^n is positive for any i, j.*
- *A Markov chain is **reducible** if it is not **irreducible**. Using our quantifier formulation, it means that it suffices to find transient states so that:*

$$\exists i, j \in S_M : \nexists n \in \mathbb{N} : q_{ij}^{(n)} > 0$$