Notes on Spectral Gap

Saurav Jadhav

1 Introduction

The spectral gap answers the question of how long it takes for the Markov chain to converge to equilibrium. This concept has significant practical implications. When using the Monte Carlo Markov Chain (MCMC) for sampling, we want the generated samples to follow the desired distribution. The convergence of the Markov chain is required to ensure this outcome. Hence, stopping sampling early may result in a sample that does not follow the desired distribution. Conversely, running the sampler for longer than required might result in bearing some extra computation cost as well as time consumption. The spectral gap measures the speed at which the chain converges to its stationary (equilibrium) distribution.

2 Basic Definitions

Before we delve into the concept of spectral gap, we first define the fundamentals of a Markov chain.

A Markov chain is a stochastic process that satisfies the memoryless property, wherein the future state depends only on the current state and not on the sequence of events that precede it. This fundamental property makes Markov chains particularly useful for modeling systems in which the future state is dependent only on the present state.

Let X_n denote the value of the process at time n. If, say, $X_n = i$, then the process is said to be in state i at time n.

We define the Markov chain as follows.

Definition (Markov Chain): A discrete-time stochastic process $\{X_n, n = 0, 1, 2, \ldots\}$ is called a Markov chain if it satisfies the Markov property:

$$\begin{aligned} P(X_{n+1} = j \mid X_n = i,\, X_{n-1} = i_{n-1},\, \ldots,\, X_1 = i_1,\, X_0 = i_0) \\ = P(X_{n+1} = j \mid X_n = i) \end{aligned} \tag{1}$$

The key components of a Markov chain are as follows:

- **State Space**: A finite or countable set $S = \{0, 1, 2, \ldots\}$ of all possible states.
- **Transition Probabilities**: $P_{ij} = P(X_{n+1} = j | X_n = i)$ represents the probability of transitioning from state i to state j.

- **Transition Matrix**: $\mathbf{P} = [P_{ij}]$ where each element P_{ij} is the transition probability from state i to state j.

The transition probabilities must satisfy the following constraints:

$$P_{ij} \geq 0 \qquad \forall i, j \in S \tag{2}$$

$$\sum_{j \in S} P_{ij} = 1 \qquad \forall i \in S \tag{3}$$

Equation (2) ensures that all probabilities are non-negative, whereas Equation (3) guarantees that each row of the transition matrix sums to one, reflecting the fact that the process must transition to some state.

3 Stationary Distribution

To define the convergence of a Markov chain, it is essential to introduce the concept of a stationary distribution.

A **stationary distribution** represents a probability distribution over the states of a Markov chain that remains unchanged as the chain evolves. If the chain starts in this distribution, all future state distributions will be the same.

Definition (Stationary Distribution): Let $\{X_n\}$ be a Markov chain with transition matrix P and state space S. A probability distribution $\pi = (\pi_j : j \in S)$ is called a stationary distribution for the chain if

$$\pi_j \geq 0 \quad \forall j \in S, \qquad \sum_{j \in S} \pi_j = 1,$$

and

$$\pi_j = \sum_{i \in S} \pi_i P_{ij} \quad \forall j \in S,$$

which can be written in matrix form as

$$\pi P = \pi. \tag{4}$$

If we carefully observe (4), it is a left eigenvector equation of the transition probability matrix $\mathbf{P}$. Here, π corresponds to the left eigenvector, and the corresponding eigenvalue is 1. In fact, 1 is the largest eigenvalue of an ergodic Markov chain, and the corresponding eigenvector is the stationary distribution $\boldsymbol{\pi}$. One might wonder what the other eigenvalues reveal. We will see that the second largest eigenvalue of an ergodic Markov chain plays a vital role in calculating the spectral gap.

A Markov chain is said to **converge to its stationary distribution** if, regardless of the initial distribution, the probability of being in state j after many steps approaches π_j as the number of steps goes to infinity:

$$\lim_{n \to \infty} P(X_n = j) = \pi_j. \tag{5}$$

Now, for a *finite state space* Markov chain, for equation (5) to hold, we want the MC to be **ergodic**. We will discuss in detail why this is necessary.

4 Ergodic Markov Chain

From now on, we will specifically discuss Markov chains in which the number of states is finite. An ergodic Markov chain satisfies two properties: *irreducibility* and *aperiodicity*.

A Markov chain is said to be **irreducible** if all states communicate with each other. Let us consider an example.

Consider a Markov chain consisting of three states 0,1, and 2, having the transition probability matrix (TPM)

$$P = \begin{pmatrix} \frac{2}{3} & \frac{1}{3} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{pmatrix}$$

You can verify from the above TPM that the Markov chain is irreducible. For example, you can go from state 0 to state 2 since

$$0 \to 1 \to 2$$

That is, to go from state 0 to state 1 (with probability $\frac{1}{3}$) and then go from state 1 to state 2 (with probability $\frac{1}{2}$). Similarly, you can reach from state 1 to state 0 via 2.

Now, let us define the **period** of a Markov chain as follows.

The *periodicity* of a state in a Markov chain is the largest integer d such that, starting from that state, the chain can return to it only at times that are multiples of d;

The following example illustrates this.

$$P = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

From the above TPM, it can be seen that every state communicates with every other state of the MC. Hence, the MC corresponding to the TPM is irreducible. What about the period? Starting from state 0, if you want to reach back to 0 itself, you will require at least three steps, that is,

$$0 \to 1 \to 2 \to 0$$

In fact, we can return to 0 only in steps that are multiples of three, that is, 3, 6, 9,... In other words, if we take n steps from state 0, and n is not a multiple of three, then we have not returned to state 0. Such states are periodic in nature, and the period is 3. In fact, for an irreducible Markov chain, the period of all states is the same. In this case, the periods of states 0,1, and 2 are the same, that is, 3.

Figure 1 shows the transition diagram for the above TPM. Arrows indicate the transition to that state. Mathematically, we define the period of a state as,

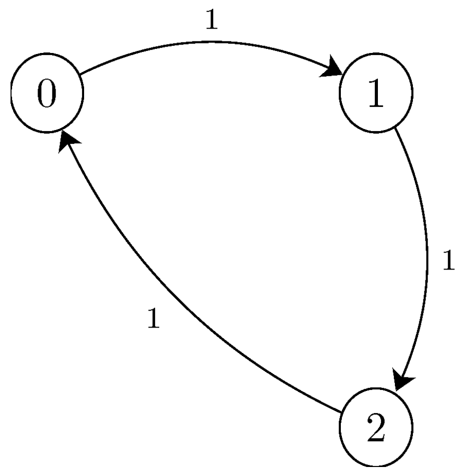

Figure 1: State Transition Diagram

The **period** of state i is the largest integer d satisfying the following property:

$$p_{ii}^{(n)} = 0, \text{ whenever } n \text{ is not divisible by } d.$$

$p_{ii}^{(n)}$ is the probability of going from state i to the same state in n steps.

The period of i is denoted by $d(i)$. If

$$p_{ii}^{(n)} = 0, \forall n > 0, \text{ then we let } d(i) = \infty.$$

- If $d(i) > 1$, we say that state i is **periodic**.
- If $d(i) = 1$, we say that state i is **aperiodic**.

In fact, $d(i)$ is the greatest common divisor (gcd) of the set of all return steps.

Returning to (5), we earlier stated that for this equation to hold, the MC must be ergodic in nature; that is, it should be both irreducible and aperiodic. One can now ask why this is the case. Assume that the MC is not irreducible, that is, not all states communicate with each other. This implies that we may have multiple sets of communicating states. This implies that the stationary distribution may not be unique. In fact, it may not exist. However, what if the MC is irreducible but not aperiodic? In this case, we have a unique stationary distribution, but the time-dependent distribution p^n of the states does not converge to π. That is, (5) is not satisfied. This is because if an MC is periodic, the distribution of the chain cycles and the chain itself oscillate.

If you solve for the above example in Figure 1, you get the unique stationary distribution, as

$$\pi_0 = \pi_1 = \pi_2 = \frac{1}{3}$$

However, consider that you are at, say, state 0. You know exactly the state in which you will be, given the number of steps n. This means that the chain cycles deterministically; that is, after three steps, you return to where you started. Hence, in this case, the chain does not converge to the stationary distribution. Therefore, to observe a unique stationary distribution and guarantee convergence, we require the MC to be ergodic, that is, both irreducible and aperiodic. One might wonder how all these things are tied to the spectral gap. We will examine this in detail when we discuss the spectral gap.

5 Reversible Markov chain

In layman's terms, a reversible Markov chain is one where, once the process has reached a stationary (long-run) behavior, the probability of moving from state i to j is the same whether you are watching the process forward in time or backward (i.e., the time-reversed chain).

Mathematically, we write it as,

A Markov chain with transition matrix P and stationary distribution π is **reversible** if and only if, for all states i and j, the following holds

$$\pi_i P_{ij} = \pi_j P_{ji} \tag{6}$$

This equation is called the **detailed balance condition**.

Again, one might ask why a reversible MC is required to study the spectral gap. If a Markov chain is reversible, the spectral gap computation is simplified. We will discuss this in detail when we discuss the spectral gap. However, we do not restrict ourselves to reversible MC. In fact, we will compute the spectral gap for a non-reversible Markov chain.

6 Spectral Gap

Now, since we have set up the context, we will see what the spectral gap is, and not only that, but also we will see why the condition of ergodicity has significant importance in studying the spectral gap.

Definition (Spectral Gap): Let P be the transition matrix of an ergodic reversible Markov chain with stationary distribution π.

$$\text{Spectral gap} := 1 - \lambda_{(2)} \tag{7}$$

where $\lambda_{(2)}$ is the second-largest eigenvalue of P.

The spectral gap is the difference between the largest and second-largest eigenvalues. This is because the largest eigenvalue is always one. Equation (7) describes the convergence of the complete Markov chain; the larger the spectral gap, the faster the convergence of the chain to its stationary distribution. Hence, for the calculation of the spectral gap in the case of an ergodic reversible Markov chain, we have a simple and meaningful equation. However, when discussing the spectral gap, we generally refer to the absolute spectral gap, as it yields the true rate of convergence, regardless of whether the eigenvalues are negative or positive.

The *absolute spectral gap* of P is defined as

$$1 - \max_{i \neq 1} |\lambda_i|.$$

Where,

$$1 = \lambda_1 > \lambda_2 \geq \cdots \geq \lambda_n \geq -1.$$

What if the Markov chain is non-ergodic? Here, we have two cases: either the MC is reducible or periodic.

We first consider the case of reducibility. If the MC is reducible, then there are multiple sets of communicating classes, and the stationary distribution is not unique. This implies that we will have various sets of irreducible states. In this case, multiple eigenvalues will be one, and the spectral gap will lose its meaning in terms of global convergence, that is, the convergence of the complete Markov chain.

What if the MC is periodic? As stated previously, when the chain is periodic, it exhibits oscillatory behavior; in this case, we may not observe the convergence of the Markov chain.

The existence of Equation (7) is because the MC is time-reversible. This connection is related to the spectral decomposition of matrices. In fact, (6) enables us to create a symmetric matrix of the TPM that can be decomposed by spectral decomposition, eventually leading to (7).

We will now generalize the concept of spectral gap for non-reversible Markov chains. For a time-irreversible ergodic Markov chain, the spectral gap is given in terms of the singular value. This is due to the decomposition of the TPM. Because when the MC is time-irreversible, the (6) will not exist for this case, and hence we won't be able to construct a symmetric matrix, and spectral decomposition won't be possible. Therefore, in this case, we use singular value decomposition, and hence, singular values become the key to the spectral gap.

Definition (Generalized Spectral Gap): For a finite, ergodic Markov chain with transition matrix P, The spectral gap is the second smallest singular value of the generator $L = I - P$.

$$\text{Spectral Gap} = \sigma_2(L) \tag{8}$$

where $\sigma_2(L)$ denotes the second smallest singular value of the matrix L, and I is an identity matrix. The smallest singular value, that is, $\sigma_1(L)$, is always zero. The above definition is not only generalized but also robust. For a non-reversible MC, the eigenvalues can be complex; however, the singular values are always non-negative and real.

Matrix L is defined in Hilbert space. To compute the spectral gap in this case, we first need to transform L to define it in Euclidean space. The transformation of L for the Euclidean space is given by

$$\tilde{L} = \Pi^{1/2} L \Pi^{-1/2} \tag{9}$$

where $\Pi = \text{diag}(\pi_1, \pi_2, \ldots, \pi_n)$, and $(\pi_1, \ldots, \pi_n)$ is the stationary distribution of the Markov chain.
Therefore, instead of finding the second smallest singular value of L, we find the same value for $\tilde{L}$.

The spectral gap defined in (8) is generalized. That is, the definition works for both reversible and non-reversible MC, and (9) allows us to calculate it.