Advanced Finite Regular Markov Chains

Prepared by Yoni Nazarathy, (Last Updated: Afternoon of October 27, 2014)

So you know a few things about Markov Chains on finite state spaces. For example you know that transition probabilities evolve by multiplication by the transition probability matrix, and you know the meaning of powers of that matrix, you know about class structure, irreducibility, transience, recurrence and periodicity. You also know about the stationary distribution. You understand the basic stuff. You've reached some level of maturity in life. Good for you!

The other thing to really know about is about the behaviour of Markov Reward Processes (MRP). These are processes of the form $\big(X_t,\, r(X_t)\big)$ where X_t is a Markov chain and $r(\cdot)$ is some function of the state space. Reward is the accumulated and possibly it's time average is taken. When dealing with *policy evaluation* for Markov Decision Processes (MDP), understanding how to analyse MRP is critical. The theory for the case of finite state spaces is closed and well understood. This is the subject of these notes.

The main use of these notes is to be an aid for following Chapter 8 of [Put94], "Average Reward and Related Criteria". That chapter relies on Appendix A of the book, which gives a concise treatment of the subject. Specifically the Drazin Inverse and the Deviation Matrix. One complication in the book and it's appendix is that the treatment is general in the sense that it supports Markov chains with several classes (not irreducible) and also periodic Markov chains. This is all nice and good, but for a first reading one may want to assume irreducibly and non-periodicity so as to understand the key concepts without complication. This is what we do in these notes. I.e. the notes summarise the results of Appendix A of [Put94] together with Section 8.2, "Markov Reward Processes and Evaluation Equations", but assume throughout an irreducible and aperiodic Markov chain.

The notes contain exercises. Some of these exercises take you through steps of proofs. As you do that, make sure you also understand the assumptions in the exercises.

Thanks to error/typo catchers: **Sophie Hautphenne, Patrick Laub and Stephen Lynch.**

Chapter 1

Basics

This material was covered in the lecture notes: "Basic Probability and Markov Chains" (and is covered in many other places also). Here we simply put down the key concepts for the purpose of notation (which differs slightly from the aforementioned notes).

1.1 Basic Definitions and Properties

Let $\{X_t,\, t = 0, 1, 2, \ldots\}$ be a sequence of random variables taking values in $\mathcal{S} = \{1, \ldots, N\}$. We say this sequence is a *Time Homogenous Finite State Space Discrete Time Markov Chain* if,

$$\mathbb{P}(X_t = j \mid X_{t-1} = s,\, X_{t-2} = j_{t-2}, \ldots, X_0 = j_0) = \mathbb{P}(X_1 = j \mid X_0 = s) := p(j \mid s).$$

The matrix, P with s, j'th entry being $p(j \mid s)$ is called the *transition probability matrix*. We have that P^m (m'th matrix power) is a matrix with entries, s, j being,

$$p^{(m)}(j \mid s) := \mathbb{P}(X_m = j \mid X_0 = s).$$

We say that the Markov chain is irreducible if for each pair of states j and s there exists an $m > 0$ such that $p^{(m)}(j \mid s) > 0$. In a finite-state space irreducible Markov chain all states are *positive recurrent*. This is typically defined to be the property that the expected return to each state (from itself) is finite.

If the greatest common divisor of $\{m : p^{(m)}(s \mid s) > 0\}$ equals 1 for state s then the state is said to have period 1. In an irreducible Markov chain, if this holds for one state, then it holds for all states. In this case we say the Markov chain is *aperiodic*.

In the remainder of these notes we deal with **finite state space** Markov chains that are **irreducible** and **aperiodic**. All statements are based on this assumption. The more general case is covered in [Put94].

1.2 The Limiting Matrix

Define the limiting matrix P^* to be with elements,

$$p^*(j \mid s) = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} p^{(t)}(j \,|\, s) = \lim_{T\to\infty} p^{(T)}(j \,|\, s). \tag{1.1}$$

The first limit is called a *Cesaro limit* it equals the second limit only in a case where the second limit exists (this is the case if the Markov chain is aperiodic). We assume this is the case. The matrix with s, j'th entry being $p^*(j \,|\, s)$ is denoted P^* and is called the *limiting matrix*.

Exercise 1. *Show that P^* satisfies the following equalities:*

$$PP^* = P^*P = P^*P^* = P^*.$$

Exercise 2. *Show that*

1. $(I - P^*)^2 = (I - P^*)$.
2. $P^*(I - P^*) = 0$.
3. P^* *is a stochastic matrix.*

The following theorem describes the *stationary distribution.*

Theorem 1. *The system of equations,*

$$\mathbf{q}'P = \mathbf{q}', \quad \mathbf{q}'\mathbf{1} = 1,$$

has a unique positive solution.

Since $P^*P = P^*$, we can write,

$$P^* = \mathbf{1}\mathbf{q}'.$$

That is P^* is an outer product of $\mathbf{1}$ and $\mathbf{q}$. It is a rank one matrix, with all rows equal the stationary distribution $\mathbf{q}'$.

Exercise 3. *Exercises* (1) *and* (2) *did not assume $P^* = \mathbf{1}\mathbf{q}'$. Carry out the computations in these exercises again under this structure of P^*.*

Chapter 2

The Generalized Inverses

An inverse of a matrix, A is a matrix A^{-1} such that $A\,A^{-1} = I$. For non-square matrices and for singular square matrices, such an inverse does not exist. But once can define the *generalized inverse* (in several ways). This is a big subject in linear algebra. In this chapter, we discuss certain generalized inverses associated with Markov chains.

2.1 The Underlying Linear Algebra

The following theorem summarizes a basic property of a stochastic, irreducible, aperiodic matrix (Markov Chain), P. It encompass what is called as the "Perron-Frobenius" theorem. There are also more general versions for irreducible and not necessarily aperiodic Markov chains. Denote by $\sigma(A)$ the spectral radius of the matrix A, this is the maximal modulus of all eigenvalues of A.

Theorem 2. *Under the finite state space, irreducible, aperiodic assumptions, the following hold:*

1. *The value* 1 *is an eigenvalue of* P *with algebraic and geometric multiplicity one and and a single linearly independent eigenvector.*
2. *There exists a non-singular matrix* W *for which,*

$$P = W^{-1} \begin{bmatrix} Q & 0 \\ 0 & 1 \end{bmatrix} W, \tag{2.1}$$

   where Q *is an* $(N-1) \times (N-1)$ *matrix with the following properties:*

   (a) *It holds that* $\sigma(Q) < 1$ *(so* 1 *is not an eigenvalue of* Q*).*
   (b) *The inverse* $(I-Q)^{-1}$ *exists.*
   (c) $\sigma(I-Q) = \sigma(I-P)$.

3. *The matrix* P^* *is unique and may be represented by,*

$$P^* = W^{-1} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} W.$$

One use Theorem 2 is to prove that the limit in (1.1) holds. We follow this proof now through a series of three straight forward exercises:

Exercise 4. *Show that,*

$$P^n = W^{-1} \begin{bmatrix} Q^n & 0 \\ 0 & 1 \end{bmatrix} W,$$

and hence,

$$\frac{1}{T} \sum_{t=0}^{T-1} P^t = W^{-1} \begin{bmatrix} \frac{1}{T} \sum_{t=0}^{T-1} Q^t & 0 \\ 0 & 1 \end{bmatrix} W.$$

Next,

Exercise 5. *Show that since $I - Q$ is non-singular,*

$$\sum_{t=0}^{T-1} Q^t = (I - Q^T)(I - Q)^{-1},$$

and hence since $\sigma(Q) < 1$, Q^T is bounded (in T) so that,

$$\lim_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T-1} Q^t = 0.$$

Finally,

Exercise 6. *Show now that,*

$$\lim_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T-1} P^t = W^{-1} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} W = P^*.$$

2.2 The Drazin Inverse

Take a matrix B that has the representation,

$$B = W^{-1} \begin{bmatrix} C & 0 \\ 0 & 0 \end{bmatrix} W,$$

where C and W are nonsingular. Then the *Drazin inverse* (or *group inverse*), denoted $B^\#$ is defined as,

$$B^\# = W^{-1} \begin{bmatrix} C^{-1} & 0 \\ 0 & 0 \end{bmatrix} W.$$

Exercise 7. *Show the following:*

1. $B^\# \, B \, B^\# = B^\#$.

2. $B \, B^\# = B^\# \, B$.

3. $B\,B^{\#}\,B = B$.

The Drazin inverse is a particular generalised inverse of B (we do not cover more definitions related to generalised inverses here).
We now study the Drazin inverse $(I-P)^{\#}$.

Theorem 3. *The following holds:*

1. *The matrix $(I - P + P^*)$ is non-singular with Z_P denoting it's inverse, i.e.,*

$$Z_P \equiv (I - P + P^*)^{-1}.$$

2. *The Drazin inverse of $(I-P)$ denoted by H_P satisfies,*

$$H_P \equiv (I-P)^{\#} = (I - P + P^*)^{-1}(I - P^*) = Z_P(I - P^*).$$

3. *It holds that,*

$$H_P = \sum_{t=0}^{\infty}(P^t - P^*).$$

We now illustrate the proof of (1) and (2) through through exercises (we skip the proof of (3)):

Exercise 8. *Prove (1) by showing the representation,*

$$I - P + P^* = W^{-1}\begin{bmatrix} I-Q & 0 \\ 0 & 1 \end{bmatrix} W.$$

Exercise 9. *Show that definition of the Drazin inverse implies,*

$$\begin{aligned}(I-P)^{\#} &= W^{-1}\begin{bmatrix} (I-Q)^{-1} & 0 \\ 0 & 0 \end{bmatrix} W \\ &= W^{-1}\begin{bmatrix} (I-Q)^{-1} & 0 \\ 0 & 1 \end{bmatrix} W - W^{-1}\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} W\end{aligned}$$

and from this (2) follows.

The matrix H_P is referred to as the *deviation matrix*. The matrix Z_P is referred to as the *fundamental matrix*. They are both vaguely referred to as generalized inverses associated with P.

Exercise 10. *Derive the following:*

1. $(I-P)H_P = H_P(I-P) = I - P^*$.
2. $H_P P^* = P^* H_P = 0$.
3. $H_P = Z_P - P^*$.
4. $Z_P P^* = P^*$.
5. $P^* = I - (I-P)(I-P)^{\#}$.

Chapter 3

The Laurent Series

For $\rho > 1$ define the *resolvent* of $P - I$ denoted by R^ρ by,

$$R^\rho = \big(\rho I + (I - P)\big)^{-1}.$$

Letting $\lambda = (1+\rho)^{-1}$, we get,

$$(I - \lambda P) = (1+\rho)^{-1}\big(\rho I + (I - P)\big).$$

When $\lambda \in [0,1)$, $\sigma(\lambda P) < 1$ so $(I - \lambda P)^{-1}$ exists. Hence for $\rho > 1$, R^ρ exists.

Exercise 11. *Show the following:*

1. $(I - \lambda P)^{-1} = (1+\rho)R^\rho$.
2. $R^\rho = \lambda(I - \lambda P)^{-1}$.

This is the Laurent series expansion for the resolvent:

Theorem 4. *For* $\rho \in \big(0, \sigma(I - P)\big)$,

$$R^\rho = \rho^{-1}P^* + \sum_{n=0}^{\infty}(-\rho)^n H_P^{n+1}.$$

Once again we supply the proof through a series of exercises.

Exercise 12. *Let* Q *be defined through* (2.1) *and set* $B = I - Q$*. Then show how to use,*

$$\rho I + I - P = W^{-1}\begin{bmatrix} \rho I + B & 0 \\ 0 & \rho \end{bmatrix} W,$$

to obtain,

$$R^\rho = \rho^{-1}W^{-1}\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} W + W^{-1}\begin{bmatrix} (\rho I + B)^{-1} & 0 \\ 0 & 0 \end{bmatrix} W. \tag{3.1}$$

Exercise 13. *Show that the first term of* (3.1) *equals* $\rho^{-1}P^*$.

Exercise 14. *Show that the second term of* (3.1) *equals,*

$$\sum_{n=0}^{\infty}(-\rho)^n W^{-1}\begin{bmatrix} (I-Q)^{-(n+1)} & 0 \\ 0 & 0 \end{bmatrix} W = \sum_{n=0}^{\infty}(-\rho)^n H_P^{n+1}.$$

Do this by first showing that,

$$(\rho I + B)^{-1} = (I + \rho B^{-1})^{-1} B^{-1}$$

Chapter 4

Evaluation of Accumulated/Discounted/Average Reward

We now show how to use some of the tools from above for evaluating the reward in a Markov Chain. While we do not mention a Markov Decision Process (in these notes), think of this as the reward obtained in an MDP with some given decision rule.

4.1 The Gain and Bias

Consider now some reward function: $r : \mathcal{S} \to \mathbf{R}$. Since we take $\mathcal{S} = \{1, \ldots, N\}$ it is convenient to denote the vector of values $r(1), \ldots, r(N)$ by $\mathbf{r}$. We are now interested in the *gain* (infinite horizon average cost):

$$g(s) = \lim_{T\to\infty} \frac{1}{T}\,\mathbb{E}_s\Big[\sum_{t=1}^{T} r(X_t)\Big] = \lim_{T\to\infty} \frac{1}{T}\,\sum_{t=1}^{T}[P^{t-1}r](s) = [P^* r](s) = \mathbf{q}'\,\mathbf{r}.$$

The fact $g(s)$ does not depend on the initial state s is because of the irreducibility assumption (holding throughout these notes). Denote the vector of constant values g, by $\mathbf{g}$

The *bias* (vector) is defined to be,

$$\mathbf{h} := H_P\,\mathbf{r},$$

where H_P is the fundamental matrix.

Exercise 15. *Show that,*

$$\mathbf{h} = \sum_{t=0}^{\infty} P^t(\mathbf{r} - \mathbf{g}),$$

and explain why,

$$h(s) = \mathbb{E}_s\Big[\sum_{t=1}^{\infty} \big(r(X_t) - g(X_t)\big)\Big].$$

The *total reward* (vector) in N time units is,

$$\mathbf{v}_{T+1} = \sum_{t=1}^{T} P^{t-1}\mathbf{r},$$

i.e. $\mathbf{v}_{T+1}(s)$ is the total reward when starting in state s.

Exercise 16. *Show that,*

$$\mathbf{v}_{T+1} = T\mathbf{g} + \mathbf{h} + o(1),$$

where $o(1)$ is a vector with components that approach 0 as $T \to \infty$.

4.2 Using the Laurent Series Expansion

We define,

$$\mathbf{v}_\lambda = (I - \lambda P)^{-1}\mathbf{r}.$$

where $\lambda \in [0,1]$. In fact, $\mathbf{v}_\lambda$ is the expected discounted cost with discount factor λ (this was shown in earlier when studying MDP).
Setting $\rho = (1-\lambda)\lambda^{-1}$ or alternatively $\lambda = (1+\rho)^{-1}$, we have

$$\mathbf{v}_\lambda = (1+\rho)\big(\rho I + [P - I]\big)^{-1}\mathbf{r}.$$

Exercise 17. *Show the following:*
Let ν denote the nonzero eigenvalue of $I - P$ with smallest modulus. Then for $0 < \rho < |\nu|$ (ρ "sufficiently small"),

$$\mathbf{v}_\lambda = (1+\rho)\Big[\rho^{-1} y_{-1} + \sum_{n=0}^{\infty} \rho^n y_n\Big],$$

where,

$$\begin{aligned} y_{-1} &= P^*\mathbf{r}, \\ y_0 &= \mathbf{g}, \\ y_n &= (-1)^n H_P^{n+1}\mathbf{r}, \qquad n = 1, 2, \ldots \end{aligned}$$

As a consequence the following holds:

Exercise 18. *Establish,*

$$\mathbf{v}_\lambda = \frac{1}{1-\lambda}\mathbf{g} + \mathbf{h} + o(1-\lambda),$$

where $o(1-\lambda)$ is a vector that converges to 0 as $\lambda \uparrow 1$.

As a consequence, the following relation between the gain and the discounted reward holds:

Exercise 19. *Establish,*

$$\mathbf{g} = \lim_{\lambda \uparrow 1} (1-\lambda)\mathbf{v}_\lambda.$$

4.3 Evaluation Equations

Computing $\mathbf{g}$ and $\mathbf{h}$ through direct evaluation of P^* and H_P can be done in very specific cases, but is otherwise inefficient. An alternative is using a system of equations (sometimes refereed to as *Poisson's equation* – although not in [Put94]).

Theorem 5. *The following holds,*

$$\mathbf{g} + (I - P)\mathbf{h} = \mathbf{r}. \tag{4.1}$$

Further, if $\mathbf{g}$ *and* $\mathbf{h}$ *are some vectors that satisfy* (4.1) *then* $P^*\mathbf{h} = \mathbf{0}$ *and* $\mathbf{h} = H_P\,\mathbf{r} + \gamma\mathbf{1}$ *for some arbitrary scalar* γ.

We establish (4.1) in this exercise.

Exercise 20. *Use now* $P^* + (I - P)H_P = I$ *to establish* (4.1).

A consequence is that (4.1) uniquely determines h up to an element of the null space of $I - P$. This (in the irreducible case) is a space of dimension 1. Thus we we can find the *relative values* $h(j) - h(k)$ by setting any component of $\mathbf{h}$ to 0 and solving (4.1).