Probabilistic Aspects of Computer Science: Markovian Models

S. Haddad[1]

September 24, 2019

[1]Professor at ENS Cachan, haddad@lsv.ens-cachan.fr, http://www.lsv.ens-cachan.fr/∼haddad/

Abstract

These lecture notes present five Markovian models: discrete time Markov chains (DTMC), continous time Markov chains (CTMC), Markov decision processes (MDP), stochastic games (SG) and probabilistic automata (PA). It is addressed for master students and tries as most as possible to be self-contained. However the basics of discrete probability (and additionally the basics of measure and integration for the study of CTMC) are required. Recommended books for french speaking students are [FOA 98, FOA 02]. There are a lot of good books for english speaking students.

Contents

Chapter 1

Discrete Time Markov Chains

1.1 Discrete event systems

Most of the probabilistic systems we study in this course are (stochastic) *Discrete Event Systems* (DES), a particular case of point processes (see [BRE 98] for more details). An execution of a DES is specified by an (a priori infinite) sequence of events $e_1, e_2, \ldots$ occurring at successive instants. Only events can change the state of the system.

Formally, the stochastic behaviour of a DES is determined by two families of random variables:

- $S_0, \ldots, S_n, \ldots$ belonging to the (discrete) state space of the system, denoted S. S_0 represents the initial state of the system and S_n $(n > 0)$ the current state after the occurrence of the n^{th} event. The occurrence of an event does not necessarily modify the state of the system, hence S_{n+1} may be equal to S_n.
- $T_0, \ldots, T_n, \ldots$ belonging to $\mathbb{R}^+$. T_0 represents the time interval before the first event and T_n $(n > 0)$ represents the time interval between the n^{th} and the $(n+1)^{th}$ event. Observe that this interval may be null (e.g. a sequence of instructions considered as instantaneous w.r.t. database transactions involving input/output operations).

When the distribution of S_0 is concentrated in a state s, one says that the DES starts in s (i.e. $\mathbf{Pr}(S_0 = s) = 1$).

A priori, there is no restriction about these families of random variables. However, most of the stochastic processes that we study cannot execute an infinite number of actions in a finite time (which is called a Zeno behaviour). Otherwise stated:

$$\sum_{n=0}^{\infty} T_n = \infty \text{ almost surely} \tag{1.1}$$

When this property holds the state of the system can be defined for all instants. Let $N(\tau)$, the random variable defined by:

$$N(\tau) \stackrel{\text{def}}{=} \inf\{n \mid \sum_{k=0}^{n} T_k > \tau\}$$

Using equation (1.1), $N(\tau)$ is defined *almost everywhere* (i.e. for almost every sample). As can be observed in figure 1.1, $N(\tau)$ presents jumps whose range is greater than 1. State $X(\tau)$ of the system at time τ, is then $S_{N(\tau)}$. $X(\tau)$ is *not* equivalent to the stochastic process, but it allows, in most of the cases, to perform the standard analyses. Figure 1.1 presents a possible *sample* of the process and illustrates the previously defined random variables. In this sample, the process is initially in state s_4 and it stays in it until time τ_0 when its state becomes s_6. At time $\tau_0 + \tau_1$, the system successively visits in zero time, states s_3 and s_{12} before reaching state s_7 where it stays

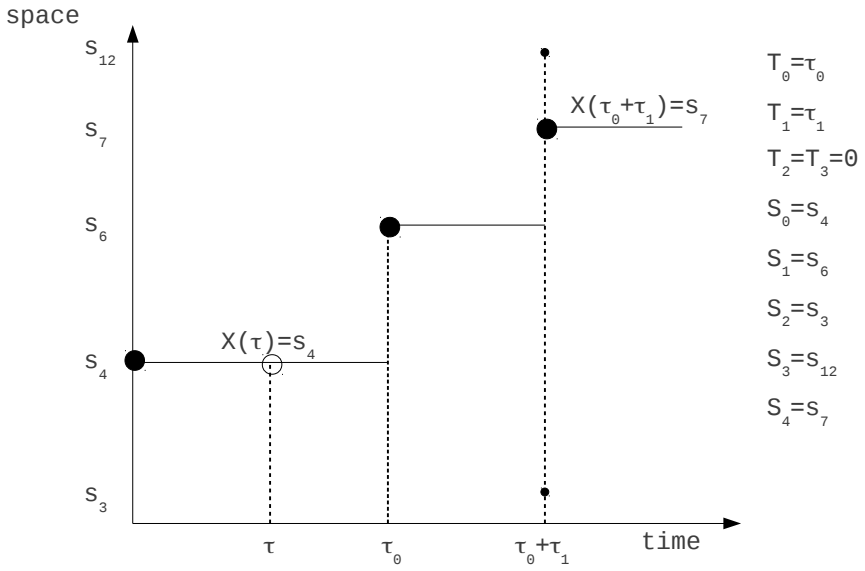

Figure 1.1: A sample of a stochastic process

some non null time interval. The observations $\{X(\tau)\}$ do not capture the vanishing states s_3 and s_{12} of this sample.

Performance evaluation of a DES is related to two kinds of analysis:

- The study of the transient behaviour, i.e. the computation of measures that depend on time elapsed since the system starts. Such a study is appropriate for systems presenting different stages and terminating systems. The two main application areas are dependability and safety analysis.
- The steady-state behaviour of the system. For numerous applications, the modeller is interested by the behaviour of the system after the intermediate stages, once it is stabilized.

Of course this requires that such a steady-state behaviour exists. Let us denote $\pi(\tau)$ the distribution of $X(\tau)$. Then the following equation expresses this requirement:

$$\lim_{\tau \to \infty} \pi(\tau) = \pi_\infty \tag{1.2}$$

where π_∞ must also be a distribution, called the *steady-state distribution*.

Transient and steady-state distributions allow to compute *performance indices*. For instance, the steady-state probability that a server is available, the probability that at time τ, a connexion is established and the mean number of clients for a service are such indices.

In order to reason in a generic way about DES, it is useful to specify functions whose domain is the set of states and whose range is $\mathbb{R}$. Then a function f can be viewed as a performance index and given some distribution π, the expression $\sum_{s \in S} \pi(s) \cdot f(s)$ represents the measure of this index.

When the range of the index is $\{0, 1\}$, it can be viewed as an *atomic proposition* satisfied by a state when the function is evaluated to 1. In the sequel, one denotes $\mathcal{P}$, the set of atomic propositions and $s \vDash \varphi$, with s a state and φ a proposition, the satisfaction of φ by s. With this notation, given some distribution π, the expression $\sum_{s \vDash \varphi} \pi(s)$ represents the measure of this index.

1.2 Renewal processes with arithmetic distribution [FEL 68]

1.2.1 The renewal theorem

The *renewal process* is a very simple case of DES: it has a single state and the time intervals between events are integers obtained by sampling i.i.d. random variables. *Renewal instants* are the instants corresponding to the occurrence of events.

In order to present renewal processes, let us discuss a simple example. Assume that a bulb is used by some lamp and that the bulb seller provides some probabilistic information about the quality of the bulb:

f_n, the probability that the bulb duration is n days

For instance $f_2 = 0.4$, $f_3 = 0.6$ and $f_n = 0$ for $n \notin \{2, 3\}$. The user wants to infer:

u_n, the probability that on day n the bulb must be changed

Continuing the example, $u_0 = 1$ since the bulb is bought on day 0, $u_1 = 0$, $u_2 = f_2 = 0.4$, $u_3 = f_3 = 0.6$. In order to determine u_4, one observes that the bulb must have been changed on the second day, so that $u_4 = u_2 f_2 = 0.16$. Similarly $u_5 = u_3 f_2 + u_2 f_3 = 0.48$.

Generalizing the reasoning about the first change (if any) before day n one obtains the following "convolution" formula:

$$u_n = u_0 f_n + \cdots + u_{n-1} f_1 \text{ when } n > 0 \tag{1.3}$$

Let us examine for the previous example different values of u_n when n increases:

$$u_{10} = 0.3696, u_{20} = 0.38867558, u_{30} = 0.38423714, u_{40} = 0.38463453, u_{50} = 0.38461546$$

Clearly (in this example) u_n seems to be a convergent sequence. So one is interested about the value of this limit. Let us develop an informal reasoning. We denote $\mu \stackrel{\text{def}}{=} \sum_{n\in\mathbb{N}} n f_n$, the expected value of the distribution $\{f_n\}$. Since the average interval between two renewal instants is μ, the average probability that n is a renewal instant is equal to μ^{-1}. In our example $\mu^{-1} = 0.38461538$. The main theorem of this section establishes this result and precises the required hypotheses. In the sequel, we note $\eta \stackrel{\text{def}}{=} \mu^{-1}$ with the convention that $\eta = 0$ when $\mu = \infty$.

Let us call $\rho_k \stackrel{\text{def}}{=} \sum_{i>k} f_i$ the probability that the duration before a new renewal instant is strictly greater than k. Observe that:

$$\mu = \sum_{i\in\mathbb{N}} i f_i = \sum_{i\in\mathbb{N}} \sum_{0\le k<i} f_i = \sum_{k\in\mathbb{N}} \sum_{i>k} f_i = \sum_{k\in\mathbb{N}} \rho_k$$

We now establish an equation fulfilled by the u_n's and the ρ_n's. Define the event L_{nk} (with $k \le n$) as:

The last renewal instant in $[0, n]$ is k.

In order for L_{nk} to be realized, k must be a renewal instant and the next renewal instant must be strictly greater than n. Using conditional probabilities, the probability of this event is exactly $u_k \rho_{n-k}$. Since (given n) exactly one such event must occur, one obtains:

$$\rho_0 u_n + \rho_1 u_{n-1} + \cdots + \rho_n u_0 = 1 \tag{1.4}$$

We are now searching about a (necessary and) sufficient condition for the existence of the limit. The next lemma establishes such a condition.

Lemma 1.1 *Let $\{u_n\}$ be any sequence that fulfills (1.4) and assume that $\limsup_{n\to\infty} u_n \le \eta$. Then $\lim_{n\to\infty} u_n$ exists and is equal to η.*

Proof

Which situations can falsify the hypothesis of the previous lemma? Let us choose a very simple distribution $f_2 = 1$ and $f_i = 0$ for $i \neq 2$. Then $\mu = 2$ implying $\eta = \frac{1}{2}$. However for all n, $u_{2n} = 1$

and $u_{2n+1} = 0$. In order to characterize such pathologic behaviours, we introduce the periodicity of a distribution.

Definition 1.2 *The* periodicity *of a distribution* $\{f_n\}$ *is defined by:* $\gcd(n \mid f_n > 0)$. *A distribution is* aperiodic *if its periodicity is 1.*

We are going to prove that aperiodicity is the single requirement in order to get the existence of the limit. To do so, we need the following characterization of aperiodicity.

Lemma 1.3 *Let* $a_1, \ldots, a_k$ *be natural integers whose* gcd *is 1. Then there exists* n_0 *such that:* $\forall n \geq n_0\ \exists \alpha_1, \ldots, \alpha_k \in \mathbb{N}\ n = a_1\alpha_1 + \ldots a_k\alpha_k$.

Proof

We also need this standard "diagonalization" lemma.

Lemma 1.4 *Let* $(x_{n,m})_{n,m\in\mathbb{N}}$ *be a bounded set of reals. Then there exists an infinite sequence of indices* $m_1 < m_2 < \cdots$ *such that for all* $n \in \mathbb{N}$ *the subsequence* $(x_{n,m_k})_{k\in\mathbb{N}}$ *is convergent.*

Proof

The next lemma is the key lemma and this is where aperiodicity is required.

Lemma 1.5 *Let* f *be an aperiodic distribution and* $(w_n)_{n\in\mathbb{N}}$ *such that for all* n, $w_n \leq w_0$ *and:*

$$w_n = \sum_{k=1}^{\infty} f_k w_{n+k} \tag{1.5}$$

Then for all n, $w_n = w_0$.

Proof

We are now in position to establish the renewal theorem.

Theorem 1.6 *Let* f *be an aperiodic distribution with a (non necessarily finite) mean* μ. *Then* $\lim_{n\to\infty} u_n = \mu^{-1}$.

Proof

1.2.2 Generalizations

We develop here several generalizations of the previous theorem.

First we show that the case of a periodic distribution can be straightforwardly reduced to the previous case. Indeed assume that distribution f has periodicity p. Then considering instant np as instant n, one recovers an aperiodic distribution. Formalizing this observation leads to a new version of the renewal theorem.

Theorem 1.7 *Let* f *be a distribution of period* p *with (non necessarily finite) mean* μ. *Then* $\lim_{n\to\infty} u_{np} = p\mu^{-1}$ *and for all* n *such that* $n \mod p \neq 0$, $u_n = 0$.

Let us consider that $\sum_{n\in\mathbb{N}} f_n \leq 1$ with the interpretation that $1 - \sum_{n\in\mathbb{N}} f_n$ is the probability that there will be no next renewal instant. In our example, this could be the case that the bulb is perfect. In order to analyze this renewal process, it is useful to introduce some generative series:

$$F(s) = \sum_{n\in\mathbb{N}} f_n s^n \qquad U(s) = \sum_{n\in\mathbb{N}} u_n s^n \text{ for } s \in [0,1]$$

When $s < 1$, Multiplying (1.3) by s^n and summing up one gets $U(s) - 1 = U(s)F(s)$ or equivalently $U(s) = \frac{1}{1-F(s)}$. Observe that $U(1)$, the mean number of renewal instants, can be either finite or infinite. The next proposition characterizes this situation.

Proposition 1.8 *The mean number of renewal instants,* $\sum_{n\in\mathbb{N}} u_n$, *is finite iff* $\sum_{n\in\mathbb{N}} f_n < 1$. *In this case,* $\sum_{n\in\mathbb{N}} u_n = \frac{1}{1-\sum_{n\in\mathbb{N}} f_n}$.

Proof
As an immediate consequence, when $\sum_{n\in\mathbb{N}} f_n < 1$, one has $\lim_{n\to\infty} u_n = 0$.

Let us continue our example. Assume that there is already a bulb with the lamp. This life duration of this specific bulb is not assumed to have the same distribution as the next ones and can be perfect. We denote $\{b_n\}$ this distribution and $B(s)$ its generative series. We say that this is a *delayed* process and we define v_n as the probability that n is a renewal instant of this process ($V(s)$ is its generative series).

Reasoning on the first renewal instant as before, one obtains the following equation:

$$v_n = b_n u_0 + b_{n-1} u_1 + \cdots + b_0 u_n$$

which can be rewritten as:

$$V(s) = B(s)U(s) = \frac{B(s)}{1 - F(s)}$$

The next theorem expresses the fact that only $B(1)$, the probability that there is another renewal instant after 0, has an impact on the result obtained for the original process.

Theorem 1.9

- *If* $\lim_{n\to\infty} u_n = \omega$ *then* $\lim_{n\to\infty} v_n = B(1)\omega$
- *If* $\sum_{n\to\infty} u_n = U(1)$ *is finite then* $\sum_{n\to\infty} v_n = B(1)U(1)$

Proof

We let the reader adapt the previous theorem for the case of a delayed periodic process.

We introduce the last generalization with our example. Assume that there is a non null probability $f_0 < 1$ that a new bulb is initially faulty. In this case, u_n can no more be considered as the probability that on day n, the bulb is changed but instead as the mean number of changes on day n.

However, there is a simple trick that allows to extend the previous theory. Suppose that n is a renewal instant, then the number of renewals at instant n follows a geometric law with parameter $1 - f_0$. Consider now a modified renewal process with $f'_0 = 0$ and $f'_i = \frac{f_i}{1-f_0}$. Let us denote u'_n the probability of a renewal instant at time n for this process. Then it is immediate that $u_n = \frac{u'_n}{1-f_0}$.

1.3 Discrete time Markov chains [KSK 76]

1.3.1 Presentation

A Discrete Time Markov Chain (DTMC) is a DES with the following features:

- The time interval between events T_n for $n \geq 1$ is the constant 1.
- The selection of the state that follows the current state only depends on that state and the transition probabilities remain constant[1] along the run:

$$\mathbf{Pr}(S_{n+1} = s_j \mid S_0 = s_{i_0}, ..., S_n = s_i) = \mathbf{Pr}(S_{n+1} = s_j \mid S_n = s_i) \stackrel{\text{def}}{=} p_{ij} \stackrel{\text{def}}{=} \mathbf{P}[i, j]$$

We indifferently use the two notations for transition probabilities. Depending on the context, we will consider that a DTMC has an initial distribution π_0 over its states. Observe that for all i, $\sum_j p_{ij} = 1$ and for all j, $p_{ij} \geq 0$. When a matrix fulfills these properties we say that it is a *transition* or a *stochastic* matrix.

The transition matrix $\mathbf{P}$ of a DTMC can be represented by a (possibly countable) oriented graph that we denote $G_{\mathbf{P}}$. It is defined as follows:

- The set of vertices is the set of the states of the DTMC;

[1] Sometimes these chains are called homogeneous DTMC.

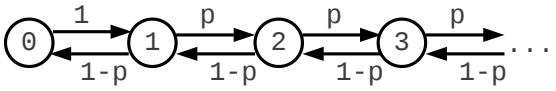

Figure 1.2: A random walk

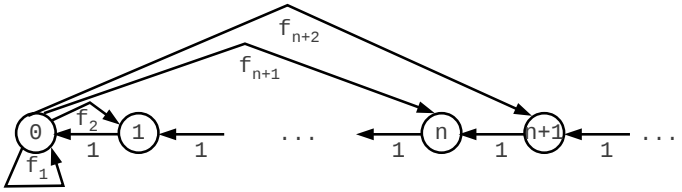

Figure 1.3: Simulating a renewal process

- There is an edge from s_i to s_j labelled by p_{ij} if $p_{ij} > 0$.

Example 1.10 (A random walk) *Figure 1.2 represents the graph of a* random walk. *The walker starts in position 0 of the path and at the next instant will go forward to position 1 with probability 1. In position $n > 0$ he goes forward to position $n+1$ with probability p or backward to position $n-1$ with probability $1-p$.*

Example 1.11 (A simulation of a renewal process) *Figure 1.3 shows that a renewal process can be seen as a particular case of* DTMC. *Being in state 0 corresponds to a renewal instant while state $n > 0$ means that the next renewal instant will occur in n time units. The single* probabilistic *state is state 0 where the selection of the time before the next renewal instant is done following the distribution $\{f_n\}$.*

Example 1.12 (A finite DTMC) *Figure 1.4 shows on the right the transition matrix and on the left its associated graph. Here there are only three states. While the graph representation could only appear to be a visual help, the structure of the graph provides useful information on the behaviour of a finite* DTMC.

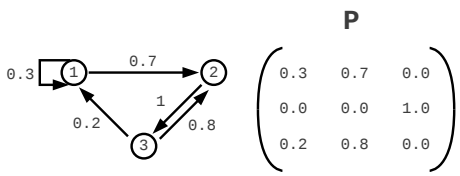

Figure 1.4: A finite DTMC

1.3.2 Transient and steady-state behaviour of a DTMC

Analyzing the transient behaviour of a DTMC does not raise any difficulty. The state changes occur at times $\{1, 2, \ldots\}$. Given an initial distribution π_0 and a transition matrix $\mathbf{P}$, one denotes π_n the distribution of S_n (i.e. the state of the chain at time n). It is given by:

$$\pi_n = \pi_0 \cdot \mathbf{P}^n$$

which can be established by an elementary recurrence.

Analyzing the asympotical behaviour requires some additional notations.

- For $n \in \mathbb{N}$, $p^n_{i,j}$ denotes the probability to reach in n steps state j from state i. These values are the items of matrix $\mathbf{P}^n$ (which is also a transition matrix).
- For $n \in \mathbb{N}$, $f^n_{i,j}$ denotes the probability to reach in n steps state j from state i *for the first time*. One denotes $f_{i,j} = \sum_{n\in\mathbb{N}} f^n_{i,j}$ the probability to reach j from i and $\mu_i = \sum_{n\in\mathbb{N}} n f^n_{i,i}$ the mean return time in i (only relevant if $f_{i,i} = 1$).

A first equation can be derived by a case decomposition w.r.t. the first time the chain reaches j from i:

$$p^n_{i,j} = \sum_{m=0}^{n} f^m_{i,j} p^{n-m}_{j,j}$$

With every state i, one associates a renewal process where the renewal instants correspond to the visits of i. One observes that $\{f^n_{i,i}\}_{n\in\mathbb{N}}$ is the distribution of the renewal time and that $p^n_{i,i}$ is the probability that n is a renewal instant. Furthermore with every pair of states (i, j), one also associates a delayed renewal process where again the renewal instants correspond to the visits of i. The distribution of the delay is given by $\{f^n_{j,i}\}_{n\in\mathbb{N}}$.

One classifies states depending on their associated renewal process.

- A state i is *transient* if $f_{i,i}$, the probability of a return is less than 1.
- A state is *null recurrent* if the probability of a return is 1 and μ_i, the mean time of a return is infinite.
- A state is *positive recurrent* if the probability of a return is 1 and the mean time of a return is finite.

In addition we say that a state is *periodic* (resp. *aperiodic*) if its associated renewal process is periodic (aperiodic). Finally a state is *ergodic* if it is positive recurrent and aperiodic.

By a straightforward application of results on renewal processes, one gets:

- A state i is transient iff $\sum_{n\in\mathbb{N}} p^n_{i,i} < \infty$. In this case, for every j, one has $\lim_{n\to\infty} p^n_{j,i} = 0$.
- A state i is null recurrent iff $\sum_{n\in\mathbb{N}} p^n_{i,i} = \infty$ and $\lim_{n\to\infty} p^n_{i,i} = 0$. In this case, for every j, one has $\lim_{n\to\infty} p^n_{j,i} = 0$.
- A state i is positive recurrent iff $\sum_{n\in\mathbb{N}} p^n_{i,i} = \infty$ and $\mu_i < \infty$. If in addition i is aperiodic, for every j, one has $\lim_{n\to\infty} p^n_{j,i} = f_{j,i}\mu_i^{-1}$.

We want to globally reason about the chain. In order to do so we introduce some properties of a DTMC.

Definition 1.13 *Let $\mathcal{C}$ be a* DTMC.

- *S', a subset of states of $\mathcal{C}$, is* closed *if for all $i \in S'$, one has $\sum_{j\in S'} p_{i,j} = 1$.*
- *Let i be a state of $\mathcal{C}$, its closure is defined by $Cl(i) = \{j \mid f_{i,j} > 0\}$.*

- $\mathcal{C}$ *is* irreducible *if for all pairs of states* i, j *one has* $f_{i,j} > 0$ *(or equivalently* $\sum_{n\in\mathbb{N}} p^n_{i,j} > 0$*).*

We let the reader prove that $Cl(i)$ is closed. A closed subset may be studied in isolation since it constitutes a DTMC. The importance of irreducibility is shown by the next theorems.

Theorem 1.14 *All states of an irreducible chain are of the same kind.*

Proof

So in the sequel, we say that an irreducible DTMC is transient (resp. null recurrent, positive recurrent, aperiodic) if its states are transient (resp. null recurrent, positive recurrent, aperiodic). Similarly the period of an irreducible DTMC is the common period of its states.

Theorem 1.15 *Let* i *be a recurrent state. Then* $Cl(i)$ *is irreducible and for all pair of states* $(j,k) \in Cl(i)$*, one has* $f_{j,k} = 1$*.*

Proof

Summarizing, a DTMC can be partitioned between the transient states say T and the irreducible subchains $C_1, C_2, \ldots$ Observe that an infinite DTMC may have only transient states as in the chain where the states are integers and $p_{i,i+1} = 1$. The periodicity is particularly relevant in case of irreducible chains.

Theorem 1.16 *Let* $\mathcal{C}$ *be an irreducible chain with periodicity* p*. Then* S*, the set of states, can be partitioned as:* $S = S_0 \uplus S_1 \uplus \ldots \uplus S_{p-1}$ *such that:*

$$\forall i \in S_k \ \forall j \in S \ p_{i,j} > 0 \Rightarrow j \in S_{(k+1) \mod p}$$

Furthermore p *is the greatest integer fulfilling this property.*

Proof

We now establish the main theorem of DTMC.

Theorem 1.17 *Let* $\mathcal{C}$ *be an irreducible* DTMC *whose states are aperiodic.*

Existence. *Assume that the states of* $\mathcal{C}$ *are positive recurrent.*
Then the limits $\lim_{n\to\infty} p^n_{j,i}$ *exist and are equal to* μ_i^{-1} *(so independent from* j*).*
Furthermore they fulfill $\sum_{i\in S} \mu_i^{-1} = 1$ *and for all* i*,* $\mu_i^{-1} = \sum_{j\in S} \mu_j^{-1} p_{j,i}$*.*

Unicity. *Conversely, assume there exist* $\{u_i\}$ *such that:*
For all i*,* $u_i \geq 0$*,* $\sum_{j\in S} u_j = 1$ *and* $u_i = \sum_{j\in S} u_j p_{j,i}$*.*
Then for all i*,* $u_i = \mu_i^{-1}$ *(which implies that states are ergodic).*

Proof

We are looking for a (more or less effective) characterization of recurrence for a state. Let S' be a subset of states, $\mathbf{P}$ restricted to states of S', denoted $\mathbf{P'}$, represents the behaviour of the chain along as it remains in S'. So $\mathbf{P'}^n[i,j]$ is the probability that at time n, the chain is in state j without ever leaving S', starting from state i. Observe that $\sum_{j\in S'} \mathbf{P'}^n[i,j]$ is the probability that at time n, the chain has never left S', starting from state i. One denotes this quantity by $pin^n_{S'}[i]$.

Proposition 1.18 *For all* i*,* $\lim_{n\to\infty} pin^n_{S'}[i]$ *exists. Denoting* $pin_{S'}[i]$ *this limit, then* $pin_{S'}$ *is the maximal solution of equation:*

$$\forall i \ x[i] = \sum_{j\in S'} \mathbf{P}[i,j]x[j] \wedge 0 \leq x[i] \leq 1 \tag{1.6}$$

Proof

Using this proposition, we get a characterization of the recurrence in the infinite case (the finite case is addressed by theorem 1.23).

Theorem 1.19 *Let $\mathcal{C}$ be an irreducible Markov chain whose state space is $\mathbb{N}$. Then 0 is recurrent iff the maximal solution of the equation:*

$$\forall i > 0 \; x[i] = \sum_{j>0} \mathbf{P}[i,j]x[j] \wedge 0 \leq x[i] \leq 1 \tag{1.7}$$

is the null vector. Otherwise stated, the null vector is the single solution of (1.7).

Proof

Finally we generalize the results of theorem 1.17 to irreducible chains whose states are (null or positive) recurrent. To this aim, we introduce the following probability: ${}_r p_{ij}^{(n)}$ represents the probability that starting from i one reaches j after n transitions without ever visiting r. We allow i to be equal to r and for the case $n = 0$, we set ${}_r p_{ij}^{(0)} \stackrel{\text{def}}{=} \mathbf{1}_{i=j}$.

One observes that ${}_r\pi_{ij}$ defined by ${}_r\pi_{ij} \stackrel{\text{def}}{=} \sum_{n\in\mathbb{N}} {}_r p_{ij}^{(n)}$ is the mean number of visits of j without visiting r. Since the chain is irreducible, the probability of a visit to r starting from j is positive. Let us consider the chain obtained by making r an *absorbing* state (i.e. $p_{rr} = 1$), all states different from r are transient. So the mean number of visits to these states is finite, i.e. ${}_r\pi_{ij} < \infty$.

Theorem 1.20 *Let $\mathcal{C}$ be an irreducible* DTMC *whose states are recurrent.*
• *Let r be an arbitrary state. Then vector $\mathbf{u}$ defined by $\mathbf{u}_i \stackrel{def}{=} {}_r\pi_{ri}$ fulfills:*

$$\mathbf{u} = \mathbf{u} \cdot \mathbf{P} \quad \textit{and for all } i \; \mathbf{u}_i > 0 \quad \textit{and} \quad \mathbf{u}_r = 1$$

• *Conversely, let $\mathbf{u} \neq 0$ such that $\mathbf{u} = \mathbf{u} \cdot \mathbf{P}$ and for all i, $\mathbf{u}_i \geq 0$. Then there exists λ such that for all i, $\mathbf{u}_i = \lambda \cdot {}_r\pi_{ri}$. Furthermore, the states of $\mathcal{C}$ are positive recurrent iff $\sum_{i\in S} \mathbf{u}_i < \infty$.*

Proof

Using the two previous theorems, we provide a useful characterization of the positive recurrence for irreducible DTMCs.

Proposition 1.21 *Let $\mathcal{C}$ be an irreducible* DTMC*, then the states of $\mathcal{C}$ are positive recurrent iff there exists $\mathbf{u}$ such that $\mathbf{u} = \mathbf{u} \cdot \mathbf{P}$ with for all i, $\mathbf{u}_i > 0$ and $\sum_{i\in S} \mathbf{u}_i < \infty$.*

Proof

The table below summarizes the characterization of the status of an irreducible DTMC.

Status	Characterization
Recurrent	$\mathbf{0}$ is the single solution of: $\forall i \in S \setminus s_0 \; \mathbf{u}_i = \sum_{j\in S\setminus\{s_0\}} p_{i,j}\mathbf{u}_j$ and $0 \leq \mathbf{u}_i \leq 1$ In this case given any $r \in S$, $\forall \mathbf{u} \geq 0 \; \mathbf{u} \cdot \mathbf{P} = \mathbf{u} \Leftrightarrow \exists \alpha > 0 \; \forall s \; \mathbf{u}[s] = \alpha \cdot {}_r\pi_{rs}$
Positive Recurrent	$\exists! \mathbf{u} > 0 \; \mathbf{u} \cdot \mathbf{P} = \mathbf{u} \wedge \sum_{i\in S} \mathbf{u}_i = 1$ *(*$\mathbf{u}$ *is the steady-state distribution when the* DTMC *is aperiodic)*
Period is p	$S = S_0 \uplus S_1 \uplus \ldots \uplus S_{p-1}$ with $\forall r < p \; \forall i \in S_r \; \forall j \in S \; p_{i,j} > 0 \Rightarrow j \in S_{r+1 \mod p}$ and p is the greatest integer fulfilling this property.

Let us apply the results we have obtained to the example 1.10 related to random walks. This chain is irreducible. So all states are of the same kind.

We first want to decide whether the states are recurrent. We apply theorem 1.19. The equation system is:

$$x_1 = px_2 \text{ and } \forall i \geq 2 \; x_i = px_{i+1} + (1-p)x_{i-1}$$

It can be rewritten as:

$$x_1 = px_2 \text{ and } \forall i \geq 2 \; x_{i+1} - x_i = \frac{1-p}{p}(x_i - x_{i-1})$$

If $x_1 = 0$ then $x_2 = 0$ and by induction $x_i = 0$ for all i. Let us consider that $x_1 > 0$. Using the first equation $x_2 - x_1 > 0$ and using the other equation for $i \geq 2$,

$$x_i = x_1 + (x_2 - x_1)\sum_{j=0}^{i-2}\left(\frac{1-p}{p}\right)^j = x_1\left(1 + \frac{1-p}{p}\sum_{j=0}^{i-2}\left(\frac{1-p}{p}\right)^j\right) = x_1\left(\sum_{j=0}^{i-1}\left(\frac{1-p}{p}\right)^j\right)$$

Thus if $p \leq \frac{1}{2}$ then the x_i's are unbounded implying that the single bounded solution is 0 and so the states of the chain are recurrent.

Otherwise the x_i's are bounded by $x_1\frac{p}{2p-1}$. So taking $x_1 = \frac{2p-1}{p}$ provides the non null maximal solution implying that the states of the chain are transient.

When $p \leq \frac{1}{2}$ we also want to know whether the states are null or positive recurrent. We apply the last assertion of theorem 1.20 to decide it. So we are looking for non null solution (unique up to a constant) of:

$$x_0 = (1-p)x_1 \text{ and } x_1 = (1-p)x_2 + x_0 \text{ and } \forall i \geq 2 \; x_i = (1-p)x_{i+1} + px_{i-1}$$

which can be rewritten as:

$$x_0 = (1-p)x_1 \text{ and } px_1 = (1-p)x_2 \text{ and } \forall i \geq 2 \; px_i + (1-p)x_i = (1-p)x_{i+1} + px_{i-1}$$

Substracting the second equation to the third one gets: $px_2 = (1-p)x_3$.
By induction, $\forall i \geq 1 \; px_i = (1-p)x_{i+1}$. So:

$$\forall i \geq 1 \; x_i = \left(\frac{p}{1-p}\right)^{i-1} x_1$$

When $p = \frac{1}{2}$, one gets for $i \geq 1$, $x_i = x_1$. So $\sum_{i\in\mathbb{N}} x_i = \infty$ and the chain is null recurrent.

When $p < \frac{1}{2}$, $\sum_{i\in S} x_i = x_1\left(1 - p + \sum_{i\in\mathbb{N}}\left(\frac{p}{1-p}\right)^i\right) = x_1(1 - p + \frac{1-p}{1-2p})$ is finite and the chain is positive recurrent. Since the chain has period 2 it is not ergodic.

1.4 Finite discrete time Markov chains [KS 60]

1.4.1 Graph analysis

In the general case, the interest of $G_{\mathbf{P}}$ is to provide a visual intuition of the behaviour of the chain. However for finite DTMC, studying this graph provides the classification of states. Let us recall that the vertices of a graph can be partitioned in *strongly connected components* (scc). This partition can be performed in linear time w.r.t. the size of the graph by the algorithm of Tarjan (see for instance [AHU 74]).

Definition 1.22 *A* scc S' *is a maximal subset of vertices such that for all* $i, j \in S'$ *there is a path from* i *to* j. *A* scc S' *is* terminal *if there is no path from* S' *to* $S \setminus S'$.

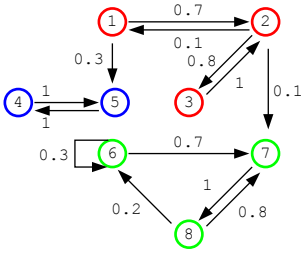

Figure 1.5: The scc of a DTMC

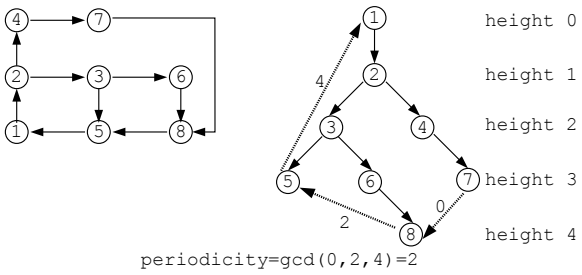

Figure 1.6: Computing the periodicity

In figure 1.5, we have highlighted the scc with colors. The blue and green scc's are terminal.

The next theorems show that $G_{\mathbf{P}}$ contains all the relevant information for classification of states.

Theorem 1.23 *In a finite* DTMC*, the transient states are the states of the non terminal* scc*'s. Every terminal* scc *is an irreducible* DTMC*. The states of terminal* scc*'s are positive recurrent.*

Proof

We now focus on periodicity in the irreducible case. So we assume that $G_{\mathbf{P}}$ is *strongly connected* (i.e. there is a single scc).

The periodicity is computed by algorithm 1 whose time complexity is linear w.r.t. the size of the graph. Let us describe its behavior. It builds an oriented tree covering the vertices by a breadth-first search. The algorithm also works with any search but here the breadth-first search is more efficient since it minimizes the height of the tree. With every discovered vertex the algorithm associates an height denoted $Height$. Every edge is labelled by an integer: the label of an edge (u, v) is defined by $Height[u] - Height[v] + 1$. So the edges of the tree have zero for label. The periodicity of the graph is then the gcd of the (non null) labels. This algorithm is illustrated in figure 1.6.

Proposition 1.24 *Algorithm 1 returns the periodicity of a (finite)* DTMC*.*

Proof

The next property on random paths of a finite DTMC is used in several contexts.

Algorithm 1: Computing the periodicity

`Periodicity`(G): an integer
Input: G, an oriented graph whose set of vertices is $\{1, \ldots, n\}$
Output: p, the periodicity of G
Data: i, j integers, $Height$ an array of size n, Q a queue

```
for i from 1 to n do Height[i] ← ∞
p ← 0; Height[0] ← 0; InsertQueue(Q, 0)
while not EmptyQueue(Q) do
    i ← ExtractQueue(Q)
    for (i, j) ∈ G do
        if Height[j] = ∞ then
            Height[j] ← Height[i] + 1
            InsertQueue(Q, j);
        else p ← gcd(p, Height[i] − Height[j] + 1)
    end
end
return p
```

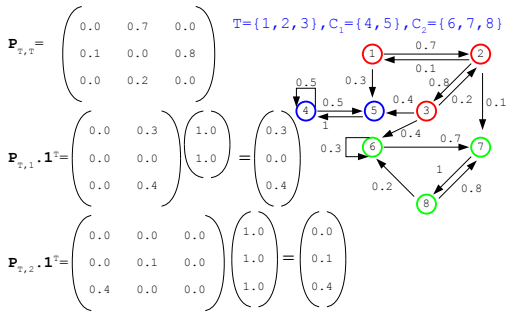

Figure 1.7: A non irreducible DTMC and some related matrices

Proposition 1.25 *Almost surely a random path ends up in a terminal* scc *and visits infinitely often all its states.*

Proof

1.4.2 Linear algebra analysis

In an ergodic chain, theorem 1.17 asserts that the steady-state distribution exists and is independent from the initial distribution. More precisely, this is the unique distribution π fulfilling the following equation:

$$\pi = \pi \cdot \mathbf{P} \tag{1.8}$$

In order to solve (1.8), a direct computation is possible (via a Gaussian elimination) enlarging the equation system with the normalisation equation $\pi \cdot \mathbf{1}^T = 1$ where $\mathbf{1}^T$ denotes the unit column vector. But iterative computations are more interesting if the state space is huge. The simplest consists in iterating $\pi \leftarrow \pi \cdot \mathbf{P}$ starting from an arbitrary distribution, (see [STE 94] for more elaborate computations).

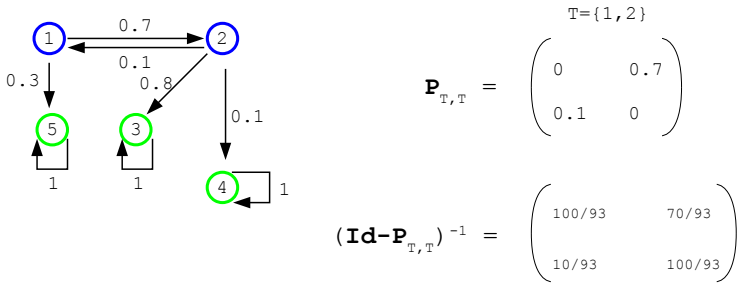

Figure 1.8: Transient behaviour of a non irreducible DTMC

Let us address a more general case. We introduce some notations to handle this issue.

Definition 1.26 *Let $\mathcal{C}$ be a finite* DTMC *whose irreducible subchains associated with terminal* scc *are denoted $\{\mathcal{C}_1, \ldots, \mathcal{C}_k\}$. For $1 \leq i \leq k$, the steady-state distribution (if it exists) of $\mathcal{C}_i$ is denoted π_i.*
The set of transient states are denoted T. $\mathbf{P}_{T,T}$ denotes the transition matrix restricted to T. For $1 \leq i \leq k$, $\mathbf{P}_{T,i}$ denotes the transition submatrix from states of T to states of $\mathcal{C}_i$.

Figure 1.7 illustrates the different matrices on some DTMC. Assuming that the terminal scc's are aperiodic, the chain also admits a steady-state distribution which here depends on the initial distribution.

Proposition 1.27 *Let $\mathcal{C}$ be a finite* DTMC *with initial distribution π_0 and whose irreducible subchains associated with terminal* scc*'s are aperiodic. Then (using the notations of definition 1.26), there exists a steady-state distribution whose expression is given by:*

$$\sum_{i=1}^{k} \left(\left(\pi_{0,i} + \pi_{0,T} \left(\mathbf{Id} - \mathbf{P}_{T,T} \right)^{-1} \cdot \mathbf{P}_{T,i} \right) \cdot \mathbf{1}^T \right) \pi_i$$

where $\pi_{0,i}$ (resp. $\pi_{0,T}$) is π_0 restricted to states of $\mathcal{C}_i$ (resp. T).
Thus $\mathbf{Id} - \mathbf{P}_{T,T}$ is invertible. Furthermore its inverse is equal to $(\sum_{n \geq 0} (\mathbf{P}_{T,T})^n)$.

Proof

Matrix $(\mathbf{Id} - \mathbf{P}_{T,T})^{-1}$ represents the mean number of visits between transient states. For instance, in the DTMC of figure 1.8 the mean number of visits of state 2 starting from state 1 is equal to $\frac{70}{93}$.

1.4.3 Convergence to the steady-state distribution

We want to analyze how fast the transient distribution of an ergodic DTMC converges to the steady-state distribution and we present two approaches.

We say that a matrix $\mathbf{M}$ (resp. a vector $\mathbf{v}$) is *positive* if for all i, j, $\mathbf{M}[i,j] > 0$ (resp. $\mathbf{v}[i] > 0$). We say that a matrix $\mathbf{M}$ (resp. a vector $\mathbf{v}$) is *non negative* if for all i, j, $\mathbf{M}[i,j] \geq 0$ (resp. $\mathbf{v}[i] \geq 0$). We say that a non negative square matrix $\mathbf{M}$ is *regular* if there exists some k such that $\mathbf{M}^k$ is positive.

Lemma 1.28 *Let $\mathbf{P}$ be the transition matrix of an ergodic* DTMC*. Then $\mathbf{P}$ is regular.*

Proof

Let π_∞ be the steady-state distribution w.r.t. transition matrix $\mathbf{P}$ of an ergodic DTMC. We introduce the steady-state square matrix Π_∞ where every row is a copy of π_∞. We let the reader prove that the following equalities hold:

- For every transition matrix $\mathbf{P}'$, $\mathbf{P}'\Pi_\infty = \Pi_\infty$;
- $\Pi_\infty \mathbf{P} = \Pi_\infty$.

The first approach is based on the steady-state matrix. The following proposition characterizes the magnitude order of the convergence rate to the steady-state distribution. In fact, the theorem is also valid if we take π_∞ as a solution of $\pi_\infty \mathbf{P} = \pi_\infty \wedge \pi_\infty \cdot \mathbf{1}^T = 1$ (there is a priori at least one solution). Thus it provides an alternative proof of the existence of a steady-state distribution for a finite irreducible DTMC.

Proposition 1.29 *Let $\mathcal{C}$ be an ergodic* DTMC *with $\mathbf{P}$ its transition matrix, π_n its distribution at time n and π_∞ its steady-state distribution. Then there exists some $0 < \lambda < 1$ such that:*

$$\|\pi_\infty - \pi_n\| = O(\lambda^n)$$

Proof

The second approach consists in studying the eigenvalues of matrix $\mathbf{P}$ and provides a more precise information on the convergence rate. It is based on the following proposition, a simplified version of the famous Perron-Frobenius theorem about non negative matrices. It also constitutes an alternative proof of the existence of a steady-state distribution for a finite irreducible DTMC.

Let us recall that given $\mathbf{M}$, a matrix, and λ, a complex value:

- λ is an *eigenvalue* of $\mathbf{M}$ if there exists a non null vector $\mathbf{v}$ such that $(\mathbf{M} - \lambda\mathbf{Id})\mathbf{v} = 0$.
- Such a vector is called an *eigenvector* associated with λ. A vector $\mathbf{v}$ is a *generalized eigenvector* w.r.t. λ if there exists some k such that $(\mathbf{M} - \lambda\mathbf{Id})^k\mathbf{v} = 0$.
- Given $\{\lambda_1, \ldots, \lambda_m\}$ the set of eigenvalues of $\mathbf{M}$, the vector space is the direct sum of $E_1, \ldots, E_m$ where E_i is the set of generalized eigenvectors w.r.t. λ_i called the *generalized eigenspace* of λ_i.

Proposition 1.30 *Let $\mathbf{M}$ be a non negative regular matrix. Then there exists λ such that:*

- *λ is an eigenvalue of $\mathbf{M}$ whose generalized eigenspace has dimension 1 and is generated by a positive vector.*
- *Every other eigenvalue λ' fulfills $|\lambda'| < \lambda$.*

Proof

Proposition 1.31 *Let $\mathcal{C}$ be an ergodic* DTMC *with $\mathbf{P}$ its transition matrix, π_n its distribution at time n and π_∞ its steady-state distribution. Then:*

$$\|\pi_\infty - \pi_n\| = O(n^{s-2}\lambda^n)$$

where s is the number of states and $\lambda < 1$ is the second largest module of eigenvalues of $\mathbf{P}$.

Proof

1.5 Proofs

1.5.1 Proofs of section 1.2.1

Proof of lemma 1.1
If $\mu = \infty$ then $\eta = 0$, and we are done.

Let us assume that μ is finite. We pick an arbitrary subsequence $u_{n_1}, u_{n_2}, \ldots$, converging toward η' ($\leq \eta$ by hypothesis). Let us select some integer $r > 0$ and some $\varepsilon > 0$.

The hypotheses ensure that there exists m such that:

$$\forall n_i \geq m \ |u_{n_i} - \eta'| \leq \varepsilon \wedge \forall 1 \leq r' \leq r \ u_{n_i - r'} - \eta \leq \varepsilon$$

We recall (1.4) for n_i:

$$\rho_0 u_{n_i} + \rho_1 u_{n_i - 1} + \cdots + \rho_{n_i} u_0 = 1$$

So:

$$\rho_0(\eta' + \varepsilon) + (\eta + \varepsilon) \sum_{r'=1}^{r} \rho_{r'} + \sum_{r'>r} \rho_{r'} \geq 1$$

When ε goes to 0, the inequation becomes:

$$\rho_0 \eta' + \eta \sum_{r'=1}^{r} \rho_{r'} + \sum_{r'>r} \rho_{r'} \geq 1$$

When r goes to ∞ (let us recall that $\rho_0 = 1$), the inequation becomes:

$$\eta' + \eta(\mu - 1) \geq 1$$

which can be rewritten as:

$$\eta' - \eta \geq 0$$

Using $\eta' \leq \eta$, one deduces that $\eta' = \eta$. Since the subsequence is arbitrary, $\lim_{n\to\infty} u_n = \eta$.

q.e.d. (lemma 1.1) ◊◊◊

Proof of lemma 1.3
Using Euclid algorithm, one obtains $y_1, \ldots, y_k \in \mathbb{Z}$ such that:

$$1 = a_1 y_1 + \cdots + a_k y_k$$

Let us note $s = a_1 + \cdots + a_k$ and $x = \sup_i |y_i|(s-1)$.
Let $n \geq xs$. One performs the Euclidian division of n by s:

$$n = qs + r = \sum_{i=1}^{k} (q + r y_i) a_i$$

and $q + r y_i$ is non negative using the hypotheses.

q.e.d. (lemma 1.3) ◊◊◊

Proof of lemma 1.4
Since the $x_{n,m}$'s are bounded, one extracts a sequence of indices $(m_k^0)_{k\in\mathbb{N}}$ such that $(x_{0,m_k^0})_{k\in\mathbb{N}}$ is convergent.

Assume that after n stages, one has extracted a sequence of indices $(m_k^n)_{k\in\mathbb{N}}$. Then one extracts from this sequence a subsequence $(m_k^{n+1})_{k\in\mathbb{N}}$ such that $(x_{n+1,m_k^{n+1}})_{k\in\mathbb{N}}$ is convergent.

Let us now consider the sequence of indices $m_k = m_k^k$. Let $n \in \mathbb{N}$, starting from the nth item, the sequence $(x_{n,m_k})_{k\in\mathbb{N}}$ is a subsequence of $(x_{n,m_k^n})_{k\in\mathbb{N}}$ and thus it is convergent.

q.e.d. (lemma 1.4) ◊◊◊

Proof of lemma 1.5
Let us note $A = \{k \mid f_k > 0\}$.
$w_0 = \sum_{k=1}^{\infty} f_k w_k \leq w_0 \sum_{k=1}^{\infty} f_k = w_0$
In order to have equality, it is necessary that $w_k = w_0$ for all $k \in A$.
Let $k \in A$,
$w_0 = w_k = \sum_{k'=1}^{\infty} f_{k'} w_{k+k'} \leq w_0 \sum_{k'=1}^{\infty} f_{k'} = w_0$
So, it is also necessary that $w_{k+k'} = w_0$ for all $k, k' \in A$.
Iterating the process, one gets $w_k = w_0$ for all k, positive linear combination of items of A.
f is aperiodic. Using lemma 1.3, there exists n_0 such that for all $n \geq n_0$, $w_n = w_0$.
So:
$w_{n_0-1} = \sum_{k=1}^{\infty} f_k w_{n_0-1+k} = w_0 \sum_{k=1}^{\infty} f_k = w_0$
Iterating this process, one concludes.

q.e.d. (lemma 1.5) ◊◊◊

Proof of theorem 1.6
Let us note $\nu = \limsup_{n\to\infty} u_n$ (between 0 and 1).
Let $(r_m)_{m\in\mathbb{N}}$ be a sequence of indices such that $\nu = \lim_{m\to\infty} u_{r_m}$. With every integer n, one associates the sequence $(u_{n,m})_{m\in\mathbb{N}}$ defined by $u_{n,m} = u_{r_m-n}$ if $n \leq r_m$ and $u_{n,m} = 0$ otherwise.
Using lemma 1.4, there exists a sequence $m_1, m_2, \ldots$ such that for all n, $(u_{n,m_k})_{k\in\mathbb{N}}$ converges to a limit denoted w_n (consistently with the notations of lemma 1.5). From definition of ν, one gets $0 \leq w_n \leq \nu$ and $w_0 = \nu$. Equation (1.3) can be rewritten as:

$$u_{n,m_k} = \sum_{i=1}^{\infty} f_i u_{n+i,m_k}$$

This equality still holds at the limit since the u_n's are bounded and $\sum_{i=1}^{\infty} f_i = 1$ hence finite. One obtains the hypotheses of lemma 1.5. Hence, for all n, $w_n = \nu$.
We now prove that $\nu \leq \eta$ in order to conclude by application of lemma 1.1.
Using (1.4), one establishes that:

$$\rho_0 u_{0,m_k} + \rho_1 u_{1,m_k} + \cdots + \rho_{r_{m_k}} u_{r_{m_k},m_k} = 1 \tag{1.9}$$

Let us use (1.9). For all fixed r,

- $\rho_0 u_{0,m_k} + \rho_1 u_{1,m_k} + \cdots + \rho_r u_{r,m_k} \leq 1$;
- $\rho_0 u_{0,m_k} + \rho_1 u_{1,m_k} + \cdots + \rho_r u_{r,m_k}$ goes to $\nu \sum_{r'=0}^{r} \rho_{r'}$ when k goes to infinity.

Hence if $\mu = \infty$ then $\nu = 0$ establishing the result.
Otherwise $\nu \sum_{r'=0}^{r} \rho_{r'} \leq 1$ for all r. Hence $\nu\mu \leq 1$.

q.e.d. (theorem 1.6) ◊◊◊

1.5.2 Proofs of section 1.2.2

Proof of proposition 1.8
Due to the non negativity of u_n and f_n,
$\lim_{s\uparrow 1} U(s) = U(1)$ (with $U(1)$ finite or infinite) and $\lim_{s\uparrow 1} F(s) = F(1)$.
Let us suppose $F(1) < 1$. Then considering the limit when $s \to 1$ of equality $U(s) = \frac{1}{1-F(s)}$,
one obtains $U(1) = \frac{1}{1-F(1)}$ as required by the proposition.
Let us suppose $U(1) < \infty$. Then considering the limit when $s \to 1$ of equality $U(s)-1 = U(s)F(s)$,
one obtains $U(1) - 1 = U(1)F(1)$. Hence $F(1) = \frac{U(1)-1}{U(1)} < 1$.

q.e.d. (proposition 1.8) ◊◊◊

Proof of theorem 1.9
Let us denote $r_k = \sum_{k'>k} b_{k'}$. Using the definition of v_n for $n > k$ one has:

$$\sum_{k'=0}^{k} b_{k'} u_{n-k'} \leq v_n \leq \sum_{k'=0}^{k} b_{k'} u_{n-k'} + r_k$$

Let $\varepsilon > 0$. Then there exists k such that $r_k \leq \varepsilon$ and there exists n_0 such that for $n \geq n_0$ one has $|\omega - u_n| \leq \varepsilon$. Consequently for all $n > n_0 + k$

$$\omega B(1) - 2\varepsilon \leq (\omega-\varepsilon)(B(1)-\varepsilon) \leq \sum_{k'=0}^{k} b_{k'} u_{n-k'} \leq v_n \leq \sum_{k'=0}^{k} b_{k'} u_{n-k'} + r_k \leq (\omega+\varepsilon)B(1)+\varepsilon \leq \omega B(1) + 2\varepsilon$$

which establishes the first assertion.

The second assertion is straightforward since $V(s) = B(s)U(s)$.

q.e.d. (theorem 1.9) ◊◊◊

1.5.3 Proofs of section 1.3.2

Proof of theorem 1.14
Let i, j be two states of the chain, there exist r and s such that $p^r_{i,j} > 0$ and $p^s_{j,i} > 0$. Furthermore,

$$p^{n+r+s}_{i,i} \geq p^r_{i,j} p^n_{j,j} p^s_{j,i} \tag{1.10}$$

So if $\sum_{n\in\mathbb{N}} p^n_{i,i}$ is finite then $\sum_{n\in\mathbb{N}} p^n_{j,j}$ is finite. If $\lim_{n\to\infty} p^n_{i,i} = 0$ then $\lim_{n\to\infty} p^n_{j,j} = 0$. Since the situation is symmetric, one deduces that i is transient (resp. null recurrent, positive recurrent) iff j is transient (resp. null recurrent, positive recurrent).

Let us examine periodicity. Assume i has period $t \geq 1$. Using (1.10), with $n = 0$, one deduces that $r + s$ is a multiple of t. So if n is not a multiple of t, then $p^n_{j,j} = 0$ which means that the period of j is a multiple of t. Since the situation is symmetric, one deduces that i and j have the same period.

q.e.d. (theorem 1.14) ◊◊◊

Proof of theorem 1.15
Let $j \in Cl(i)$. By definition, $f_{i,j} > 0$. Moreover $1 - f_{i,i} \geq f_{i,j}(1 - f_{j,i})$. Since i is recurrent one has $f_{i,i} = 1$ which implies $f_{j,i} = 1$.

Let $k \in Cl(i)$, $f_{j,k} \geq f_{j,i} f_{i,k} > 0$. So $Cl(i)$ is irreducible. Since $Cl(i)$ is irreducible, all states are recurrent and "replaying" the beginning of the proof with k instead of i establishes $f_{j,k} = 1$.

q.e.d. (theorem 1.15) ◊◊◊

Proof of theorem 1.16
Let us pick some arbitrary state i and define S_k, with $0 \leq k < p$ as the subset of states j for which there is a path from i to j with length equal to $lp + k$ for some l. Since the chain is irreducible, these subsets cover S. By definition of S_k, $\forall i \in S_k \; \forall j \in S \; p_{i,j} > 0 \Rightarrow j \in S_{(k+1) \mod p}$. It remains to prove that their intersection is empty.

Assume there exists $j \in S_k \cap S_{k'}$ with $k \neq k'$. Since from j the chain reaches i, there exists at least a path from i to i whose length is not a multiple of p, leading to a contradiction.

Assume now there is some $p' > p$ fulfilling this property. Then the periodicity of the renewal process associated with i is a multiple of p' leading to another contradiction.

q.e.d. (theorem 1.16) ◊◊◊

Proof of theorem 1.17
Using theorem 1.9 on delayed renewal process:
$\lim_{n\to\infty} p^n_{j,i} = f_{j,i}\mu_i^{-1} = \mu_i^{-1}$ since $\mathcal{C}$ is irreducible.
We assume that S is countable and identify it to $\mathbb{N}$ (the case of a finite S is even simpler and left to the reader). One has:

$$p^{n+1}_{k,i} = \sum_{j\in S} p^n_{k,j}p_{j,i} \geq \sum_{m\leq M} p^n_{k,m}p_{m,i}$$

Consequently, letting n goes to infinity: $\mu_i^{-1} \geq \sum_{m\leq M} \mu_m^{-1}p_{m,i}$
and letting M goes to infinity: $\mu_i^{-1} \geq \sum_{j\in S} \mu_j^{-1}p_{j,i}$
We now prove that $\sum_{j\in S} \mu_j^{-1}$ is finite.

$$1 = \sum_{j\in S} p^n_{i,j}$$

Consequently, letting n goes to infinity: $1 \geq \sum_{m\leq M} \mu_m^{-1}$
and letting M goes to infinity: $1 \geq \sum_{j\in S} \mu_j^{-1}$
Summing $\mu_i^{-1} \geq \sum_{j\in S} \mu_j^{-1}p_{j,i}$ over i, one obtains:

$$\sum_{i\in S} \mu_i^{-1} \geq \sum_{i\in S}\sum_{j\in S} \mu_j^{-1}p_{j,i} = \sum_{j\in S} \mu_j^{-1}\sum_{i\in S} p_{j,i} = \sum_{j\in S} \mu_j^{-1}$$

One has equality of sums, so also equality of terms: $\mu_i^{-1} = \sum_{j\in S} \mu_j^{-1}p_{j,i}$
By iteration: $\mu_i^{-1} = \sum_{j\in S} \mu_j^{-1}p^n_{j,i}$
This equality holds at the limit since $\sum_{j\in S} \mu_j^{-1}$ is finite and $p^n_{j,i}$ is bounded by 1:

$$\mu_i^{-1} = \sum_{j\in S} \mu_j^{-1}\mu_i^{-1}$$

Since $\mu_i^{-1} > 0$, $\sum_{j\in S} \mu_j^{-1} = 1$
Conversely, assume there exist $\{u_i\}$ such that $u_i \geq 0$, $\sum_{i\in S} u_i = 1$ and for all i, $u_i = \sum_{j\in S} u_j p_{j,i}$.
Let us pick some $u_i > 0$, by iteration of the last equation: $u_i = \sum_{j\in S} u_j p^n_{j,i}$
Since the states of the chain are aperiodic, $\lim_{n\to\infty} p^n_{j,i}$ exists for all j. Moreover the equality holds at the limit since $\sum_{i\in S} u_i = 1$ and $p^n_{j,i} \leq 1$. If for all j, $p^n_{j,i}$ goes to 0 then, using the equality, one obtains $u_i = 0$ leading to a contradiction.

So i is positive recurrent and then ergodic (since aperiodic). So all states are ergodic implying that the limits of $p^n_{j,i}$ are μ_i^{-1}. The (limit) equation can be written as:

$$u_i = \sum_{j\in S} u_j\mu_i^{-1} = \mu_i^{-1}$$

q.e.d. (theorem 1.17) ◊◊◊

Proof of proposition 1.18
By definition the sequence $pin^n_{S'}[i]$ is decreasing and it is lower bounded by 0 so it is convergent.
$\sum_{j\in S'} \mathbf{P}'^{n+1}[i,j] = \sum_{j,j'\in S'} \mathbf{P}[i,j']\mathbf{P}'^n[j',j]$
So:
$pin^{n+1}_{S'}[i] = \sum_{j'\in S'} \mathbf{P}[i,j']pin^n_{S'}[j']$
This equality holds at the limit since $\sum_{j'\in S'} \mathbf{P}[i,j'] \leq 1$ and $pin^n_{S'}[i]$ are bounded by 1.
Substituting j' by j:
$pin_{S'}[i] = \sum_{j\in S'} \mathbf{P}[i,j]pin_{S'}[j]$
Otherwise stated, the $pin_{S'}[i]$'s are a solution of (1.6).

Let x be a solution of this equation. One has $\forall i\ x[i] \leq pin^0_{S'}[i] = 1$.
We prove by induction that the equality holds for all n.
$x[i] = \sum_{j\in S'} \mathbf{P}[i,j]x[j] \leq \sum_{j\in S'} \mathbf{P}[i,j]pin^n_{S'}[i] = pin^{n+1}_{S'}[i]$
Letting n goes to infinity, the maximality is obtained.

q.e.d. (proposition 1.18) $\Diamond\Diamond\Diamond$

Proof of theorem 1.19
Let x denote the maximal solution of (1.7). Then $x[i]$ represents the probability to not return to 0 starting from i.

If state 0 is recurrent the return probability is 1. Since there exists a non null probability to reach from 0 state i (i being arbitrary), $x[i]$ must be null.

If the maximal solution is null then the probability to stay in states of $\mathbb{N}^*$ is null whatever the initial state. So the probability to return in 0 is equal to 1.

q.e.d. (theorem 1.19) $\Diamond\Diamond\Diamond$

Proof of theorem 1.20
Let $\mathbf{u}$ defined by $\mathbf{u}_i \equiv {}_r\pi_{ri}$. One has ${}_rp^{(0)}_{rr} = 1$ and ${}_rp^{(n)}_{rr} = 0$ for $n > 0$. So $\mathbf{u}_r = 1$. Since the chain is irreducible, there exists a non null probability to reach an arbitrary state i from r implying $\mathbf{u}_i > 0$.
Using definition of ${}_rp^{(n)}_{ij}$, one obtains for $i \neq r$: ${}_rp^{(n+1)}_{ri} = \sum_{j\in\mathbb{N}} {}_rp^{(n)}_{rj} p_{ji}$.
Summing over n, one obtains:
$\sum_{n\geq 1} {}_rp^{(n)}_{ri} = \sum_{n\geq 0} \sum_{j\in\mathbb{N}} {}_rp^{(n)}_{rj} p_{ji}$
Since ${}_rp^{(0)}_{ri} = 0$, one deduces: ${}_r\pi_{ri} = \sum_{j\in\mathbb{N}} {}_r\pi_{rj} p_{ji}$.

In case $i = r$, one observes that $\sum_{j\in\mathbb{N}} {}_rp^{(n)}_{rj} p_{jr}$ is the probability of a first return to r at the $n+1^{th}$ transition. Summing over n, one deduces that $\sum_{j\in\mathbb{N}} {}_r\pi_{rj} p_{jr}$ is the probability of a return to r. Since r is recurrent, this quantity is equal to 1. This finally proves that $\mathbf{u} = \mathbf{u}\cdot\mathbf{P}$.

Assume now that there exists $\mathbf{u}$ such that $\mathbf{u} = \mathbf{u}\cdot\mathbf{P}$ and for all i, $\mathbf{u}_i \geq 0$. Using the first equation if $\mathbf{u}_i = 0$ for i arbitrary then $\mathbf{u}_j = 0$ for all j such that $p_{ji} > 0$. By induction, one deduces that $\mathbf{u}_j = 0$ for all j that allows to reach k. Since the chain is irreducible, either $\mathbf{u}$ is null, or all its components are strictly positive.

In case $\mathbf{u}$ is strictly positive, one assumes w.l.o.g. (by applying a multiplicative factor) that $\mathbf{u}_r = 1$. Hence for $i \neq r$,
$\mathbf{u}_i = p_{ri} + \sum_{j\neq r} \mathbf{u}_j p_{ji} = p_{ri} + \sum_{j\neq r} \left(p_{rj} + \sum_{k\neq r} \mathbf{u}_k p_{kj} \right) p_{ji} = p_{ri} + {}_rp^{(2)}_{ri} + \sum_{k\neq r} \mathbf{u}_k \cdot {}_rp^{(2)}_{ki}$
Proceeding by induction, one obtains:
$\mathbf{u}_i = p_{ri} + {}_rp^{(2)}_{ri} + \cdots + {}_rp^{(n)}_{ri} + \sum_{j\neq r} \mathbf{u}_j \cdot {}_rp^{(n)}_{ji}$
Letting n go to infinity, one deduces that $\mathbf{u}_i \geq {}_r\pi_{ri}$. So $\{\mathbf{u}_i - {}_r\pi_{ri}\}$ is also a non negative solution of the equation system. Since it is null for component r, it is null for all components so that $\mathbf{u}_i = {}_r\pi_{ri}$ for all i.

Let us compute the mean return time to r. The probability that the return time is greater or equal than n (for $n \geq 1$) is equal to $\sum_{j\in S} {}_rp^{(n-1)}_{rj}$. Thus the mean return time is: $\sum_{n\geq 1} \sum_{j\in S} {}_rp^{(n-1)}_{rj} = \sum_{j\in S} {}_r\pi_{rj}$ which proves the last assertion of the theorem.

q.e.d. (theorem 1.20) $\Diamond\Diamond\Diamond$

Proof of proposition 1.21
Assume that the chain is positive recurrent. Then theorem 1.20 allows to conclude.

Assume there exists $\mathbf{u}$ such that $\mathbf{u} = \mathbf{u}\cdot\mathbf{P}$ (and thus $\mathbf{u} = \mathbf{u}\cdot\mathbf{P}^k$ for any k), for all i, $\mathbf{u}_i > 0$ and $\sum_{i\in S} \mathbf{u}_i$ is finite.
Let us suppose that there exists x, a non null solution of:

$$\forall i > 0\ x[i] = \sum_{j>0} \mathbf{P}[i,j]x[j] \wedge 0 \leq x[i] \leq 1$$

and by iteration

$$\forall k\ \forall i > 0\ x[i] = \sum_{j>0} \mathbf{P}^k[i,j]x[j] \wedge 0 \leq x[i] \leq 1$$

By hypothesis, there is some i_0 with $x[i_0] > 0$ and some k such that $\mathbf{P}^k[0, i_0] > 0$.

By a weighted sum of the previous equalities one gets:
$\sum_{i>0} \mathbf{u}[i]x[i] = \sum_{i>0} \mathbf{u}[i] \left(\sum_{j>0} \mathbf{P}^k[i,j]x[j]\right)$
Thus:
$\sum_{i>0} \mathbf{u}[i]x[i] = \sum_{j>0} \left(\sum_{i>0} \mathbf{u}[i]\mathbf{P}^k[i,j]\right) x[j] = \sum_{j>0}(1 - \mathbf{u}[0]\mathbf{P}^k[0,j])\mathbf{u}[j]x[j] \leq \sum_{i>0} \mathbf{u}[j]x[j]$
Since the sum is finite all terms must be equal but $(1 - \mathbf{u}[0]\mathbf{P}^k[0, i_0])\mathbf{u}[i_0]x[i_0] < \mathbf{u}[i_0]x[i_0]$. Thus applying theorem 1.19, we deduce that the chain is recurrent and by an application of theorem 1.20 that it is positive recurrent.

q.e.d. (proposition 1.21) ◊◊◊

1.5.4 Proofs of section 1.4.1

Proof of theorem 1.23
Let i belonging to a non terminal scc. Then there is a path from i to j belonging to a terminal scc. The probability to follow this path is non null and it is a lower bound to the probability of non returning to i. So i is transient.

Let S' be a terminal scc. Since there is no way to exit S', S' constitutes a DTMC. It is irreducible by definition of scc.

Let $\mathbf{P}$ be its transition matrix. Pick some state $i \in S'$ and observe that for all n, $\sum_{j \in S'} p^n_{i,j} = 1$. Since this is a finite sum, there exists at least some j such that $p^n_{i,j}$ does not converge to 0 when n goes to infinity. This implies that j is positive recurrent. Since all states of S' are of the same kind, they are all positive recurrent.

q.e.d. (theorem 1.23) ◊◊◊

Proof of proposition 1.24
Let p be the periodicity of the graph, p' the gcd of the edge labels and r the root of the tree.

Given two paths with same source and destination, the difference between the lengths of these paths must be a multiple of p, using the same argument applied in the proof of theorem 1.16.

- Let (u, v) be an edge with non null label. Let us denote σ_u, the path from r to u along the tree and σ_v the path from r to v along the tree. The length of σ_u is $Height[u]$, the one of σ_v is $Height[v]$. With the edge (u, v), one obtains another path $\sigma_u(u, v)$ from r to v. The difference between the lengths of the two paths is: $Height[u] - Height[v] + 1$ which must be a multiple of p. Since (u, v) is arbitrary, one deduces that $p|p'$.
- Let us partition the states in $S'_0, \ldots, S'_{p'-1}$ with $s \in S'_i$ iff $Height[s] \mod p' = i$. By construction, an edge of the tree joins a vertex of S'_i to a vertex of $S'_{i+1 \mod p'}$. An edge (u, v) out of the tree joins $u \in S'_{Height[u] \mod p'}$ to $v \in S'_{Height[v] \mod p'}$ but $Height[u] - Height[v] + 1 \mod p' = 0$. So, $S'_{Height[v] \mod p'} = S'_{Height[u]+1 \mod p'}$. By theorem 1.16, $p' \leq p$. Since $p|p'$, one obtains $p = p'$.

q.e.d. (proposition 1.24) ◊◊◊

Proof of proposition 1.25
Every infinite path ends up in a scc.

Let us fix some non terminal scc, S'. There is an edge (u, v) with $u \in S'$ and $v \notin S'$. So given any state $w \in S'$, there is a path from w to v with length say l_w and probability say $\varepsilon_w > 0$. Define $l = \max(l_w)$ and $\varepsilon = \min(\varepsilon_w)$. We partition the paths ending in S' depending on their first entry

in S'. After the entry, we split the suffix of the path in segments of length l. There is a probability at least ε to leave S' during a segment. Thus the probability to remain in S' after n segments is at most $(1-\varepsilon)^n$ and goes to 0 when n goes to infinity. This proves that with probability 1 a random path ends up in a terminal scc.

Let us fix some terminal scc, S'. When entering S' the probability to visit once every state is 1 (since states are recurrent). By induction, the probability to visit every state at least n times is 1. Letting n go to infinity yields the result.

q.e.d. (proposition 1.25) ◊◊◊

1.5.5 Proofs of section 1.4.2

Proof of proposition 1.27
The distribution is given by $\pi = \sum_{i=1}^{k} \Pr(\text{to reach } \mathcal{C}_i) \cdot \pi_i$. It remains to compute the probability to reach a terminal scc. One evaluates this quantity starting from an arbitrary state and then one "unconditions" it w.r.t. the initial distribution:

$$\Pr(\text{to reach } \mathcal{C}_i) = \sum_{s \in S} \pi_0(s) \cdot \pi'_{\mathcal{C}_i}(s)$$

where $\pi'_{\mathcal{C}_i}(s) = \Pr(\text{to reach } \mathcal{C}_i \mid S_0 = s)$.

When state s belongs to $\mathcal{C}_i$, then $\pi'_{\mathcal{C}_i}(s) = 1$ and $\pi'_{\mathcal{C}_j}(s) = 0$ for $j \neq i$.

Let us consider the transient states. The probability for a transient state s to reach $\mathcal{C}_i$ can be decomposed w.r.t. to the length of the path from s to $\mathcal{C}_i$ along the transient states. Let $n+1$ be the length of such paths. Then the probability of these paths is exactly $\left((\mathbf{P}_{T,T})^n \cdot \mathbf{P}_{T,i} \cdot \mathbf{1}^T\right)[s]$. Summing over n gives the desired formula once we have proven that:

$$\sum_{n \geq 0} (\mathbf{P}_{T,T})^n = (\mathbf{Id} - \mathbf{P}_{T,T})^{-1}$$

We observe that $(\sum_{n \geq 0} (\mathbf{P}_{T,T})^n)[i,j]$ corresponds to the mean number of visits of j starting from i and is finite due to the definition of transient states. So the sum is convergent. Now for every n_0:

$$(\sum_{n \leq n_0} (\mathbf{P}_{T,T})^n)(\mathbf{Id} - \mathbf{P}_{T,T}) = \mathbf{Id} - (\mathbf{P}_{T,T})^{n_0+1}$$

Since $\lim_{n \to \infty} (\mathbf{P}_{T,T})^n = 0$ letting n_0 go to infinity establishes the result.

q.e.d. (proposition 1.27) ◊◊◊

1.5.6 Proofs of section 1.4.3

Proof of lemma 1.28
We denote $s \stackrel{\text{def}}{=} |S|$.
Let us pick some state i. Since i is ergodic, by definition there is a n_0 such that for all $n \geq n_0$, $p_{ii}^n > 0$. Furthermore for all j, there are $m, m' \leq s-1$ such that $p_{ij}^m > 0$ and $p_{ji}^{m'} > 0$. So for all j, j' and $n \geq n_0 + 2(s-1)$, one has $p_{jj'}^n > 0$.

q.e.d. (lemma 1.28) ◊◊◊

Proof of proposition 1.29
Since the chain is ergodic there is a positive integer k such that $\mathbf{P}^k$ is positive. Thus there is some $0 < \delta < 1$ such that for all i, j one has $\mathbf{P}^k[i,j] \geq \delta \Pi_\infty[i,j]$. Define $\theta = 1 - \delta$ $(0 < \theta < 1)$ and matrix $\mathbf{Q}$ by:

$$\mathbf{Q} = \frac{1}{\theta}\mathbf{P}^k - \frac{1-\theta}{\theta}\Pi_\infty$$

One observes that $\mathbf{Q}$ is a transition matrix and fulfills:

$$\mathbf{P}^k = \theta\mathbf{Q} + (1-\theta)\Pi_\infty$$

Let us prove by induction that:

$$\forall n\ \mathbf{P}^{kn} = \theta^n\mathbf{Q}^n + (1-\theta^n)\Pi_\infty$$

First we observe that $\Pi_\infty\mathbf{Q} = \frac{1}{\theta}\Pi_\infty\mathbf{P}^k - \frac{1-\theta}{\theta}\Pi_\infty\Pi_\infty = \frac{1}{\theta}\Pi_\infty - \frac{1-\theta}{\theta}\Pi_\infty = \Pi_\infty$.
Then:

$$\mathbf{P}^{kn+k} = (\theta^n\mathbf{Q}^n + (1-\theta^n)\Pi_\infty)(\theta\mathbf{Q} + (1-\theta)\Pi_\infty)$$

$$= \theta^{n+1}\mathbf{Q}^{n+1} + ((1-\theta^n)\theta + (1-\theta)\theta^n + (1-\theta^n)(1-\theta))\Pi_\infty = \theta^{n+1}\mathbf{Q}^{n+1} + (1-\theta^{n+1})\Pi_\infty$$

We rewrite the equation as:

$$\forall n\ \mathbf{P}^{kn} - \Pi_\infty = \theta^n(\mathbf{Q}^n - \Pi_\infty)$$

Multiplying by $\mathbf{P}^j$ with $0 \leq j < k$:

$$\forall n\ \forall j < k\ \mathbf{P}^{kn+j} - \Pi_\infty = \theta^n(\mathbf{Q}^n\mathbf{P}^j - \Pi_\infty)$$

Multiplying by π_0:

$$\forall n\ \forall j < k\ \pi_{kn+j} - \pi_\infty = \theta^n(\pi_0\mathbf{Q}^n\mathbf{P}^j - \pi_\infty)$$

Thus the result is established with $\lambda = \theta^{\frac{1}{k}}$.

q.e.d. (proposition 1.29) ◊◊◊

Proof of proposition 1.30
We denote k the integer such that $\mathbf{M}^k$ is positive.
Let us define λ by $\lambda \equiv \sup(\alpha \mid \exists\mathbf{v} \geq 0 \wedge \mathbf{v} \neq 0 \wedge \mathbf{M}\mathbf{v} \geq \alpha\mathbf{v})$.
Let $m \equiv \min_i(\max_j \mathbf{M}[i,j])$.
If $m = 0$ then there is a null row and this row is also null for $\mathbf{M}^k$ yielding a contradiction.
Then $\mathbf{M}\mathbf{1}^T \geq m\mathbf{1}^T$. So $\lambda > 0$.
Let us pick $\lambda_n \uparrow \lambda$ and $\mathbf{v}_n \geq 0$ such that $\|\mathbf{v}_n\|_1 = 1$ and $\mathbf{M}\mathbf{v}_n \geq \lambda_n\mathbf{v}_n$.
Then by compactness of $\{\mathbf{v} \mid \|\mathbf{v}\|_1 = 1\}$, there exists $\mathbf{v} \geq 0$ such that $\|\mathbf{v}\|_1 = 1$ and $\mathbf{M}\mathbf{v} \geq \lambda\mathbf{v}$.
Suppose that $\mathbf{M}\mathbf{v} \neq \lambda\mathbf{v}$.
Then vector $\mathbf{v}' \stackrel{\text{def}}{=} \mathbf{M}^k\mathbf{v}$ fulfills $\mathbf{M}\mathbf{v}' \geq \lambda\mathbf{v}'$, and by positivity of $\mathbf{M}^k$ fulfills for all i, $\mathbf{M}\mathbf{v}'[i] > \lambda\mathbf{v}'[i]$.
Thus there exists some $\varepsilon > 0$ such that $\mathbf{M}\mathbf{v}' \geq (\lambda+\varepsilon)\mathbf{v}'$ yielding a contradiction. So $\mathbf{M}\mathbf{v} = \lambda\mathbf{v}$.
Since $\mathbf{M}^k\mathbf{v} = \lambda^k\mathbf{v}$, $\mathbf{v}$ is a positive vector.
Assume there exists $\mathbf{w}$ another λ-eigenvector independent from $\mathbf{v}$.
Then $\alpha \stackrel{\text{def}}{=} \max_i -\frac{\mathbf{w}[i]}{\mathbf{v}[i]}$ verifies $\mathbf{w} + \alpha\mathbf{v} \geq 0$ and there exists i with $(\mathbf{w}+\alpha\mathbf{v})[i] = 0$.
But $\mathbf{M}^k(\mathbf{w}+\alpha\mathbf{v}) = \lambda^k(\mathbf{w}+\alpha\mathbf{v})$. So $\mathbf{w}+\alpha\mathbf{v}$ should be a positive vector yielding a contradiction.
Assume there exists $\mathbf{w}$ a vector such that $(\mathbf{M}-\lambda\mathbf{Id})\mathbf{w} \neq 0$ and $(\mathbf{M}-\lambda\mathbf{Id})^2\mathbf{w} = 0$.
Due to the previous paragraph, $(\mathbf{M}-\lambda\mathbf{Id})\mathbf{w} = \alpha\mathbf{v}$ for some $\alpha \neq 0$.
W.l.o.g. we assume that $\alpha > 0$.
There exists a β such that $\mathbf{w}+\beta\mathbf{v} \geq 0$.
Furthermore, $(\mathbf{M}-\lambda\mathbf{Id})(\mathbf{w}+\beta\mathbf{v}) = \alpha\mathbf{v}$, i.e. $\mathbf{M}(\mathbf{w}+\beta\mathbf{v}) = \lambda(\mathbf{w}+\beta\mathbf{v}) + \alpha\mathbf{v}$.
So there exists some $\varepsilon > 0$ such that $\mathbf{M}(\mathbf{w}+\beta\mathbf{v}) \geq (\lambda+\varepsilon)(\mathbf{w}+\beta\mathbf{v})$ yielding a contradiction. We have finally proved the first item of the proposition.
To prove the second item, let $\lambda' \neq \lambda$ be an eigenvalue and $\mathbf{v}'$ be an associated eigenvector.
Define $\mathbf{v}''$ as the vector of modules of components of $\mathbf{v}'$.
$\mathbf{M}\mathbf{v}'' \geq |\mathbf{M}\mathbf{v}'| = |\lambda'|\mathbf{v}''$ *(with a slight abuse of notation)*
We claim that the equality holds only if all (non null) components of $\mathbf{v}'$ have the same argument.
Indeed $\mathbf{M}\mathbf{v}' = \lambda'\mathbf{v}'$ implies $\mathbf{M}^k\mathbf{v}' = \lambda'^k\mathbf{v}'$.
So $\mathbf{M}^k\mathbf{v}'' \geq |\lambda'^k|\mathbf{v}''$. Assuming $\mathbf{M}\mathbf{v}'' = |\lambda'|\mathbf{v}''$ implies $\mathbf{M}^k\mathbf{v}'' = |\lambda'|^k\mathbf{v}''$.
Since $\mathbf{M}^k$ is positive, equality holds only if all (non null) components of $\mathbf{v}'$ have the same argument.

On the other hand, $\mathbf{M}\mathbf{v}'' \geq |\lambda'|\mathbf{v}''$ implies $|\lambda'| \leq \lambda$.
If $|\lambda'| = \lambda$ then:

1. Since $\mathbf{M}\mathbf{v}'' \geq \lambda\mathbf{v}''$ and $\mathbf{v}''$ is non negative, $\mathbf{M}\mathbf{v}'' = \lambda\mathbf{v}''$. So $\mathbf{v}''$ is an eigenvector of λ.
2. Using the previous claim, all (non null) components of $\mathbf{v}'$ have the same argument. Then $\mathbf{v}''$ is also an eigenvector of λ' yielding a contradiction.

So $|\lambda'| < \lambda$.

q.e.d. (proposition 1.30) $\Diamond\Diamond\Diamond$

Proof of proposition 1.31
We denote $s = |S|$.

Let $\mathbf{v} \geq 0$. Then $\mathbf{P}\mathbf{v}[i] = \sum_j \mathbf{P}[i,j]\mathbf{v}[j] \leq \left(\sum_j \mathbf{P}[i,j]\right) \max(\mathbf{v}[j]) = \max(\mathbf{v}[j])$. So if $\mathbf{P}\mathbf{v} = \lambda\mathbf{v}$ then $\lambda \leq 1$. On the other hand, $\mathbf{P}\mathbf{1} = \mathbf{1}$. So 1 is the eigenvalue with largest module.

Let us decompose the vector space $\mathbb{R}^s$ into the generalized eigenspaces corresponding to the left generalized eigenvectors: $E = \bigoplus_{k=1}^{K} E_k$ with E_k corresponding to the eigenvalue λ_k sorted by decreasing module. Since $\mathbf{P}$ is regular, one applies proposition 1.30: $\lambda_1 = 1$, $\dim(E_1) = 1$, E_1 is generated by v_1 a positive vector and for all $k > 1$, $|\lambda_k| < 1$. W.l.o.g. we assume that $\|v_1\|_1 = 1$ so that $v_1 = \pi_\infty$. One can select a basis of E_k such that $\mathbf{P}$, restricted to E_k, becomes an upper triangular matrix $\mathbf{T}_k$ whose diagonal is a sequence of λ_k: $\mathbf{T}_k = \lambda_k\mathbf{Id} + \mathbf{N}_k$ where $\mathbf{N}_k$ fulfills $(\mathbf{N}_k)^{s-1} = 0$ (since $\dim(E_1) = 1$ and $\dim(E_k) \leq s-1$ for $k \geq 2$).

Let us denote $\mathbf{B}$ the matrix corresponding to the basis of $\mathbb{R}^s$ obtained as the union of the previous bases. Then: $\mathbf{P} = \mathbf{B}^{-1}\mathbf{T}\mathbf{B}$ where $\mathbf{T}$ is the block-diagonal matrix whose blocks are $\mathbf{T}_k$'s. So $\pi_n = \pi_0\mathbf{B}^{-1}\mathbf{T}^n\mathbf{B}$. Otherwise stated, $\pi_n\mathbf{B}^{-1} = \pi_0\mathbf{B}^{-1}\mathbf{T}^n$. Define $\pi'_n = \pi_n\mathbf{B}^{-1}$. The equality can be rewritten $\pi'_n = \pi'_0\mathbf{T}^n$. Given a vector $\mathbf{v}$, one denotes $proj_k(\mathbf{v})$ the subvector $\mathbf{v}$ related to E_k. So $\mathbf{v} = \sum_{k=1}^{K} proj_k(\mathbf{v})$.

$$\pi'_n = \sum_{k=1}^{K} proj_k(\pi'_n) = \sum_{k=1}^{K} proj_k(\pi'_0)(\lambda_k\mathbf{Id} + \mathbf{N}_k)^n$$

$$= proj_1(\pi'_0) + \sum_{k=2}^{K}\sum_{i=0}^{s-2} \binom{n}{i} (\lambda_k)^{n-i} proj_k(\pi'_0)(\mathbf{N}_k)^i$$

So:

$$\|\pi'_n - proj_1(\pi'_0)\| \leq \sum_{k=2}^{K}\sum_{i=0}^{s-2} \binom{n}{i} |\lambda_k|^{n-i} \|proj_k(\pi'_0)(\mathbf{N}_k)^i\| \leq C' n^{s-2} |\lambda_2|^n$$

for some constant C'.
Applying $\mathbf{B}$:

$$\|\pi_n - proj_1(\pi'_0)\mathbf{B}\| \leq C n^{s-2} |\lambda_2|^n$$

for some constant C.
This also establishes that π_n converges toward αv_1 and since $\|\pi_n\|_1 = \|v_1\|_1 = 1$, $\alpha = 1$.

q.e.d. (proposition 1.31) $\Diamond\Diamond\Diamond$

Chapter 2

Continuous Time Markov Chains

2.1 Renewal processes with non arithmetic distribution [FEL 71]

2.1.1 Limits, measures and integration

In this paragraph, we state some results (most of them elementary) about limits of functions, measures and integration useful for establishing the renewal theorem.

Lemma 2.1 *Let $(x_n)_{n\in\mathbb{N}}$ be a bounded sequence of reals such that there exists l, limit of every convergent subsequence. Then $(x_n)_{n\in\mathbb{N}}$ converges toward l.*

Proof

The following lemma is nothing but a rewriting of lemma 1.4 once one observes that $\overline{\mathbb{R}}$ is compact.

Lemma 2.2 *Let $(x_n)_{n\in\mathbb{N}}$ be a sequence of reals and f_m be a sequence of functions from $\mathbb{R}$ to $\mathbb{R}$. Then there exists a sequence of indices $m_1 < m_2 < \cdots$ such that for all $n \in \mathbb{N}$ the sequence $\{f_{m_k}(x_n)\}_{k\in\mathbb{N}}$ converges in $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$ (in $\mathbb{R}$ if for all x_n, $\sup_m |f_m(x_n)| < \infty$).*

The notion of *equicontinuity* is similar to uniform continuity but it is defined for a whole family of functions.

Definition 2.3 *Let $\{f_n\}_{n\in\mathbb{N}}$ be a family of real functions, this family is* equicontinuous *if:*

$$\forall \varepsilon > 0\ \exists \delta > 0\ \forall n\ \forall x, x'\ |x - x'| \leq \delta \Rightarrow |f_n(x) - f_n(x')| \leq \varepsilon$$

Proposition 2.4 *Let $\{f_n\}_{n\in\mathbb{N}}$ be an equicontinuous sequence of functions such that there exists B with for all x and n, $|f_n(x)| \leq B$. Then there exists a subsequence that converges, uniformly over every bounded interval, toward a uniformly continuous function.*

Proof

Let us recall that a (real) measure μ associates with every measurable set E of $\mathbb{R}$ (and in particular with every interval) a (finite or infinite) positive value $\mu\{E\}$ such that:

1. $\mu\{\emptyset\} = 0$;
2. for every countable family of disjoint sets E_i, one has $\mu\{\biguplus_{i\in\mathbb{N}} E_i\} = \sum_{i\in\mathbb{N}} \mu\{E_i\}$.

Here we are only interested in *locally finite* measures, i.e. such that for every bounded interval I, one has $\mu\{I\} < \infty$. A point a is a *continuity point* of μ if it fulfills $\lim_{a'\uparrow a} \mu\{[b, a']\} = \mu\{[b, a]\}$ with $b < a$ (this definition is independent of b). The set of discontinuity points, also called *atoms* is countable and a, discontinuity point, fulfills $\mu\{a\} > 0$. A *continuity interval* is an interval whose bounds are continuity points (by convention $-\infty, +\infty$ are considered as continuity points).

Definition 2.5 *A sequence of measures* $\{\mu_n\}_{n\in\mathbb{N}}$ converges *toward a measure* μ *iff for every* I*, bounded continuity interval of* μ*, one has:*

$$\lim_{n\to\infty} \mu_n\{I\} = \mu\{I\}$$

Proposition 2.6 *Let* $\{\mu_n\}_{n\in\mathbb{N}}$ *be a sequence of measures fulfilling* $\sup_n(\mu_n\{I\}) < \infty$ *for every bounded interval* I*. Then:*

- $\{\mu_n\}_{n\in\mathbb{N}}$ *has a subsequence which converges toward a measure.*
- *If every convergent subsequence* $\{\mu_n\}_{n\in\mathbb{N}}$ *converges toward a (fixed) measure* μ *then* $\{\mu_n\}_{n\in\mathbb{N}}$ *converges toward* μ*.*

Proof

This convergence of measures is not the only possible one. As shown by proposition 2.8, it is appropriate to obtain limits of integrals.

Notations. $\mathcal{C}_f$ denotes the set of continuous functions u whose support (i.e. $u^{-1}(\mathbb{R}^*)$) is bounded.

Lemma 2.7 *Let* μ *and* μ' *be two measures fulfilling* $\int u(x)\mu\{dx\} = \int u(x)\mu'\{dx\}$ *for all* $u \in \mathcal{C}_f$*. Then:* $\mu = \mu'$*.*

Proof

Proposition 2.8 *Let* $\{\mu_n\}_{n\in\mathbb{N}}$ *be a sequence of measures such that (1) it converges to* μ*, and (2) it fulfills* $\sup(\mu_n\{I\}) < \infty$ *for every bounded interval* I*.*
Then for all $u \in \mathcal{C}_f$*:*

$$\lim_{n\to\infty} \int u(x)\mu_n\{dx\} = \int u(x)\mu\{dx\}$$

Proof

For technical reasons, we need a stronger definition of integrable function than the usual ones (Riemann and Lebesgue integration).

Definition 2.9 *A function* z *from* $\mathbb{R}^+$ *to* $\mathbb{R}^+$ *is* directly Riemann integrable *if denoting* $m_{kh} = \inf(z(x) \mid kh \le x < (k+1)h)$ *and* $M_{kh} = \sup(z(x) \mid kh \le x < (k+1)h)$ *with* $h \in \mathbb{R}^+$ *and* $k \in \mathbb{N}$*, one has:*

$$\lim_{h\to 0} h\sum_{k\in\mathbb{N}} m_{kh} = \lim_{h\to 0} h\sum_{k\in\mathbb{N}} M_{kh} \text{ (and is finite)}$$

For instance, let function f be defined by: for every $n \in \mathbb{N}^*$ and every $x \in [n, n+\frac{1}{n^2}]$, $f(x) \stackrel{\text{def}}{=} 1$ and $f(x) \stackrel{\text{def}}{=} 0$ otherwise. Then f is Riemann integrable but not directly Rieman integrable. For functions with bounded support direct and standard Riemann integration are the same ones. We give below another sufficient condition.

Proposition 2.10 *A Lebesgue integrable function from* $\mathbb{R}^+$ *to* $\mathbb{R}^+$ *which is non increasing is directly Riemann integrable.*

Proof

2.1.2 The renewal theorem

In the sequel, we will use without mentioning it the equivalence between the distribution functions and the probability measures on $\mathbb{R}$. For a measure μ concentrated on $\mathbb{R}^+$, one defines $\mu(x)$ by $\mu(x) \stackrel{\text{def}}{=} \mu\{[0,x]\}$.

The convolution of two distributions F, G denoted $F \star G$ is the distribution of the sum of two independent random variables following distributions F and G. It is defined by:

$$F \star G(x) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} F(x-y)G\{dy\}$$

This convolution is also defined for an integrable function and a measure. When the random variables are positive it can be rewritten:

$$F \star G(x) = \int_{0}^{\infty} F(x-y)G\{dy\} = \int_{0}^{x} F(x-y)G\{dy\}$$

One denotes $F^{k\star}$, F convoluted $k-1$ times with itself. $F^{0\star}$ consists in the Dirac distribution concentrated in 0 i.e. $F^{0\star}(0) = 1$.

Suppose that a renewal process follows distribution F. Then $F^{k\star}$ represents the distribution of the k^{th} renewal instant (starting with 0). The measure $U \stackrel{\text{def}}{=} \sum_{n=0}^{\infty} F^{n\star}$ is related to the number of renewal instants: given E a measurable set of $\mathbb{R}^+$, $U(E)$ is the mean number of renewal instants in E. We are interested about the asymptotic behaviour of U and more precisely we want to prove a statement similar to the discrete case: $\lim_{t\to\infty} U([t, t+\delta]) = \frac{\delta}{\mu}$ where μ is the mean value of F. We start by showing that U is locally finite.

Lemma 2.11 *Let F be a distribution whose support is included in $]0,\infty[$ (i.e. $F(0) = 0$). Then for all x, $U(x)$ is finite. More precisely for all $h \geq 0$, there exists C_h such that $U\{I\} \leq C_h$ for all interval I with length h.*

Proof

An *increasing point* of a measure μ is a real x such that for every open interval I including x, $\mu(I) > 0$. An atom is an increasing point but the converse is not necessarily true. A measure is *arithmetic* if there exists a real $\lambda > 0$ such that the measure is concentrated on atoms $k\lambda$ for $k \in \mathbb{Z}$. The greatest λ fulfilling this property is called the *period* of the measure. We have studied renewal theory in the case of an arithmetic distribution in paragraph 1.2. Here we are interested by the non arithmetic case. The next lemma establishes the main property of non arithmetic distributions. It presents similarities with lemma 1.3 for the discrete case.

Lemma 2.12 *Let F be a non arithmetic distribution whose support is included in $]0,\infty[$ and let Σ, be the set of increasing points of distributions $F, F^{2\star}, F^{3\star}, \ldots$ (included in those of $U = \sum_{i\in\mathbb{N}} F^{i\star}$). Then Σ is* asymptotically dense *in $\mathbb{R}^+$, that is to say:*

$$\forall \varepsilon > 0 \ \exists x_\varepsilon \ \forall x \geq x_\varepsilon \ \Sigma \cap [x, x+\varepsilon] \neq \emptyset$$

Proof

Using the previous lemma, one establishes another lemma similar to lemma 1.5 for the discrete case.

Lemma 2.13 *Let F be a non arithmetic distribution whose support is included in $]0,\infty[$ and let g be a bounded and uniformly continuous function fulfilling for all x, $g(x) \leq g(0)$ and:*

$$g(x) = \int_{0}^{\infty} g(x-y)F\{dy\}$$

Then for all x, one has $g(x) = g(0)$.

Proof

Let us consider that at every renewal instant one buys a product whose value evolves as time elapses and is given by function z (by convention z is null on $\mathbb{R}^{-*}$). An example of function z could be $z(x) = \mathbf{1}_{x\leq\delta}$ meaning that the product price is 1 until δ time unit elapses and then the price becomes 0 (a sharp amortization of the price). One is interested to study the function

$Z(x)$ which is defined as the expectation of the value at time x of all products that have been bought. Reasoning about the first renewal instant that occurs after 0, one gets the following renewal equation:

$$Z(x) = z(x) + \int_0^x Z(x-y)F\{dy\} \tag{2.1}$$

This lemma relates Z with U.

Lemma 2.14 *Let F be a distribution whose support is included in $]0,\infty[$ and let z be a function bounded on bounded intervals.*
Then the function Z defined by $Z(x) \stackrel{def}{=} \int_0^x z(x-y)U\{dy\}$ is the single solution, bounded on bounded intervals, of (2.1).

Proof

In the sequel we say that $U \star z$ is *the* solution of the renewal equation associated with z.

Lemma 2.15 *Let F be a non arithmetic distribution whose support is included in $]0,\infty[$. Let z be a continuous function whose support is included in $[0,h]$. Let Z be the corresponding solution of the renewal equation. Then Z is uniformly continuous and for every $a \geq 0$ one has:*

$$\lim_{x\to\infty} Z(x+a) - Z(x) = 0$$

Proof

The next proposition establishes that a property of the asympotic behaviour of measure U entails a similar property for the solution Z of the renewal equation.

Proposition 2.16 *Let F be a distribution whose support is included in $]0,\infty[$ and U be the measure defined by $U \stackrel{def}{=} \sum_{k\in\mathbb{N}} F^{k\star}$. Let us suppose that there exists $\eta > 0$ and an increasing sequence of instants t_n going to ∞ such that:*

$$\lim_{n\to\infty} U(t_n) - U(t_n - h) = h\eta \text{ for all } \ h > 0$$

Then for every function z directly Riemann integrable, the solution Z of the renewal equation fulfills:

$$\lim_{n\to\infty} Z(t_n) = \eta \int_0^\infty z(x)dx$$

Proof

We are now in position to establish the renewal theorem.

Theorem 2.17 *Let F be a non arithmetic distribution whose support in included $\mathbb{R}^{+*}$ with (finite or infinite) expectation μ and let U be the measure defined by $U \stackrel{def}{=} \sum_{k\in\mathbb{N}} F^{k\star}$. Then:*

$$\lim_{t\to\infty} U(t) - U(t-h) = \frac{h}{\mu} \text{ for all } h > 0$$

Proof

The following theorem allows to study the asymptotic behaviour of rewards for renewal instants. Its omitted proof is immediately obtained by substituting t_n by x in the proof of proposition 2.16.

Theorem 2.18 *Let z be a directly Riemann integrable function, F be a non arithmetic distribution whose support is included in $\mathbb{R}^{+*}$ with (finite or infinite) expectation μ and Z be the corresponding solution of the renewal equation. Then:*

$$\lim_{x\to\infty} Z(x) = \mu^{-1} \int_0^\infty z(y)dy$$

We also provide an intuitive justification of this theorem. Let us consider a large x as the current instant (see Figure 2.1). Assume that there has been approximatively in the past one renewal instant uniformly distributed per interval $[x-i\mu, x-(i+1)\mu]$. Then: $Z(x) \approx \frac{1}{\mu}\int_0^\mu z(y)dy + \frac{1}{\mu}\int_\mu^{2\mu} z(y)dy + \cdots$ yielding the equality of the theorem.

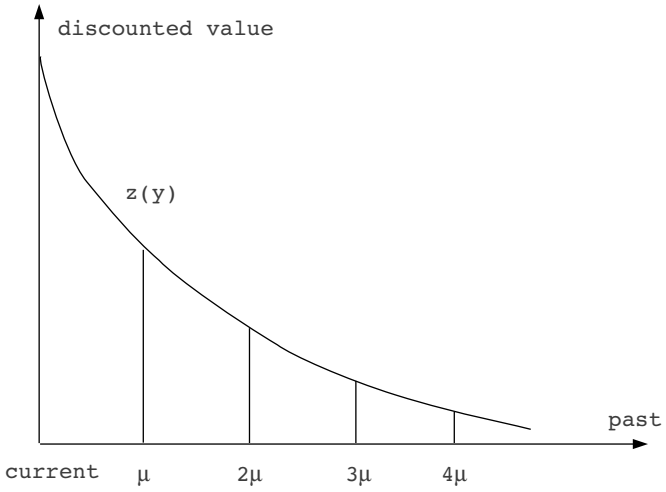

Figure 2.1: An interpretation of Theorem 2.18

2.1.3 Generalizations

One straightforwardly generalizes the previous results when F has 0 for atom, i.e. $0 < p \stackrel{\text{def}}{=} F(0) < 1$. Indeed, one reduces it to the previous case by considering that the renewal instants are produced by the following mechanism:

- Choose a number of repetitions of renewal at the same instant with a Bernoulli law whose parameter is p.
- Choose the next renewal instant with distribution G defined by $G(0) \stackrel{\text{def}}{=} 0$ and $G(x) \stackrel{\text{def}}{=} (1-p)^{-1}(F(x)-p)$ for $x > 0$.

Let us denote $U \stackrel{\text{def}}{=} \sum_{i\in\mathbb{N}} F^{i\star}$ and $V \stackrel{\text{def}}{=} \sum_{i\in\mathbb{N}} G^{i\star}$. Using the above reasoning, one has $U = (1-p)^{-1}V$. So, the results are still applicable.

One also generalizes the results with a delayed renewal process, where the first renewal instant is not necessarily 0 but is randomly chosen by a distribution G. Afterwards, the process behaves as a standard renewal process. Let $V(t)$, be the number of renewal instants until t, then V fulfills: $V = G \star U$ where U is the number of renewal instants of the standard process. So for $h > 0$, one has:

$$V(t+h) - V(t) = \int_0^{t+h} U(t+h-y) - U(t-y)G\{dy\}$$

Let t_0 be such that $1 - G(t_0) \leq \varepsilon$ and let t_1 be such that for all $t \geq t_1$,
one has $|U(t+h) - U(t) - h/\mu| \leq \varepsilon$.
Then for all $t \geq t_0 + t_1$, one gets:
$|V(t+h) - V(t) - h/\mu|$
$\leq \int_0^{t_0} |U(t+h-y) - U(t-y) - h\mu|G\{dy\} + \int_{t_0}^{t+h} |U(t+h-y) - U(t-y)|G\{dy\} + h/\mu \int_{t_0}^{\infty} G\{dy\}$
$\leq (1 + C_h + h/\mu)\varepsilon$

The same reasoning for the delayed solutions of the renewal equation can be performed with the restriction that they must be *bounded*. This yields to the following theorem which includes the previous generalizations.

Theorem 2.19 *Let G be a distribution whose support is included in $\mathbb{R}^+$ and F be a non arithmetic distribution whose support is included in $\mathbb{R}^+$ with a (finite or infinite) expectation μ. Let U be the measure defined by $U \stackrel{def}{=} \sum_{k\in\mathbb{N}} F^{k\star}$ and $V \stackrel{def}{=} G \star U$. Then:*

$$\lim_{t\to\infty} V(t) - V(t+h) = \frac{h}{\mu} \text{ for all } h > 0$$

Moreover, let z be a directly Riemann integrable function and Z be the corresponding solution of the renewal equation. If Z is bounded then:

$$\lim_{x\to\infty} (G \star Z)(x) = \mu^{-1} \int_0^\infty z(y)dy$$

2.2 Continuous time Markov chains [CIN 75]

2.2.1 Presentation

A Continuous Time Markov Chain (CTMC) is a DES with the following features:

- the time interval between events T_n is a random variable whose distribution is the exponential one and whose rate only depends on state S_n. More formally:

$$\mathbf{Pr}(T_n \leq \tau \mid S_0 = s_{i_0}, ..., S_n = s_i, T_0 \leq \tau_0, ..., T_{n-1} \leq \tau_{n-1}) =$$

$$\mathbf{Pr}(T_n \leq \tau \mid S_n = s_i) \stackrel{\text{def}}{=} 1 - e^{-\lambda_i \cdot \tau}$$

- The selection of the state that follows the current state only depends on that state and the transition probabilities remain constant[1] along the run:

$$\mathbf{Pr}(S_{n+1} = s_j \mid S_0 = s_{i_0}, ..., S_n = s_i, T_0 \leq \tau_0, ..., T_n \leq \tau_n) =$$

$$\mathbf{Pr}(S_{n+1} = s_j \mid S_n = s_i) \stackrel{\text{def}}{=} \mathbf{P}[i,j] \stackrel{\text{def}}{=} p_{ij}$$

The DTMC defined by transition matrix $\mathbf{P}$ is called the *embedded chain*. It observes the state changes of the CTMC without taking into account the time elapsed. A state of the CTMC is *absorbing* if it is absorbing w.r.t. the embedded DTMC. The chain is said *irreducible* if the embedded chain is irreducible.

The choice of the exponential distribution is fundamental since the distribution of the remaining time after τ, is the same as the original one ($\tau' > \tau$):

$$\mathbf{Pr}(T > \tau' \mid T > \tau) = \frac{\mathbf{Pr}(T > \tau')}{\mathbf{Pr}(T > \tau)} = \frac{e^{-\lambda\tau'}}{e^{-\lambda\tau}} = e^{-\lambda(\tau'-\tau)} = \mathbf{Pr}(T > \tau' - \tau)$$

Furthermore let X (resp. Y independent from X) follow an exponential distribution with rate λ (resp. μ). Then: $\mathbf{Pr}(\min(X,Y) > \tau) = e^{-\lambda\tau}e^{-\mu\tau} = e^{-(\lambda+\mu)\tau}$. So $\min(X,Y)$ follows an exponential distribution with rate $\lambda + \mu$. Last $\mathbf{Pr}(X < Y) = \int_0^\infty e^{-\mu\tau}\lambda e^{-\lambda\tau}d\tau = \frac{\lambda}{\lambda+\mu}$.

A CTMC has also an oriented graph representation defined as follows:

- The set of vertices is the set of the states of the CTMC;
- There is an edge from s_i to s_j labelled by $\lambda_i p_{ij}$ if $p_{ij} > 0$ and $s_i \neq s_j$. This choice of labels will be justified in section 2.2.2.

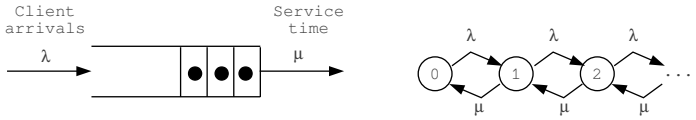

Figure 2.2: A single-server queue

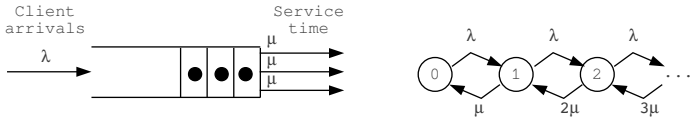

Figure 2.3: An infinite-server queue

Example 2.20 (A single-server queue) *Figure 2.2 represents the behaviour of a client queue in front of a service. Interarrival times of client are i.i.d. and their common distribution is an exponential one with rate λ. The service time has also an exponential distribution with rate μ. One client is served at a time. The states of the infinite* CTMC *are characterized by the number of clients. The exit rate of state 0 is λ (the arrival of a client) while the exit rate of state $n \geq 1$ is $\lambda + \mu$. Here we use the fact that if X (resp. Y) has an exponential distribution with rate λ (resp. μ) and X and Y are independent then $\min(X, Y)$ has an exponential distribution with rate $\lambda + \mu$. Its is also known that $\mathbf{Pr}(X < Y) = \frac{\lambda}{\lambda+\mu}$. Thus for $n \geq 1$, $p_{n,n+1} = \frac{\lambda}{\lambda+\mu}$ and $p_{n,n-1} = \frac{\mu}{\lambda+\mu}$.*

Example 2.21 (An infinite-server queue) *Figure 2.3 represents the behaviour of a client queue in front of a multi-threaded service. While the parameters of the system are the same as the previous ones, the clients are served simultaneously, each one by its own server. Then the end of a service in state n follows an exponential distribution with rate $n\mu$. So its exit rate is $\lambda + n\mu$, $p_{n,n+1} = \frac{\lambda}{\lambda+n\mu}$ and $p_{n,n-1} = \frac{n\mu}{\lambda+n\mu}$.*

Example 2.22 (A tandem queue) *Figure 2.4 represents a more complex system when clients are successively served by two servers, one with rate μ and the other with rate δ. This system*

[1]Sometimes these chains are called homogeneous CTMC.

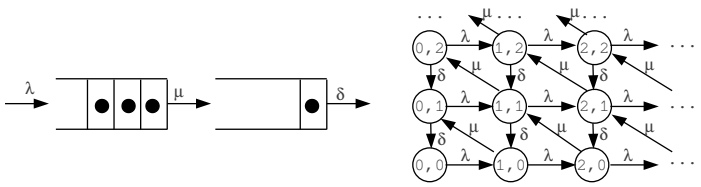

Figure 2.4: A tandem queue

is known as a tandem queue. Thus the states of the CTMC *are indexed by a pair of integers corresponding to the number of clients in each queue. We leave it to the reader to check that the infinite graph on the right part of the figure is the* CTMC *associated with this system.*

2.2.2 Transient behaviour of a CTMC

In a CTMC, due to the memoryless property of the exponential distribution, the evolution of the DES only depends on its current state. Let $\pi_{ij}(\tau)$ denote the probability that at time τ, the CTMC is in state s_j knowing that at time 0 the CTMC is in state s_i. Then by the memoryless property, $\pi_{ij}(\tau)$ is also the probability that at time $\Delta + \tau$, the CTMC is in state s_j knowing that at time Δ the CTMC is in state s_i. Thus the following equation is satisfied:

$$\pi_{ij}(\Delta + \tau) = \sum_k \pi_{ik}(\Delta)\pi_{kj}(\tau) \tag{2.2}$$

We now study the transient behaviour of a CTMC and first its smoothness. Let us define matrix $\mathbf{Q}$ by: $q_{ij} \stackrel{\text{def}}{=} \lambda_i \cdot p_{ij}$ for $i \neq j$ and $q_{ii} \stackrel{\text{def}}{=} (p_{ii} - 1)\lambda_i (= -\sum_{j \neq i} q_{ij})$. Matrix $\mathbf{Q}$ is called the *infinitesimal generator* of the CTMC. The next proposition shows that the behaviour of the CTMC fulfills a *backward* differential equation system related to $\mathbf{Q}$.

Proposition 2.23 *Let $\mathcal{C}$ be a* CTMC *with $\mathbf{P}$ its transition matrix and λ its rate vector. The family of functions $\{\pi_{ij}\}_{ij}$ satisfies the following properties:*

- *For all $i \neq j$, $\lim_{\tau \downarrow 0} \pi_{ii}(\tau) = 1$ and $\lim_{\tau \downarrow 0} \pi_{ij}(\tau) = 0$;*
- *For all i, j, π_{ij} is differentiable and fulfills: $\frac{d\pi_{ij}(\tau)}{d\tau} = \sum_k q_{ik}\pi_{kj}(\tau)$.*

Proof

Introducing matrix Π whose item $\Pi[i, j]$ is π_{ij}, the previous equation can be rewritten:

$$\frac{d\Pi}{d\tau} = \mathbf{Q} \cdot \Pi \tag{2.3}$$

Looking at the proof, it could be wrongly (why?) deduced that by another use of (2.2), one would obtain $\frac{d\Pi}{d\tau} = \Pi \cdot \mathbf{Q}$. In fact we need additional constraints in order to get such a *forward* differential equation system.

For instance observe that the time divergence expressed by (1.1) may be falsified. Let us consider an infinite CTMC whose set of states is $\mathbb{N}$ and whose transitions are: $i \xrightarrow{2^i} i + 1$. Then the mean time of an execution is 2, implying that almost surely an infinite execution lasts a finite time! The next lemma exhibits a condition that excludes such pathologic behaviours and provides a useful bound on the probability that n events take place in a time interval.

Lemma 2.24 *Let $\mathcal{C}$ be a* CTMC *with $\mathbf{P}$ its transition matrix and λ its rate vector. Define $\alpha \stackrel{def}{=} \sup_i(\lambda_i)$ and assume that $\alpha < \infty$. Then for all Δ:*

$$\mathbf{Pr}(\sum_{m \leq n} T_m \leq \Delta) \leq 1 - e^{-\alpha\Delta} \sum_{0 \leq m \leq n} \frac{(\alpha\Delta)^m}{m!}$$

In particular:

$$\lim_{n \to \infty} \mathbf{Pr}(\sum_{m \leq n} T_m \leq \Delta) = 0$$

Proof

As a consequence when α is finite, (1.1) holds. We also establish that the set of functions π_{ij} follows a forward differential equation system.

Proposition 2.25 *Let $\mathcal{C}$ be a* CTMC *with* $\mathbf{P}$ *its transition matrix and* λ *its rate vector such that for* $\alpha \stackrel{def}{=} \sup_k(\lambda_k)$ *is finite. Then the family of functions* $\{\pi_{ij}\}$ *fulfills:*

$$\frac{d\pi_{ij}(\tau)}{d\tau} = \sum_k \pi_{ik}(\tau) q_{kj}$$

which can be rewritten as:

$$\frac{d\Pi}{d\tau} = \Pi \cdot \mathbf{Q} \tag{2.4}$$

Proof

W.r.t. a state point of view (i.e. forgetting the events) a CTMC is indifferently specified by its infinitesimal generator $\mathbf{Q}$ or by its transition matrix $\mathbf{P}$ and its exit rate vector λ.

2.2.3 Steady-state behaviour of a CTMC

The classification of states (transient, null recurrent or positive recurrent) has the same definition as for the discrete case (see section 1.3.2). One observes that *the recurrent character only depends on* $\mathbf{P}$ (but not the distinction between null or positive recurrence). So states are recurrent in the CTMC iff they are recurrent in the embedded DTMC. Using renewal theory, we establish a first characterization of positive recurrence.

Theorem 2.26 *Let $\mathcal{C}$ be a* CTMC *(with* $\mathbf{P}$*,* λ *and* $\mathbf{Q}$ *defined as usual). Let i be a recurrent state and D_i be the mean time between two visits of i. Then:*

$$\lim_{\tau \to \infty} \pi_{ii}(\tau) = \frac{1}{\lambda_i D_i}$$

Thus i is positive recurrent iff $\lim_{\tau\to\infty} \pi_{ii}(\tau) > 0$.

Proof

With this characterization, we achieve our study of irreducibility.

Proposition 2.27 *Let i, j be two states of an irreducible recurrent* CTMC*. Then i is positive recurrent iff j is positive recurrent.*

Proof

We now present two theorems similar to theorems 1.20 and 1.17 for the discrete case.

Theorem 2.28 *Let $\mathcal{C}$ be an irreducible* CTMC *(with* $\mathbf{P}$*,* λ *and* $\mathbf{Q}$ *defined as usual) whose states are recurrent. Let* $\mathbf{v}$ *be a non negative and non null solution de* $\mathbf{v} = \mathbf{v} \cdot \mathbf{P}$ *(unique up to a multiplicative factor). Then vector* $\mathbf{u}$ *defined by* $\mathbf{u}_i \stackrel{def}{=} \frac{\mathbf{v}_i}{\lambda_i}$ *fulfills:*

$$\mathbf{u} \cdot \mathbf{Q} = 0$$

Conversely, let $\mathbf{u}' \neq 0$ *be such that* $\mathbf{u}' \cdot \mathbf{Q} = 0$ *and for all* i $\mathbf{u}'_i \geq 0$*, then there exists* α *such that* $\mathbf{u}'_i = \alpha \cdot \mathbf{u}_i$

Proof

The above equation is called (global) balance equation. The following theorem establishes a necessary and sufficient condition for the existence and unicity of a steady-state distribution. Additionally it proves that when the CTMC is recurrent then time diverges almost surely (why?).

Theorem 2.29 *Let $\mathcal{C}$ be an irreducible* CTMC *(with* $\mathbf{P}$*,* λ *and* $\mathbf{Q}$ *defined as usual) whose states are recurrent. If the states are null recurrent then the transient distribution* $\pi(\tau)$ *converges to 0 when time* τ *goes to infinity. Otherwise it converges to a steady-state distribution* $\mathbf{u}$ *which is the single solution of* $\mathbf{u} \cdot \mathbf{Q} = 0 \wedge \mathbf{u} \cdot \mathbf{1}^T = 1 \wedge \mathbf{u} \geq 0$*. Furthermore for any state i,* $u_i = \frac{1}{D_i \lambda_i}$ *where D_i is the mean return time to state i.*

Moreover, let $\mathbf{v} \neq 0$ *be such that for all* i $\mathbf{v}_i \geq 0$ *and* $\mathbf{v} = \mathbf{v} \cdot \mathbf{P}$ *(unique up to a multiplicative factor). Then the states are positive recurrent iff* $s \stackrel{def}{=} \sum_{i \in \mathbb{N}} \frac{\mathbf{v}_i}{\lambda_i}$ *is finite.*

Proof

We now provide a useful characterization of the positive recurrence for irreducible CTMCs.

Proposition 2.30 *Let $\mathcal{C}$ be an irreducible* CTMC*, (with $\mathbf{P}$, λ and $\mathbf{Q}$ defined as usual) such that* $\sup(\lambda_s \mid s \in S)$ *is finite. Then the states of $\mathcal{C}$ are positive recurrent iff there exists $\mathbf{u}$ such that $\mathbf{u} \cdot \mathbf{Q} = 0$ with for all i, $\mathbf{u}_i > 0$ and $\sum_{i \in S} \mathbf{u}_i < \infty$.*

Proof

Observe that we have required that $\sup(\lambda_s \mid s \in S)$ is finite. Indeed without an additional condition, this proposition is false. Let us examine the example of the transient random walk with $p > \frac{1}{2}$. This random walk admits a positive vector $\mathbf{v}$ that verifies $\mathbf{v} \cdot \mathbf{P} = \mathbf{v}$ (but $\sum_i \mathbf{v}_i = \infty$). Define $\lambda_i \stackrel{\text{def}}{=} 2^i \frac{\mathbf{v}_i}{\mathbf{v}_0}$. Then $\mathbf{u}_i \stackrel{\text{def}}{=} \frac{\mathbf{v}_i}{\lambda_i} = 2^{-i}\mathbf{v}_0$ fulfills $\mathbf{u} \cdot \mathbf{Q} = 0$ and $\sum_i \mathbf{u}_i < \infty$.

The table below summarizes the characterization of the status of an irreducible CTMC.

Status	Characterization
Recurrent	The embedded DTMC is recurrent.
Positive Recurrent	(1) The embedded DTMC is recurrent *(implied by (2) when* $\sup(\lambda_i \mid i \in S) < \infty$*)* and (2) $\exists \mathbf{u} > 0 \ \mathbf{u} \cdot \mathbf{Q} = 0 \wedge \sum_{i \in S} \mathbf{u}_i = 1$ *($\mathbf{u}$ is the steady-state distribution)*

We illustrate these characterizations with the analysis of the infinite-server queue. First we discriminate between recurrence and transience. So we are looking for a positive non null bounded solution of:

$$x_1 = \frac{\lambda}{\mu + \lambda} x_2 \text{ and } \forall i \geq 2 \ x_i = \frac{\lambda}{i\mu + \lambda} x_{i+1} + \frac{i\mu}{i\mu + \lambda} x_{i-1}$$

It can be rewritten as:

$$x_1 = \frac{\lambda}{\mu + \lambda} x_2 \text{ and } \forall i \geq 2 \ x_{i+1} - x_i = \frac{i\mu}{\lambda}(x_i - x_{i-1})$$

By induction:

$$\forall i \geq 1 \ x_{i+1} - x_i > 0 \text{ and } \forall i \geq i_0 \stackrel{\text{def}}{=} \left\lceil \frac{\lambda}{\mu} \right\rceil \ x_{i+1} - x_i \geq x_i - x_{i-1}$$

Thus $\forall i \geq i_0 \ x_i \geq (i - i_0)(x_{i_0} - x_{i_0 - 1})$ implying that the x_i's are unbounded. So the CTMC is recurrent.

Let us state the *global balance equation* $\mathbf{x} \cdot \mathbf{Q} = 0$:

$$\lambda x_0 = \mu x_1 \text{ and } \forall i \geq 1 \ \lambda x_i + i\mu x_i = (i+1)\mu x_{i+1} + \lambda x_{i-1}$$

Observe that the first equation can be subtracted to the second one yielding $\lambda x_1 = 2\mu x_2$ and by induction, one gets the following *local balance equations*:

$$\forall i \geq 0 \ \lambda x_i = (i+1)\mu x_{i+1}$$

Let $\rho \stackrel{\text{def}}{=} \frac{\lambda}{\mu}$. For $i \geq 0$, $x_i = x_0 \frac{\rho^i}{i!}$. Thus the CTMC is positive recurrent and $\pi_\infty(i) = e^{-\rho} \frac{\rho^i}{i!}$.

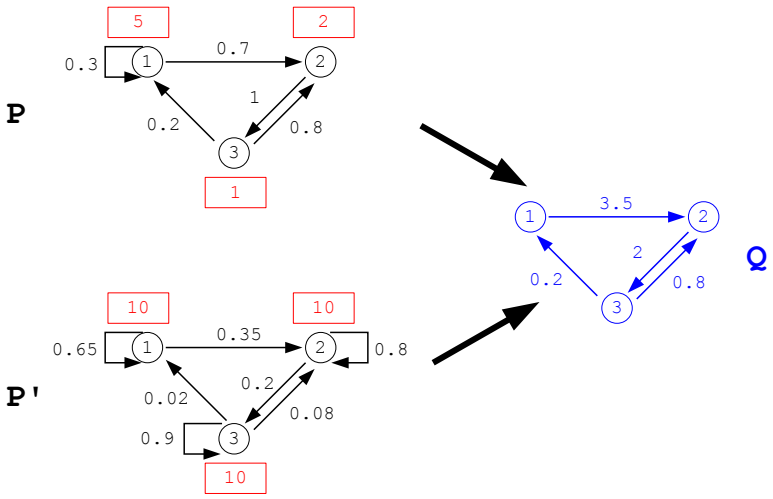

Figure 2.5: Illustration of the uniformization technique

2.3 Finite continuous time Markov chains

We now exploit the results of the previous sections in order to provide algorithms for computing the transient and the steady-state distributions of a CTMC.

2.3.1 Transient analysis of finite CTMC

In a finite CTMC, the condition of proposition 2.25 is obviously satisfied. So the forward differential equation system (2.4) could be the basis of an approximate computation of the transient distribution.

However there is a more efficient and accurate technique to compute the transient distribution. As a first step, given an arbitrary CTMC $\mathcal{C}$, we show how to obtain a *uniform* CTMC $\mathcal{C}$ with the same infinitesimal generator. A uniform CTMC is a CTMC with constant exit rates.

Let us pick an arbitrary value $\mu \geq \sup_i(\lambda_i)$. Then for every state s_i, its exit rate in $\mathcal{C}'$ λ'_i, is equal to μ. The transition matrix $\mathbf{P'}$ is defined by:

$$\forall i \neq j \;\; p'_{ij} = \frac{\lambda_i}{\mu} p_{ij} \text{ and } p'_{ii} = 1 - \sum_{j \neq i} p'_{ij}$$

Since for all $j \neq i$, $0 \leq p'_{ij} \leq p_{ij}$ one gets $p_{ii} \leq p'_{ii} \leq 1$. So $\mathbf{P'}$ is a transition matrix. Now:

$$\forall i \neq j \;\; q'_{ij} = p'_{ij}\lambda'_i = \frac{\lambda_i}{\mu} p_{ij}\mu = p_{ij}\lambda_i = q_{ij}$$

Thus the infinitesimal generator is unchanged.

Figure 2.5 illustrates the uniformization of a chain. First one chooses $\mu = 10 \geq \max(1, 2, 5)$. Then the p'_{ij}'s are defined accordingly: for instance, $p'_{12} = \frac{5}{10}0.7 = 0.35$.

The expression of the transient distribution $\pi(\tau)$ is obtained as follows. We decompose this probability depending on the number of state changes.

$$\pi(\tau) = \sum_{n\in\mathbb{N}} \mathbf{Pr}(\sum_{m<n} T_m \leq \tau < \sum_{m\leq n} T_m)\pi(0)(\mathbf{P}')^n$$

The transient distribution of the states knowing that there have been n state changes is given by the behaviour of the embedded chain and is equal to $\pi(0)\cdot(\mathbf{P}')^n$. On the other hand,

$$\mathbf{Pr}(\sum_{m<n} T_m \leq \tau < \sum_{m\leq n} T_m)$$

is the probability of exactly n state changes in $[0,\tau]$ when T_m follows an exponential distribution with parameter μ. Using the proof of lemma 2.24, one obtains:

$$\mathbf{Pr}(\sum_{m<n} T_m \leq \tau < \sum_{m\leq n} T_m) = \left(1 - e^{-\mu\tau} \sum_{0\leq m<n} \frac{(\mu\tau)^m}{m!}\right) - \left(1 - e^{-\mu\tau} \sum_{0\leq m<n+1} \frac{(\mu\tau)^m}{m!}\right)$$

$$= e^{-\mu\tau}\frac{(\mu\tau)^n}{n!}$$

So we obtain:

$$\pi(\tau) = \pi(0)\cdot\left(e^{-\mu\tau}\sum_{n\geq 0}\frac{(\mu\tau)^n(\mathbf{P}')^n}{n!}\right)$$

From an implementation point of view, this infinite sum is not a problem as it converges very quickly. For instance, the summation can be stopped as soon as the required precision is greater than $\frac{e^{-\mu\tau}\cdot(\mu\tau)^n}{n!}$ (see [FOX 88] for more information on the convergence of this sum).

This technique has been introduced in [JEN 53] and is known as the *uniformization* technique.

2.3.2 Steady-state analysis of finite CTMC

Let us examine the asympotic behaviour of a finite CTMC. The easiest way to do it consists in exploiting the embedded chain. As observed during the presentation of the uniformization technique, it is not unique. Let us focus on a DTMC obtained with a choice of $\mu > \max_i(\lambda_i)$. In this case, every state s_i fulfills $p'_{ii} > 0$. Thus every terminal scc of this DTMC is ergodic. Using proposition 1.27 this implies that it has a steady-state distribution. This distribution measures in the original CTMC the steady-state probability of occurrence of states. Since the chain is uniform, the mean sojourn time in states is identical ($\frac{1}{\mu}$). Thus using theorem 2.29, it is also the steady-state distribution of the CTMC.

In the particular case of an ergodic chain, using again theorem 2.29 this distribution is obtained by solving the following equation (X is the unknown distribution).

$$X\cdot\mathbf{Q} = 0 \quad \text{and} \quad X\cdot\mathbf{1}^T = 1$$

2.4 Proofs

2.4.1 Proofs of section 2.1.1

Proof of lemma 2.1
Suppose there exists $\varepsilon > 0$ such that $\forall n\ \exists n' > n\ |x_{n'} - l| \geq \varepsilon$. Iterating this property one extracts a sequence $(x_{n_r})_{r\in\mathbb{N}}$ distant of at least ε from l. Since the initial sequence is bounded one extracts from sequence $(x_{n_r})_{r\in\mathbb{N}}$ a convergent subsequence thus converging toward l, yielding a contradiction.

q.e.d. (lemma 2.1) ◊◊◊

Proof of proposition 2.4
Let us pick some countable and dense set of reals $\{a_i\}_{i\in\mathbb{N}}$. Using lemma 2.2, there exists a subsequence $\{f_{n_r}\}_{r\in\mathbb{N}}$ which converges for all a_i.

Let us fix a bounded interval I. Let ε et δ corresponding to the definition of equicontinuity. Using density, there exists a finite subset of indices S such that for all $x \in I$, there exists a_i with $i \in S$ and $|x - a_i| \leq \delta$. On the other hand, there exists r_0 such that for all $r, r' \geq r_0$ and $i \in S$, $|f_{n_r}(a_i) - f_{n_{r'}}(a_i)| \leq \varepsilon$. So for all $x \in I$, $|f_{n_r}(x) - f_{n_{r'}}(x)| \leq 3\varepsilon$. This proves that $\{f_{n_r}(x)\}_{r\in\mathbb{N}}$ is a Cauchy sequence and thus converges to a limit denoted $f(x)$. The proof also establishes that the convergence is uniform over interval I.

The uniform continuity is obtained by letting r go to infinity in the implication:

$$|x - x'| \leq \delta \Rightarrow |f_{n_r}(x) - f_{n_r}(x')| \leq \varepsilon$$

q.e.d. (proposition 2.4) $\Diamond\Diamond\Diamond$

Proof of proposition 2.6
One picks a point r which is not an atom for every measure μ_n, and a countable dense set $\{a_i\}_{i\in\mathbb{N}}$ of $\mathbb{R} \setminus \{r\}$. One introduces functions f_n defined on $\mathbb{R} \setminus \{r\}$ by:

- $f_n(x) \stackrel{\text{def}}{=} \mu_n([r, x])$ if $x > r$;
- $f_n(x) \stackrel{\text{def}}{=} \mu_n([x, r])$ if $x < r$.

One applies lemma 2.2 to these sequences of functions and points. Let f_{n_s} be the convergent subsequence, one defines function g by:

- $g(a_i) \stackrel{\text{def}}{=} \lim_{s\to\infty} f_{n_s}(a_i)$;
- $g(x) \stackrel{\text{def}}{=} \inf(g(a_i) \mid a_i \geq x)$ for $x > r$;
- $g(x) \stackrel{\text{def}}{=} \inf(g(a_i) \mid a_i \leq x)$ for $x < r$.

The function g is defined over $\mathbb{R}\setminus\{r\}$, non increasing over $]-\infty, r[$ and non decreasing over $]r, +\infty[$. So it admits left and right limits $g(x^-), g(x^+)$ for all x.

One introduces function h equal to g over its continuity points and such that:

- for every discontinuity point $x < r$, $h(x) \stackrel{\text{def}}{=} g(x^-)$;
- for every discontinuity point $x > r$, $h(x) \stackrel{\text{def}}{=} g(x^+)$.

Then one defines a measure μ by

- $\forall a < r\ \mu\{[a, r[\} \stackrel{\text{def}}{=} h(a) - g(r^-)$;
- $\forall b > r\ \mu\{]r, b]\} \stackrel{\text{def}}{=} h(b) - g(r^+)$;
- $\mu\{r\} = g(r^+) + g(r^-)$.

Let us prove that $\{\mu_{n_s}\}_{s\in\mathbb{N}}$ converges to μ. We establish the case of a continuity interval $[a, b]$ with $-\infty < a < r < b < \infty$ and let the other cases to the reader. Let $\varepsilon > 0$, there exists:
$a_{i_1} \leq a \leq a_{i_2} < r < a_{i_3} \leq b \leq a_{i_4}$ such that:
$g(a_{i_1}) \geq h(a) \geq g(a_{i_2}) \geq g(a_{i_1}) - \varepsilon$ and $g(a_{i_3}) \leq h(b) \leq g(a_{i_4}) \leq g(a_{i_3}) + \varepsilon$.
On the other hand, there exists s_0 such that for all $s \geq s_0$, and $1 \leq j \leq 4$ one has:
$|g(a_{i_j}) - f_{n_s}(a_{i_j})| \leq \varepsilon$
So,
$\mu\{[a, b]\} \leq \mu\{[a_{i_2}, a_{i_3}]\} + 2\varepsilon \leq \mu_{n_s}\{[a_{i_2}, a_{i_3}]\} + 4\varepsilon \leq \mu_{n_s}\{[a, b]\} + 4\varepsilon$

Similarly it can be proven that:
$\mu\{[a,b]\} \geq \mu_{n_s}\{[a,b]\} - 4\varepsilon$
Assume that every convergent subsequence of $\{\mu_n\}_{n\in\mathbb{N}}$ converges toward μ but $\{\mu_n\}_{n\in\mathbb{N}}$ does not converge toward μ. There exists $\varepsilon > 0$ and I, a continuity interval of μ, such that:
$\forall n\ \exists n' > n\ |\mu_{n'}\{I\} - \mu\{I\}| \geq \varepsilon$
Iterating this property, one obtains a subsequence $(\mu_{n_s})_{s\in\mathbb{N}}$ with $\mu_{n_s}\{I\}$ distant from at least ε from $\mu\{I\}$. Since the subsequence fulfills the hypotheses of the proposition, one extracts from $\{\mu_{n_s}\}_{s\in\mathbb{N}}$ a convergent subsequence which must converges to μ, yielding a contradiction.

q.e.d. (proposition 2.6) $\Diamond\Diamond\Diamond$

Proof of lemma 2.7
Let I be a bounded interval, its indicator function can be obtained as the increasing limit of functions of $\mathcal{C}_f$. The monotone convergence theorem allows us to conclude.

q.e.d. (lemma 2.7) $\Diamond\Diamond\Diamond$

Proof of proposition 2.8
Let $u \in \mathcal{C}_f$ with $M = \sup(|u(x)|)$. Pick some $\varepsilon > 0$. Let I be a bounded continuity interval of μ containing the support of u. Define $B \stackrel{\text{def}}{=} \sup(\mu_n\{I\})$.

Since u is uniformly continuous, one partitions I in continuity intervals $I_1, \ldots, I_k$ such that for all $x \in I_j$ $|u(x) - u_j| \leq \varepsilon$ given some arbitrary $u_j \in u(I_j)$. Define v by $v(x) \stackrel{\text{def}}{=} u_j$ for $x \in I_j$ and $v(x) \stackrel{\text{def}}{=} 0$ for $x \in I^c$.

By hypothesis, there exists n_0 such that for all $n \geq n_0$ and j, one has $|\mu_n\{I_j\} - \mu\{I_j\}| \leq \varepsilon/k$.
So:
$\left|\int u(x)\mu\{dx\} - \int u(x)\mu_n\{dx\}\right| = \left|\sum_{j=1}^k \int_{I_j} u(x)\mu\{dx\} - \int_{I_j} u(x)\mu_n\{dx\}\right|$
$= \left|\sum_{j=1}^k \int_{I_j} (u(x) - v(x))\mu\{dx\} + \int_{I_j} v(x)\mu\{dx\} - \int_{I_j} v(x)\mu_n\{dx\} + \int_{I_j} (v(x) - u(x))\mu_n\{dx\}\right|$
$\leq \sum_{j=1}^k \int_{I_j} |u(x) - v(x)|\,\mu\{dx\} + M\varepsilon + \sum_{j=1}^k \int_{I_j} |v(x) - u(x)|\mu_n\{dx\}$
$\leq (\mu\{I\} + M + B)\varepsilon$

q.e.d. (proposition 2.8) $\Diamond\Diamond\Diamond$

Proof of lemma 2.10
A non negative function z is Lebesgue integrable if:
$\sup\left(\sum_{i\in\mathbb{N}} \inf\left(z(x) \mid x \in E_i\right) \cdot \mu(E_i) \mid \{E_i\}_{i\in\mathbb{N}} \text{ countable and measurable partition of } \mathbb{R}^+\right) < \infty$
with μ the Lebesgue measure on the real line.
So, $\lim_{h\to 0} h \sum_{k\in\mathbb{N}} m_{kh}$ is finite.
Assume that z is non increasing. Then $h\sum_{k\in\mathbb{N}} M_{kh} \leq hM_{0h} + h\sum_{k\in\mathbb{N}} m_{kh}$. So the two limits converge and are equal.

q.e.d. (proposition 2.10) $\Diamond\Diamond\Diamond$

2.4.2 Proofs of section 2.1.2

Proof of lemma 2.11
Define $U_n \stackrel{\text{def}}{=} \sum_{k=0}^n F^{k\star}$. One observes that:

$$\int_0^x (1 - F(x-y))\, U_n\{dy\} = 1 - F^{(n+1)\star}(x) \leq 1$$

Pick $\tau > 0$ and $\eta > 0$ such that $1 - F(\tau) \geq \eta$. Then:

$$\eta(U_n(x) - U_n(x-\tau)) = \eta \int_{x-\tau}^x U_n\{dy\} \leq \int_{x-\tau}^x (1 - F(x-y))\, U_n\{dy\} \leq 1$$

Letting n go to ∞, one concludes that $U\{I\}$ is bounded by η^{-1} on every interval I with length τ and that $U\{I\}$ is bounded by $C_h \stackrel{\text{def}}{=} \eta^{-1}(1 + \lfloor \frac{h}{\tau} \rfloor)$ on every interval I with length h.
In particular $U(x) \leq C_x$ is finite.

q.e.d. (lemma 2.11) $\Diamond\Diamond\Diamond$

Proof of lemma 2.12
One observes that if a and b belong to Σ then $a + b \in \Sigma$ (proof left to the reader).

Let us suppose that $\delta \stackrel{\text{def}}{=} \inf(b - a \mid a < b \in \Sigma) > 0$. Pick some pair $a, b \in \Sigma$ such that $h \stackrel{\text{def}}{=} b - a < 2\delta$. Let $n \in \mathbb{N}$ be such that $nb \geq (n+1)a$, then we claim that:

$$\Sigma \cap [na, (n+1)a] = \{na + kh \mid na \leq na + kh \leq (n+1)a\}$$

Since $na + kh = (n-k)a + kb$, one deduces that: $\{na + kh \mid na \leq na + kh \leq (n+1)a\} \subseteq \Sigma \cap [na, (n+1)a]$. Assume that there exists $c \in \Sigma \cap [na, (n+1)a] \setminus \{na + kh \mid na \leq na + kh \leq (n+1)a\}$. Thus there exists k such that $na + kh < c < na + (k+1)h$ and so either $c - (na + kh) \leq \frac{b-a}{2} < \delta$ or $na + (k+1)h - c \leq \frac{b-a}{2} < \delta$ yielding in both cases a contradiction.

Since $(n+1)a \in \Sigma$, one deduces that a and b are multiples of h. Let c be another increasing point and let n be chosen such that $nh \geq c$. Then $na \leq na + c \leq nb$. So, c is also a multiple of h.

Thus we have established that for all $\varepsilon > 0$, there exists $a < b$ two points of Σ with $b - a < \varepsilon$. As previously observed, for n_0 large enough, $\bigcup_{n \geq n_0} [na, nb] = [n_0 a, \infty[$. This concludes the proof.

q.e.d. (lemma 2.12) $\Diamond\Diamond\Diamond$

Proof of lemma 2.13
By induction, one obtains that $g(x) = \int_0^\infty g(x - y) F^{k\star}\{dy\}$. So the equality is possible only if $g(-y) = g(0)$ for all $y \in \Sigma$, defined as in lemma 2.12. Since Σ is asympotically dense and g is uniformly continuous one deduces that $\lim_{y \to \infty} g(-y) = g(0)$.

There exists $\delta > 0$ such that $F(\delta) < 1$. Observe that $\sum_{i<n} T_i \leq k\delta$ implies that $|\{i \mid i < n \wedge T_i \geq \delta\}| \leq k$ or equivalently $|\{i \mid i < n \wedge T_i < \delta\}| \geq n - k$. So, for all k and for all n:

$$F^{n\star}(k\delta) \leq \tbinom{n}{n-k} F(\delta)^{n-k}$$

Hence for all k, $\lim_{n \to \infty} F^{n\star}(k\delta) = 0$.

Let x and $\varepsilon > 0$ be arbitrarily chosen, there exists y_0 such that $\forall y \geq y_0 \; g(x - y) \geq g(0) - \varepsilon$.
Let n be such that $F^{n\star}(y_0) \leq \varepsilon$. One obtains:

$$g(x) = \int_0^\infty g(x - y) F^{n\star}\{dy\} \geq \int_{y_0}^\infty g(x - y) F^{n\star}\{dy\} \geq (g(0) - \varepsilon)(1 - \varepsilon)$$

Letting ε go to 0, $g(x) \geq g(0)$. Since $g(x) \leq g(0)$, one deduces that $g(x) = g(0)$.

q.e.d. (lemma 2.13) $\Diamond\Diamond\Diamond$

Proof of lemma 2.14
By construction, $U_n(x)$ defined as in lemma 2.11 converges to $U(x)$. Define $H_n \stackrel{\text{def}}{=} U_n \star z$, $H_n(x)$ goes to $(U \star z)(x)$ uniformly on every bounded interval:

$$|(U \star z)(x) - H_n(x)| \leq \sup_{y \leq x}(|z(y)|)(U(x) - U_n(x))$$

One observes that $H_{n+1} = z + F \star H_n$.
Since the convergence of H_n is uniform on every bounded interval, one deduces that $F \star H_n$ converges to $F \star (U \star z)$. So $U \star z = z + F \star (U \star z)$ which proves that Z is a solution of (2.1).

Let V be the difference of two solutions of 2.1 bounded on every bounded interval. V fulfills $V = F \star V$. By induction $V = F^{k\star} \star V$. Since $F^{k\star}$ converges to 0 (U is finite) and V is bounded, letting k go to infinity, one gets $V = 0$.

q.e.d. (lemma 2.14) $\Diamond\Diamond\Diamond$

Proof of lemma 2.15
First, we suppose that z has a continuous derivative.

Let us consider the interval $[0, M]$ with $M \geq x + \delta$. For all y, one has:
$\delta^{-1}(z(x-y) - z(x+\delta-y)) = z'(x+\delta_x - y)$ with $0 < \delta_x < \delta$
z' is uniformly continuous in $[0, M]$. So let $\varepsilon > 0$. There exists $\delta > 0$ such that:
$|z'(u) - z'(v)| \leq \varepsilon$ when $|u - v| \leq \delta$

So:
$|\int_0^\infty \delta^{-1}(z(x-y) - z(x+\delta-y))U\{dy\} - \int_0^\infty z'(x-y)U\{dy\}|$
$= |\int_0^M z'(x-y+\delta_x) - z'(x-y)U\{dy\}| \leq U(M)\varepsilon.$

Hence $Z(x)$ has a derivative $\int_0^\infty z'(x-y)U\{dy\}$ which fulfills the renewal equation corresponding to z'.

Let v be a continuous function whose support is included in $[0, h]$ and V be the corresponding solution of the renewal equation. The support of function $v(x+\delta) - v(x)$ is included in an interval of length $h + 2\delta$ and so: $|V(x+\delta) - V(x)| \leq C_{h+2\delta} \sup |v(x+\delta) - v(x)|$. This establishes that V is uniformly continuous since v is uniformly continuous. Applying this reasoning to z' establishes that Z' is uniformly continuous. Moreover, Z' is bounded by $\sup_x(|z'(x)|)C_h$.

So, $\eta \stackrel{\text{def}}{=} \limsup_{x\to\infty} Z'(x)$ is finite. Let us pick a sequence t_n such that $Z'(t_n)$ converges to η. The family of functions $\zeta_n(x) \stackrel{\text{def}}{=} Z'(t_n + x)$ is equicontinuous. Using proposition 2.4, one extracts a subsequence $Z'(t_{n_r} + x)$ which converges, uniformly on every bounded interval, towards a limit ζ uniformly continuous (and bounded).

ζ_n fulfills equation:

$$\zeta_n(x) = z'(t_n + x) + \int_0^\infty \zeta_n(x-y)F\{dy\}$$

Letting n go to infinity and using the dominated convergence theorem yields:

$$\zeta(x) = \int_0^\infty \zeta(x-y)F\{dy\}$$

ζ fulfills the hypotheses of lemma 2.13 (since $\zeta(0) = \eta$). So for all x, $Z'(t_{n_r} + x) \to \eta$ uniformly on every bounded interval. Fix some arbitrary a. Since $Z(t_{n_r} + a) - Z(t_{n_r}) = Z'(t_{n_r} + x)a$ for some $x \in [0, a]$, one deduces that $\lim_{r\to\infty} Z(t_{n_r} + a) - Z(t_{n_r}) = a\eta$. Z is bounded implying $\eta = 0$. The same reasoning yields: $\liminf_{x\to\infty} Z'(x)$=0. So $\lim_{x\to\infty} Z'(x)$ exists and is equal to 0. Using the mean value theorem, the result is proved for a continuously derivable function z.

Every continuous function with support in $[0, h]$ can be approximated within $\varepsilon > 0$ by a continuously derivable fonction z_1, with support in $[0, h]$. Let Z_1 be the corresponding solution of the renewal equation. Since $|z(x) - z_1(x)| \leq \varepsilon$, one has $|Z(x) - Z_1(x)| \leq C_h\varepsilon$. Fix some arbitrary a. For x large enough, $|Z_1(x + a) - Z_1(x)| \leq \varepsilon$. So $|Z(x + a) - Z(x)| \leq (2C_h + 1)\varepsilon$ which yields the result.

q.e.d. (lemma 2.15) ◊◊◊

Proof of proposition 2.16
Let $h > 0$ be fixed then:
$\lim_{n\to\infty} U(t_n) - U(t_n - h) = h\eta$ and $\lim_{n\to\infty} U(t_n) - U(t_n - 2h) = 2h\eta$
Thus: $\lim_{n\to\infty} U(t_n - h) - U(t_n - 2h) = h\eta$
By induction for any k, $\lim_{n\to\infty} U(t_n - kh) - U(t_n - (k+1)h) = h\eta$

Let z_{kh} be the indicator function of interval $[kh, kh + h[$ and Z_{kh} be the corresponding solution of the renewal equation. For a given z, m_{kh} and M_{kh} are defined as in the direct Riemman integration.

$Z_{kh}(x) = U(x - kh) - U(x - (k+1)h) \leq C_h$ for all k and all x.
So, the infinite sum of fonctions $Z_h^m \stackrel{\text{def}}{=} \sum_k m_{kh}Z_{kh}$ and $Z_h^M \stackrel{\text{def}}{=} \sum_k M_{kh}Z_{kh}$ are finite and constitute a frame for Z.

For all $\varepsilon > 0$, there exists k_0 such that one has $\sum_{k \geq k_0} M_{kh} \leq \varepsilon$.
Pick n_0 such that for all $n \geq n_0$ and all $k < k_0$, $|U(t_n - kh) - U(t_n - (k+1)h) - h\eta| \leq \varepsilon / k_0$.
Then:
$|\sum_k M_{kh} Z_{kh}(t_n) - \sum_k h\eta M_{kh}| \leq$
$|\sum_{k<k_0} M_{kh} Z_{kh}(t_n) - \sum_{k<k_0} h\eta M_{kh}| + |\sum_{k \geq k_0} M_{kh} Z_{kh}(t_n)| + |h\eta \sum_{k \geq k_0} M_{kh}|$
$\leq (1 + C_h + h\eta)\varepsilon$
One concludes that $\lim_{n\to\infty} Z_h^M(t_n) = \eta \sum_k h M_{kh}$.
By a similar reasoning $\lim_{n\to\infty} Z_h^m(t_n) = \eta \sum_k h m_{kh}$.
So, every limit l of convergent subsequence of $Z(t_n)$ fulfills:

$$\eta \sum_k h m_{kh} \leq l \leq \eta \sum_k h M_{kh}$$

Letting h go to 0, one obtains $l = \eta \int_0^\infty z(x)dx$. Since the $Z(t_n)$'s are bounded, one concludes that $\lim_{n\to\infty} Z(t_n)$ exists (lemma 2.1) and is equal to $\eta \int_0^\infty z(x)dx$.

q.e.d. (proposition 2.16) ◊◊◊

Proof of theorem 2.17
Let M be an arbitrary measure. Denote by M_t the measure defined by $M_t\{I\} \stackrel{\text{def}}{=} M\{I + t\}$ where I is an interval and $I + t$ is the translation by t of interval I.

Let I be a bounded interval, we know by lemma 2.11 that $\sup_{t\in\mathbb{R}} U_t\{I\}$ is finite. Applying proposition 2.6, one deduces that there exists a sequence $t_k \to \infty$ such that $U_{t_k}\{I\}$ converges toward a measure V.

Let z be a continuous function whose support is included in $[0, a]$ and Z be the corresponding solution of the renewal equation. Let $x \geq 0$, using proposition 2.8, one gets:

$$\lim_{k\to\infty} Z(t_k+x+a) = \lim_{k\to\infty} \int_{t_k+x}^{t_k+x+a} z(t_k+x+a-y)U\{dy\} = \lim_{k\to\infty} \int_0^a z(a-y)U_{t_k+x}\{dy\} = \int_0^a z(a-y)V_x\{dy\}$$

Since $\lim_{k\to\infty} Z(t_k + x + a) = \lim_{k\to\infty} Z(t_k + a)$ (lemma 2.15), one deduces that V and V_x are equal on continuous functions with bounded support. Using lemma 2.7, $V = V_x$. So measure V is invariant by translation. We let the reader prove that $V\{I\}$ is proportional to the length of I for every bounded interval I. Let γ be the proportionality factor, one deduces that:
$\lim_{k\to\infty} U(t_k + h) - U(t_k) = h\gamma$ (every interval is a continuity interval of V).

Using proposition 2.16, one deduces that for every z directly Riemann integrable function and Z the corresponding solution of the renewal equation, one has:

$$\lim_{k\to\infty} Z(t_k) = \gamma \int_0^\infty z(y)dy$$

Observe that function $z(x) = 1 - F(x)$ for $x \geq 0$ is non negative and non increasing. Z the solution of the renewal equation is the constant 1 (proof left to the reader). Moreover $\int_0^\infty z(y)dy$ is equal to μ. If $\mu < \infty$ then z is directly Riemmann integrable (proposition 2.10) and so $\mu\gamma = 1$. If $\mu = \infty$, one truncates z and concludes that $\gamma \int_0^a z(y)dy \leq 1$ for all a which implies $\gamma = 0$.

So γ is independent from the convergent subsequence and using proposition 2.6, the result is established.

q.e.d. (theorem 2.17) ◊◊◊

2.4.3 Proofs of section 2.2.2

Proof of proposition 2.23
In order to have a state change, at least one event must occur. So:
$\pi_{ii}(\tau) \geq e^{-\lambda_i \tau}$ and thus $\lim_{\tau\downarrow 0} \pi_{ii}(\tau) = 1$
Since for $j \neq i$,
$\pi_{ii}(\tau) + \pi_{ij}(\tau) \leq 1$ one obtains $\lim_{\tau\downarrow 0} \pi_{ij}(\tau) = 0$

One has (see (2.2)):
$\pi_{ij}(\tau + d\tau) = \sum_k \pi_{ik}(\tau)\pi_{kj}(d\tau)$
Since $\sum_k \pi_{ik}(\tau) \leq 1$ and for all k, $0 \leq \pi_{kj}(\tau) \leq 1$:
$\lim_{d\tau\downarrow 0} \pi_{ij}(\tau + d\tau) = \sum_k \pi_{ik}(\tau) \lim_{d\tau\downarrow 0} \pi_{kj}(d\tau) = \pi_{ij}(\tau)$
Thus π_{ij} is right-continuous and so measurable. We introduce a "renewal equation" (justified by measurability of π_{ij}). It is based on a case decomposition w.r.t. the (possible) occurrence of the first event in $[0, \tau]$:

$$\pi_{ij}(\tau) = \mathbf{1}_{i=j} e^{-\lambda_i \tau} + \sum_k \lambda_i p_{ik} \int_0^\tau e^{-\lambda_i(\tau - x)} \pi_{kj}(x) dx = e^{-\lambda_i \tau} \left(\mathbf{1}_{i=j} + \sum_k \lambda_i p_{ik} \int_0^\tau e^{\lambda_i x} \pi_{kj}(x) dx \right)$$

Every integral is a continuous function of τ. Furthermore, the infinite sum of functions is normally convergent (bounded by $\sum_k \lambda_i p_{ik} \tau e^{\lambda_i \tau}$). So the infinite sum is continuous implying the continuity of π_{ij}.
Since the π_{ij}'s are continuous, any integral is differentiable and its derivative is equal to $e^{\lambda_i \tau} \pi_{kj}(\tau)$. Due to the normal convergence of the sum of derivatives (bounded by $\sum_k \lambda_i p_{ik} e^{\lambda_i \tau}$), the infinite sum is differentiable implying the differentiability of π_{ij}.
Let us compute the derivative of π_{ij}:

$$\frac{d\pi_{ij}(\tau)}{d\tau} = e^{-\lambda_i \tau} \left(\sum_k \lambda_i p_{ik} e^{\lambda_i \tau} \pi_{kj}(\tau) \right) - \lambda_i \pi_{ij}(\tau) = \sum_k q_{ik} \pi_{kj}(\tau)$$

q.e.d. (proposition 2.23) ◊◊◊

Proof of lemma 2.24
Let us consider the i.i.d random variables T'_n following an exponential distribution with rate α. By hypothesis,

$$\mathbf{Pr}(\sum_{m \leq n} T_m \leq \Delta) \leq \mathbf{Pr}(\sum_{m \leq n} T'_m \leq \Delta)$$

We prove by induction that g_n, the density function of $\sum_{m \leq n} T'_m$, is the following one:

$$g_n(x) = \alpha e^{-\alpha x} \frac{(\alpha x)^n}{n!}$$

The basis case $n = 0$ follows from the definition of the exponential distribution. Now:

$$g_{n+1}(x) = \int_0^x g_n(x - \tau) g_0(\tau) d\tau = \int_0^x \alpha e^{-\alpha(x-\tau)} \frac{(\alpha(x-\tau))^n}{n!} \alpha e^{-\alpha\tau} d\tau$$

$$= \alpha e^{-\alpha x} \int_0^x \alpha \frac{(\alpha(x-\tau))^n}{n!} d\tau = \alpha e^{-\alpha x} \frac{(\alpha x)^{n+1}}{n+1!}$$

The reader can check (by derivation) that the corresponding distribution is then defined by:

$$\mathbf{Pr}(\sum_{m \leq n} T'_m \leq \Delta) = 1 - e^{-\alpha\Delta} \sum_{0 \leq m \leq n} \frac{(\alpha\Delta)^m}{m!}$$

q.e.d. (lemma 2.24) ◊◊◊

Proof of proposition 2.25
Let us express $\pi_{ij}(\tau + d\tau)$ conditionally w.r.t. $\pi_{ik}(\tau)$.

$$\pi_{ij}(\tau + d\tau) = \sum_k \sum_{n \in \mathbb{N}} \pi_{ik}(\tau) \mathbf{Pr} \left(\sum_{m \leq n} T_m \leq d\tau < \sum_{m \leq n+1} T_m \wedge S_n = s_j \mid S_0 = s_k \right)$$

Thus considering $n = 0$, $n = 1$ and $n \geq 2$:

$$\frac{1}{d\tau}\left|\pi_{ij}(\tau+d\tau) - \pi_{ij}(\tau)(1-(1-p_{jj})(1-e^{-\lambda_j d\tau})) - \sum_{k\neq j}\pi_{ik}(\tau)p_{kj}(1-e^{-\lambda_k d\tau})\right|$$

$$\leq \frac{2}{d\tau}\sum_k \pi_{ik}(\tau)\mathbf{Pr}(T_0+T_1 \leq d\tau \mid S_0 = s_k)$$

$$\leq \frac{2}{d\tau}(1-e^{-\alpha d\tau}(1+\alpha d\tau))$$

(here we have used lemma 2.24)

$$\leq \frac{2}{d\tau}(1-(1-\alpha d\tau)(1+\alpha d\tau)) = 2\alpha^2 d\tau$$

So

$$\lim_{d\tau\to 0}\frac{1}{d\tau}\left(\pi_{ij}(\tau+d\tau) - \pi_{ij}(\tau)(1-(1-p_{jj})(1-e^{-\lambda_j d\tau})) - \sum_{k\neq j}\pi_{ik}(\tau)p_{kj}(1-e^{-\lambda_k d\tau})\right) = 0$$

Now

$$\frac{d\pi_{ij}(\tau)}{d\tau} = \lim_{d\tau\to 0}\frac{1}{d\tau}\left(\pi_{ij}(\tau)(p_{jj}-1)(1-e^{-\lambda_j d\tau}) + \sum_{k\neq j}\pi_{ik}(\tau)p_{kj}(1-e^{-\lambda_k d\tau})\right)$$

$$= \pi_{ij}(\tau)(p_{jj}-1)\lambda_j + \sum_{k\neq j}\pi_{ik}(\tau)\lambda_k p_{kj}$$

(to invert the sum and the limit we have used the dominated convergence theorem since $\sum_{k\neq j}\pi_{ik} \leq 1$ and $\frac{1-e^{-\lambda_k d\tau}}{d\tau} \leq \alpha$)

q.e.d. (proposition 2.25) ◊◊◊

2.4.4 Proofs of section 2.2.3

Proof of theorem 2.26

Let $z_i(\tau)$ be the (non increasing) probability to stay in i during at least τ time units: $z_i(\tau) \stackrel{\text{def}}{=} e^{-\lambda_i\tau}$. Furthermore $\int_0^\infty z_i(\tau)d\tau = \frac{1}{\lambda_i}$ is finite.

$\pi_{ii}(\tau)$ fulfills the renewal equation:

$$\pi_{ii}(\tau) = z_i(\tau) + \int_0^\tau \pi_{ii}(\tau-y)F\{dy\}$$

where F is the distribution of the return time to i. Since F is the convolution the sojourn time in i, a continuous distribution and an arbitrary distribution, F is continuous, hence non arithmetic. Using renewal theorem 2.18:

$$\lim_{\tau\to\infty}\pi_{ii}(\tau) = \frac{1}{D_i}\int_0^\infty e^{-\lambda_i\tau}d\tau = \frac{1}{\lambda_i D_i}$$

Let i be a transient state. For all $\epsilon > 0$, there is an integer n_0 such that the probability of n_0 visits to i is less than ε. Let us define the (possibly infinite) random variable $T_{i,n}$ of time entrance in i at the n^{th} visit. There is a time d such that for all $n < n_0$, $\mathbf{Pr}(d \leq T_{i,n} < \infty) \leq \frac{\varepsilon}{n_0}$. In addition there is a time d' such that the probability to stay in i at least d' is less or equal than $\frac{\varepsilon}{n_0}$. So $\pi_{i,i}(\tau) \leq \mathbf{Pr}(\{\exists u \geq \tau\ X(u) = i\}) \leq 3\varepsilon$ for $\tau \geq d + d'$ which implies that $\lim_{\tau\to\infty}\pi_{ii}(\tau) = 0$.

Let i be a recurrent state then it is null recurrent iff $D_i = \infty$. We have thus established the characterization.

q.e.d. (theorem 2.26) ◊◊◊

Proof of proposition 2.27
Assume that i is positive recurrent. There is a path from j to i and vice versa. So given an arbitrary $\delta > 0$, $\pi_{ji}(\delta) > 0$ and $\pi_{ij}(\delta) > 0$.
$\pi_{jj}(\tau + 2\delta) \geq \pi_{ji}(\delta)\pi_{ii}(\tau)\pi_{ij}(\delta)$ implies: $\lim_{\tau\to\infty} \pi_{jj}(\tau + 2\delta) > 0$.
So j is positive recurrent.

q.e.d. (proposition 2.27) ◊◊◊

Proof of theorem 2.28
The existence and unicity of vector $\mathbf{v}$ is proved by theorem 1.20.
One has: $\forall i\ (p_{ii} - 1)\mathbf{v}_i + \sum_{j\neq i} p_{ji}\mathbf{v}_j = 0$
Thus: $\forall i\ (p_{ii} - 1)\lambda_i\mathbf{u}_i + \sum_{j\neq i} \lambda_j p_{ji}\mathbf{u}_j = 0$
Which yields: $\forall i\ q_{ii}\mathbf{u}_i + \sum_{j\neq i} q_{ji}\mathbf{u}_j = 0$
The other statement of the theorem is obtained by observing that the transformation of equations can be done in the converse direction.

q.e.d. (theorem 2.28) ◊◊◊

Proof of theorem 2.29
Let r be some state of the CTMC. Let G be the distribution (which depends on the initial distribution of the CTMC) of the time to reach r and $Y_r(\tau)$ the probability to be in r at time τ *after a visit in* r. $Y_r = G \star Z_r$. Applying theorem 2.19 related to the delayed renewal process, one obtains:

$$\lim_{\tau\to\infty} Y_r(\tau) = \frac{1}{\lambda_r D_r}$$

Since the probability to reach r is 1, this limit is also the limit of the probability to be in r at time τ when $\tau \to \infty$. So the transient distribution has a limit independent from the initial distribution.
If $D_r = \infty$, (i.e. r is null recurrent) then the limit of the transient distribution is null.
Otherwise observe that $D_r = \sum_i \frac{{}_r\pi_{ri}}{\lambda_i}$, or equivalently $1 = \frac{1}{D_r}\sum_i \frac{{}_r\pi_{ri}}{\lambda_i}$.
We know by theorem 1.20 that whatever s, $({}_s\pi_{si})_i$ is an invariant vector of the embedded DTMC and that all these vectors are proportional. So:

$$\frac{{}_r\pi_{ri}}{D_r} = \frac{{}_s\pi_{si}}{D_s} \text{ and then } \frac{{}_r\pi_{ri}}{D_r} = \frac{{}_i\pi_{ii}}{D_i} = \frac{1}{D_i}$$

Thus:

$$1 = \sum_i \frac{1}{D_i\lambda_i}$$

which shows that the limits $\left(\frac{1}{\lambda_i D_i}\right)_i$ constitute a steady-state distribution. Moreover applying theorem 2.28, it is the single solution of $\mathbf{u}\cdot\mathbf{Q} = 0 \wedge \mathbf{u}\cdot\mathbf{1}^T = 1 \wedge \mathbf{u} \geq 0$.
Conversely let $\mathbf{v} \neq 0$ be such that for all i, $\mathbf{v}_i \geq 0$ and $\mathbf{v} = \mathbf{v}\cdot\mathbf{P}$. Then for some $\alpha > 0$, $v_i = \alpha\ {}_r\pi_{ri}$. Since $D_r = \sum_i \frac{{}_r\pi_{ri}}{\lambda_i}$, it is finite iff $\sum_i \frac{v_i}{\lambda_i}$ is finite.

q.e.d. (theorem 2.29) ◊◊◊

Proof of proposition 2.30
Assume that the chain is positive recurrent. Then theorem 2.29 allows to conclude.

Assume there exists $\mathbf{u}$ such that $\mathbf{u}\cdot\mathbf{Q} = 0$ for all i, $\mathbf{u}_i > 0$ and $\sum_{i\in S}\mathbf{u}_i$ is finite. Define $\mathbf{v}_i \stackrel{\text{def}}{=} \mathbf{u}_i\lambda_i$. Then $\mathbf{v}\cdot\mathbf{P} = \mathbf{v}$ and $\sum_{i\in S}\mathbf{v}_i \leq \sup_i(\lambda_i)\sum_{i\in S}\mathbf{u}_i$ is finite.

So applying proposition 1.21, one obtains that the embedded DTMC is positive recurrent and so the CTMC is recurrent. Applying now theorem 2.29, one finally gets that it is positive recurrent.

q.e.d. (proposition 2.30) ◊◊◊

Chapter 3

Markov Decision Processes [PUT 94]

3.1 Presentation

The previous models of these lecture notes are purely probabilistic. In this chapter (and in the following one), we consider models that present both non deterministic and probabilistic features. There are several interests of such simultaneous features. Let us present two examples.

Example 3.1 (The spinner game) *In this game, the player has to compose a five-digit number whose digits are randomly chosen by a spinner during five rounds. After every round (except the last one), the player chooses in which position he inserts the current digit. The goal of the player is to obtain the largest number as possible. In figure 3.1, the spinner has successively output 3 placed by the player in the fifth position and 6 placed by the player in the second position.*

Example 3.2 (Management of a stock) *The manager of a stock in a warehouse with fixed capacity decides at the beginning of every month, which additional stock he will order. Then the monthly commands randomly arrive following some distribution. If the commands exceed the inventory the commands are lost. Every unit of a stock has a monthly cost while selling it provides a benefit. The aim of the manager is to maximize the expected profit during a year including the value of the stock at the end of the year. In figure 3.2, the activity of two months is presented with an excess of the demands during the second month.*

An MDP is a transition system. In this chapter, we only consider a finite number of states and transitions. The dynamic of the system is defined as follows. Non deterministically, one chooses an action which is enabled in the current state. Then one selects randomly the next state. The corresponding distribution depends on the current state and on the selected action. As seen in the previous examples MDP's have been introduced to specify optimization problems. So there is a numerical reward associated with every pair of (current) state and (selected) action. In addition,

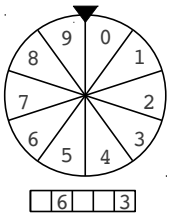

Figure 3.1: The spinner game

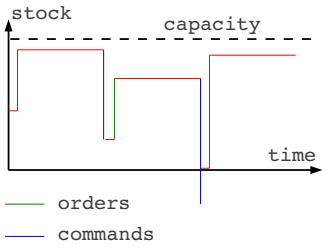

Figure 3.2: Management of a stock

when one considers optimization problems with finite horizon, there is a reward associated with every (terminal) state.

Definition 3.3 *A* Markov decision process *(*MDP*)* $\mathcal{M} = (S, \{A_s\}_{s\in S}, p, r, rend)$ *is defined by:*

- S*, the finite set of states;*
- *For every state* s*,* A_s*, the finite set of actions enabled in* s*.*
  We denote $A \stackrel{def}{=} \bigcup_{s\in S} A_s$*, the whole set of actions.*
- p*, a mapping from* $\{(s,a) \mid s \in S, a \in A_s\}$ *to the set of distributions over* S*. The conditional probability transition* $p(s'|s,a)$ *denotes the probability to go from* s *to* s' *if* a *is selected.*
- r*, a mapping from* $\{(s,a) \mid s \in S, a \in A_s\}$ *to* $\mathbb{R}$*.* $r(s,a)$ *is the reward associated with the selection of* a *in state* s*.*
- $rend$*, a mapping from* S *to* $\mathbb{R}$*.* $rend(s)$ *is the reward obtained when ending in state* s*.*

Example 3.4 (A simple MDP**)** *Figure 3.3 depicts a* MDP *with two states* s_1 *and* s_2*. In* s_1 *actions* a *and* b *are enabled while in* s_2 *only action* a *is possible. An edge from a state to another one is labelled by (1) the action that has triggered the transition, (2) the probability that this transition is selected given the chosen action and, (3) the reward associated with the source state and the transition. For instance, the transition labelled by* $(a, 0.7, 5)$ *means that when* a *is chosen in state* s_1*, the probability that the next state is* s_2*,* $p(s_2|s_1,a)$*, is equal to* 0.7 *and the reward* $r(s_1,a)$ *is equal to 5. The outgoing edge from a state with no destination is labelled by its terminal reward. There are some redundant information in the graph and this suggests that the reward could also depend on the destination state. We discuss this topic later on.*

A history is a possible finite or infinite execution of the MDP.

Definition 3.5 *Given an* MDP $\mathcal{M}$*, a* history *is a finite or infinite sequence alternating states and actions* $\sigma = (s_0, a_0, \ldots, s_i, a_i, \ldots)$*.* $\lg(\sigma)$ *denotes the number of actions of* σ*. One requires that for all* $0 \leq i < \lg(\sigma)$*,* $p(s_{i+1}|s_i,a_i) > 0$*.*

We now introduce the three main criteria related to optimization problems specified by MDP's. The total reward of a finite history consists in the sum of the rewards of the selected actions and the reward of the final state. The discounted reward of an infinite history consists in the sum of the rewards of the selected actions discounted by a multiplicative factor relative to the time. Since the rewards are bounded, this infinite sum is well defined. The average reward of an infinite history is the limit of the average reward of its finite sub-histories. Since this limit does not necessarily exist, we consider both the liminf and limsup which are finite due to the boundedness of rewards.

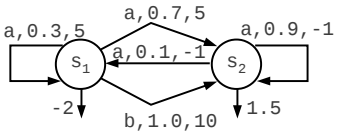

Figure 3.3: A simple MDP

Definition 3.6 *Let σ be an history of an* MDP $\mathcal{M}$ *and* $0 < \lambda < 1$. *Then:*

- *When* $\lg(\sigma) < \infty$, *the* total reward *of* σ *is* $u(\sigma) \stackrel{def}{=} \sum_{0 \leq i < \lg(\sigma)} r(s_i, a_i) + rend(s_{\lg(\sigma)})$.
  We also denote $v(\sigma) \stackrel{def}{=} \sum_{0 \leq i < \lg(\sigma)} r(s_i, a_i)$ *the* pure total reward *which does not take into account the final reward.*
- *When* $\lg(\sigma) = \infty$, *the* discounted reward *of* σ *w.r.t.* λ *is* $v_\lambda(\sigma) \stackrel{def}{=} \sum_{0 \leq i} r(s_i, a_i)\lambda^i$.
- *When* $\lg(\sigma) = \infty$, *the* lim sup average reward *of* σ *is* $g_+(\sigma) \stackrel{def}{=} \limsup_{n \to \infty} \frac{1}{n} \sum_{0 \leq i < n} r(s_i, a_i)$.
- *When* $\lg(\sigma) = \infty$, *the* lim inf average reward *of* σ *is* $g_-(\sigma) \stackrel{def}{=} \liminf_{n \to \infty} \frac{1}{n} \sum_{0 \leq i < n} r(s_i, a_i)$.

In order to obtain a stochastic process, we need to fix the non deterministic features of the MDP. This is done via (1) decision rules that select at some time instant the next action depending on the history of the execution, and (2) policies which specify which decision rules should be used at any time instant. Different classes of decision rules and policies are defined depending on two criteria: (1) the information used in the history and (2) the way the selection is performed (deterministically or randomly).

Definition 3.7 *Given an* MDP $\mathcal{M}$ *and* $t \in \mathbb{N}$, *a* decision rule d_t *associates with every history* σ *of length* $t < \infty$, *a distribution* $d_t(\sigma)$ *over* $A_{s_{\lg(\sigma)}}$.

- *The set of all (and is also called* history-dependent randomized*) decision rules at time* t *is denoted* D_t^{HR} *and is also called* history-dependent randomized *decision rules.*
- *The subset of* history-dependent deterministic *decision rules at time* t, D_t^{HD} *consists in selecting a single action. In this case* $d_t(\sigma) \in A_{s_{\lg(\sigma)}}$.
- *The subset of* Markovian randomized *decision rules at time* t, D_t^{MR} *(also denoted* D^{MR}*) only depends on the final state of the history. So one denotes* $d_t(s)$ *the distribution that depends on* s.
- *The subset of* Markovian deterministic *decision rules at time* t, D_t^{MD} *(also denoted* D^{MD}*) only depends on the final state of the history and selects a single action. So one denotes* $d_t(s)$ *this action belonging to* A_s.

Given a Markovian decision rule d, the vector $\mathbf{r}_d$, defined by $\mathbf{r}_d[s] \stackrel{\text{def}}{=} \sum_{a \in A_s} d(a) r(s, a)$, represents the immediate expected reward obtained by d.

Definition 3.8 *Given an* MDP $\mathcal{M}$ *and* $t \in \mathbb{N}$, *a* policy *(also called a* strategy*)* $\boldsymbol{\pi}$ *is a finite or infinite sequence of decision rules* $\boldsymbol{\pi} = (d_0, \ldots, d_t, \ldots)$ *such that* d_t *is a decision rule at time* t.
The set of policies such that for all t, $d_t \in D_t^K$ *is denoted* Π^K.
When decisions d_t *are Markovian and all equal to some* d, $\boldsymbol{\pi}$ *is said* stationary *and denoted* d^∞.
The set of stationary randomized (resp. deterministic) policies is denoted Π^{SR} *(resp.* Π^{SD}*).*

Once a policy is chosen, an MDP becomes a DTMC whose states are histories. Assuming an initial distribution we denote X_n the random state of the MDP at time n and Y_n corresponding to the chosen action at time n. Observe that $\{X_n\}_{n\in\mathbb{N}}$ is not generally a DTMC. However when a stationary policy d^∞ is chosen, the states of the DTMC are those of the MDP and the transition matrix $\mathbf{P}_d$ is defined by:

$$\mathbf{P}_d[s,s'] \stackrel{\text{def}}{=} \sum_{a\in A_s} d(s)(a)p(s'|s,a)$$

Given a policy $\boldsymbol{\pi}$, we indicate the probabilities (resp. the expectations) induced by such a policy by $\mathbf{Pr}^{\boldsymbol{\pi}}$ (resp. $\mathbf{E}^{\boldsymbol{\pi}}$). When the policy is clear from the context, we omit the superscript. We are now in position to express the rewards produced by a policy.

Definition 3.9 *Let $\boldsymbol{\pi}$ be a policy of an* MDP *$\mathcal{M}$, $t\in\mathbb{N}$ and $0<\lambda<1$. Then:*

- *The* total (expected) reward *at time t of $\boldsymbol{\pi}$ is $u_t^{\boldsymbol{\pi}} \stackrel{def}{=} \sum_{0\le i<t}\mathbf{E}^{\boldsymbol{\pi}}(r(X_i,Y_i)) + \mathbf{E}^{\boldsymbol{\pi}}(rend(X_t))$.*
- *The* pure total (expected) reward *at time t of $\boldsymbol{\pi}$ is $v_t^{\boldsymbol{\pi}} \stackrel{def}{=} \sum_{0\le i<t}\mathbf{E}^{\boldsymbol{\pi}}(r(X_i,Y_i))$.*
- *The* discounted (expected) reward *of $\boldsymbol{\pi}$ w.r.t. λ is $v_\lambda^{\boldsymbol{\pi}} \stackrel{def}{=} \sum_{0\le i}\lambda^i\mathbf{E}^{\boldsymbol{\pi}}(r(X_i,Y_i))$.*
- *The* lim sup average (expected) reward *of $\boldsymbol{\pi}$ is $g_+^{\boldsymbol{\pi}} \stackrel{def}{=} \limsup_{n\to\infty}\frac{1}{n}\sum_{0\le i<n}\mathbf{E}^{\boldsymbol{\pi}}(r(X_i,Y_i))$.*
- *The* lim inf average (expected) reward *of $\boldsymbol{\pi}$ is $g_-^{\boldsymbol{\pi}} \stackrel{def}{=} \liminf_{n\to\infty}\frac{1}{n}\sum_{0\le i<n}\mathbf{E}^{\boldsymbol{\pi}}(r(X_i,Y_i))$.*

One observes that the initial distribution X_0 is independent from the policy and the first decision applies after the initial state is selected. So we focus on the vector of rewards indexed by the initial state. The scalar reward is then obtained by the sum of its components weighted by the initial distribution. So in the sequel, $\mathbf{u}_t^{\boldsymbol{\pi}}$, $\mathbf{v}_t^{\boldsymbol{\pi}}$, $\mathbf{v}_\lambda^{\boldsymbol{\pi}}$, $\mathbf{g}_+^{\boldsymbol{\pi}}$, $\mathbf{g}_-^{\boldsymbol{\pi}}$ denote such reward vectors. We also denote *optimal* vectors $\mathbf{u}_t^*$, $\mathbf{v}_t^*$, $\mathbf{v}_\lambda^*$, $\mathbf{g}_+^*$, $\mathbf{g}_-^*$ with the following definition: $\mathbf{u}_t^*[s] \stackrel{\text{def}}{=} \sup_{\boldsymbol{\pi}}(\mathbf{u}_t^{\boldsymbol{\pi}}[s])$ (the other definitions are similar).

In the sequel we will look for (almost) optimal policies w.r.t. some of the above rewards. The next result shows that we can safely restrict ourselves to Markovian policies. This will simplify both the notations and the theoretical developments.

Theorem 3.10 *Let $\boldsymbol{\pi}\in\Pi^{HR}$ be a policy of an* MDP *$\mathcal{M}$. Then there exists a policy $\boldsymbol{\pi}'\in\Pi^{MR}$ such that for all $n\in\mathbb{N}$, $s_0,s\in S$ and $a\in A_s$:*

$$\mathbf{Pr}^{\boldsymbol{\pi}'}(X_n=s,Y_n=a \mid X_0=s_0) = \mathbf{Pr}^{\boldsymbol{\pi}}(X_n=s,Y_n=a \mid X_0=s_0)$$

Proof

Let us consider the generalization of the model which consists in allowing the reward to be also dependent on the destination state specified by $r(s,a,s')$. Let us fix some current state s, some time instant t and some Markovian policy $\boldsymbol{\pi}=(d_0,d_1,\ldots)$. In the expression of rewards $\mathbf{E}^{\boldsymbol{\pi}}(r(X_i,Y_i))$ should be replaced by $\mathbf{E}^{\boldsymbol{\pi}}(r(X_i,Y_i,X_{i+1}))$ but this value is given by:

$$\sum_{a\in A_s} d_t(s)(a)\sum_{s'\in S'} p(s'|s,a)r(s,a,s')$$

So defining $r(s,a)$ by $r(s,a)\stackrel{\text{def}}{=}\sum_{s'\in S'}p(s'|s,a)r(s,a,s')$ and forgetting the individual rewards do not modify the optimization problem. Summarizing one allows modelling rewards to depend on the destination state while theoretical developments assume wlog that rewards are independent from the destination state.

3.2 Finite horizon analysis

Let us design on example 3.4 a procedure to compute an optimal policy for the total reward for time horizon 2. We are going to solve the problem not only for horizon 2 but also for horizons 0 and 1.

At horizon 0, there is no decision to take and the total reward is given by *rend*: -2 for state s_1 and 1.5 for state s_2.

At horizon 1, in state s_2 a single action a is possible and since we know the optimal values at horizon 0, we get as expected value $-1 + 0.9 * 1.5 + 0.1 * -2 = 0.15$. In state s_1, we consider successively actions a with expected value $5 + 0.7 * 1.5 + 0.3 * -2 = 5.45$ and b with expected value $10 + 1 * 1.5 = 11.5$. So the optimal decision is b with associated value 11.5.

At horizon 2, in state s_2 a single action a is possible and since we know the optimal values at horizon 1, we get as expected value $-1 + 0.9 * 0.15 + 0.1 * 11.5 = 0.285$. In state s_1, we consider successively actions a with expected value $5 + 0.7 * 0.15 + 0.3 * 11.5 = 8.555$ and b with expected value $10 + 1 * 0.15 = 10.15$. So the optimal decision is b with associated value 10.15.

state \ time	0	1	2
s_1	-2	a: 5.45 **b: 11.5**	a: 8.555 **b: 10.15**
s_2	1.5	**a: 0.15**	**a: 0.285**

Algorithm 2: Computing an optimal policy for the total expected reward

`TotalReward`$(\mathcal{M}, n)$
Input: $\mathcal{M}$, an MDP and n, a finite horizon

Output: *optval*, the optimal value given a state and an horizon $\leq n$
Output: *optdec*, the optimal decision given a state and positive horizon $\leq n$

Data: i integer, s, s' states, a action, $temp, best$ reals

for $s \in S$ **do** $optval[s, 0] \leftarrow rend(s)$
for i **from** 1 **to** n **do**
  for $s \in S$ **do**
    $best \leftarrow -\infty$
    for $a \in A_s$ **do**
      $temp \leftarrow r(s, a)$
      for $s' \in S$ **do** $temp \leftarrow temp + p(s'|s, a)optval[s', i-1]$
      if $best < temp$ **then** $best \leftarrow temp$; $optdec[s, i] \leftarrow a$
    end
    $optval[s, i] \leftarrow temp$
  end
end

Algorithm 2 generalizes this procedure. Arrays *optval* and *optdec* indexed by pairs of states and horizons respectively contain the optimal values and optimal decisions. They are filled by increasing horizons until the researched value. For time horizon 0, *optval* is filled with the *rend*'s values. Then at successive time horizons, *optdec* and *optval* are filled state per state comparing the reward obtained by choosing the enabled actions assuming that the optimal value at the next time unit is the one computed at the previous step.

This algorithm performs in polynomial time w.r.t. the size of the MDP and n. Hence it is pseudo-polynomial w.r.t. the size of the problem. The next proposition establishes its correction.

Proposition 3.11 *Algorithm 2 returns an optimal policy for the total expected reward.*

Proof

3.3 Discounted reward analysis

3.3.1 Characterization of optimality

Let $\boldsymbol{\pi} = (d_0, \ldots, d_n, \ldots)$ be some Markovian policy. Then summing over all instants, its discounted expected reward is:

$$\mathbf{v}_\lambda^{\boldsymbol{\pi}} = \sum_{i \in \mathbb{N}} \lambda^i \left(\prod_{0 \leq j < i} \mathbf{P}_{d_j} \right) \mathbf{r}_{d_i}$$

In case of a stationary policy d^∞, this reward can be rewritten as:

$$\mathbf{v}_\lambda^{\boldsymbol{\pi}} = \sum_{i \in \mathbb{N}} (\lambda \mathbf{P}_d)^i \, \mathbf{r}_d$$

Observing that the maximal modulus of an eigenvalue of a stochastic matrix is 1, for every $0 < \lambda < 1$, $\mathbf{Id} - \lambda \mathbf{P}_d$ is invertible and its inverse is $\sum_{i \in \mathbb{N}} (\lambda \mathbf{P}_d)^i$. So:

$$\mathbf{v}_\lambda^{\boldsymbol{\pi}} = (\mathbf{Id} - \lambda \mathbf{P}_d)^{-1} \, \mathbf{r}_d$$

and consequently

$$\mathbf{v}_\lambda^{\boldsymbol{\pi}} = \mathbf{r}_d + \lambda \mathbf{P}_d \mathbf{v}_\lambda^{\boldsymbol{\pi}} \tag{3.1}$$

Let us define $\mathbf{v}_\lambda^*[s] \stackrel{\text{def}}{=} \sup(\mathbf{v}_\lambda^{\boldsymbol{\pi}}[s] \mid \boldsymbol{\pi} \in \Pi^{MR})$. Observe that $\mathbf{v}_\lambda^*$ is a real vector over S. In order to compute this vector and (if possible) an associated policy, we introduce a transformation on such vectors.

Definition 3.12 *L is a mapping from $\mathbb{R}^S$ to $\mathbb{R}^S$ defined by:*

$$L(\mathbf{v})[s] \stackrel{def}{=} \max(r(s,a) + \lambda \sum_{s' \in S} p(s'|s,a)\mathbf{v}[s'] \mid a \in A_s)$$

Observe first that given $\mathbf{v}$ by picking an optimal action for each state s, one obtains a Markovian deterministic rule d such that $L(\mathbf{v}) = \mathbf{r}_d + \lambda \mathbf{P}_d \mathbf{v}$. Furthermore let d' be a Markovian randomized rule. Then $\mathbf{r}_{d'}[s] + \lambda \mathbf{P}_{d'} \mathbf{v}[s] = \sum_{a \in A_s} d'(s)(a) \left(r(s,a) + \lambda \sum_{s' \in S} p(s'|s,a)\mathbf{v}[s'] \right)$. So for all s, $\mathbf{r}_{d'}[s] + \lambda \mathbf{P}_{d'} \mathbf{v}[s] \leq \mathbf{r}_d[s] + \lambda \mathbf{P}_d \mathbf{v}[s]$. This leads to an alternative definition of L.

$$L(\mathbf{v}) = \sup(\mathbf{r}_d + \lambda \mathbf{P}_d \mathbf{v} \mid d \in D^{MR})$$

We establish some useful properties of this operator and in particular that an hypothetic fixed point would lead to the optimal vector.

Proposition 3.13 *Let $\mathbf{v} \in \mathbb{R}^S$. Then:*

- *If $\mathbf{v} \leq L(\mathbf{v})$ then $\mathbf{v} \leq \mathbf{v}_\lambda^*$*
- *If $\mathbf{v} \geq L(\mathbf{v})$ then $\mathbf{v} \geq \mathbf{v}_\lambda^*$*
- *If $\mathbf{v} = L(\mathbf{v})$ then $\mathbf{v} = \mathbf{v}_\lambda^*$*

Proof

The next proposition shows that L has indeed a fixed point. In fact, we present here a particular case of the Banach fixed-point theorem.

Proposition 3.14 *Let $\mathbf{v}_0$ be an arbitrary vector and define inductively $\mathbf{v}_{n+1} \stackrel{def}{=} L(\mathbf{v}_n)$. Then:*

- *L is Lipschitz-continuous with Lipschitz constant equal to λ.*
- *For all n, $\|\mathbf{v}_{n+1} - \mathbf{v}_n\|_\infty \leq \lambda^n \|\mathbf{v}_1 - \mathbf{v}_0\|_\infty$*

- *For all* n, $\|\mathbf{v}_\lambda^* - \mathbf{v}_n\|_\infty \leq \frac{\lambda^n}{1-\lambda}\|\mathbf{v}_1 - \mathbf{v}_0\|_\infty$

Thus $\lim_{n\to\infty} \mathbf{v}_n = \mathbf{v}_\lambda^*$ *and* $\mathbf{v}_\lambda^*$ *is a solution of* $\mathbf{v} = L(\mathbf{v})$.

Proof

In example 3.4, starting with $\mathbf{v}_0 \stackrel{\text{def}}{=} \mathbf{0}$ and $\lambda = 0.5$, one converges very quickly to the fixed point $\mathbf{v}_\lambda^* = (9.523809524, -0.952380952)$. For instance, $\mathbf{v}_3 = (9.525, -0.9525)$.

An immediate consequence of the previous result is the existence of an optimal stationary deterministic policy for the discounted reward problem.

Theorem 3.15 *Every deterministic stationary policy* d^∞ *whose associated decision rule* d *fulfills* $\mathbf{v}_\lambda^* = \mathbf{r}_d + \lambda \mathbf{P}_d \mathbf{v}_\lambda^*$ *is an optimal policy. There is at least one such policy.*

Proof

In example 3.4, there are exactly two decision rules d and d' defined by $d(s_1) = b$ and $d'(s_1) = d(s_2) = d'(s_2) = a$. It can be checked that for $\lambda = 0.5$, d fulfills the requirement of the theorem while $\mathbf{r}_{d'} + 0.5\mathbf{P}_{d'}\mathbf{v}_\lambda^* = (6.095238095, -0.95238095)$.

We want to analyze the dependency of optimal policies w.r.t. λ. In particular we are interested in *Blackwell optimal policies*.

Definition 3.16 *A policy* $\boldsymbol{\pi}$ *is Blackwell optimal if there exists* $0 \leq \lambda_0 < 1$ *such that* $\boldsymbol{\pi}$ *is optimal for every* $\lambda \in [\lambda_0, 1[$.

Theorem 3.17 *There exist* $k \in \mathbb{N}$, $0 = \lambda_0 < \lambda_1 < \cdots < \lambda_k < \lambda_{k+1} = 1$ *and* $d_0, \ldots, d_k$ *deterministic rules such that:*

$$\forall 0 \leq i \leq k \ \forall \lambda \in [0, 1[\qquad \lambda \in [\lambda_i, \lambda_{i+1}] \Rightarrow d_i^\infty \text{ is an optimal policy for } \lambda$$

In particular d_k^∞ *is a Blackwell optimal policy.*

Proof

We now present three different ways to compute the optimal vector (or an almost optimal one) and an associated deterministic stationary policy: value iteration, policy iteration and linear programming. These procedures are presented in increasing order of complexity design.

3.3.2 Value iteration approach

The value iteration approach is directly based on proposition 3.14. Algorithm 3 implements such an approach. It starts with the null vector but could start with an arbitrary vector. When it stops at iteration n, one knows that $\|\mathbf{v}_{n+1} - \mathbf{v}_n\|_\infty \leq \frac{\varepsilon(1-\lambda)}{2\lambda}$ and so $\|\mathbf{v}_{n+1} - \mathbf{v}_\lambda^*\|_\infty \leq \frac{\varepsilon}{2}$. The interesting issue is the guarantee of the expected reward provided by the stationary policy associated with *optdec*.

Proposition 3.18 *Let* d *be the decision rule computed by algorithm 3. Then:* $\|\mathbf{v}_\lambda^{d^\infty} - \mathbf{v}_\lambda^*\|_\infty \leq \varepsilon$.

Proof

3.3.3 Policy iteration approach

In the value iteration approach, the expected discounted reward due to the current policy is never evaluated. Otherwise stated, the variables *optval* and *oldval* are only approximations of the reward induced by the decision rule specified by the variable *optdec*. Unlike value iteration approach, the policy iteration approach maintains the exact value of the rewards associated with the current decision rule and tries to improve this reward by substituting another decision rule.

Algorithm 4 implements this principle. Given the current decision rule d specified by *optdec*, the first step of the iteration consists to build matrix $\mathbf{Id} - \lambda\mathbf{P}_d$ (represented by variable **Md**) and vector

Algorithm 3: Value iteration for the discounted reward

`DiscountedReward`$(\mathcal{M}, \lambda, \varepsilon)$
Input: $\mathcal{M}$ an MDP, λ the discount factor, ε the precision

Output: *optval*, the almost optimal value array indexed by states
Output: *optdec*, the almost optimal decision array indexed by states

Data: *oldval* array, s, s' states, a action, *temp*, *best* reals, *stop* boolean

for $s \in S$ **do** $optval[s] \leftarrow 0$
repeat
  $oldval \leftarrow optval$
  for $s \in S$ **do**
    $best \leftarrow -\infty$
    for $a \in A_s$ **do**
      $temp \leftarrow r(s,a)$
      for $s' \in S$ **do** $temp \leftarrow temp + \lambda p(s'|s,a) oldval[s']$
      if $best < temp$ **then** $best \leftarrow temp$; $optdec[s] \leftarrow a$
    end
    $optval[s] \leftarrow temp$
  end
  $stop \leftarrow$ **true**
  for $s \in S$ **do if** $|optval[s] - oldval[s]| > \frac{\varepsilon(1-\lambda)}{2\lambda}$ **then** $stop \leftarrow$ **false**
until *stop*

$\mathbf{r}_d$ (represented by variable **rd**). Then one solves the linear equation system $(\mathbf{Id} - \lambda \mathbf{P}_d)\mathbf{v}_\lambda^{d^\infty} = \mathbf{r}_d$ (by a call to LinearSolve) and stores the result in *optval*. Finally one computes d' such that $L(\mathbf{v}_\lambda^{d^\infty}) = (\mathbf{r}_{d'} + \lambda \mathbf{P}_{d'})\mathbf{v}_\lambda^{d^\infty}$ trying to choose d if possible.

If the algorithm terminates then $L(\mathbf{v}_\lambda^{d^\infty}) = (\mathbf{r}_d + \lambda \mathbf{P}_d)\mathbf{v}_\lambda^{d^\infty} = \mathbf{v}_\lambda^{d^\infty}$ (by equation 3.1). So $\mathbf{v}_\lambda^{d^\infty}$ is the optimal value and d is an optimal decision rule. It remains to prove that the algorithm terminates.

Proposition 3.19 *Let $\mathbf{v}_n$ and $\mathbf{v}_{n+1}$ be two successive rewards computed by algorithm 4. Then: $\mathbf{v}_{n+1} \neq \mathbf{v}_n$ and $\mathbf{v}_{n+1} \geq \mathbf{v}_n$. As there is a finite number of deterministic decision rules, algorithm 4 terminates.*

Proof

Independently of the fact that policy iteration finds the optimal solution, it is interesting to compare policy and value iterations with respect to the convergence. Indeed, solving the linear equation system in an iteration of algorithm 4 leads to a complexity of $O(|S|^3)$ compared to a complexity of $O(|S|^2)$ for an iteration of algorithm 3. The next proposition shows that policy iteration always converges at least as quick as value iteration when starting with the same initial reward.

Proposition 3.20 *Let d_0 be an initial decision rule. Let $\{\mathbf{u}_n\}$ and $\{\mathbf{v}_n\}$ be respectively the sequences of rewards obtained by value and policy iterations starting with $\mathbf{v}_\lambda^{d_0^\infty}$. Then:*

$$\forall n \ \mathbf{u}_n \leq \mathbf{v}_n \leq \mathbf{v}_\lambda^*$$

Proof

3.3.4 Basics of linear programming [CHV 83]

A linear program is the specification of an optimization problem where both constraints and goal (also called objective) are expressed by linear expressions related to the variables of the problem.

Algorithm 4: Policy iteration for the discounted reward

`DiscountedReward`$(\mathcal{M}, \lambda)$
Input: $\mathcal{M}$ an MDP, λ the discount factor

Output: *optval*, the optimal value array indexed by states
Output: *optdec*, the optimal decision array indexed by states

Data: s, s' states, a action, $temp, best$ reals, *stop* boolean, **Md** matrix, **rd** vector

for $s \in S$ **do** $optdec[s] \leftarrow$ some $a \in A_s$
repeat
  $stop \leftarrow$ **true**
  for $s \in S$ **do**
    $\mathbf{rd}[s] \leftarrow r(s, optdec[s])$
    for $s' \in S$ **do**
      if $s = s'$ **then** $\mathbf{Md}[s, s'] \leftarrow 1 - \lambda p(s'|s, optdec[s])$
      else $\mathbf{Md}[s, s'] \leftarrow -\lambda p(s'|s, optdec[s])$
    end
  end
  $optval \leftarrow$ `LinearSolve`$(\mathbf{Md}, \mathbf{rd})$
  for $s \in S$ **do**
    $best \leftarrow optval[s]$
    for $a \in A_s$ **do**
      $temp \leftarrow r(s, a)$
      for $s' \in S$ **do** $temp \leftarrow temp + \lambda p(s'|s, a) optval[s']$
      if $best < temp$ **then** $best \leftarrow temp$; $optdec[s] \leftarrow a$; $stop \leftarrow$ **false**
    end
  end
until *stop*

There are different ways to express such problems: general, canonic or standard ones. They are all equivalent with linear time reductions between them.

We present the standard form as it is more convenient to design resolution algorithms. Such a problem is specified by a (constraint) matrix $\mathbf{A}$ with dimension $m \times n$, a (constraint) column vector $\mathbf{b}$ with dimension m and a (goal) row vector $\mathbf{c}$ with dimension n. $\mathbf{x}$ is a vector of variables with dimension n. It can be expressed by:

$$\texttt{Maximize } \mathbf{c} \cdot \mathbf{x} \texttt{ such that } \mathbf{A}\mathbf{x} = \mathbf{b} \wedge \mathbf{x} \geq 0$$

Observe that choosing $\mathbf{c}' = -\mathbf{c}$, one transforms a minimization problem in a maximization problem and vice versa. There are three possible outputs.

1. The set of feasible solutions (i.e. $\mathbf{x}$ fulfilling the constraints) is empty.
2. The problem is unbounded, i.e. there exists a sequence of feasible solutions $\{\mathbf{x}_n\}$ such that $\lim_{n\to\infty} \mathbf{c} \cdot \mathbf{x}_n = \infty$.
3. The problem admits an optimal value v, i.e. for all feasible solution $\mathbf{x}$, $\mathbf{c} \cdot \mathbf{x} \leq v$ and for all $\varepsilon > 0$ there exists a feasible solution $\mathbf{x}$ with $\mathbf{c} \cdot \mathbf{x} \geq v - \varepsilon$.

The simplex algorithm

The most well known algorithm, the simplex algorithm, proceeds in several steps that we partially present below.

First step. One modifies the problem in order to fulfill $rank(\mathbf{A}) = m$. By a variant of Gauss elimination, this is performed in polynomial time. At the end of the first step there are two possible results: either there is no feasible solution or the rank of the matrix is equal to its number of lines (i.e. constraints) and so $m \leq n$.

One denotes I, the set of row indices and J, the set of column indices. One considers a dynamic partition of indices of $J = B \uplus N$ with $|B| = m$ and $|N| = n - m$. One denotes $\mathbf{A}_B$ (resp. $\mathbf{A}_N$) the submatrix of A constituted by columns of indices belonging to B (resp. N). We introduce the same notations with $\mathbf{c}$ et $\mathbf{x}$. The next definition is a key ingredient for the study of linear programming.

Definition 3.21 (Basis of a linear program) *A* basis B *is a subset of* m *indices of* J *such that* $\mathbf{A}_B$ *is invertible and* $\mathbf{A}_B^{-1}\mathbf{b}$ *is non negative. There is a solution associated with basis* B *which is defined by:*

$$\forall j \in B\ \mathbf{x}[j] \stackrel{def}{=} (\mathbf{A}_B^{-1}\mathbf{b})[j] \text{ and } \forall j \notin B\ \mathbf{x}[j] \stackrel{def}{=} 0$$

and it is called a basic feasible solution.

Second step. It consists in looking for an initial basic feasible solution. The principle of this research is to build another linear problem which has an obvious initial basic feasible solution and whose resolution has two possible outputs: either it detects that there is no feasible solution or it returns an initial basic feasible solution.

Third step. It consists in improving the current basis (and basic feasible solution) by substituting an index of the current basis by an index out of the basis with a value of the objective function at least as good as before. This is always possible unless: (1) either the current basic feasible solution is optimal which is detected by the satisfaction of equation $\mathbf{c}_N - \mathbf{c}_B\mathbf{A}_B^{-1}\mathbf{A}_N \leq 0$ or (2) the algorithm detects that the problem is unbounded. Furthermore with an appropriate exchange strategy for indices, the algorithm never produces twice the same basis. So it must terminate. Observe that when the linear program admits an optimal value, it is reached by a feasible solution and moreover by a basic feasible solution.

Duality

Duality is a key concept in mathematics (and also in computer science). We illustrate its use for linear programming. Assume that we have a linear combination $\mathbf{y}$ of the row vectors of $\mathbf{A}$,

$$\mathbf{d} \stackrel{\text{def}}{=} \mathbf{y}\mathbf{A} \left(= \sum_{i \in I} \mathbf{y}[i]\mathbf{A}[i,-] \right) \text{ such that } \mathbf{d} \geq \mathbf{c}$$

Then for all feasible solution $\mathbf{x}$,

$$\mathbf{c} \cdot \mathbf{x} \leq \mathbf{d} \cdot \mathbf{x} = \sum_{i \in I} \mathbf{y}[i](\mathbf{A}[i,-] \cdot \mathbf{x}) = \sum_{i \in I} \mathbf{y}[i]\mathbf{b}[i]$$

Otherwise stated, $\sum_{i \in I} \mathbf{y}[i]\mathbf{b}[i]$ is an upper bound of the optimal value. Looking for the smallest possible upper bound, one obtains a *dual* problem:

$$\texttt{Minimize } \mathbf{y} \cdot \mathbf{b} \texttt{ such that } \mathbf{y}\mathbf{A} \geq \mathbf{c} \wedge \mathbf{y} \in \mathbb{R}^I$$

The dual problem can be defined for the general formulation of linear programs and it is routine to check that the dual of the dual is the primal linear problem.

The next proposition is the fundamental result related to duality.

Proposition 3.22 *Let* $\mathbf{P}$ *be a linear problem and* $\mathbf{D}$ *be its dual. Then the following relations hold:*

- *If* $\mathbf{P}$ *is unbounded then* $\mathbf{D}$ *does not admit a feasible solution.*
- *If* $\mathbf{D}$ *is unbounded then* $\mathbf{P}$ *does not admit a feasible solution.*
- $\mathbf{P}$ *admits an optimal solution if and only if* $\mathbf{D}$ *admits an optimal solution. In that case, the optimal values are equal.*

We only develop the proof of the last item as it gives additional informations that we will use later on.

Assume that $\mathbf{P}$ has an optimal solution and let B be a basis corresponding to an optimal basic feasible solution. We denote $f(B)$ the (optimal) value associated with B. Let us recall equations fulfilled by this basis:

$$f(B) = \mathbf{c}_B \mathbf{A}_B^{-1} \mathbf{b} \tag{3.2}$$

(this equation is fulfilled by every basis)

$$\mathbf{c}_N - \mathbf{c}_B \mathbf{A}_B^{-1} \mathbf{A}_N \leq 0 \tag{3.3}$$

(this equation is fulfilled by every optimal basis)

We are going to produce a solution of the dual problem. Define $\mathbf{y} \stackrel{\text{def}}{=} \mathbf{c}_B \mathbf{A}_B^{-1}$. On the one hand $\mathbf{c}_B = \mathbf{y}\mathbf{A}_B$ and, on the other hand, equation 3.3 can be rewritten as: $\mathbf{c}_N \leq \mathbf{y}\mathbf{A}_N$. So $\mathbf{c} \leq \mathbf{y}\mathbf{A}$. This shows that $\mathbf{y}$ is a solution of $\mathbf{D}$. Observe that the constraints of $\mathbf{D}$ indexed by B are fulfilled with equality. This is called the *slackness property.*

Equation 3.2 can be rewritten as: $f(B) = \mathbf{y} \cdot \mathbf{b}$. Since the value associated with $\mathbf{y}$ is the optimal value of $\mathbf{P}$ and so a lower bound of the possible values for $\mathbf{D}$, $\mathbf{y}$ is an optimal solution and the optimal values of the two problems are equal.

One can solve either the primal or the dual problem. As we maintain a basis, it appears that the main criterion is the number of rows m. The dual problem should be transformed in a standard form. This is done in two steps. First every real variable is considered as the difference of two non negative variables and then an additional non negative variable is added per constraint in order to transform inequalities into equalities. This leads to a problem whose dimension is $n \times (2m + n)$. Since $m \leq n$, it is usually more efficient to solve the primal problem.

3.3.5 Linear programming approach

In this section, we prove that a linear program specification of the optimal reward and decision rule is possible leading to a polynomial time algorithm [RTV 97]. This is not the case for the policy iteration algorithm since the number of deterministic decision rules is exponential (equal to $\prod_{s\in S}|A_s|$).

The starting point of this approach is one statement of proposition 3.13: any $\mathbf{v}$ that fulfills $\mathbf{v} \geq L(\mathbf{v})$ is an upper bound of $\mathbf{v}^*_\lambda$ which also fulfills this inequation. So one could constraint the set of solutions by:

$$\text{for all } d \text{ deterministic decision rule } \mathbf{v} \geq \mathbf{r}_d + \lambda \mathbf{P}_d \mathbf{v}$$

Since the number of rules is exponential, we use an alternative equivalent specification.

Primal Linear Program

$$\text{Minimize } \sum_{s\in S} \alpha_s \mathbf{v}[s]$$

$$\text{subject to } \forall s \in S \ \forall a \in A_s \mathbf{v}[s] - \sum_{s'\in S} \lambda p(s'|s,a)\mathbf{v}[s'] \geq r(s,a)$$

Here the variables are the components of vector $\mathbf{v}$ while the α_s's are arbitrary constants that fulfill: $\forall s\ 0 < \alpha_s$ and $\sum_{s\in S}\alpha_s = 1$. We choose $\{\alpha_s\}_{s\in S}$ to be a distribution since this distribution will have an interpretation in the dual problem.

In order to improve the efficiency of this approach, we build the dual of this problem. Indeed the dual has $|S|$ constraints while the primal has $\sum_{s\in S}|A_s|$ constraints.

Dual Linear Program

$$\text{Maximize } \sum_{s\in S}\sum_{a\in A_s} r(s,a)x(s,a)$$

$$\text{subject to } \forall s \in S \ \sum_{a\in A_s} x(s,a) - \sum_{s'\in S}\sum_{a\in A_{s'}} \lambda p(s|s',a)x(s',a) = \alpha_s$$

$$\forall s \in S \ \forall a \in A_s \ x(s,a) \geq 0$$

Here the variables are the $x(s,a)$'s. A feasible solution of the dual linear program x fulfills for all s, $\sum_{a\in A_s} x(s,a) \geq \alpha_s > 0$.

We now introduce mappings between Markovian decision rules and solutions of the dual linear program.

Definition 3.23 *Let d be a Markovian decision rule. Then x_d is defined by:*

$$x_d(s,a) \stackrel{def}{=} d(s)(a) \sum_{s'\in S} \alpha_{s'} \sum_{n\in\mathbb{N}} \lambda^n (\mathbf{P}_d)^n[s',s]$$

Proposition 3.24 *Let d be a Markovian decision rule. Then:*

- *For all s, $\sum_{a\in A_s} x_d(s,a) > 0$ and $d(s)(a) = \frac{x_d(s,a)}{\sum_{a'\in A_s} x_d(s,a')}$;*
- *For all s, a, $x_d(s,a)$ is the average discounted number of times that action a is selected in state s knowing that the initial distribution is given by $\{\alpha_s\}$;*
- *x_d is a feasible solution of the dual linear program;*
- *$\sum_{s\in S}\sum_{a\in A_s} r(s,a)x_d(s,a)$ is the expected discounted reward of policy d^∞ knowing that the initial distribution is given by $\{\alpha_s\}$.*

Proof

Definition 3.25 *Let x be a feasible solution of the dual linear program. Then the decision rule d_x is defined by:*

$$d_x(s)(a) \stackrel{def}{=} \frac{x(s,a)}{\sum_{a \in A_s} x(s,a)}$$

Proposition 3.26 *Let d be a Markovian decision rule. Then $d_{x_d} = d$.*
Let x be a feasible solution of the dual linear program. Then $x_{d_x} = x$.

Proof

The last proposition shows that there is a one-to-one correspondence between Markovian decision rules and feasible solutions of the dual linear program. Since the objective value of a feasible solution is the expected average reward of the associated stationary policy, solving the linear program yields an optimal strategy.

Most of the algorithms for linear programming return a basic feasible solution. In the case of the above dual linear program, the rank of the constraint matrix is already the number of lines, $|S|$ (prove it). We establish another correspondence between deterministic decision rules and basic feasible solutions of the dual linear program.

Proposition 3.27 *Let d be a Markovian deterministic decision rule. Then x_d is a basic feasible solution of the dual linear program.*
Let x be a basic feasible solution of the dual linear program. Then d_x is a Markovian deterministic decision rule.

Proof

3.4 Average reward analysis

3.4.1 More results on finite DTMC's

In order to analyze the average reward criterion, we need to investigate a little bit further the long run behaviour of a finite DTMC. Proposition 1.27 establishes that if the terminal scc's are aperiodic then $\mathbf{P}^n$ converges to a matrix whose row indexed by s is the steady-state distribution when the chain starts from s. We want to get rid of the aperiodicity requirement. So we introduce alternative notions of convergence.

Definition 3.28 *Let $\{u_n\}_{n \in \mathbb{N}}$ be a sequence of reals. Then:*

- *$\{u_n\}_{n \in \mathbb{N}}$ is* Cesaro convergent *to a limit l if $\lim_{n \to \infty} \frac{1}{n+1} \sum_{i \leq n} u_i = l$. One denotes it by $u_n \to_c l$.*
- *$\{u_n\}_{n \in \mathbb{N}}$ is* Abel convergent *to a limit l if for all $0 \leq \lambda < 1$, $u(\lambda) \stackrel{def}{=} \sum_{n \in \mathbb{N}} u_n \lambda^n$ exists and $\lim_{\lambda \uparrow 1} (1-\lambda) u(\lambda) = l$. One denotes it by $u_n \to_a l$.*

Observe the analogy of these definitions with the discounted and average rewards. We will also use these definitions for sequences of vectors and matrices. The next lemma establishes a relation between the convergence notions.

Lemma 3.29 *Let $\{u_n\}_{n \in \mathbb{N}}$ be a sequence of reals.*

- *If $u_n \to l$ then $u_n \to_c l$.*
- *If $u_n \to_c l$ then $u_n \to_a l$.*

Proof

We are now in position to "generalize" proposition 1.27 (this is not really a generalization since the conclusion is weaker).

Theorem 3.30 *Let* $\mathbf{P}$ *be a stochastic matrix. Then* $\{\mathbf{P}^n\}$ *is Cesaro convergent to a stochastic matrix. One denotes its limit* $\mathbf{P}^*$ *and one has:*

$$\mathbf{P}^*\mathbf{P} = \mathbf{P}\mathbf{P}^* = \mathbf{P}^*\mathbf{P}^* = \mathbf{P}^*$$

Proof

The item $\mathbf{P}^*[i,j]$ has a probabilistic interpretation: it is the mean number of visits of state j per time unit starting from state i. We introduce two matrices related to the *rate* of the (Cesaro-) convergence toward $\mathbf{P}^*$, the *fundamental* and *deviation* matrices.

Theorem 3.31 *Let* $\mathbf{P}$ *be a stochastic matrix. Then* $\mathbf{Id} - \mathbf{P} + \mathbf{P}^*$ *is invertible and its inverse called the fundamental matrix and denoted* $\mathbf{Z}$ *fulfills:*

$$\sum_{i=0}^{n} (\mathbf{P} - \mathbf{P}^*)^i \to_c \mathbf{Z}$$

Proof

The deviation matrix $\mathbf{D}$ is defined by $\mathbf{D} \stackrel{\text{def}}{=} \mathbf{Z} - \mathbf{P}^*$. The aperiodic case gives more information about the meaning of these matrices. In this case, one knows $\mathbf{P}^n \to \mathbf{P}^*$. Due to the property of $\mathbf{P}^*$, for $n \geq 1$ one has $\mathbf{P}^n - \mathbf{P}^* = (\mathbf{P} - \mathbf{P}^*)^n$, implying that the greatest module of matrix $\mathbf{P} - \mathbf{P}^*$ is smaller than 1 and thus the fundamental matrix $\mathbf{Z}$ is simply $\mathbf{Id} + \sum_{n\geq 1}(\mathbf{P}^n - \mathbf{P}^*)$. Thus the deviation matrix $\mathbf{D}$ is $\sum_{n\in\mathbb{N}}(\mathbf{P}^n - \mathbf{P}^*)$. So $\mathbf{D}[s,s']$ is the limit when n goes to ∞ of the difference between:

1. the mean number of visits of s' starting from s until time n;
2. the mean number of visits of s' starting from the steady-state distribution reached when the initial state is s until time n.

Observe that when s' is recurrent and reachable from s, the limits of these mean number of visits are infinite but their difference converges to a finite value.

Theorem 3.32 *Let* $\mathbf{P}$ *be a stochastic matrix. Its deviation matrix* $\mathbf{D}$ *fulfills:*

- $\mathbf{D} = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \sum_{k=0}^{i-1} (\mathbf{P}^k - \mathbf{P}^*) = \lim_{\lambda\uparrow 1} \sum_{n=0}^{\infty} \lambda^n (\mathbf{P}^n - \mathbf{P}^*)$
- $\mathbf{P}^*\mathbf{D} = \mathbf{D}\mathbf{P}^* = (\mathbf{Id} - \mathbf{P})\mathbf{D} + \mathbf{P}^* - \mathbf{Id} = \mathbf{D}(\mathbf{Id} - \mathbf{P}) + \mathbf{P}^* - \mathbf{Id} = 0$

Proof

In case of aperiodic chains all the equalities above have a probabilistic proof. $\mathbf{P}^*\mathbf{D} = 0$ since $(\mathbf{P}^*\mathbf{D})[s,s']$ is the difference of twice the same quantity: the mean number of visits of s' starting from the steady-state distribution reached when the initial state is s. $((\mathbf{Id} - \mathbf{P})\mathbf{D})[s,s'] = \mathbf{D}[s,s'] - \mathbf{P}\mathbf{D}[s,s']$. As the second quantity is the the limit when n goes to ∞ of the difference between:

1. the mean number of visits of s' starting from s until time n *excluding time 0*;
2. the mean number of visits of s' starting from the steady-state distribution reached when the initial state is s until time n *excluding time 0*.

So $\mathbf{D}[s,s'] - \mathbf{P}\mathbf{D}[s,s']$ is the difference between (1) the probability of an *initial* visit of s' starting from s and (2) the probability of an *initial* visit of s' starting from the steady-state distribution reached when the initial state is s. This is exactly $(\mathbf{Id} - \mathbf{P}^*)[s,s']$.

3.4.2 Characterization of optimality

In order to characterize optimal policies for the average reward criterion, we first establish an upper bound on the optimal average reward (compare to the second statement of proposition 3.13).

Proposition 3.33 *Assume there exists two vectors* $\mathbf{g}, \mathbf{h}$ *over states such that for all* $d \in D^{MD}$*:*

- $\mathbf{g} \geq \mathbf{P}_d \mathbf{g}$
- $\mathbf{g} + \mathbf{h} \geq \mathbf{P}_d \mathbf{h} + \mathbf{r}_d$

Then $\mathbf{g} \geq \mathbf{g}_+^*$

Proof

The next proposition weakens the conditions to obtain an upper bound.

Proposition 3.34 *Assume there exists two vectors* $\mathbf{g}, \mathbf{h}$ *over states such that for all* $d \in D^{MD}$*, for all* $s \in S$*:*

- *either* $\mathbf{g}[s] > \sum_{s' \in S} \mathbf{P}_d[s, s'] \mathbf{g}[s']$
- *or* $\mathbf{g}[s] = \sum_{s' \in S} \mathbf{P}_d[s, s'] \mathbf{g}[s']$ *and* $\mathbf{g}[s] + \mathbf{h}[s] \geq \sum_{s' \in S} \mathbf{P}_d[s, s'] \mathbf{h}[s'] + \mathbf{r}_d[s]$

Then $\mathbf{g} \geq \mathbf{g}_+^*$

Proof

Let us focus on stationary policies. We first study the average reward triggered by such a policy. Accordingly with the previous section, given d a decision rule, $\mathbf{P}_d^*$ is the Cesaro limit of $\{\mathbf{P}_d^n\}$, $\mathbf{Z}_d$ is the fundamental matrix of $\{\mathbf{P}_d\}$ and $\mathbf{D}_d$ is the deviation matrix of $\{\mathbf{P}_d\}$.

Proposition 3.35 *Let* $d \in D^{MD}$*, then:*

$$\mathbf{g}_-^{d^\infty} = \mathbf{g}_+^{d^\infty} = \mathbf{P}_d^* \mathbf{r}_d = \lim_{\lambda \uparrow 1} (1 - \lambda) \mathbf{v}_\lambda^{d^\infty}$$

Proof

We now denote $\mathbf{g}^{d^\infty} \stackrel{\text{def}}{=} \mathbf{g}_-^{d^\infty} = \mathbf{g}_+^{d^\infty}$. We want to determine more precisely the convergence of $(1 - \lambda)\mathbf{v}_\lambda^{d^\infty}$ towards $\mathbf{g}^{d^\infty}$. Let us define $\rho \stackrel{\text{def}}{=} \frac{1-\lambda}{\lambda}$.

Proposition 3.36 *Let* $d \in D^{MD}$*, assume that* $\frac{\|\mathbf{D}_d\|}{1+\|\mathbf{D}_d\|} < \lambda < 1$ *then:*

$$\mathbf{v}_\lambda^{d^\infty} = \frac{1}{1 - \lambda} \left(\mathbf{P}_d^* \mathbf{r}_d - \sum_{n=1}^{\infty} (-\rho \mathbf{D}_d)^n \mathbf{r}_d \right)$$

Proof

We only need a "first-order Taylor" development of $\mathbf{v}_\lambda^{d^\infty}$.

Corollary 3.37 *Let* $d \in D^{MD}$*, then:*

$$\mathbf{v}_\lambda^{d^\infty} = \frac{1}{1 - \lambda} \mathbf{P}_d^* \mathbf{r}_d + \mathbf{D}_d \mathbf{r}_d + O(1 - \lambda)$$

Proof

The next theorem is the characterization we looked for.

Theorem 3.38 *Consider the following equation system where the variables are vectors* $\mathbf{g}$ *and* $\mathbf{h}$*.*

$$\forall s \in S \ \mathbf{g}[s] = \max_{a \in A_s} \left(\sum_{s' \in S} p(s'|s, a) \mathbf{g}[s'] \right) \tag{3.4}$$

$$\forall s \in S \ \mathbf{g}[s] + \mathbf{h}[s] = \max_{a \in B_s} \left(\sum_{s' \in S} p(s'|s,a)\mathbf{h}[s'] \right) + \mathbf{r}_d[s] \ \textit{where} \ B_s \stackrel{def}{=} \arg\max_{a \in A_s} \left(\sum_{s' \in S} p(s'|s,a)\mathbf{h}[s'] \right) \tag{3.5}$$

Then:

- *Let* $(\mathbf{g}_0, \mathbf{h}_0)$ *be a solution of this system. Then* $\mathbf{g}_0 = \mathbf{g}_+^* = \mathbf{g}_-^*$.
- *Let* d^∞ *be a Blackwell optimal policy. Then* $(\mathbf{P}_d^* \mathbf{r}_d, \mathbf{D}_d \mathbf{r}_d)$ *is a solution of this system.*

Proof

Corollary 3.39 *Every Blackwell optimal policy is optimal w.r.t. the average reward criterion.*

In the sequel, we only consider the policy iteration and linear programming approaches.

3.4.3 Policy iteration approach

Algorithm 5: Policy iteration for the average reward

`AverageReward`($\mathcal{M}$)
Input: $\mathcal{M}$ an MDP

Output: $\mathbf{x}$, the optimal value array indexed by states
Output: *optdec*, the optimal decision array indexed by states

Data: s, s' states, a action, $temp, best, tempbis, bestbis$ reals, *stop* boolean
Data: $\mathbf{Md}$ matrix, $\mathbf{rd}, \mathbf{y}, \mathbf{z}$ vectors

for $s \in S$ **do** $optdec[s] \leftarrow$ some $a \in A_s$
repeat
  $stop \leftarrow$ **true**
  for $s \in S$ **do**
    $\mathbf{rd}[s] \leftarrow r(s, optdec[s])$
    for $s' \in S$ **do**
      if $s = s'$ **then** $\mathbf{Md}[s,s'] \leftarrow 1 - p(s'|s, optdec[s])$
      else $\mathbf{Md}[s,s'] \leftarrow -p(s'|s, optdec[s])$
    end
  end
  $(\mathbf{x}, \mathbf{y}, \mathbf{z}) \leftarrow$ `LinearSolve`$\left(\begin{pmatrix} \mathbf{Md} & 0 & 0 \\ \mathbf{Id} & \mathbf{Md} & 0 \\ 0 & \mathbf{Id} & \mathbf{Md} \end{pmatrix}, \begin{pmatrix} 0 \\ \mathbf{rd} \\ 0 \end{pmatrix} \right)$
  for $s \in S$ **do**
    $best \leftarrow \mathbf{x}[s]$; $bestbis \leftarrow \mathbf{x}[s] + \mathbf{y}[s]$
    for $a \in A_s$ **do**
      $temp \leftarrow r(s,a)$; $tempbis \leftarrow r(s,a)$
      for $s' \in S$ **do**
        $temp \leftarrow temp + \lambda p(s'|s,a)\mathbf{x}[s']$
        $tempbis \leftarrow tempbis + \lambda p(s'|s,a)\mathbf{y}[s']$
      end
      if $best < temp$ **or** $(best = temp$ **and** $bestbis < tempbis)$ **then**
        $best \leftarrow temp$; $bestbis \leftarrow tempbis$; $optdec[s] \leftarrow a$; $stop \leftarrow$ **false**
      end
    end
  end
until *stop*

As seen for the discounted reward, the policy approach is based on two key items.

- Computing the reward provided by a stationary policy d^∞. Here we are going to compute both the reward $\mathbf{P}_d^*\mathbf{r}_d$ but also the first term of the above Taylor development $\mathbf{D}_d\mathbf{r}_d$.
- Designing a rule that either identifies an optimal stationary policy or provides a way to *improve* it. As we will see later, the improvement is more elaborated than for the discounted case.

Contrary to the discounted reward, one cannot specify the average reward as the unique solution of a linear system, the main reason being that $\mathbf{Id} - \mathbf{P}_d$ is not invertible. Here the trick consists in introducing additional variables and equations to the linear equation system in order (1) to also compute the *deviation reward* $\mathbf{D}_d\mathbf{r}_d$ and (2) to confine the non deterministic part of the solution into irrelevant variables. The correctness of the proposition is based on the properties of the deviation matrix.

Proposition 3.40 *Let d be a decision rule and consider the following equation system where the variables are vectors $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$.*

$$(\mathbf{Id} - \mathbf{P}_d)\mathbf{x} = \mathbf{0} \tag{3.6}$$
$$\mathbf{x} + (\mathbf{Id} - \mathbf{P}_d)\mathbf{y} = \mathbf{r}_d \tag{3.7}$$
$$\mathbf{y} + (\mathbf{Id} - \mathbf{P}_d)\mathbf{z} = \mathbf{0} \tag{3.8}$$

Then:

- *Vectors $\mathbf{P}_d^*\mathbf{r}_d$, $\mathbf{D}_d\mathbf{r}_d$ and $-\mathbf{D}_d^2\mathbf{r}_d$ are solutions of this system.*
- *Any $(\mathbf{x}, \mathbf{y}, \mathbf{z})$ solution of this system fulfills $\mathbf{x} = \mathbf{P}_d^*\mathbf{r}_d$ and $\mathbf{y} = \mathbf{D}_d\mathbf{r}_d$.*

Proof

We illustrate this characterization on example 3.4 with the two policies d and d' already described. First we compute $\mathbf{P}_d$ and $\mathbf{P}_{d'}$.

$$\mathbf{Id} - \mathbf{P}_d = \begin{pmatrix} 1 & -1 \\ -0.1 & 0.1 \end{pmatrix} \text{ and } \mathbf{Id} - \mathbf{P}_{d'} = \begin{pmatrix} 0.7 & -0.7 \\ -0.1 & 0.1 \end{pmatrix}$$

The range of $\mathbf{Id} - \mathbf{P}_d$ is $\alpha(1, -0.1)$. So $\mathbf{x} = \alpha(1, -0.1) + (10, -1)$ for some α. Furthermore $\mathbf{x}$ is in the kernel of $\mathbf{Id} - \mathbf{P}_d$. So we get $\alpha + 10 = -0.1\alpha - 1$ yielding $\alpha = -10$ and $\mathbf{x} = (0, 0)$.

The range of $\mathbf{Id} - \mathbf{P}_{d'}$ is $\alpha(0.7, -0.1)$. So $\mathbf{x} = \alpha(0.7, -0.1) + (5, -1)$ for some α. Furthermore $\mathbf{x}$ is in the kernel of $\mathbf{Id} - \mathbf{P}_{d'}$. So we get $0.7\alpha + 5 = -0.1\alpha - 1$ yielding $\alpha = -\frac{15}{2}$ and $\mathbf{x} = (-\frac{1}{4}, -\frac{1}{4})$.

The following proposition is the main ingredient of algorithm 5 based on policy iteration. Observe that using the assertions of the proposition, when one substitutes d' for d, the average reward is not decreased and due to the strict increasing of the discounted reward, one cannot encounter twice the same policy. More precisely the proposition yields two procedures: (1) to decide whether a policy is Blackwell optimal and (2) compute for a policy which is not Blackwell optimal a better policy for the discounted reward given that λ the discount factor is close enough to 1. The proof is mainly based on corollary 3.37 which provides a Taylor development for the discounted reward in the neighborhood of 1.

Proposition 3.41 *Let d be a decision rule and s be a state. Define:*

$$\mathsf{Improve}(d, s) \stackrel{def}{=} \{a \in A_s \mid (\mathbf{P}_d^*\mathbf{r}_d)[s] < \sum_{s' \in S} p(s'|s, a)(\mathbf{P}_d^*\mathbf{r}_d)[s']\} \quad \cup$$

$$\{a \in A_s \mid (\mathbf{P}_d^*\mathbf{r}_d[s] = \sum_{s' \in S} p(s'|s, a)(\mathbf{P}_d^*\mathbf{r}_d)[s'] \wedge ((\mathbf{P}_d^* + \mathbf{D}_d)\mathbf{r}_d)[s] < r(s, a) + \sum_{s' \in S} p(s'|s, a)(\mathbf{D}_d\mathbf{r}_d)[s']\}$$

Then the following assertions hold:

- *If for all s, $\mathsf{Improve}(d, s) = \emptyset$ then d^∞ is average optimal.*

- *Otherwise let d' be any policy such that for all s,*
  (1) $\mathsf{Improve}(d,s) = \emptyset$ *implies* $d'(s) = d(s)$
  and (2) $\mathsf{Improve}(d,s) \neq \emptyset$ *implies* $d'(s) \in \mathsf{Improve}(d,s)$.
  Then $\mathbf{P}_d^* \mathbf{r}_d \leq \mathbf{P}_{d'}^* \mathbf{r}_{d'}$
  and there exists some $0 < \lambda_0 < 1$ *such that for all* $\lambda_0 < \lambda < 1$, $\mathbf{v}_\lambda^{d^\infty} < \mathbf{v}_\lambda^{d'^\infty}$.

Proof

3.4.4 Linear programming approach

Let us compile some results in order to specify the linear program associated with the average reward. Using proposition 3.33, for every pair of vectors $(\mathbf{g}, \mathbf{h})$ such that for all $d \in D^{MD}$, $\mathbf{g} \geq \mathbf{P}_d \mathbf{g}$ and $\mathbf{g} + \mathbf{h} \geq \mathbf{P}_d \mathbf{h} + \mathbf{r}_d$ one gets: $\mathbf{g} \geq \mathbf{g}_+^*$.

Using theorem 3.38 and the proof of proposition 3.34, one gets that given any Blackwell optimal policy d^∞, as soon as M is large enough, then $(\mathbf{P}_d^* \mathbf{r}_d, \mathbf{D}_d \mathbf{r}_d + M \mathbf{P}_d^* \mathbf{r}_d)$ is a solution of such a system. Thus one concludes that any optimal solution of the following linear program has its $\mathbf{g}$ component equal to the optimal expected average reward.

Primal Linear Program

$$\text{Minimize} \sum_{s \in S} \alpha_s \mathbf{g}[s]$$

$$\text{subject to } \forall s \in S \ \forall a \in A_s,$$

$$\mathbf{g}[s] - \sum_{s' \in S} p(s'|s,a) \mathbf{g}[s'] \geq 0 \text{ and}$$

$$\mathbf{g}[s] + \mathbf{h}[s] - \sum_{s' \in S} p(s'|s,a) \mathbf{h}[s'] \geq r(s,a)$$

Here the variables are the components of vectors $\mathbf{g}$ and $\mathbf{h}$ while the α_s's are arbitrary positive constants. Let us build the dual program.

Dual Linear Program

$$\text{Maximize} \sum_{s \in S} r(s,a) \mathbf{x}[s,a]$$

$$\text{subject to } \forall s \in S$$

$$\sum_{a \in A_s} \mathbf{x}[s,a] - \sum_{s' \in S} \sum_{a \in A_{s'}} p(s|s',a) \mathbf{x}[s',a] = 0 \text{ and}$$

$$\sum_{a \in A_s} (\mathbf{x}[s,a] + \mathbf{y}[s,a]) - \sum_{s' \in S} \sum_{a \in A_{s'}} p(s|s',a) \mathbf{y}[s',a] = \alpha_s$$

$$\forall s \in S \ \forall a \in A_s \ \mathbf{x}[s,a] \geq 0 \wedge \mathbf{y}[s,a] \geq 0$$

The next proposition shows how to recover optimal average deterministic stationary policies from a basic optimal solution of the dual linear program.

Proposition 3.42 *Let $(\mathbf{x}, \mathbf{y})$ be a basic optimal solution of the dual linear program. Then every deterministic d^∞ that fulfills the following requirements for every $s \in S$ is average optimal.*

- *either* $\mathbf{x}[s, d(s)] > 0$
- *or* $\sum_{a \in A_s} \mathbf{x}[s,a] = 0$ *and* $\mathbf{y}[s, d(s)] > 0$

Furthermore there is at least one such policy.

Proof

3.5 Proofs

3.5.1 Proofs of section 3.1

Proof of theorem 3.10
Given $\boldsymbol{\pi}$ be an arbitrary policy, let us define Markovian policy $\boldsymbol{\pi}' = (d'_0, d'_1, \ldots)$ by:

$$d'_n(s)(a) \stackrel{\text{def}}{=} \mathbf{Pr}^{\boldsymbol{\pi}}(Y_n = a \mid X_n = s, X_0 = s_0)$$

This is a partially defined strategy since $\mathbf{Pr}^{\boldsymbol{\pi}}(X_n = s, X_0 = s_0)$ could be equal to 0. However due to the equality of the proposition that we are going to inductively establish in this case $\mathbf{Pr}^{\boldsymbol{\pi}'}(X_n = s, X_0 = s_0)$ is also equal to 0.

For $n = 0$, the equality $\mathbf{Pr}^{\boldsymbol{\pi}'}(X_n = s, Y_n = a \mid X_0 = s_0) = \mathbf{Pr}^{\boldsymbol{\pi}}(X_n = s, Y_n = a \mid X_0 = s_0)$ is only relevant for $s = s_0$ and holds by definition of $\boldsymbol{\pi}'$.

Assume that the equality holds up to n. Then:
$\mathbf{Pr}^{\boldsymbol{\pi}'}(X_{n+1} = s \mid X_0 = s_0) = \sum_{s' \in S} \sum_{a \in A_{s'}} \mathbf{Pr}^{\boldsymbol{\pi}'}(X_n = s', Y_n = a \mid X_0 = s_0) p(s|s', a)$
$= \sum_{s' \in S} \sum_{a \in A_{s'}} \mathbf{Pr}^{\boldsymbol{\pi}}(X_n = s', Y_n = a \mid X_0 = s_0) p(s|s', a) = \mathbf{Pr}^{\boldsymbol{\pi}}(X_{n+1} = s \mid X_0 = s_0)$
Now:
$\mathbf{Pr}^{\boldsymbol{\pi}'}(X_{n+1} = s, Y_{n+1} = a \mid X_0 = s_0)$
$= d'_{n+1}(s)(a) \mathbf{Pr}^{\boldsymbol{\pi}'}(X_{n+1} = s \mid X_0 = s_0)$
$= \mathbf{Pr}^{\boldsymbol{\pi}}(Y_{n+1} = a \mid X_{n+1} = s, X_0 = s_0) \mathbf{Pr}^{\boldsymbol{\pi}'}(X_{n+1} = s \mid X_0 = s_0)$
$= \mathbf{Pr}^{\boldsymbol{\pi}}(X_{n+1} = s, Y_{n+1} = a \mid X_0 = s_0)$

q.e.d. (theorem 3.10) ◊◊◊

3.5.2 Proofs of section 3.2

Proof of proposition 3.11
We prove by induction on the time horizon that the policy computed by the algorithm is optimal. In order to give more intuition about the proof we index in a backward way the decision rules.

Assume that $\boldsymbol{\pi}_{n-1} \stackrel{\text{def}}{=} d_{n-1}, \ldots, d_1$, the policy computed by the algorithm for time horizon $n-1$ is optimal and let d_n be the decision rule computed at the n^{th} iteration. Pick an arbitrary policy $\boldsymbol{\pi}'_n \stackrel{\text{def}}{=} d'_n, \ldots, d'_1$ and denote $\boldsymbol{\pi}'_{n-1} \stackrel{\text{def}}{=} d'_{n-1}, \ldots, d'_1$.
Let $s \in S$,

$$\mathbf{u}_n^{\boldsymbol{\pi}_n}[s] = r(s, d_n(s)) + \sum_{s' \in S} p(s'|s, d_n(s)) \mathbf{u}_{n-1}^{\boldsymbol{\pi}_{n-1}}[s'] \geq r(s, d'_n(s)) + \sum_{s' \in S} p(s'|s, d'_n(s)) \mathbf{u}_{n-1}^{\boldsymbol{\pi}_{n-1}}[s]$$

(due to the iterative step of the algorithm)

$$\geq r(s, d'_n(s)) + \sum_{s' \in S} p(s'|s, d'_n(s)) \mathbf{u}_{n-1}^{\boldsymbol{\pi}'_{n-1}}[s] = \mathbf{u}_n^{\boldsymbol{\pi}'_n}[s]$$

(due to the inductive hypothesis)

q.e.d. (proposition 3.11) ◊◊◊

3.5.3 Proofs of section 3.3

Proof of proposition 3.13
Let $\mathbf{v} \leq L(\mathbf{v})$. By definition, there is a decision rule d such that $L(\mathbf{v}) = \mathbf{r}_d + \lambda \mathbf{P}_d \mathbf{v}$.
Thus $\mathbf{v} - \lambda \mathbf{P}_d \mathbf{v} \leq \mathbf{r}_d$. Applying the non negative matrix $(\mathbf{Id} - \lambda \mathbf{P}_d)^{-1}$ to the inequality yields:
$\mathbf{v} \leq (\mathbf{Id} - \lambda \mathbf{P}_d)^{-1} \mathbf{r}_d = \mathbf{v}^{d^\infty} \leq \mathbf{v}^*_\lambda$

Let $\mathbf{v} \geq L(\mathbf{v})$. Let $\boldsymbol{\pi} \stackrel{\text{def}}{=} (d_0, \ldots, d_n, \ldots)$ be a Markovian policy.
$\mathbf{v} \geq L(\mathbf{v}) \geq \mathbf{r}_{d_0} + \lambda \mathbf{P}_{d_0}\mathbf{v}$. By induction for $n \geq 0$,

$$\mathbf{v} \geq \sum_{0 \leq i < n} \lambda^i \left(\prod_{0 \leq j < i} \mathbf{P}_{d_j} \right) \mathbf{r}_{d_i} + \lambda^n \left(\prod_{0 \leq j < n} \mathbf{P}_{d_j} \right) \mathbf{v}$$

On the other hand,

$$\mathbf{v}_\lambda^{\boldsymbol{\pi}} = \sum_{i \in \mathbb{N}} \lambda^i \left(\prod_{0 \leq j < i} \mathbf{P}_{d_j} \right) \mathbf{r}_{d_i}$$

Let us define $B \stackrel{\text{def}}{=} \max(\max_s(|\mathbf{v}[s]|), \max_{s,a}(|r(s,a)|)$. Then for all $s \in S$ and $n \in \mathbb{N}$:

$$\mathbf{v}[s] - \mathbf{v}_\lambda^{\boldsymbol{\pi}}[s] \geq -\lambda^n B(1 + \sum_{i \in \mathbb{N}} \lambda^i)$$

Letting n go to ∞, one gets: $\mathbf{v} \geq \mathbf{v}_\lambda^{\boldsymbol{\pi}}$. Since $\boldsymbol{\pi}$ is arbitrary, one obtains: $\mathbf{v} \geq \mathbf{v}_\lambda^*$.
The last assertion is a consequence of the previous ones.

q.e.d. (proposition 3.13) ◊◊◊

Proof of proposition 3.14
Let $\mathbf{v}$ and $\mathbf{v}'$ be two vectors. Let d be a decision rule such that $L(\mathbf{v}) = \mathbf{r}_d + \lambda \mathbf{P}_d \mathbf{v}$. Then:

$$L(\mathbf{v})[s] - L(\mathbf{v}')[s] \leq \lambda \left(\mathbf{P}_d(\mathbf{v} - \mathbf{v}')\right)[s] \leq \lambda \|\mathbf{v} - \mathbf{v}'\|_\infty$$

So: $\|L(\mathbf{v}) - L(\mathbf{v}')\|_\infty \leq \lambda \|\mathbf{v} - \mathbf{v}'\|_\infty$
This proves the Lipschitz-continuity of L.
Let $\{\mathbf{v}_n\}$ be defined as in the proposition. For all m, n with $m \geq n$:

$$\|\mathbf{v}_m - \mathbf{v}_n\|_\infty \leq \sum_{n \leq i < m} \|\mathbf{v}_{i+1} - \mathbf{v}_i\|_\infty \leq \lambda^n (\sum_{i \in \mathbb{N}} \lambda^i) \|\mathbf{v}_1 - \mathbf{v}_0\|_\infty$$

This proves that $\{v_n\}$ is a Cauchy sequence thus convergent to some $\mathbf{v}_\infty$.
More precisely: $\|\mathbf{v}_\infty - \mathbf{v}_n\|_\infty \leq \frac{\lambda^n}{1-\lambda} \|\mathbf{v}_1 - \mathbf{v}_0\|_\infty$
By continuity of L, $\mathbf{v}_\infty = L(\mathbf{v}_\infty)$. So by proposition 3.13, $\mathbf{v}_\infty = \mathbf{v}_\lambda^*$.

q.e.d. (proposition 3.14) ◊◊◊

Proof of theorem 3.15
Observe that the discounted reward associated with policy d^∞ is: $(\mathbf{Id} - \lambda \mathbf{P}_d)^{-1} \mathbf{r}_d$.
Thus if d fulfills $\mathbf{v}_\lambda^* = \mathbf{r}_d + \lambda \mathbf{P}_d \mathbf{v}_\lambda^*$ then $\mathbf{v}_\lambda^* = (\mathbf{Id} - \lambda \mathbf{P}_d)^{-1} \mathbf{r}_d$.
We know that $\mathbf{v}_\lambda^* = L(\mathbf{v}_\lambda^*)$. Let $d \in D^{MD}$ be a decision rule associated with $L(\mathbf{v}_\lambda^*)$.
Then d fulfills $\mathbf{v}_\lambda^* = \mathbf{r}_d + \lambda \mathbf{P}_d \mathbf{v}_\lambda^*$.

q.e.d. (theorem 3.15) ◊◊◊

Proof of theorem 3.17
Observe that the discounted reward $\mathbf{v}_\lambda^{d^\infty}$ associated with policy d^∞ fulfills: $\mathbf{v}_\lambda^{d^\infty} = (\mathbf{Id} - \lambda \mathbf{P}_d)^{-1} \mathbf{r}_d$.
So every item of this vector is a rational fraction of λ with poles outside $[0, 1[$. We consider $\mathbf{v}_x^{d^\infty}[s]$ as a function of x.
Thus $Zero \stackrel{\text{def}}{=} \{\lambda \mid \exists d, d' \in D^{MD}\ \exists s \in S\ \mathbf{v}_x^{d^\infty}[s] \neq \mathbf{v}_x^{d'^\infty}[s] \wedge \mathbf{v}_\lambda^{d^\infty}[s] = \mathbf{v}_\lambda^{d'^\infty}[s]\}$ is finite.

Let $I \stackrel{\text{def}}{=}]a, b[$ be an interval such that $Zero \cap I = \emptyset$. Pick an arbitrary $c \in I$ and let d be an optimal decision rule w.r.t. to c. We claim that d is optimal for the whole interval I. Otherwise, due to the continuity of $\mathbf{v}_x^{d^\infty}[s]$, there should exist $\lambda \in I$, d' and s with $\mathbf{v}_x^{d^\infty}[s] \neq \mathbf{v}_x^{d'^\infty}[s] \wedge \mathbf{v}_\lambda^{d^\infty}[s] = \mathbf{v}_\lambda^{d'^\infty}[s]$.
Furthermore again by continuity d is also optimal at a and b (when $b \neq 1$).
So the decomposition of $[0, 1[$ in appropriate subintervals is obtained by considering $[0, 1[\backslash Zero$.

q.e.d. (theorem 3.17) ◊◊◊

Proof of proposition 3.18

$$\|\mathbf{v}_\lambda^{d^\infty} - \mathbf{v}_{n+1}\|_\infty \leq \|\mathbf{v}_\lambda^{d^\infty} - (\mathbf{r}_d + \lambda \mathbf{P}_d \mathbf{v}_{n+1})\|_\infty + \|(\mathbf{r}_d + \lambda \mathbf{P}_d \mathbf{v}_{n+1}) - \mathbf{v}_{n+1}\|_\infty$$
$$= \lambda \|\mathbf{P}_d \mathbf{v}_\lambda^{d^\infty} - \mathbf{P}_d \mathbf{v}_{n+1}\|_\infty + \lambda \|\mathbf{P}_d \mathbf{v}_{n+1} - \mathbf{P}_d \mathbf{v}_n\|_\infty \leq \lambda \|\mathbf{v}_\lambda^{d^\infty} - \mathbf{v}_{n+1}\|_\infty + \lambda \|\mathbf{v}_{n+1} - \mathbf{v}_n\|_\infty$$

So

$$\|\mathbf{v}_\lambda^{d^\infty} - \mathbf{v}_{n+1}\|_\infty \leq \frac{\lambda}{1-\lambda} \|\mathbf{v}_{n+1} - \mathbf{v}_n\|_\infty \leq \frac{\varepsilon}{2}$$

Thus:

$$\|\mathbf{v}_\lambda^{d^\infty} - \mathbf{v}_\lambda^*\|_\infty \leq \|\mathbf{v}_\lambda^{d^\infty} - \mathbf{v}_{n+1}\|_\infty + \|\mathbf{v}_{n+1} - \mathbf{v}_\lambda^*\|_\infty \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

q.e.d. (proposition 3.18) ◊◊◊

Proof of proposition 3.19
Let d_n and d_{n+1} be the decision rules associated with $\mathbf{v}_n$ and $\mathbf{v}_{n+1}$. One has:

$$\mathbf{r}_{d_{n+1}} + \lambda \mathbf{P}_{d_{n+1}} \mathbf{v}_n \geq \mathbf{r}_{d_n} + \lambda \mathbf{P}_{d_n} \mathbf{v}_n = \mathbf{v}_n$$

with at least one strict inequality.
Thus:

$$\mathbf{r}_{d_{n+1}} \geq (\mathbf{Id} - \lambda \mathbf{P}_{d_{n+1}}) \mathbf{v}_n$$

Applying $(\mathbf{Id} - \lambda \mathbf{P}_{d_{n+1}})^{-1}$ $(= \sum_{i \in N} (\lambda \mathbf{P}_{d_{n+1}})^i)$

$$\mathbf{v}_{n+1} \geq \mathbf{v}_n$$

Moreover since $(\mathbf{Id} - \lambda \mathbf{P}_{d_{n+1}})^{-1} \geq \mathbf{Id}$, the strict inequality is preserved.

q.e.d. (proposition 3.19) ◊◊◊

Proof of proposition 3.20
By proposition 3.19, $\mathbf{v}_n$ is an increasing sequence which converges to $\mathbf{v}_\lambda^*$. So the second inequality is fulfilled. Let us prove the first inequation by induction. The basis case is included in the hypotheses of the proposition. Let us call du_n (resp. dv_n) the decision rule corresponding to the n^{th} iteration of the value (resp. policy) iteration algorithm.

$\mathbf{v}_{n+1} = \mathbf{r}_{dv_{n+1}} + \lambda \mathbf{P}_{dv_{n+1}} \mathbf{v}_{n+1} \geq \mathbf{r}_{dv_{n+1}} + \lambda \mathbf{P}_{dv_{n+1}} \mathbf{v}_n$
since $\mathbf{v}_{n+1} \geq \mathbf{v}_n$
$\mathbf{r}_{dv_{n+1}} + \lambda \mathbf{P}_{dv_{n+1}} \mathbf{v}_n \geq \mathbf{r}_{du_{n+1}} + \lambda \mathbf{P}_{du_{n+1}} \mathbf{v}_n$
since $\mathbf{r}_{dv_{n+1}} + \lambda \mathbf{P}_{dv_{n+1}} \mathbf{v}_n = L(\mathbf{v}_n)$
$\mathbf{r}_{du_{n+1}} + \lambda \mathbf{P}_{du_{n+1}} \mathbf{v}_n \geq \mathbf{r}_{du_{n+1}} + \lambda \mathbf{P}_{du_{n+1}} \mathbf{u}_n = \mathbf{u}_{n+1}$
since $\mathbf{v}_n \geq \mathbf{u}_n$

q.e.d. (proposition 3.20) ◊◊◊

Proof of proposition 3.24
$(\mathbf{P}_d)^n[s', s]$ is the probability (under policy d^∞), that the state at time n is s knowing that at time 0 the state is s'.
Thus $\sum_{a \in A_s} x_d(s,a) = \sum_{s' \in S} \alpha_{s'} \sum_{n \in \mathbb{N}} \lambda^n (\mathbf{P}_d)^n[s', s]$ is the discounted number of visits of s knowing that the initial distribution is $\{\alpha_s\}$.
So $\sum_{a \in A_s} x_d(s,a) \geq \alpha_s > 0$ and $d(s)(a) = \frac{x_d(s,a)}{\sum_{a' \in A_s} x_d(s,a')}$.
Since at every visit of s, a is selected with probability $d(s)(a)$, $x_d(s,a)$ is the average discounted number of times that action a is selected in state s knowing that the initial distribution is given by $\{\alpha_s\}$.
Multiplying by rewards and summing over states and actions, $\sum_{s \in S} \sum_{a \in A_s} r(s,a) x_d(s,a)$ is the expected discounted reward of policy d^∞ knowing that the initial distribution is given by $\{\alpha_s\}$.

A visit to state s is either done at time 0, or done at time n due to a visit at some state s' at time $n-1$ followed by selection of action $a \in A_{s'}$ with probability $d(s')(a)$ and random choice of s with probability $p(s|s',a)$. In the latter case, the discounted value of the visit of s is the discounted value of visit of s' multiplied by λ. So:

$$\sum_{a\in A_s} x_d(s,a) = \alpha_s + \lambda \sum_{s'\in S} \left(\sum_{a'\in A_{s'}} x_d(s',a') \right) \sum_{a\in A_{s'}} d(s')(a)p(s|s',a)$$

$$= \alpha_s + \lambda \sum_{s'\in S} \left(\sum_{a'\in A_{s'}} x_d(s',a') \right) \sum_{a\in A_{s'}} \frac{x_d(s',a)}{\sum_{a'\in A_{s'}} x_d(s',a')} p(s|s',a) = \alpha_s + \lambda \sum_{s'\in S} \sum_{a\in A_{s'}} x_d(s',a)p(s|s',a)$$

Consequently x_d is a feasible solution of the dual linear program.

q.e.d. (proposition 3.24) ◊◊◊

Proof of proposition 3.26
The result $d_{x_d} = d$ is included in the first statement of proposition 3.24.

Let x be a feasible solution of the dual linear program. Define $y(s) \stackrel{\text{def}}{=} \sum_{a\in A_s} x(s,a)$.
$\alpha_s = y(s) - \lambda \sum_{s'\in S} \sum_{a\in A_{s'}} p(s|s',a)x(s',a) = y(s) - \lambda \sum_{s'\in S} \sum_{a\in A_{s'}} p(s|s',a)d_x(s')(a)y(s')$
Which leads in a vectorial notation to: $\alpha = y(\mathbf{Id} - \lambda \mathbf{P}_{d_x})$
So $y = \alpha(\mathbf{Id} - \lambda \mathbf{P}_{d_x})^{-1} = \alpha(\sum_{n\in\mathbb{N}} (\lambda \mathbf{P}_{d_x})^n)$
Thus $y(s) = \sum_{s'\in S} \alpha_{s'} \sum_{n\in\mathbb{N}} \lambda^n (\mathbf{P}_{d_x})^n[s',s] = \sum_{a\in A_s} x_{d_x}(s,a)$
Since vector $(x_{d_x}(s,a))_{a\in A_s}$ is proportional to vector $d_x(s)$ which is proportional to vector $(x(s,a))_{a\in A_s}$, one concludes that $x_{d_x} = x$.

q.e.d. (proposition 3.26) ◊◊◊

Proof of proposition 3.27
Let d be a Markovian deterministic decision rule. Then x_d has exactly $|S|$ non null components corresponding to the columns $(s, d(s))$. The submatrix corresponding to these variables is $(\mathbf{Id} - \lambda \mathbf{P}_d)$. As it is invertible, x_d is a basic feasible solution.

Let x be a basic feasible solution of the dual linear program. For all s, $\sum_{a\in A_s} x(s,a) > 0$. Since x is basic there is at most $|S|$ non null components. This implies that for all s, there is exactly a single a such that $x(s,a) > 0$. Since $d_x(s)$ is proportional to $(x(s,a))_{a\in A_s}$, d_x is a Markovian deterministic decision rule.

q.e.d. (proposition 3.27) ◊◊◊

3.5.4 Proofs of section 3.4

Proof of lemma 3.29
Let $\{u_n\}_{n\in\mathbb{N}}$ be such that $u_n \to l$. Then for all ε there exists n_0 such that for all $n \geq n_0$, $|u_n - l| \leq \varepsilon$
Let $n \geq n_0$,
$|\frac{1}{n} \sum_{k<n} (u_n - l)| \leq \frac{1}{n} |\sum_{k<n_0} (u_n - l)| + \frac{n-n_0}{n} \varepsilon \leq 2\varepsilon$
for n large enough.

Let $\{u_n\}_{n\in\mathbb{N}}$ be such that $u_n \to_c l$.
Define $s_n = \sum_{i\leq n} u_n$ and $v_n = \frac{s_n}{n+1}$.
By hypothesis, $v_n \to l$.
$\sum_{k=0}^{n} \lambda^k u_k = (1-\lambda) \sum_{k=0}^{n-1} \lambda^k (k+1) v_k + \lambda^n (n+1) v_n$
Using the Cauchy criterion, the series $\sum_{k=0}^{n-1} \lambda^k (k+1) v_k$ is convergent. So:
$u(\lambda) = (1-\lambda) \sum_{k=0}^{\infty} \lambda^k (k+1) v_k$
For all ε there exists n_0 such that for all $n \geq n_0$, $|v_n - l| \leq \varepsilon$
$(1-\lambda)u(\lambda) \leq (1-\lambda)^2 \left(\sum_{k<n_0} \lambda^k (k+1)(v_k - l - \varepsilon) + (l+\varepsilon) \sum_{k=0}^{\infty} \lambda^k (k+1) \right)$

$= (1-\lambda)^2 \left(\sum_{k<n_0} \lambda^k (k+1)(v_k - l - \varepsilon)\right) + (l+\varepsilon)$
Thus: $\limsup_{\lambda\uparrow 1}(1-\lambda)u(\lambda) \leq l + \varepsilon$
By a similar reasoning: $\liminf_{\lambda\uparrow 1}(1-\lambda)u(\lambda) \geq l - \varepsilon$
Letting ε go to 0, one gets: $\lim_{\lambda\uparrow 1}(1-\lambda)u(\lambda) = l$

q.e.d. (lemma 3.29) ◊◊◊

Proof of theorem 3.30
Let $\tilde{\mathbf{P}}_n \stackrel{\text{def}}{=} \frac{1}{n}\sum_{0\leq i<n} \mathbf{P}^i$ for $n > 0$.
$\tilde{\mathbf{P}}_n$ is a stochastic matrix thus the sequence $\{\tilde{\mathbf{P}}_n\}$ is bounded.
Pick a sequence of indices $n_0 < n_1 < \cdots$ such that $\mathbf{L} \stackrel{\text{def}}{=} \lim_{k\to\infty} \tilde{\mathbf{P}}_{n_k}$ exists.
Observe that: $\tilde{\mathbf{P}}_n\mathbf{P} = \mathbf{P}\tilde{\mathbf{P}}_n = \tilde{\mathbf{P}}_n + \frac{1}{n}(\mathbf{P}^n - \mathbf{Id})$
Applying these equalities to n_k letting k go to ∞ yields: $\mathbf{LP} = \mathbf{PL} = \mathbf{L}$
Let $\mathbf{L}'$ be another limit of a subsequence of $\{\tilde{\mathbf{P}}_n\}$. Then: $\mathbf{PL}' = \mathbf{L}'\mathbf{P} = \mathbf{L}'$.
By iteration, $\mathbf{P}^n\mathbf{L}' = \mathbf{L}'\mathbf{P}^n = \mathbf{L}'$ for all n.
By linear combination, $\tilde{\mathbf{P}}_n\mathbf{L}' = \mathbf{L}'\tilde{\mathbf{P}}_n = \mathbf{L}'$ for all n.
Applying this equality for n_k and letting k go to ∞ yields $\mathbf{L}'\mathbf{L} = \mathbf{L}\mathbf{L}' = \mathbf{L}'$.
Swapping $\mathbf{L}$ and $\mathbf{L}'$ yields $\mathbf{L}\mathbf{L}' = \mathbf{L}'\mathbf{L} = \mathbf{L}$. Thus $\mathbf{L}' = \mathbf{L}$.
So $\tilde{\mathbf{P}}_n$ is convergent and the limit is stochastic since the (finite) sum of the items of every row of a stochastic matrix is 1 and this equality is preserved by limit operations.

q.e.d. (theorem 3.30) ◊◊◊

Proof of theorem 3.31
Let k be the number of terminal scc's and let us order the states per terminal scc's followed by transient states. Then $\mathbf{P}$ can be described as:

$$\begin{pmatrix} \mathbf{P}_1 & 0 & \dots & 0 & 0 \\ 0 & \mathbf{P}_2 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & \mathbf{P}_k & 0 \\ \mathbf{P}_{T,1} & \mathbf{P}_{T,2} & \dots & \mathbf{P}_{T,k} & \mathbf{P}_{T,T} \end{pmatrix}$$

where $\mathbf{P}_i$ is the stochastic matrix associated with the i^{th} terminal scc,
$\mathbf{P}_{T,i}$ is the submatrix of the transitions from the transient states to the i^{th} terminal scc
and $\mathbf{P}_{T,T}$ is the substochastic matrix associated with the transient states.
Recall that $\mathbf{P}^*[i,j]$ represent the mean number of visits to j starting from i per time unit. Since the total mean number of visits of transient states is finite, one can rewrite $\mathbf{P}^*$ as:

$$\begin{pmatrix} \mathbf{P}^*_1 & 0 & \dots & 0 & 0 \\ 0 & \mathbf{P}^*_2 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & \mathbf{P}^*_k & 0 \\ \mathbf{P}^*_{T,1} & \mathbf{P}^*_{T,2} & \dots & \mathbf{P}^*_{T,k} & 0 \end{pmatrix}$$

where $\mathbf{P}^*_i$ is a matrix where all rows are identical and represent the single distribution solution of $\mathbf{x}\mathbf{P}_i = \mathbf{x}$.
So $\mathbf{Id} - \mathbf{P} + \mathbf{P}^*$ can be rewritten as:

$$\begin{pmatrix} \mathbf{Id} - \mathbf{P}_1 + \mathbf{P}^*_1 & 0 & \dots & 0 & 0 \\ 0 & \mathbf{Id} - \mathbf{P}_2 + \mathbf{P}^*_2 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & \mathbf{Id} - \mathbf{P}_k + \mathbf{P}^*_k & 0 \\ -\mathbf{P}_{T,1} + \mathbf{P}^*_{T,1} & -\mathbf{P}_{T,1} + \mathbf{P}^*_{T,2} & \dots & -\mathbf{P}_{T,k} + \mathbf{P}^*_{T,k} & \mathbf{Id} - \mathbf{P}_{T,T} \end{pmatrix}$$

We already know that $\mathbf{Id} - \mathbf{P}_{T,T}$ is invertible. So to prove that $\mathbf{Id} - \mathbf{P} + \mathbf{P}^*$ is invertible it only remains to prove that for all i, $\mathbf{Id} - \mathbf{P}_i + \mathbf{P}_i^*$ is invertible.

Assume (by contradiction) that $\mathbf{x}$ is a non null vector such that $(\mathbf{Id} - \mathbf{P}_i + \mathbf{P}_i^*)\mathbf{x} = 0$.
Let π be the unique invariant distribution of $\mathbf{P}_i$.
$\pi(\mathbf{Id} - \mathbf{P}_i + \mathbf{P}_i^*)\mathbf{x} = 0$ but π is invariant by $\mathbf{P}_i^*$, so $\pi\mathbf{x} = 0$.
As every row of $\mathbf{P}_i^*$ is equal to π, one gets $\mathbf{P}_i^*\mathbf{x} = 0$.
So $(\mathbf{Id} - \mathbf{P}_i)\mathbf{x} = 0$ meaning that $\mathbf{x}$ is a right eigenvector of $\mathbf{P}_i$ for eigenvalue 1.
Since $\mathbf{P}_i$ is an irreducible chain the dimension of eigenspace corresponding to 1 is 1.
So $\mathbf{x} = \alpha\mathbf{1}^T$ but since $\pi\mathbf{x} = 0$ this entails that α is null, a contradiction.

Since $\mathbf{P}^*\mathbf{P} = \mathbf{P}\mathbf{P}^* = \mathbf{P}^*\mathbf{P}^* = \mathbf{P}^*$ we claim that $(\mathbf{P} - \mathbf{P}^*)^n = \mathbf{P}^n - \mathbf{P}^*$.
By induction: $(\mathbf{P} - \mathbf{P}^*)^{n+1} = (\mathbf{P}^n - \mathbf{P}^*)(\mathbf{P} - \mathbf{P}^*) = \mathbf{P}^{n+1} - \mathbf{P}^* + \mathbf{P}^* - \mathbf{P}^*$
Let $\mathbf{\Delta} = \mathbf{P} - \mathbf{P}^*$.
For all $i \geq 1$, $\mathbf{Id} - \mathbf{\Delta}^i = (\mathbf{Id} - \mathbf{\Delta}) \sum_{0 \leq k < i} \mathbf{\Delta}^k$
Averaging over i yields:

$$\mathbf{Id} - \frac{1}{n}\sum_{i=1}^{n} \mathbf{\Delta}^i = (\mathbf{Id} - \mathbf{\Delta})\frac{1}{n}\sum_{i=1}^{n}\sum_{0 \leq k < i} \mathbf{\Delta}^k$$

Otherwise stated

$$(\mathbf{Id} - \mathbf{\Delta})^{-1}(\mathbf{Id} - \frac{1}{n}\sum_{i=1}^{n} \mathbf{\Delta}^i) = \frac{1}{n}\sum_{i=1}^{n}\sum_{0 \leq k < i} \mathbf{\Delta}^k$$

In order to conclude, it remains to prove that: $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \mathbf{\Delta}^i = 0$
$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \mathbf{\Delta}^i = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}(\mathbf{P}^i - \mathbf{P}^*) = \lim_{n\to\infty}(\frac{1}{n}\sum_{i=1}^{n} \mathbf{P}^i) - \mathbf{P}^* = 0$

q.e.d. (theorem 3.31) $\Diamond\Diamond\Diamond$

Proof of theorem 3.32

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{k=0}^{i-1}(\mathbf{P}^k - \mathbf{P}^*) = \mathbf{Id} - \mathbf{P}^* + \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{i-1}(\mathbf{P} - \mathbf{P}^*)^k$$

$$= \mathbf{Id} - \mathbf{P}^* + \frac{1}{n}\sum_{i=1}^{n}\sum_{k=0}^{i-1}(\mathbf{P} - \mathbf{P}^*)^k - \mathbf{Id} = -\mathbf{P}^* + \frac{1}{n}\sum_{i=1}^{n}\sum_{k=0}^{i-1}(\mathbf{P} - \mathbf{P}^*)^k$$

Letting n go to ∞ one gets: $\mathbf{D} = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}\sum_{k=0}^{i-1}(\mathbf{P}^k - \mathbf{P}^*)$

Let us consider the series $\mathbf{H}_\alpha \stackrel{\text{def}}{=} \sum_{n=0}^{\infty} \alpha^n(\mathbf{P} - \mathbf{P}^*)^n$ for $0 \leq \alpha < 1$. Since $(\mathbf{P} - \mathbf{P}^*)^n = \mathbf{P}^n - \mathbf{P}^*$, $\{\|\mathbf{P}^n - \mathbf{P}^*\|\}$ is bounded and the series is convergent. Consequently $\mathbf{H}_\alpha$ is convergent and it is the inverse of $\mathbf{Id} - \alpha(\mathbf{P} - \mathbf{P}^*)$.
So: $\mathbf{Id} = \mathbf{H}_\alpha(\mathbf{Id} - \alpha(\mathbf{P} - \mathbf{P}^*)) = \mathbf{H}_\alpha(\mathbf{Id} - \mathbf{P} + \mathbf{P}^*) + (1 - \alpha)\mathbf{H}_\alpha(\mathbf{P} - \mathbf{P}^*)$
Since $\{\mathbf{P}^n - \mathbf{P}^*\}$ is Cesaro convergent to zero, it is also Abel convergent i.e., $\lim_{\alpha\uparrow 1}(1-\alpha)\mathbf{H}_\alpha = 0$.
So $\mathbf{Z} = \lim_{\alpha\uparrow 1} \mathbf{H}_\alpha$.
$\sum_{n=0}^{\infty} \alpha^n(\mathbf{P}^n - \mathbf{P}^*) = \mathbf{H}_\alpha - \mathbf{P}^*$
So $\lim_{\alpha\uparrow 1} \sum_{n=0}^{\infty} \alpha^n(\mathbf{P}^n - \mathbf{P}^*) = \mathbf{Z} - \mathbf{P}^* = \mathbf{D}$
Since $\mathbf{P}\mathbf{P}^* = \mathbf{P}^*\mathbf{P}$, $\mathbf{P}\mathbf{D} = \mathbf{D}\mathbf{P}$ and $\mathbf{P}^*\mathbf{D} = \mathbf{D}\mathbf{P}^*$.
• $\mathbf{D}\mathbf{P}^* = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}\sum_{k=0}^{i-1}(\mathbf{P}^k - \mathbf{P}^*)\mathbf{P}^* = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}\sum_{k=0}^{i-1}(\mathbf{P}^* - \mathbf{P}^*) = 0$
• $\mathbf{D}(\mathbf{Id} - \mathbf{P}) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}\sum_{k=0}^{i-1}(\mathbf{P}^k - \mathbf{P}^*)(\mathbf{Id} - \mathbf{P})$
$= \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}\sum_{k=0}^{i-1}(\mathbf{P}^k - \mathbf{P}^{k+1}) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}(\mathbf{Id} - \mathbf{P}^n) = \mathbf{Id} - \mathbf{P}^*$

q.e.d. (theorem 3.32) $\Diamond\Diamond\Diamond$

Proof of proposition 3.33
One first observes that by convexity the inequations hold for every $d \in D^{MR}$.
Let $\boldsymbol{\pi} = (d_1, d_2, \ldots)$ be a Markovian policy. Using the second inequation with d_k, one gets:

$$\mathbf{g} \geq \mathbf{r}_{d_k} + (\mathbf{P}_{d_k} - \mathbf{Id})\mathbf{h}$$

Then one applies the first inequation with d_{k-1} getting:

$$\mathbf{g} \geq \mathbf{P}_{d_{k-1}}\mathbf{g} \geq \mathbf{P}_{d_{k-1}}\mathbf{r}_{d_k} + \mathbf{P}_{d_{k-1}}(\mathbf{P}_{d_k} - \mathbf{Id})\mathbf{h}$$

Applying iteratively the first inequation with $\mathbf{P}_{d_{k-2}}, \ldots, \mathbf{P}_{d_1}$ one obtains:

$$\mathbf{g} \geq \mathbf{P}_{d_1} \ldots \mathbf{P}_{d_{k-1}}\mathbf{r}_{d_k} + \mathbf{P}_{d_1} \ldots \mathbf{P}_{d_{k-1}}(\mathbf{P}_{d_k} - \mathbf{Id})\mathbf{h}$$

Summing this inequation for k from 1 to n, one gets

$$n\mathbf{g} \geq \mathbf{v}_n^{\boldsymbol{\pi}} + (\mathbf{P}_{d_1} \ldots \mathbf{P}_{d_{n-1}}\mathbf{P}_{d_n} - \mathbf{Id})\mathbf{h}$$

Since the last term is bounded by $2\|h\|$, dividing by n and letting n go to ∞ yields:

$$\mathbf{g} \geq \limsup_{n\to\infty} \frac{1}{n}\mathbf{v}_n^{\boldsymbol{\pi}} = \mathbf{g}_+^{\boldsymbol{\pi}}$$

Since $\boldsymbol{\pi}$ is arbitrary, the result follows.

q.e.d. (proposition 3.33) $\Diamond\Diamond\Diamond$

Proof of proposition 3.34
Let $\mathbf{g}, \mathbf{h}$ be a solution of this system.
We claim that $\mathbf{g}, \mathbf{h} + M\mathbf{g}$ for M large enough is a solution of the system of proposition 3.33.
An equation that could not be fulfilled is an equation of the following kind:
$\mathbf{g}(s) + (\mathbf{h}[s] + M\mathbf{g}[s]) \geq \sum_{s' \in S} \mathbf{P}_d[s, s'](\mathbf{h}[s'] + M\mathbf{g}[s']) + \mathbf{r}_d[s]$
for which $\mathbf{g}[s] > \sum_{s' \in S} \mathbf{P}_d[s, s']\mathbf{g}[s']$
but as $M\mathbf{g}[s]$ occurs on the left side and $\sum_{s' \in S} \mathbf{P}_d[s, s']M\mathbf{g}[s']$ occurs on the right side it is enough to take a value of M large enough to satisfy such an equation. Since there are only a finite number of equations, choosing M as the maximal value provides such a solution.

q.e.d. (proposition 3.34) $\Diamond\Diamond\Diamond$

Proof of proposition 3.35
Observe that the average reward triggered by policy d^∞ is given by: $\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1}(\mathbf{P}_d)^i\mathbf{r}_d$
As $\{(\mathbf{P}_d)^n\}$ is Cesaro convergent to $\mathbf{P}_d^*$, one concludes that: $\mathbf{g}_-^{d^\infty} = \mathbf{g}_+^{d^\infty} = \mathbf{P}_d^*\mathbf{r}_d$
Using lemma 3.29, $\mathbf{P}_d$ is Abel convergent to $\mathbf{P}_d^*$.
Thus: $\lim_{\lambda\uparrow 1}(1-\lambda)\mathbf{v}_\lambda^{d^\infty} = \lim_{\lambda\uparrow 1}(1-\lambda)\sum_{i=0}^{\infty}(\lambda\mathbf{P}_d)^i\mathbf{r}_d = \mathbf{P}_d^*\mathbf{r}_d$

q.e.d. (proposition 3.35) $\Diamond\Diamond\Diamond$

Proof of proposition 3.36
One knows that $\mathbf{v}_\lambda^{d^\infty}$ is the single solution of $\mathbf{r}_d - (\mathbf{Id} - \lambda\mathbf{P}_d)\mathbf{v}_\lambda^{d^\infty} = 0$.
So we prove that the expression of the proposition fulfills this equation.

$$\mathbf{r}_d - \frac{1}{1-\lambda}(\mathbf{Id} - \lambda\mathbf{P}_d)\left(\mathbf{P}_d^*\mathbf{r}_d - \sum_{n=1}^{\infty}(-\rho\mathbf{D}_d)^n\mathbf{r}_d\right) = \mathbf{r}_d - \mathbf{P}_d^*\mathbf{r}_d + \frac{\mathbf{Id} - \lambda\mathbf{P}_d}{1-\lambda}\sum_{n=1}^{\infty}(-\rho\mathbf{D}_d)^n\mathbf{r}_d$$

(using $\mathbf{P}_d\mathbf{P}_d^* = \mathbf{P}_d^*$*)*

$$= (\mathbf{Id} - \mathbf{P}_d^*)\mathbf{r}_d + \frac{\lambda(\mathbf{Id} - \mathbf{P}_d) + (1-\lambda)\mathbf{Id}}{1-\lambda}\sum_{n=1}^{\infty}(-\rho\mathbf{D}_d)^n\mathbf{r}_d$$

$$= (\mathbf{Id} - \mathbf{P}_d^*)\mathbf{r}_d - (\mathbf{Id} - \mathbf{P}_d^*)\sum_{n=0}^{\infty}(-\rho\mathbf{D}_d)^n\mathbf{r}_d + \sum_{n=1}^{\infty}(-\rho\mathbf{D}_d)^n\mathbf{r}_d$$

(using $(\mathbf{Id} - \mathbf{P}_d)\mathbf{D}_d = \mathbf{Id} - \mathbf{P}_d^*$*)*

$$= (\mathbf{Id} - \mathbf{P}_d^*)\mathbf{r}_d - (\mathbf{Id} - \mathbf{P}_d^*)\mathbf{r}_d - \sum_{n=1}^{\infty}(-\rho\mathbf{D}_d)^n\mathbf{r}_d + \sum_{n=1}^{\infty}(-\rho\mathbf{D}_d)^n\mathbf{r}_d = 0$$

(using $\mathbf{P}_d^*\mathbf{D}_d = 0$*)*

q.e.d. (proposition 3.36) ◊◊◊

Proof of corollary 3.37
Using formula of proposition 3.36 we get:

$$\mathbf{v}_\lambda^{d^\infty} = \frac{1}{1-\lambda}\mathbf{P}_d^*\mathbf{r}_d + \frac{1}{\lambda}\mathbf{D}_d\mathbf{r}_d + O(1-\lambda)$$

which can be rewritten as:

$$\mathbf{v}_\lambda^{d^\infty} = \frac{1}{1-\lambda}\mathbf{P}_d^*\mathbf{r}_d + \mathbf{D}_d\mathbf{r}_d + \frac{1-\lambda}{\lambda}\mathbf{D}_d\mathbf{r}_d + O(1-\lambda) = \frac{1}{1-\lambda}\mathbf{P}_d^*\mathbf{r}_d + \mathbf{D}_d\mathbf{r}_d + O(1-\lambda)$$

q.e.d. (proposition 3.37) ◊◊◊

Proof of theorem 3.38
Let $(\mathbf{g}_0, \mathbf{h}_0)$ be a solution of this system. It is immediate that $(\mathbf{g}_0, \mathbf{h}_0)$ fulfill the requirements of proposition 3.34. So $\mathbf{g}_0 \geq \mathbf{g}_+^*$.

Let us define a deterministic decision rule d by choosing some $d(s) \in B_s$. The equation system can be rewritten

$$\mathbf{g}_0 = \mathbf{P}_d\mathbf{g}_0 \text{ and } \mathbf{g}_0 + \mathbf{h}_0 = \mathbf{P}_d\mathbf{h}_0 + \mathbf{r}_d$$

Using the second equation, one gets:

$$\mathbf{g}_0 = \mathbf{r}_d + (\mathbf{P}_d - \mathbf{Id})\mathbf{h}_0$$

Then one applies the first equation getting:

$$\mathbf{g}_0 = \mathbf{P}_d\mathbf{g}_0 = \mathbf{P}_d\mathbf{r}_d + \mathbf{P}_d(\mathbf{P}_d - \mathbf{Id})\mathbf{h}_0$$

Applying iteratively the first equation one obtains:

$$\mathbf{g}_0 = \mathbf{P}_d^k\mathbf{r}_d + \mathbf{P}_d^{k-1}(\mathbf{P}_d - \mathbf{Id})\mathbf{h}_0$$

Summing these equations for k from 1 to n, one gets

$$n\mathbf{g}_0 = \mathbf{u}_n^{d^\infty} + (\mathbf{P}_d^n - \mathbf{Id})\mathbf{h}_0$$

Since the last term is bounded by $2\|\mathbf{h}_0\|$, dividing by n and letting n go to ∞ yields:

$$\mathbf{g}_0 = \lim_{n\to\infty}\frac{1}{n}\mathbf{u}_n^{d^\infty} = \mathbf{g}_+^{d^\infty} = \mathbf{g}_-^{d^\infty}$$

Since $\mathbf{g}_0 \geq \mathbf{g}_+^*$, the result follows.

Let d^∞ be a Blackwell optimal policy. Consider $\lambda \in [\lambda_0, 1[$ an interval where d^∞ is optimal.
Then using characterization of theorem 3.15, one gets for all $s \in S$ and $a \in A_s$:
$\mathbf{v}_\lambda^{d^\infty}[s] \geq r(s,a) + \lambda\sum_{s'\in S} p(s'|s,a)\mathbf{v}_\lambda^{d^\infty}[s']$
Using corollary 3.37
$\frac{1}{1-\lambda}(\mathbf{P}_d^*\mathbf{r}_d)[s] + (\mathbf{D}_d\mathbf{r}_d)[s] + O(1-\lambda)$

$\geq r(s,a) + \lambda \sum_{s' \in S} p(s'|s,a) \left(\frac{1}{1-\lambda} (\mathbf{P}_d^* \mathbf{r}_d)[s'] + (\mathbf{D}_d \mathbf{r}_d)[s'] \right) + O(1-\lambda)$
$= r(s,a) + \sum_{s' \in S} p(s'|s,a) \left(\frac{1}{1-\lambda} (\mathbf{P}_d^* \mathbf{r}_d)[s'] + (\mathbf{D}_d \mathbf{r}_d)[s'] \right) - \sum_{s' \in S} p(s'|s,a) (\mathbf{P}_d^* \mathbf{r}_d)[s'] + O(1-\lambda)$
Thus:
$\frac{1}{1-\lambda} \left((\mathbf{P}_d^* \mathbf{r}_d)[s] - \sum_{s' \in S} p(s'|s,a) (\mathbf{P}_d^* \mathbf{r}_d)[s'] \right) +$
$(\mathbf{D}_d \mathbf{r}_d)[s] - r(s,a) - \sum_{s' \in S} p(s'|s,a) (\mathbf{D}_d \mathbf{r}_d - \mathbf{P}_d^* \mathbf{r}_d)[s'] + O(1-\lambda) \geq 0$
So one must have:
$(\mathbf{P}_d^* \mathbf{r}_d)[s] - \sum_{s' \in S} p(s'|s,a) (\mathbf{P}_d^* \mathbf{r}_d)[s'] \geq 0$
and when the equality holds one must have:
$(\mathbf{D}_d \mathbf{r}_d)[s] - r(s,a) - \sum_{s' \in S} p(s'|s,a) (\mathbf{D}_d \mathbf{r}_d - \mathbf{P}_d^* \mathbf{r}_d)[s'] \geq 0$
which can be rewritten (due to equality) as:
$(\mathbf{D}_d \mathbf{r}_d)[s] - r(s,a) - \sum_{s' \in S} p(s'|s,a) (\mathbf{D}_d \mathbf{r}_d)[s'] + (\mathbf{P}_d^* \mathbf{r}_d)[s] \geq 0$
This establishes that equations (3.4) and (3.5) hold with inequalities instead of equalities.

To prove that the equalities hold, let us choose for state s, action $d(s)$.
Then equation (3.4) holds due to equality $\mathbf{P}_d^* = \mathbf{P}_d \mathbf{P}_d^*$ (see theorem 3.30).
And equation (3.5) holds due to equality $\mathbf{P}_d^* + \mathbf{D}_d = \mathbf{Id} + \mathbf{P}_d \mathbf{D}_d$ (see theorem 3.32).

q.e.d. (proposition 3.38) ◊◊◊

Proof of proposition 3.40
Let us check that $\mathbf{P}_d^* \mathbf{r}_d$, $\mathbf{D}_d \mathbf{r}_d$ and $-\mathbf{D}_d^2 \mathbf{r}_d$ are solutions of this system.

- $(\mathbf{Id} - \mathbf{P}_d) \mathbf{P}_d^* \mathbf{r}_d = (\mathbf{P}_d^* - \mathbf{P}_d^*) \mathbf{r}_d = \mathbf{0}$
- $\mathbf{P}_d^* \mathbf{r}_d + (\mathbf{Id} - \mathbf{P}_d) \mathbf{D}_d \mathbf{r}_d = (\mathbf{P}_d^* + (\mathbf{Id} - \mathbf{P}_d) \mathbf{D}_d) \, \mathbf{r}_d = \mathbf{r}_d$
- $\mathbf{D}_d \mathbf{r}_d - (\mathbf{Id} - \mathbf{P}_d) \mathbf{D}_d^2 \mathbf{r}_d = (\mathbf{Id} - (\mathbf{Id} - \mathbf{P}_d) \mathbf{D}_d) \, \mathbf{D}_d \mathbf{r}_d = \mathbf{P}_d^* \mathbf{D}_d \mathbf{r}_d = \mathbf{0}$

Let $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$ be a solution of this system. $\mathbf{P}_d \mathbf{x} = \mathbf{x}$ entails $\mathbf{P}_d^* \mathbf{x} = \mathbf{x}$.
So $\mathbf{x} = \mathbf{P}_d^* \mathbf{x} = \mathbf{P}_d^* \mathbf{r}_d - \mathbf{P}_d^* (\mathbf{Id} - \mathbf{P}_d) \mathbf{y} = \mathbf{P}_d^* \mathbf{r}_d$
$\mathbf{0} = \mathbf{P}_d^* \left(\mathbf{y} + (\mathbf{Id} - \mathbf{P}_d) \mathbf{z} \right) = \mathbf{P}_d^* \mathbf{y}$
Thus using second equation of the system:
$\mathbf{r}_d - \mathbf{P}_d^* \mathbf{r}_d = (\mathbf{Id} - \mathbf{P}_d) \mathbf{y} = (\mathbf{Id} - \mathbf{P}_d + \mathbf{P}_d^*) \mathbf{y}$
which can be rewritten as:
$\mathbf{y} = (\mathbf{Id} - \mathbf{P}_d + \mathbf{P}_d^*)^{-1} (\mathbf{Id} - \mathbf{P}_d^*) \mathbf{r}_d = (\mathbf{D}_d + \mathbf{P}_d^*) (\mathbf{Id} - \mathbf{P}_d^*) \mathbf{r}_d = \mathbf{D}_d \mathbf{r}_d$

q.e.d. (proposition 3.40) ◊◊◊

Proof of proposition 3.41
Assume that for all s, $\mathsf{Improve}(d,s) = \emptyset$. Observing that choosing $a = d(s)$, leads to equality in the two equations, one concludes that $(\mathbf{P}_d^* \mathbf{r}_d, \mathbf{D}_d \mathbf{r}_d)$ fulfill the equations of theorem 3.38. So d^∞ is an optimal policy.

Assume that for at least one s, $\mathsf{Improve}(d,s) \neq \emptyset$ and consider d' fulfilling the requirements of the proposition. Define policy $\boldsymbol{\pi} \stackrel{\text{def}}{=} (d', d, d, \ldots)$.

If $d'(s) = d(s)$ then for all λ, $\mathbf{v}_\lambda^{\boldsymbol{\pi}}[s] = \mathbf{v}_\lambda^{d^\infty}[s]$.

On the other hand,
$\mathbf{v}_\lambda^{\boldsymbol{\pi}} = \mathbf{r}_{d'} + \lambda \mathbf{P}_{d'} \mathbf{v}_\lambda^{d^\infty} = \mathbf{r}_{d'} + \lambda \mathbf{P}_{d'} (\frac{1}{1-\lambda} \mathbf{P}_d^* \mathbf{r}_d + \mathbf{D}_d \mathbf{r}_d + O(1-\lambda))$
$= \mathbf{r}_{d'} + \frac{\lambda}{1-\lambda} \mathbf{P}_{d'} \mathbf{P}_d^* \mathbf{r}_d + \lambda \mathbf{P}_{d'} \mathbf{D}_d \mathbf{r}_d + O(1-\lambda)$
$= \mathbf{r}_{d'} + \frac{1}{1-\lambda} \mathbf{P}_{d'} \mathbf{P}_d^* \mathbf{r}_d - \mathbf{P}_{d'} \mathbf{P}_d^* \mathbf{r}_d + \mathbf{P}_{d'} \mathbf{D}_d \mathbf{r}_d + O(1-\lambda)$
So:

$$\mathbf{v}_\lambda^{\boldsymbol{\pi}} - \mathbf{v}_\lambda^{d^\infty} = \frac{1}{1-\lambda} (\mathbf{P}_{d'} - \mathbf{Id}) \mathbf{P}_d^* \mathbf{r}_d + (\mathbf{r}_{d'} + \mathbf{P}_{d'} \mathbf{D}_d \mathbf{r}_d - \mathbf{P}_{d'} \mathbf{P}_d^* \mathbf{r}_d - \mathbf{D}_d \mathbf{r}_d) + O(1-\lambda) \tag{3.9}$$

So if $a \stackrel{\text{def}}{=} d'(s) \neq d(s)$ then

- either $(\mathbf{P}_d^* \mathbf{r}_d)[s] < \sum_{s' \in S} p(s'|s,a) (\mathbf{P}_d^* \mathbf{r}_d)[s']$ and using the term where $\frac{1}{1-\lambda}$ occurs in equation (3.9), one concludes that for λ enough close to 1, $\mathbf{v}_\lambda^{\boldsymbol{\pi}}[s] > \mathbf{v}_\lambda^{d^\infty}[s]$.

- or $(\mathbf{P}_d^*\mathbf{r}_d)[s] = \sum_{s' \in S} p(s'|s,a)(\mathbf{P}_d^*\mathbf{r}_d)[s']$ and $((\mathbf{P}_d^*+\mathbf{D}_d)\mathbf{r}_d)[s] < r(s,a)+\sum_{s' \in S} p(s'|s,a)(\mathbf{D}_d\mathbf{r}_d)[s']$.
  Due to the former equation, the constant term in equation (3.9) becomes:
  $(\mathbf{r}_{d'} + \mathbf{P}_{d'}\mathbf{D}_d\mathbf{r}_d - (\mathbf{P}_d^* + \mathbf{D}_d)\mathbf{r}_d)[s]$
  So one concludes that for λ close enough to 1, $\mathbf{v}_\lambda^{\boldsymbol{\pi}}[s] > \mathbf{v}_\lambda^{d^\infty}[s]$.

Summarizing for λ close enough to 1, $\mathbf{v}_\lambda^{\boldsymbol{\pi}} > \mathbf{v}_\lambda^{d^\infty}$ which can be rewritten as:

$$\mathbf{r}_{d'} + \lambda \mathbf{P}_{d'} \mathbf{v}_\lambda^{d^\infty} > \mathbf{v}_\lambda^{d^\infty}$$

or as:

$$\mathbf{v}_\lambda^{d'^\infty} = (\mathbf{Id} - \lambda \mathbf{P}_{d'})^{-1} \mathbf{r}_{d'} > \mathbf{v}_\lambda^{d^\infty}$$

Multiplying by $1-\lambda$ and letting λ go to 1, one gets:

$$\mathbf{g}^{d'^\infty} \geq \mathbf{g}^{d^\infty}$$

q.e.d. (proposition 3.41) $\Diamond\Diamond\Diamond$

Proof of proposition 3.42
Due to the second equation for all s, $\sum_{a \in A_s}(\mathbf{x}[s,a] + \mathbf{y}[s,a]) > 0$. So there is at least one policy d fulfilling the requirements of the proposition.

We define the set $S_{\mathbf{x}} \stackrel{\text{def}}{=} \{s \in S \mid \sum_{a \in A_s} \mathbf{x}[s,a] > 0\}$ and we introduce the optimal solution $(\mathbf{g}, \mathbf{h})$ of the primal linear program built from $(\mathbf{x}, \mathbf{y})$ (see section 3.3.4).

Let us prove that $S_{\mathbf{x}}$ is closed in the DTMC specified by $\mathbf{P}_d$.
Let $s \notin S_{\mathbf{x}}$, one gets: $0 = \sum_{a \in A_s} \mathbf{x}[s,a] = \sum_{s' \in S} \sum_{a \in A_{s'}} \mathbf{x}[s',a] p(s|s',a) \geq 0$
So for all $s' \in S$ and $a \in A_{s'}$: $\mathbf{x}[s',a]p(s|s',a) = 0$
Choosing $s' \in S_{\mathbf{x}}$ and $a = d(s')$ yields: $\mathbf{P}_d[s',s] \stackrel{\text{def}}{=} p(s|s',d(s')) = 0$

Let us prove that all states of $S \setminus S_{\mathbf{x}}$ are transient in the DTMC specified by $\mathbf{P}_d$.
Otherwise since $S_{\mathbf{x}}$ is closed, there exists a terminal scc S' included in $S \setminus S_{\mathbf{x}}$. Since $(\mathbf{x}, \mathbf{y})$ is a basic solution, the columns of the matrix of constraints indexed by $\{\mathbf{y}[s,d(s)]\}_{s \in S'}$ are linearly independent.

Let us analyze the components of these columns. Those corresponding to the equations involving only $\mathbf{x}$ are null. Those corresponding to the equations where α_s occurs with $s \notin S'$ are also null since $\mathbf{P}_d[s',s] = 0$ for all $s' \in S'$ (S' is a terminal scc). So these columns restricted to equations where α_s occurs with $s \in S'$ are also linearly independent. But summing these columns one gets for the component related to α_s with $s \in S'$:

$$\mathbf{y}[s,d(s)] \left(1 - \sum_{s' \in S'} p(s'|s,d(s))\right) = 0$$

yielding a contradiction.
So for all $s \in S$ and $s' \notin S_{\mathbf{x}}$, $\mathbf{P}_d^*[s,s'] = 0$.
Using the slackness property applied to $\mathbf{x}[s,d(s)]$ with $s \in S_{\mathbf{x}}$, one gets:

$$\mathbf{g}[s] + \mathbf{h}[s] - \sum_{s' \in S} p(s'|s,a)\mathbf{h}[s'] = r(s,d(s))$$

So for all $s \in S$:

$$\mathbf{g}^{d^\infty}[s] = (\mathbf{P}_d^*\mathbf{r}_d)[s] = \sum_{s' \in S_{\mathbf{x}}} \mathbf{P}_d^*[s,s'] r(s',d(s'))$$

$$= \sum_{s' \in S_{\mathbf{x}}} \mathbf{P}_d^*[s,s'] \left(\mathbf{g}[s'] + \mathbf{h}[s'] - \sum_{s'' \in S} p(s''|s',d(s'))\mathbf{h}[s']\right)$$

$$= (\mathbf{P}_d^*\mathbf{g})[s] + (\mathbf{P}_d^*(\mathbf{h} - \mathbf{P}_d\mathbf{h}))[s] = (\mathbf{P}_d^*\mathbf{g})[s]$$

So in order to conclude it remains to prove that $\mathbf{g} = \mathbf{P}_d^* \mathbf{g}$.
We are going to prove a stronger statement: $\mathbf{g} = \mathbf{P}_d \mathbf{g}$.

For all $s \in S$ and $a \in A_s$, $\mathbf{g}[s] - \sum_{s' \in S} p(s'|s,a)\mathbf{g}[s'] \geq 0$
So: $\sum_{s \in S} \sum_{a \in A_s} \mathbf{x}[s,a] \left(\mathbf{g}[s] - \sum_{s' \in S} p(s'|s,a)\mathbf{g}[s']\right) \geq 0$
On the other hand:
$\sum_{s \in S} \sum_{a \in A_s} \mathbf{x}[s,a] \left(\mathbf{g}[s] - \sum_{s' \in S} p(s'|s,a)\mathbf{g}[s']\right)$
$= \sum_{s \in S} \left(\sum_{a \in A_s} \mathbf{x}[s,a] - \sum_{s' \in S} \sum_{a \in A_{s'}} p(s|s',a)\mathbf{x}[s',a]\right) \mathbf{g}[s] = 0$
So for all $s \in S$, and $a \in A_s$, $\mathbf{x}[s,a] \left(\mathbf{g}[s] - \sum_{s' \in S} p(s'|s,a)\mathbf{g}[s']\right) = 0$
Thus for $s \in S_x$ as $\mathbf{x}[s, d(s)] > 0$, we deduce that: $\mathbf{g}[s] - \sum_{s' \in S} p(s'|s,d(s))\mathbf{g}[s'] = 0$

For $s \notin S_x$ one has $\mathbf{y}[s, d(s)] > 0$. So using the slackness property applied to variable $\mathbf{y}[s, d(s)]$ and looking at the corresponding inequation in the primal linear program one also obtains:
$\mathbf{g}[s] - \sum_{s' \in S} p(s'|s,d(s))\mathbf{g}[s'] = 0$

q.e.d. (proposition 3.42) $\Diamond\Diamond\Diamond$

Chapter 4

Stochastic Games

4.1 Presentation

In MDP, the agent who selects a strategy (also called a policy) may be viewed as a player in a game whose opponent is chance. Since a random player does not aim at minimizing the gain of the agent, MDP are also called games with one and half players. In this chapter, we generalize this idea by considering games with two players or with two and half players.

Let us first give a general overview. Whatever the variant that we study in this chapter, we assume that the two players, denoted Max and Min have opposite objectives. Such games are called *zero-sum games* meaning that, given any play, the sum of the rewards of the players is null. One may alternatively formalize it as a single reward r which maps plays to values and that Max (respectively Min) tries to maximize (respectively minimize). Denote $\mathcal{G}$ the structure and the protocol of the game equipped with the reward r. At this point we could imagine two (among other) ways of playing in $\mathcal{G}$. Either Max announces its strategy σ and then Min selects a strategy τ or they do it in the reverse order. Observe that the first (respectively second) way of playing is the least (respectively most) favourable for Max among all possible ways of revealing the strategies. Let us denote h the random play in $\mathcal{G}$, once the strategies are fixed and $\mathbf{E}_{\mathcal{G}}^{\sigma,\tau}(r(h))$ the expected reward of the play triggered by these strategies. Then with the first way of playing, there are strategies for Max that ensure an expected reward as close as possible to:

$$val_{\downarrow}(\mathcal{G}) = \sup_{\sigma} \inf_{\tau} \mathbf{E}_{\mathcal{G}}^{\sigma,\tau}(r(h))$$

In addition $val_{\downarrow}(\mathcal{G})$ is the greatest value fulfilling this property. With the second way of playing, there are strategies for Min that ensure an expected reward as close as possible to:

$$val_{\uparrow}(\mathcal{G}) = \inf_{\tau\in} \sup_{\sigma} \mathbf{E}_{\mathcal{G}}^{\sigma,\tau}(r(h))$$

In addition $val_{\uparrow}(\mathcal{G})$ is the smallest value fulfilling this property. By definition, $val_{\downarrow}(\mathcal{G}) \leq val_{\uparrow}(\mathcal{G})$. An interesting question is whether these values are equal. In such a case, one says that the game has a value or that it is *determined*. This important issue will be handled in Section 4.2. We will focus on another important issue for games that are determined: does some player have an optimal strategy and in the positive case does there exist "simple" optimal strategies? For instance, a *mixed* strategy includes random choices while a *pure* strategy only performs deterministic choices. The other kind of restrictions we are looking for, is related to the memory required to manage an optimal strategy. If the structure of the game is a graph whose vertices may be visited several times, a *memoryless* strategy only depends on the current vertex and so is less computationally costly.

Example 4.1 (The spinner game revisited) *In this game, the player has to compose a five-digit number whose digits are randomly chosen by a spinner during five rounds. After every round*

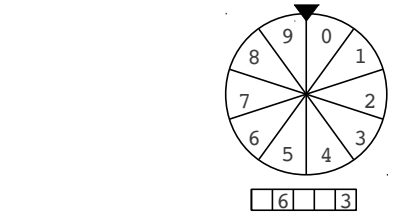

Figure 4.1: The spinner game revisited

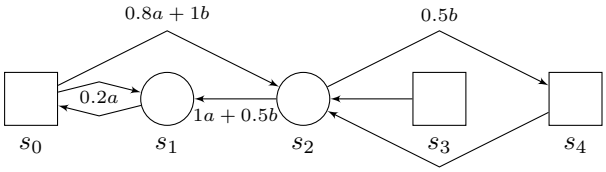

Figure 4.2: A simple SG

(except the last one), the player chooses in which position he inserts the current digit. The goal of the player is to obtain the largest number as possible. In figure 3.1, the spinner has successively output 3 placed by the player in the fifth position and 6 placed by the player in the second position.

However at any time but at most once, the TV presenter may switch the current digit with a previous one when their value difference is at most 2. The goal of the presenter is to obtain the smallest number as possible.

A stochastic game (SG) is a finite transition system where any state belongs to either player Max or Min. The dynamic of the system is defined as follows. The player owning the current state chooses (possibly randomly) an enabled *action*. Then the environment randomly selects the next state. The distribution depends on the current state and the selected action. There are several ways to define rewards that will be introduced later on.

Definition 4.2 *A* SG $\mathcal{G} \stackrel{def}{=} (S, \{A_s\}_{s\in S}, p)$ *is defined by:*

- $S = S_{\mathsf{Min}} \uplus S_{\mathsf{Max}}$*, the finite set of states;*
- *For every state* s*,* A_s*, the finite set of actions enabled in* s*.* $A \stackrel{def}{=} \bigcup_{s\in S} A_s$ *is the whole set of actions.*
- p*, a mapping from* $\{(s,a) \mid s \in S, a \in A_s\}$ *to the set of distributions over* S*.* $p(s'|s,a)$ *denotes the probability to go from* s *to* s' *if* a *is selected.*

A *history* $h \stackrel{\text{def}}{=} s_0 a_0 \dots s_i a_i \dots$ is a finite or infinite sequence alternating states and actions such that when s_{i+1} is defined $p(s_{i+1}|s_i, a_i) > 0$.

Example 4.3 (A simple SG) *A stochastic game is depicted as a labelled graph (see Figure 4.2). States of player* Max *are represented by circles (◯). States of player* Min *are represented by squares (□). An edge* (s, s') *is labelled by* $\sum_{a\in A_s} p(s'|s,a)a$ *(when non null). When a state has a single enabled action with Dirac distribution (i.e. wihout randomness) as for* s_1*,* s_3 *and* s_4*, we omit the label on the single outgoing edge.*

In order to obtain a stochastic process, one needs to fix the non deterministic features of the SG. A *strategy* of a player P is a mapping from histories ending in a state $s \in S_{\mathsf{P}}$ to a distribution over A_s. Classes of strategies are defined depending on two criteria.

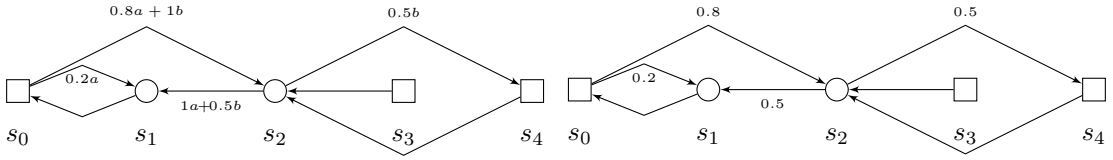

Figure 4.3: From SG to DTMC.

- the information used in the history. When a strategy only depends on the last state, it is called *memoryless*;
- the way the selection is performed. When a strategy deterministically selects its actions, it is called *pure*.

The DTMC $\mathcal{G}^{\sigma,\tau}$ is the behaviour of the SG $\mathcal{G}$ once strategies σ and τ of respectively Max and Min are chosen. Its states are information used in strategies. One denotes h the random infinite history and $\mathbf{Pr}^{\sigma,\tau}_{\mathcal{G},s}$ (resp. $\mathbf{E}^{\sigma,\tau}_{\mathcal{G},s}$) the probability measure (the expectation operator) in $\mathcal{G}^{\sigma,\tau}$ when starting in s.

Example 4.4 *Consider the SG depiceted in Figure 4.2. Let σ be the pure memoryless strategy of* Max *that selects b in s_2 and τ be the strategy of* Min *that selects a in s_0. Then $\mathcal{G}^{\sigma,\tau}$ is depicted in Figure 4.3.*

We now introduce several kinds of SG depending on the specification of rewards. In the next definition we specify for $h = s_0 a_0 s_1 \ldots$, an infinite history, its reward depending on the game. For some of the games we need to introduce $\mathit{Inf}(h) = \{s \mid \forall i\ \exists j > i\ s_j = s\}$, the set of states occurring infinitely often in h.

Definition 4.5 *Let $\mathcal{G}$ be a game. Then:*

- *Let r be a mapping from $\{(s,a) \mid s \in S, a \in A_s\}$ to $[0,1]$ and λ be a real in $]0,1[$. Then $\mathcal{G}$ is a* discounted *game if the reward of h is defined by: $r(h) = \sum_{n\in\mathbb{N}} \lambda^n r(s_n, a_n)$;*
- *Let r be a mapping from $\{(s,a) \mid s \in S, a \in A_s\}$ to $[0,1]$. Then $\mathcal{G}$ is a* mean payoff *game if the reward of h is defined by: $r(h) = \liminf_{n\to\infty} \frac{1}{n} \sum_{i<n} r(s_i, a_i)$;*
- *Let pri be a mapping from S to $\mathbb{N}$. Then $\mathcal{G}$ is a* parity *game if the reward of h is defined by: $r(h) = \mathbb{1}_{pri(h)\ \text{is even}}$ where $pri(h) = \max(pri(s) \mid s \in \mathit{Inf}(h))$;*
- *Let pri be an injective mapping from S to $\mathbb{N}$. Then $\mathcal{G}$ is a* parity *game if the reward of h is defined by: $r(h) = r(s_{\max})$ where $s_{\max}(h) = \arg\max(pri(s) \mid s \in \mathit{Inf}(h))$.*

Observe that priority SG extend parity SG. Moreover the requirement that numerical rewards must belong to $[0,1]$ is only introduced for convenience and will be sometimes relaxed as it can be recovered by an affine transformation.

We now introduce the main problems we adress in this chapter.

Determinacy problem. Let s be a state of a SG $\mathcal{G}$. Define $m_s = \sup_\sigma \inf_\tau \mathbf{E}^{\sigma,\tau}_{\mathcal{G},s}(r(h))$ and $M_s = \inf_\tau \sup_\sigma \mathbf{E}^{\sigma,\tau}_{\mathcal{G},s}(r(h))$ By construction, $m_s \leq M_s$. Does $m_s = M_s$? For all games we consider, the answer is yes. It is called the *value* of s in $\mathcal{G}$ and denoted $val_{\mathcal{G}}(s)$.

Existence of optimal strategies. Does there exist σ (resp. τ) such that $\inf_\tau \mathbf{E}^{\sigma,\tau}_{\mathcal{G},s}(r(h)) = val_{\mathcal{G}}(s)$ (resp. $\sup_\sigma \mathbf{E}^{\sigma,\tau}_{\mathcal{G},s}(r(h)) = val_{\mathcal{G}}(s)$)? The answer is still yes.

Classes of optimal strategies. How can σ and τ be chosen? For all games we consider, there exist pure and memoryless optimal strategies.

Computational problems. What is the complexity of the associated decision problems like $val_{\mathcal{G}}(s) \geq \nu$ for some input threshold ν? For most of the games we consider, these problems belong to $\mathsf{NP} \cap \mathsf{coNP}$. Furthermore, we establish polynomial time reductions between problems implying that their complexity are "equivalent".

4.2 Pure memoryless determinacy

4.2.1 Discounted games

Consider $\mathcal{G}$ a discounted game and let L be the mapping from $\mathbb{R}^S$ to $\mathbb{R}^S$ defined by:

- When $s \in S_{\mathsf{Max}}$,

$$L(\mathbf{v})[s] \stackrel{\text{def}}{=} \max \left(r(s,a) + \lambda \sum_{s' \in S} p(s'|s,a)\mathbf{v}[s'] \mid a \in A_s \right)$$

- When $s \in S_{\mathsf{Min}}$,

$$L(\mathbf{v})[s] \stackrel{\text{def}}{=} \min \left(r(s,a) + \lambda \sum_{s' \in S} p(s'|s,a)\mathbf{v}[s'] \mid a \in A_s \right)$$

L "selects" the best decision rule for the owner of s in a game that stops at time 1 including a terminal reward $\lambda\mathbf{v}$. Using a proof similar to the one for discounted rewards in MDP, one shows that L is Lipschitz-continuous with Lipschitz constant equal to $\lambda < 1$. Thus L admits a unique fixed-point denoted $\mathbf{v}^*_\lambda$.
Let σ^* be a pure memoryless strategy of player Max that selects in $s \in S_{\mathsf{Max}}$ some a_s such that:

$$r(s,a_s) + \lambda \sum_{s' \in S} p(s'|s,a_s)\mathbf{v}^*_\lambda[s'] = \mathbf{v}^*_\lambda[s]$$

Let τ^* be a pure memoryless strategy of player Min that selects in $s \in S_{\mathsf{Min}}$ some a_s such that:

$$r(s,a_s) + \lambda \sum_{s' \in S} p(s'|s,a_s)\mathbf{v}^*_\lambda[s'] = \mathbf{v}^*_\lambda[s]$$

The next two lemmas show that the items of $\mathbf{v}^*_\lambda$ are the values of the game and imply pure memoryless determinacy.

Lemma 4.6 *Let $\mathbf{v}_n$ be the infimum of the expected discounted rewards up to time n in $\mathcal{G}^{\sigma^*,\tau}$ against an arbitrary strategy τ of player* Min*. Then:*

$$\mathbf{v}_n[s] \geq \mathbf{v}^*_\lambda[s] - \frac{\lambda^n}{1-\lambda}$$

Proof of Lemma 4.6
We establish the proof by induction on n. The basis case is a consequence that expected rewards for discounted games are bounded by $\frac{\lambda}{1-\lambda}$.
• Inductive step when $s \in S_{\mathsf{Max}}$.

$$\begin{aligned}
\mathbf{v}_{n+1}[s] = & \quad r(s,a_s) + \lambda \sum_{s' \in S} p(s'|s,a_s)\mathbf{v}_n[s'] \\
\geq & \quad r(s,a) + \lambda \sum_{s' \in S} p(s'|s,a)(\mathbf{v}^*_\lambda[s'] - \frac{\lambda^n}{1-\lambda}) \\
= & \quad r(s,a_s) + \lambda \sum_{s' \in S} p(s'|s,a_s)\mathbf{v}^*_\lambda[s'] - \frac{\lambda^{n+1}}{1-\lambda} \\
= & \quad \mathbf{v}^*_\lambda[s] - \frac{\lambda^{n+1}}{1-\lambda}
\end{aligned}$$

• Inductive step when $s \in S_{\mathsf{Min}}$. Let a be the action selected by τ.

$$\begin{aligned}
\mathbf{v}_{n+1}[s] = & \quad r(s,a) + \lambda \sum_{s' \in S} p(s'|s,a)\mathbf{v}_n[s'] \\
\geq & \quad r(s,a) + \lambda \sum_{s' \in S} p(s'|s,a)(\mathbf{v}^*_\lambda[s'] - \frac{\lambda^n}{1-\lambda}) \\
= & \quad r(s,a) + \lambda \sum_{s' \in S} p(s'|s,a)\mathbf{v}^*_\lambda[s'] - \frac{\lambda^{n+1}}{1-\lambda} \\
\geq & \quad \mathbf{v}^*_\lambda[s] - \frac{\lambda^{n+1}}{1-\lambda}
\end{aligned}$$

q.e.d. (Lemma 4.6) ◊◊◊

The proof of the second lemma is very close to the previous one and so is omitted.

Lemma 4.7 *Let $\mathbf{w}_n$ be the supremum of the expected discounted rewards up to time n in $\mathcal{G}^{\sigma,\tau^*}$ against an arbitrary strategy σ of player* Max. *Then:*

$$\mathbf{w}_n[s] \leq \mathbf{v}^*_\lambda[s] + \frac{\lambda^n}{1-\lambda}$$

The following theorem is a direct consequence of the lemmas.

Theorem 4.8 *Any discounted game $\mathcal{G}$ is determined with value $\mathbf{v}^*_\lambda$ and optimal strategies σ^* for* Max *and τ^* for* Min.

4.2.2 Mean Payoff games

In order to establish pure memoryless determinacy of mean payoff games we proceed as for MDP establishing the existence of Blackwell strategies. Let $\mathcal{G}$ be a mean payoff game and $\mathcal{G}_\lambda$ the discounted version with discount λ.

Pick some increasing sequence $\{\lambda_n\}_{n\in\mathbb{N}}$ such that $\lim_{n\to\infty} \lambda_n = 1$. Let σ_n and τ_n be pure memoryless optimal strategies for $\mathcal{G}_{\lambda_n}$. Since there are only finite such strategies, some strategies σ^* and τ^* must occur *simultaneously* infinitely often. By considering a subsequence, one assumes that σ^* and τ^* are optimal for all $\mathcal{G}_{\lambda_n}$.

Lemma 4.9 *There exists n_0 such that for all $\lambda \geq \lambda_{n_0}$, σ^* and τ^* are optimal in $\mathcal{G}_\lambda$.*

Proof of Lemma 4.9
We prove it by contradiction.

Assume there exists some increasing sequence $\{n_k\}_{k\in\mathbb{N}}$ and $\lambda_{n_k} < \mu_k < \lambda_{n_{k+1}}$ such that for all k, there exist $s \in S$ and pure memoryless strategies σ and τ (optimal for μ_k) fulfilling:

- either $\mathbf{E}^{\sigma,\tau}_{\mathcal{G}_{\mu_k},s}(r(h)) > \mathbf{E}^{\sigma^*,\tau}_{\mathcal{G}_{\mu_k},s}(r(h))$;
- or $\mathbf{E}^{\sigma,\tau}_{\mathcal{G}_{\mu_k},s}(r(h)) < \mathbf{E}^{\sigma,\tau^*}_{\mathcal{G}_{\mu_k},s}(r(h))$.

For pure memoryless strategies σ and τ, $\mathbf{E}^{\sigma,\tau}_{\mathcal{G}_\lambda,s}(r(h))$ is a rational function of λ. So define:

$$f_s(\lambda) = \prod_{\mathbf{E}^{\sigma,\tau}_{\mathcal{G}_\lambda,s}(r(h)) \neq \mathbf{E}^{\sigma',\tau'}_{\mathcal{G}_\lambda,s}(r(h))} \mathbf{E}^{\sigma,\tau}_{\mathcal{G}_\lambda,s}(r(h)) - \mathbf{E}^{\sigma',\tau'}_{\mathcal{G}_\lambda,s}(r(h))$$

Then some f_s would have an infinite number of zeroes.

q.e.d. (Lemma 4.9) ◊◊◊

We are now in position to prove the pure memoryless strategies of mean payoff games.

Theorem 4.10 *Any mean payoff game $\mathcal{G}$ is determined with optimal strategies σ^* for* Max *and τ^* for* Min.

Proof of Theorem 4.10
Let us denote the random history $h = s_0 a_0 s_1 \ldots$
Consider the MDP $\mathcal{G}^{\tau^*}$ obtained by using strategy τ^* for player Min. σ^* is a Blackwell policy in $\mathcal{G}^{\tau^*}$. So it is optimal for mean payoff reward:

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{i<n} \mathbf{E}^{\sigma,\tau^*}_{\mathcal{G},s}(r(s_i,a_i)) \leq \lim_{n\to\infty} \frac{1}{n} \sum_{i<n} \mathbf{E}^{\sigma^*,\tau^*}_{\mathcal{G},s}(r(s_i,a_i))$$

Using a similar reasoning, one gets for all s and τ:

$$\liminf_{n\to\infty} \frac{1}{n} \sum_{i<n} \mathbf{E}^{\sigma^*,\tau}_{\mathcal{G},s}(r(s_i,a_i)) \geq \lim_{n\to\infty} \frac{1}{n} \sum_{i<n} \mathbf{E}^{\sigma^*,\tau^*}_{\mathcal{G},s}(r(s_i,a_i))$$

Thus:

$$\begin{aligned}
&\sup_\sigma \inf_\tau (\liminf_{n\to\infty} \frac{1}{n} \sum_{i<n} \mathbf{E}^{\sigma,\tau}_{\mathcal{G},s}(r(s_i,a_i))) && \geq \inf_\tau (\liminf_{n\to\infty} \frac{1}{n} \sum_{i<n} \mathbf{E}^{\sigma^*,\tau}_{\mathcal{G},s}(r(s_i,a_i))) \\
&= \lim_{n\to\infty} \frac{1}{n} \sum_{i<n} \mathbf{E}^{\sigma^*,\tau^*}_{\mathcal{G},s}(r(s_i,a_i)) && \geq \sup_\sigma (\limsup_{n\to\infty} \frac{1}{n} \sum_{i<n} \mathbf{E}^{\sigma,\tau^*}_{\mathcal{G},s}(r(s_i,a_i))) \\
&\geq \inf_\tau \sup_\sigma (\limsup_{n\to\infty} \frac{1}{n} \sum_{i<n} \mathbf{E}^{\sigma,\tau}_{\mathcal{G},s}(r(s_i,a_i))) &&
\end{aligned}$$

q.e.d. (Theorem 4.10) $\diamondsuit\diamondsuit\diamondsuit$

4.2.3 Priority games

In order to establish pure memoryless determinacy of priority games, we recall an elementary result of analysis about non decreasing 1-Lipschitz functions (see Figure 4.4 for an example of such functions).

Lemma 4.11 *Let f be a non decreasing function from $[0,1]$ to $[0,1]$ that is 1-Lipschitz: $|f(x) - f(x')| \leq |x - x'|$. Then the set of fixed points of f is a non empty interval denoted $[a,b]$. Furthermore denoting $f^\infty(x) = \lim_{n\to\infty} f^{(n)}(x)$:*

- *for all $x < a$, $f^\infty(x) = a$ and $f(x) > x$;*
- *for all $a \leq x \leq b$, $f^\infty(x) = x$;*
- *for all $b < x$, $f^\infty(x) = b$ and $f(x) < x$.*

Proof of Lemma 4.11
Let $I = \{x \mid x = f(x)\}$. Since $f(0) \geq 0$ and $f(1) \leq 1$, using the intermediate value theorem I is not empty. Since f is continuous, I is closed.
Let $x < z < y$ with $x = f(x)$ and $y = f(y)$. Then:

- $f(z) - f(x) \leq z - x$ implying $f(z) \leq z$;
- $f(y) - f(z) \leq y - z$ implying $f(z) \geq z$.

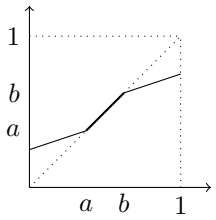

Figure 4.4: A non decreasing 1-Lipschitz function.

Thus $f(z) = z$.

Let $x < a$ then $f(a) - f(x) \leq a - x$ implying $f(x) \geq x$ but since $f(x) \neq x$, $f(x) > x$. Since f is non decreasing, $f(x) \leq f(a) = a$. By induction $f^n(x) \leq a$ and by continuity $f^\infty(x) \leq a$. Again by continuity, $f(f^\infty(x)) = f^\infty(x)$. Thus $f^\infty(x) \geq a$.

The case $x > b$ is handled similarly.

q.e.d. (Lemma 4.11) $\Diamond\Diamond\Diamond$

A state s of a stochastic game is *absorbing* if $A_s = \{a\}$ for some a and $p(s|s,a) = 1$. Observe that the priority of an absorbing state is irrelevant. A state s is *vanishing* if for all s' and $a \in A_{s'}$, $p(s|s',a) = 0$. A state is *relevant* if it is neither absorbing nor vanishing. The proof of the pure memoryless determinacy is done by induction on the number of relevant states. So we introduce a construction for the inductive step.

Let $\mathcal{G}$ be a stochastic game with s the relevant state with maximal priority. We define the game $\mathcal{G}'$ as follows.

- Add an absorbing state $\tilde{s}$ with reward v.
- Redirect all incoming transitions in s to $\tilde{s}$:
  $p'(\tilde{s}|s',a) = p(s|s',a)$ and $p'(s|s',a) = 0$.

Since s is vanishing in $\mathcal{G}'$, it has less relevant states than $\mathcal{G}$. This construction is illustrated in Figure 4.5.

Theorem 4.12 *Any priority game $\mathcal{G}$ is determined with pure memoryless optimal strategies.*

Proof of Theorem 4.12
By induction on the number of relevant states.

Basis case. When there is no relevant state, all strategies are memoryless. The value of an absorbing state s is $r(s)$. The value of a vanishing state s belonging to Max (resp. Min) is:

$$\max_{a \in A_s} \sum_{s'} p(s'|s,a) r(s') \qquad (\text{resp. } \min_{a \in A_s} \sum_{s'} p(s'|s,a) r(s'))$$

and an action corresponding to a pure optimal strategy is some:

$$\arg\max_{a \in A_s} \sum_{s'} p(s'|s,a) r(s') \qquad (\text{resp. } \arg\min_{a \in A_s} \sum_{s'} p(s'|s,a) r(s'))$$

Inductive step. Let $\mathcal{G}$ be a stochastic game with s the relevant state with maximal priority. We consider all rewards for s and denote $\mathcal{G}_v$ the game $\mathcal{G}$ with $r(s) = v$ and $\mathcal{G}'_v$ the reduced game for which induction applies.

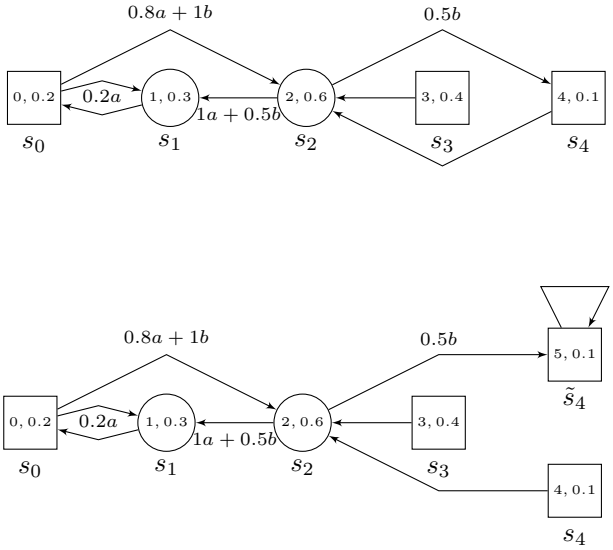

Figure 4.5: Reducing the number of relevant states.

One denotes $val_{\mathcal{G}'_v}(t)$ by $f_t(v)$, the value of state t in $\mathcal{G}'_v$. Let $v < v'$. Then:

$$\begin{aligned}
\mathbf{E}^{\sigma,\tau}_{\mathcal{G}'_v,t}(r(h)) = \quad & \mathbf{Pr}^{\sigma,\tau}_{\mathcal{G}'_v,t}(s_{max}(h) \neq s)\mathbf{E}^{\sigma,\tau}_{\mathcal{G}'_v,t}(r(h)|s_{max}(h) \neq s) + v\mathbf{Pr}^{\sigma,\tau}_{\mathcal{G}'_v,t}(s_{max}(h) = s) \\
\leq \quad & \mathbf{E}^{\sigma,\tau}_{\mathcal{G}'_{v'},t}(r(h)|s_{max}(h) \neq s) + v'\mathbf{Pr}^{\sigma,\tau}_{\mathcal{G}'_v,t}(s_{max}(h) = s) \\
= \quad & \mathbf{E}^{\sigma,\tau}_{\mathcal{G}'_{v'},t}(r(h))
\end{aligned}$$

Thus f_t is non decreasing and 1-Lipschitz.

• **A first property of $\mathcal{G}'_v$ and $\mathcal{G}_v$.**

Let σ_v be a pure memoryless optimal strategy of Max in $\mathcal{G}'_v$. Assume $v < f_s(v)$. Then there exists $\varepsilon > 0$ such that given any strategy τ of Min: the probability to reach $\tilde{s}$ from s in $\mathcal{G}'^{\sigma_v,\tau}_v$ is bounded by $1 - \varepsilon$.

Otherwise by a family of strategies τ_n reaching $\tilde{s}$ with probability at least $1 - \frac{1}{n}$ Min can ensure that $f_s(v) \leq v$.

So when $v < f_s(v)$, for all τ the probability to visit infinitely often s in $\mathcal{G}^{\sigma_v,\tau}_v$ is null.

• **A second property of $\mathcal{G}'_v$ and $\mathcal{G}_v$.**

Let σ_v be a pure memoryless optimal strategy of Max in $\mathcal{G}'_v$. Assume $v \leq f_s(v)$. Let Div be the event: h does not reach $\tilde{s}$. Then for all strategy τ of Min: *(when defined)* $\mathbf{E}^{\sigma_v,\tau}_{\mathcal{G}'_v,s}(r(h)|Div) \geq f_s(v)$

$f_s(v) \leq \mathbf{E}^{\sigma_v,\tau}_{\mathcal{G}'_v,s}(r(h)) = \mathbf{Pr}^{\sigma_v,\tau}_{\mathcal{G}'_v,s}(Div)\mathbf{E}^{\sigma_v,\tau}_{\mathcal{G}'_v,s}(r(h)|Div) + (1 - \mathbf{Pr}^{\sigma_v,\tau}_{\mathcal{G}'_v,s}(Div))v$
So $\mathbf{E}^{\sigma_v,\tau}_{\mathcal{G}'_v,s}(r(h)|Div) \geq f_s(v)$.
Let R_n be the event: h visits s exactly n times.

Thus if $v \leq f_s(v)$ then for all strategy τ of Min: *(when defined)* $\mathbf{E}^{\sigma_v,\tau}_{\mathcal{G}_v,s}(r(h)|R_n) \geq f_s(v)$.

• **A first lower bound.**

Let σ_v be a pure memoryless optimal strategy of Max in $\mathcal{G}'_v$.
If $v \leq f_s(v)$ then for all strategy τ of Min:
$f_s(v) \leq \mathbf{E}^{\sigma_v,\tau}_{\mathcal{G}_v,s}(r(h))$.

Let R_∞ be the event: h visits s infinitely often.

$$\mathbf{E}^{\sigma_v,\tau}_{\mathcal{G}_v,s}(r(h)) = \sum_n \mathbf{Pr}^{\sigma_v,\tau}_{\mathcal{G}_v,s}(R_n)\mathbf{E}^{\sigma_v,\tau}_{\mathcal{G}_v,s}(r(h)|R_n) + \mathbf{Pr}^{\sigma_v,\tau}_{\mathcal{G}_v,s}(R_\infty)v$$

Recall that $\mathbf{E}^{\sigma_v,\tau}_{\mathcal{G}_v,s}(r(h)|R_n) \geq f_s(v)$.
Now:

- either $f_s(v) = v$ and thus $\mathbf{E}^{\sigma_v,\tau}_{s}(r(h)) \geq f_s(v)$;
- or $f_s(v) > v$ and implying $\mathbf{Pr}^{\sigma_v,\tau}(R_\infty) = 0$ implying $\mathbf{E}^{\sigma_v,\tau}_{s}(r(h)) \geq f_s(v)$.

• **A second lower bound.**

There exists a pure memoryless strategy σ of Max in $\mathcal{G}_v$ such that:

1. σ is optimal in $\mathcal{G}'_{f^\infty(v)}$;
2. for all τ, $\mathbf{E}^{\sigma,\tau}_{\mathcal{G}_v,s}(r(h)) \geq f^\infty_s(v)$;
3. for all t, for all τ, $\mathbf{E}^{\sigma,\tau}_{\mathcal{G}_v,t}(r(h)) \geq f_t(f^\infty_s(v))$.

Proof of 1,2: Case $f_s(v) \leq v$.
A pure memoryless optimal strategy $\sigma_{f^\infty_s(v)}$ in $\mathcal{G}'_{f^\infty_s(v)}$ ensures for s a value $f^\infty_s(v)$ in $\mathcal{G}_{f^\infty_s(v)}$ thus also in $\mathcal{G}_v$.

Proof of 1,2: Case $v < f_s(v)$.
A pure memoryless optimal strategy σ_v in $\mathcal{G}'_v$ ensures for s a value $f_s(v)$ in $\mathcal{G}_v$. Since for all τ $\mathbf{Pr}^{\sigma_v,\tau}_{\mathcal{G}_v,s}(R_\infty) = 0$, σ_v ensures a value $f_s(v)$ in $\mathcal{G}_{v'}$ for any v'. Let us note $a = f^\infty_s(v)$ the least fixed point of f_s. Observe that $v < f_s(v)$ is equivalent to $v < a$. There is a finite number of pure memoryless strategies. Consider a strategy σ such for all $\varepsilon > 0$ there is some $a - \varepsilon < v < a$ with $\sigma_v = \sigma$. Thus σ ensures for s a value a in all $\mathcal{G}_{v'}$. Since σ is optimal in $\mathcal{G}'_{v'}$ for v' as close as possible to a, σ is optimal in $\mathcal{G}'_a$.

Proof of 3.
Since σ is optimal in $\mathcal{G}'_{f^\infty_s(v)}$, for all τ, $f_t(f^\infty_s(v)) \leq \mathbf{E}^{\sigma,\tau}_{\mathcal{G}'_{f^\infty_s(v)},t}(r(h))$ Let R be the event h reaches $\tilde{s}$. Then:

$$\begin{aligned}
\mathbf{E}^{\sigma,\tau}_{\mathcal{G}'_{f^\infty_s(v)},t}(r(h)) = \quad & (1 - \mathbf{Pr}^{\sigma,\tau}_{\mathcal{G}'_{f^\infty_s(v)},t}(R))\mathbf{E}^{\sigma,\tau}_{\mathcal{G}'_{f^\infty_s(v)},t}(r(h)|R^c) + \mathbf{Pr}^{\sigma,\tau}_{\mathcal{G}'_{f^\infty_s(v)},t}(R) f^\infty_s(v) \\
\leq \quad & (1 - \mathbf{Pr}^{\sigma,\tau}_{\mathcal{G}_v,t}(R))\mathbf{E}^{\sigma,\tau}_{\mathcal{G}_v,t}(r(h)|R^c) + \mathbf{Pr}^{\sigma,\tau}_{\mathcal{G}_v,t}(R)\mathbf{E}^{\sigma,\tau}_{\mathcal{G}_{f^\infty_s(v)},t}(r(h)|R) \\
= \quad & \mathbf{E}^{\sigma,\tau}_{\mathcal{G}_v,t}(r(h))
\end{aligned}$$

By a similar reasoning, one gets:

There exists a pure memoryless strategy τ of Min in $\mathcal{G}_v$ such that:

- τ is optimal in $\mathcal{G}'_{f^\infty_s(v)}$;
- for all σ, $\mathbf{E}^{\sigma,\tau}_{\mathcal{G}_v,s}(r(h)) \leq f^\infty_s(v)$.
- for all t, for all σ, $\mathbf{E}^{\sigma,\tau}_{\mathcal{G}_v,t}(r(h)) \leq f_t(f^\infty_s(v))$.

which concludes the proof.

q.e.d. (Theorem 4.12) ◊◊◊

4.3 Computational issues

4.3.1 Complexity results

Using the results on MDP, one straightforwardly gets the following result. In the next subsection, we establish by a polynomial time reduction the same result for parity games.

Theorem 4.13 *Let $\mathcal{G}$ be a mean payoff or discounted game and a value v. The decision problem whether $val_{\mathcal{G}}(s) \geq v$ belongs to* NP ∩ coNP.

Proof of Theorem 4.13
This problem is in NP.
Guess a pure memoryless strategy σ of Max. Build the MDP $\mathcal{G}^{\sigma}$. Minimize (in polynomial time) the objective o. Answer yes if $o \geq v$.
This problem is in coNP.
Guess a pure memoryless strategy τ of Min. Build the MDP $\mathcal{G}^{\tau}$. Maximize (in polynomial time) the objective o. Answer no if $o < v$.

q.e.d. (Theorem 4.13) ◊◊◊

While the exact complexity of these problems is a longstanding open issue, there is a particular case where it is known to be solvable in polynomial time.

Theorem 4.14 *Let $\mathcal{G}$ be a discounted game where λ is represented in unary. Then the values and the optimal strategis can be computed in polynomial time.*

Proof of Theorem 4.14

The algorithm proceeds by first computing two values n and d, and then iterating n times the operator L and d-rounding the final value.

> $\mathbf{v} \leftarrow \mathbf{0}$
> **For** i **from** 1 to n **do** $\mathbf{v} \leftarrow L(\mathbf{v})$
> **For** $s \in S$ **do** $val_{\mathcal{G}}[s] \leftarrow [\mathbf{v}[s]]_d$ *(where $[x]_d$ is the rational $\frac{c}{d}$ closest to x)*

Computation of d.
Let σ and τ be some pure memoryless optimal policies and the DTMC $\mathcal{G}^{\sigma,\tau}$. Denote:

- $\mathbf{P}$ its transition matrix
- $\mathbf{r}$ defined by $\mathbf{r}[s] = r(s, a_s)$ where a_s is the action selected by the owner of s

Then $val_{\mathcal{G}}[s] = ((\mathbf{Id} - \lambda\mathbf{P})^{-1}\mathbf{r})[s]$. So:

- one computes β the product of the denominators of the probabilities and rewards occurring in $\mathcal{G}$ and λ in polynomial time;
- one rewrites all values (including $1 - \lambda x$ for appropriate x's) as $\frac{\alpha}{\beta}$;
- one computes d the product of the α's in polynomial time.

By analysis of the Cramer's rule for solving linear equations, one deduces that any $val_{\mathcal{G}}[s]$ can be written as $\frac{c}{d}$ for some c.
Computation of n.
L the contracting operator fulfills $\|val_{\mathcal{G}} - L^n(\mathbf{0})\| \leq \frac{\lambda^n}{1-\lambda}$. Thus:
$[L^n(\mathbf{0})[s]]_d = val_{\mathcal{G}}[s]$ when $\frac{\lambda^n}{1-\lambda} < \frac{1}{2d}$, i.e. $n \log_2(\frac{1}{\lambda}) > \log_2(\frac{1}{1-\lambda}) + \log_2(d) + 1$
Write $\lambda = \frac{p}{q}$. Then $\log_2(\frac{1}{\lambda}) \geq \log_2(1 + \frac{1}{p}) \geq \frac{1}{p}$ and $\log_2(\frac{1}{1-\lambda}) \leq \log_2(q)$.

So any value of n greater than $p(\log_2(q) + \log_2(d) + 1)$ implies $\frac{\lambda^n}{1-\lambda} < \frac{1}{2d}$.

Complexity analysis.

- $\log_2(d) + 1$ is equivalent to the size of the representation of d so polynomial w.r.t. the size of the problem.
- p is polynomial w.r.t. the size of the problem when λ is specified in unary.
- The operations involve numbers whose denominators are bounded by d^n and numerators by nd^n so performed in polynomial time.

q.e.d. (Theorem 4.14) ◊◊◊

4.3.2 Polynomial time reductions

In this part, we show that whatever the objective among discounted reward, mean payoff or parity, the complexity of solving these games is essentially the same. More precisely, we establish polynomial time reductions between the problems. When solving problem a with the help of problem b, polynomial time reduction is a polynomial time algorithm for solving an instance of a that can use the solutions of one or several instances of b in constant time (as an oracle).

Theorem 4.15 *There is polynomial time reduction from the problem of computing the optimal strategies in mean payoff games to the problem of computing the optimal strategies in discounted games.*

Proof of Theorem 4.15
Let $\mathcal{G}$ be a mean payoff game. The proof is based on the computation of an appropriate λ_∞ such that optimal strategies are Blackwell strategies. The search of an appropriate λ_∞ is done by an analysis of the possible zeroes of $\mathbf{E}^{\sigma,\tau}_{\mathcal{G}_\lambda,s}(r(h)) - \mathbf{E}^{\sigma',\tau'}_{\mathcal{G}_\lambda,s}(r(h))$:

$$\mathbf{E}^{\sigma,\tau}_{\mathcal{G}_\lambda,s}(r(h)) - \mathbf{E}^{\sigma',\tau'}_{\mathcal{G}_\lambda,s}(r(h)) = (\mathbf{Id} - \lambda\mathbf{P})^{-1}\mathbf{r} - (\mathbf{Id} - \lambda\mathbf{P}')^{-1}\mathbf{r}'$$

for some $\mathbf{P}$, $\mathbf{P}'$, $\mathbf{r}$, $\mathbf{r}'$ with items occurring in $\mathcal{G}$;
Let M be the product of denominators occurring in values of $\mathcal{G}$. and $X = 1 - \lambda$ with X in $]0, \frac{1}{2}]$. The coefficients of $\mathbf{Id}-(1-X)\mathbf{P}$, $\mathbf{Id}-(1-X)\mathbf{P}'$, $\mathbf{r}$ and $\mathbf{r}'$ can be written as $aX+b$ with numerators of a and b bounded by M and denominator M. Looking for zeroes one may omit the common denominator.

$$(\mathbf{Id} - (1-X)\mathbf{P})^{-1}\mathbf{r} - (\mathbf{Id} - (1-X)\mathbf{P}')^{-1}\mathbf{r}' = \frac{N}{D} - \frac{N'}{D'}$$

with $N, D, N', D' \in \mathbb{Z}[X]$.
Using Cramer's rule the coefficients of $ND' - N'D$ are bounded by:

$$R = 2n(n!)^4 M^{2n}$$

Let $P \in \mathbb{Z}[X]$ whose coefficients are bounded by R. Then the smallest (if any) root of P in $]0, \frac{1}{2}]$ is at least $\frac{1}{2R}$. Thus an upper bound of λ_∞ is $1 - \frac{1}{2R+1}$. Since R has a polynomial size w.r.t. the size of $\mathcal{G}$, computing λ_∞ can performed in polynomial time.

q.e.d. (Theorem 4.15) ◊◊◊

A reachability objective where w.l.o.g. the target states are absorbing is a very particular case of parity objective: parity 2 for target states and parity 1 for other states. However it is enough for the reduction of discounted games.

Theorem 4.16 *There is a polynomial time reduction from the problem of computing the optimal strategies and game values in discounted games to the problem of computing the optimal strategies and game values in reachability games.*

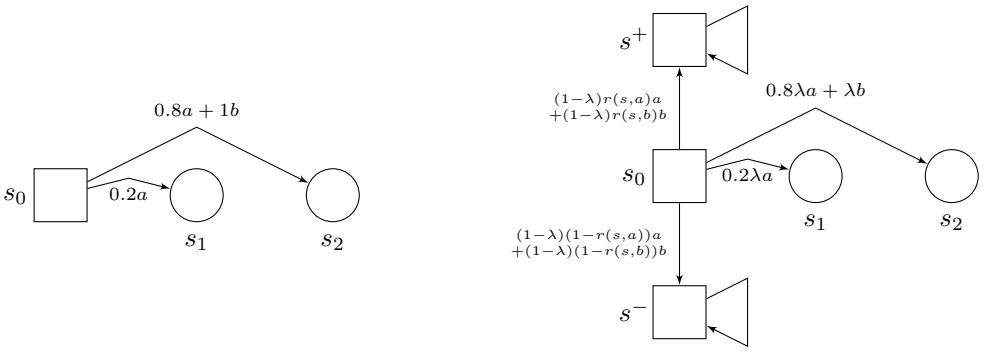

Figure 4.6: From discounted games to reachability games.

Proof of Theorem 4.16
Let $\mathcal{G}$ a discounted game with discount λ. One builds a reachability game $\mathcal{G}_\lambda$ with additional states s^+ and s^- and reachability target s^+ as illustrated in Figure 4.6. Then by a straightforward examination, for all s, σ and τ:

$$\mathbf{E}^{\sigma,\tau}_{\mathcal{G},s}(r(h)) = (1-\lambda)\mathbf{Pr}^{\sigma,\tau}_{\mathcal{G}_\lambda,s}(h \text{ reaches } s^+)$$

which concludes the proof.

q.e.d. (Theorem 4.16) $\diamondsuit\diamondsuit\diamondsuit$

The reduction from parity games to mean payoff games is much more intricate than the previous ones. It proceeds in two steps:

- Computing the states s for which $val_{\mathcal{G}}(s) \in \{0,1\}$;
- Reducing the parity game to a mean-payoff game once these states are computed.

Let us call a game *pure* is there is no random choice inside: once an action is selected, there is a single target state reached with probability 1. So in the graph representation of a pure game, there is a bijection between actions and egdes. Thus we will sometimes omit to label edges with actions when it is not necessary. We also sometimes consider that a strategy selects the next state instead of the action that leads to this state.

In order to show that computing the states s for which $val_{\mathcal{G}}(s) \in \{0,1\}$ can be reduced to solving to mean payoff games, one also proceeds in two steps:

- Reducing this problem to a pure parity game;
- Reducing a pure parity game to a pure mean-payoff game.

We present the different stages in the reverse order starting by the reduction between pure games.

Theorem 4.17 *There is a polynomial time reduction from the problem of computing the optimal strategies an game values in pure parity games to the problem of computing the optimal strategies and game values in pure mean payoff games.*

Proof of Theorem 4.17
Let $\mathcal{G}$ a pure parity game. One builds a reachability game $\mathcal{G}'$ by defining rewards as follows.

$$\text{When } pri(s) = x \text{ in } \mathcal{G}, r(s,a) = (-m)^x \text{ in } \mathcal{G}' \text{ with } m = |S|$$

Such a construction is illustrated in Figure 4.7.

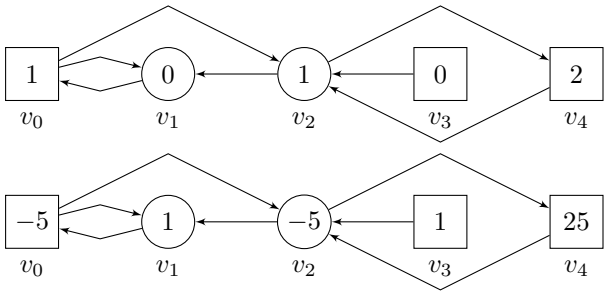

Figure 4.7: From pure parity games to pure mean payoff games.

Observe that the value of a pure parity game $\mathcal{G}$ belongs to $\{0, 1\}$. We claim that the mean payoff game $\mathcal{G}'$ fulfills $val_{\mathcal{G}'}(s) > 0$ iff $val_{\mathcal{G}}(s) = 1$.

• Let σ be a pure optimal strategy of Player Max in $\mathcal{G}$ and τ' a pure optimal strategy of Player Min in $\mathcal{G}'$. $\mathcal{G}^{\sigma,\tau'}$ is a graph where any vertex has exactly one successor. From s one reaches a circuit. Let p be the maximal priority occurring in the circuit. If p is even then $\mathbf{E}^{\sigma,\tau'}_{\mathcal{G}',s}(r(h)) \geq m^p - (m-1)m^{p-1} > 0$. Thus:

$$val_{\mathcal{G}}(s) = 1 \text{ implies } val_{\mathcal{G}'}(s) > 0$$

• Let τ be a pure optimal strategy of Player Min in $\mathcal{G}$ and σ' a pure optimal strategy of Player Max in $\mathcal{G}'$. $\mathcal{G}^{\sigma',\tau}$ is a graph where any vertex has exactly one successor. From s one reaches a circuit. Let p be the maximal priority occurring in the circuit. If p is odd then $\mathbf{E}^{\sigma',\tau}_{\mathcal{G}',s}(r(h)) \leq -m^p + (m-1)m^{p-1} < 0$; Thus:

$$val_{\mathcal{G}}(s) = 0 \text{ implies } val_{\mathcal{G}'}(s) < 0$$

q.e.d. (Theorem 4.17) ◊◊◊

We now reduce the problem of computing states s in a parity game $\mathcal{G}$ for which $val_{\mathcal{G}}(s) = 1$ to the value problem in a pure parity game $\mathcal{G}'$. Game $\mathcal{G}'$ is built as follows. We call the former problem "the value 1 problem". Let $p_{\max}$ be the maximal priority assumed to be even w.l.o.g. For all $s \in S$ with $pri(s) = p$ and $a \in A_s$:

- Add to S_{Max}: $\tilde{s}^q_a$ with $q \geq p-1$ and q even and $\hat{s}^q_a$ with $q \geq p$ and q odd;
- Add to S_{Min}: s_a and $\hat{s}^q_a$ with $q \geq p$ and q even.

The priority of the new states are the following ones: $pri(s_a) = pri(\tilde{s}^q_a) = p$ and $pri(\hat{s}^q_a) = q$. The set of edges (i.e. actions) is:

- (s, s_a) and $(s_a, \tilde{s}^q_a)$ for all $\tilde{s}^q_a$;
- $(\tilde{s}^q_a, \hat{s}^q_a)$ and $(\tilde{s}^q_a, \hat{s}^{q+1}_a)$ *when defined*;
- $(\hat{s}^q_a, s')$ when $p(s'|s, a) > 0$.

The construction of $\mathcal{G}'$ is illustrated in Figure 4.8. In a pure parity parity game $\mathcal{G}'$ the winning set of player Max (resp. Min) is the set $\{s \mid val_{\mathcal{G}'}(s) = 1\}$ (resp. $\{s \mid val_{\mathcal{G}'}(s) = 0\}$). Let E' (resp. O') the winning set of player Max (resp. Min) in $\mathcal{G}'$ and $E = E' \cap S$ and $O = O' \cap S$. The next two lemmas establish the correctness of this construction.

Lemma 4.18 *For all $s \in E$, $val_{\mathcal{G}}(s) = 1$.*

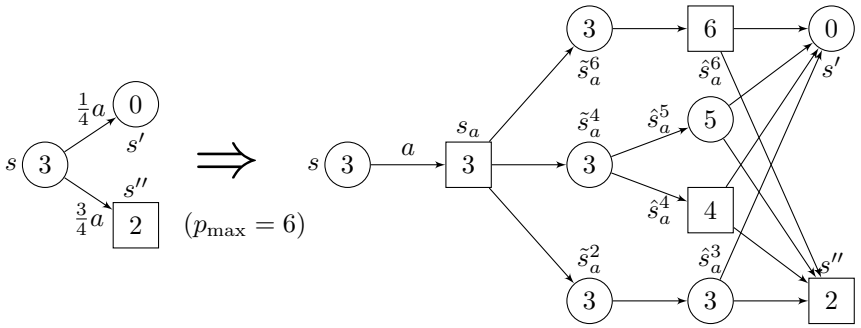

Figure 4.8: From parity games to pure parity games.

Proof of 4.18
Let σ be a pure memoryless optimal strategy of Max in $\mathcal{G}'$. We claim that in the MDP $\mathcal{G}^\sigma$, one never leaves E. Assume by contradiction that there exists $s \in E$ and $a \in A_s^\sigma$ such that $p(s'|s,a) > 0$ and $s' \in O$. In $\mathcal{G}'$ (after possibly selecting a),

- in s_a, Min could select $\tilde{s}_a^{p_{\max}}$;
- and in $\hat{s}_a^{p_{\max}}$ Min could select $s' \in O$, a contradiction.

Let τ be a pure memoryless optimal strategy of Min in the MDP $\mathcal{G}^\sigma$. Consider $\mathcal{M}$ the Markov chain $\mathcal{G}^{\sigma,\tau}$ restricted to E. Assume there exists $\mathcal{C}$ a terminal s.c.c. of $\mathcal{M}$ whose maximal priority is odd, say $2r+1$ for state s_0. Let τ' be (partially) defined as follows. For all $s \in \mathcal{C} \cap S_{\mathsf{Min}}$, $\tau'(s) = \tau(s)$. Let $\mathcal{C}^\bullet = \{s_a \mid s \in \mathcal{C} \cap S \ a \in A_s \text{ is selected by } \sigma \text{ or } \tau\}$. For all $s_a \in \mathcal{C}^\bullet$:

- $\tau'(s_a) = \tilde{s}_a^{2r}$;
- if $\sigma(\tilde{s}_a^{2r}) = \hat{s}_a^{2r}$ then $\tau'(\hat{s}_a^{2r}) = s'$ with s' minimizing the distance to s_0 in $\mathcal{G}^{\sigma,\tau}$.

Consider in $\mathcal{G}'$ the set of states $S^* = \mathcal{C} \cup \mathcal{C}^\bullet \cup \{s_a^{2r}, \sigma(s_a^{2r}) \mid s_a \in \mathcal{C}^\bullet\}$. Observe that for all $t \in S^*$, $pri(t) \leq 2r+1$. Every state in S^* has exactly one successor defined by σ or τ' still in S^*. Consider any circuit in the induced graph:

- either some state $\hat{s}_a^{2r+1}$ occurs in the circuit;
- or s_0 occurs in the circuit.

Thus $S^* \cap E' = \emptyset$ which contradicts the definition of $\mathcal{M}$.

q.e.d. (4.18) $\Diamond\Diamond\Diamond$

Lemma 4.19 *For all $s \in O$, $val_{\mathcal{G}}(s) < 1$.*

Proof of 4.19
Let τ be a pure memoryless optimal strategy of Min in $\mathcal{G}'$ and consider the MDP $\mathcal{G}^\tau$. Let σ be a pure memoryless optimal strategy of Max in $\mathcal{G}^\tau$ and consider the DTMC $\mathcal{G}^{\sigma,\tau}$.
Let $\mathcal{H}$ be the graph over S', the set of vertices, defined by:

- If $s \in S_{\mathsf{Max}}$ (resp. $t \in S'_{\mathsf{Min}}$) then $(s, s_{\sigma(s)})$ (resp. $(t, \tau(t))$ is an edge;
- for other t, any edge (t,t') of $\mathcal{G}'$ is an edge.

Let $s_0 \in O$ belonging to a terminal s.c.c. $\mathcal{C}$ of $\mathcal{H}$. By construction, $\mathcal{C} \subseteq O'$ and the maximal priority in $\mathcal{C}$ is odd.

We prove by induction that for all s reachable from s_0 in $\mathcal{G}^{\sigma,\tau}$, $s \in \mathcal{C}$. Let $a \in A_s$ be selected either by σ or τ. Then in s_a, τ does not select $\tilde{s}_a^{p_{\max}}$. Otherwise $\hat{s}_a^{p_{\max}}$ would belong to $\mathcal{C}$. Let $\tilde{s}_a^{2\ell}$ be selected by τ. Then $\hat{s}_a^{2\ell+1}$ belongs to $\mathcal{C}$ and so all s' with $p(s'|s,a) > 0$ belongs to $\mathcal{C}$. Thus in $\mathcal{G}^{\sigma,\tau}$, s_0 belongs to a terminal s.c.c. with all states in O.

We claim that for all $s \in O$, there is a positive probability in $\mathcal{G}^{\sigma,\tau}$ to reach a state $s' \in O$ such that s' belongs to a terminal s.c.c. $\mathcal{C}$ of $\mathcal{H}$. We prove it by induction on the length of a path from s along O' to some $s' \in O$ of a terminal s.c.c. $\mathcal{C}$ of $\mathcal{H}$. Assume the path starts by $ss_a\tilde{s}_a^r\hat{s}_a^\ell s'$. for some a selected either by σ or τ, and some r and some ℓ. Then $p(s'|s,a) > 0$.

Thus, for all $s \in O$ there is a positive probability in $\mathcal{G}^{\sigma,\tau}$ to reach a terminal s.c.c. with all states in O.

Assume there exists $\mathcal{C}$ a terminal s.c.c. of $\mathcal{G}^{\sigma,\tau}$ with all states in O whose maximal priority is even, say $2r$ for state s_0. Let σ' be (partially) defined as follows. For all $s \in \mathcal{C} \cap S_{\mathsf{Max}}$, $\sigma'(s) = \sigma(s)$. Let $\mathcal{C}^\bullet = \{s_a \mid s \in \mathcal{C} \cap S \ a \in A_s$ is selected by σ or $\tau\}$. For all $s_a \in \mathcal{C}^\bullet$:

- If $\tau(s_a) = \tilde{s}_a^{2\ell}$ with $\ell \geq r$ then $\sigma'(\tilde{s}_a^{2\ell}) = \hat{s}_a^{2\ell}$;
- If $\tau(s_a) = \tilde{s}_a^{2\ell}$ with $\ell < r$ then $\sigma'(\tilde{s}_a^{2\ell}) = \hat{s}_a^{2\ell+1}$ and $\sigma'(\hat{s}_a^{2\ell+1}) = s'$ with s' minimizing the distance to s_0 in $\mathcal{G}^{\sigma,\tau}$.

Consider in $\mathcal{G}'$ the set of states $S^* = \mathcal{C} \cup \mathcal{C}^\bullet \cup \{\tau(s_a), \sigma'(\tau(s_a)) \mid s_a \in \mathcal{C}^\bullet\}$. Every state in S^* has exactly one successor defined by σ' or τ still in S^*. Consider the maximal priority of any circuit in the induced graph:

- either its is 2ℓ for some $\ell \geq r$ and state $\hat{s}_a^{2\ell}$;
- or it is $2r$ with s_0 occuring in the circuit.

Thus $S^* \cap O' = \emptyset$ which contradicts the definition of $\mathcal{C}$.

q.e.d. (4.19) ◊◊◊

As an immediate consequence of the lemmas, one gets:

Theorem 4.20 *There is a polynomial time reduction from the problem of computing the optimal strategies for the value 1 problem in parity games to the problem of computing the optimal strategies in pure parity games.*

Before buliding the last step of the reduction from parity games to mean payoff games, one needs an auxilliary lemma about DTMC.

Lemma 4.21 *Let $\mathcal{M}$ be an irreducible Markov chain with m states and minimum positive transition probability δ. Then for all $s \in S$,*

$$\pi_\infty(s) \stackrel{def}{=} \lim_{n\to\infty} \frac{1}{n} \sum_{i<n} \mathbf{Pr}(X_i = s) \geq \frac{1}{m}\delta^{m-1}$$

Proof of 4.21
Consider s_0, a state with maximal Cesaro-limit probability $\pi_\infty(s_0) \geq \frac{1}{m}$. In the $\mathcal{M}$, there is a path of length $\ell \leq m-1$ from s_0 to s. Thus:

$$\mathbf{Pr}(X_{i+\ell} = s) \geq \delta^\ell \mathbf{Pr}(X_i = s_0) \geq \delta^{m-1}\mathbf{Pr}(X_i = s_0)$$

implying:

$$\pi_\infty(s) \geq \pi_\infty(s_0)\delta^{m-1} \geq \frac{1}{m}\delta^{m-1}$$

q.e.d. (4.21) ◊◊◊

Theorem 4.22 *There is a polynomial time reduction from the problem of computing the optimal strategies in parity games to the problem of computing the optimal strategies in mean payoff games.*

Proof of 4.22
Let $\mathcal{G}$ be a parity game with m states and δ minimal positive probability: Define:

$$S_i = \{s \mid val_{\mathcal{G}} = i\} \text{ for } i \in \{0, 1\}$$

By Theorems 4.20 and 4.17, these sets and their corresponding optimal strategies can be computed using a reduction to pure mean payoff games. $\mathcal{G}'$ the mean payoff game with same structure as $\mathcal{G}$ is defined by:

- For all $s \in S_1$ and $a \in A_s$, $r(s, a) = 1$;
- For all $s \in S_0$ and $a \in A_s$, $r(s, a) = -1$;
- For all $s \notin S_0 \cup S_1$ with $p = pri(s)$ and $a \in A_s$, $r(s, a) = (\frac{-2m}{\delta^{m-1}})^p$;

Observe that this reduction is performed in polynomial time. We claim that:

$$\boxed{\text{For all } s \in S,\ val_{\mathcal{G}'}(s) = 2val_{\mathcal{G}}(s) - 1}$$

Let us prove that $val_{\mathcal{G}'}(s) \geq 2val_{\mathcal{G}}(s) - 1$.
Let σ (resp. τ) be a pure optimal strategy of Player Max (resp. Min) in $\mathcal{G}$ and σ' (resp. τ') be a pure optimal strategy of Player Max (resp. Min) in $\mathcal{G}'$.

$\mathbf{Pr}_{\mathcal{G},s}^{\sigma,\tau'}(h \text{ reaches } S_0) \leq 1 - val_{\mathcal{G}}(s)$ Otherwise combining τ and τ', Min would ensure a value strictly less than $val_{\mathcal{G}}(s)$.

• Let $\mathcal{C}$ be a terminal s.c.c. of $\mathcal{G}^{\sigma,\tau'}$. Then:

- either $S_1 \cap \mathcal{C} \neq \emptyset$ implying $\mathcal{C} \subseteq S_1$ and thus $val_{\mathcal{G}'}(t) = 1$ for all $t \in \mathcal{C}$;
- either $S_0 \cap \mathcal{C} \neq \emptyset$ implying $\mathcal{C} \subseteq S_0$ and thus $val_{\mathcal{G}'}(t) = -1$ for all $t \in \mathcal{C}$;
  Let us denote $\mathcal{C}_0$ the union of these s.c.c.
- or $\mathcal{C} \cap (S_0 \cup S_1) = \emptyset$.

In the latter case, all states $t \in \mathcal{C}$ fulfill $0 < val_{\mathcal{G}}(t) < 1$. Thus $z \in \mathcal{C}$, a vertex with maximal priority, fulfills $p \stackrel{\text{def}}{=} pri(z)$ is even.
When $p = 0$, for all $t \in \mathcal{C}$, $r(t, a) = 1$. So one immediately gets $\mathbf{E}_{\mathcal{G},t}^{\sigma,\tau'}(r(h)) = 1$.
When $p > 0$, the contribution of z to the mean payoff reward is at least:

$$\frac{1}{m}\delta^{m-1}(\frac{2m}{\delta^{m-1}})^p = 2(\frac{2m}{\delta^{m-1}})^{p-1}$$

The accumulated contribution of all $t \in \mathcal{C} \setminus \{z\}$ is at least: $-(\frac{2m}{\delta^{m-1}})^{p-1}$. So for all $t \in \mathcal{C}$, $\mathbf{E}_{\mathcal{G},t}^{\sigma,\tau'}(r(h)) \geq (\frac{2m}{\delta^{m-1}})^{p-1} \geq 1$. Thus:

$$\begin{aligned}
val_{\mathcal{G}'}(s) \geq & \quad -\mathbf{Pr}_{\mathcal{G},s}^{\sigma,\tau'}(h \text{ reaches } \mathcal{C}_0) + (1 - \mathbf{Pr}_{\mathcal{G},s}^{\sigma,\tau'}(h \text{ reaches } \mathcal{C}_0)) \\
= & \quad 1 - 2\mathbf{Pr}_{\mathcal{G},s}^{\sigma,\tau'}(h \text{ reaches } \mathcal{C}_0) \\
\geq & \quad 1 - 2\mathbf{Pr}_{\mathcal{G},s}^{\sigma,\tau'}(h \text{ reaches } \mathcal{S}_0) \\
\geq & \quad 1 - 2(1 - val_{\mathcal{G}}(s)) \\
= & \quad 2val_{\mathcal{G}}(s) - 1
\end{aligned}$$

One gets $val_{\mathcal{G}'}(s) \leq 2val_{\mathcal{G}}(s) - 1$ by a similar reasoning about $\mathcal{G}^{\sigma',\tau}$.

q.e.d. (4.22) ◊◊◊

Chapter 5

Probabilistic Automata

5.1 Presentation

Let us consider a MDP modelling a reachability problem, i.e. given some target state, we are looking for a policy that maximizes the probability to reach this state. Such a problem can be solved with a linear programming approach and thus it is computable in polynomial time.

Assume now that we are required to give an *a priori* policy. In order to model it by an MDP, this entails that for all $s, s' \in S$ one has $A_s = A_{s'} \stackrel{\text{def}}{=} A$. W.r.t. the reachability problem, this is not a restriction since it consists in adding a new absorbing state whatever the action and to add transitions (with probability 1) to this state when an action was not originally allowed in some state. Another modification of the problem setting which is much more important, is that the considered policies are finite. Let us explain why in some context this restriction is meaningful. Suppose that you plan to go in a foreign country for your holidays, you must choose which train or plane you will use, you must rent an house or a room in an hotel, you must buy tickets for some exhibitions, etc. All these actions have a cost and your budget is finite whence the restriction to a finite sequence of actions over A which can also be seen as a finite word of A^*.

This leads to the definition of probabilistic automata first introduced in [RAB 63]. Observe that we adopt the usual terminology of automata theory.

Definition 5.1 *A* probabilistic automaton *(*PA*)* $\mathcal{A} = (Q, A, \{\mathbf{P}_a\}_{a \in A}, \pi_0, F)$ *is defined by:*

- Q*, the finite set of states;*
- A*, the finite alphabet;*
- *For all* $a \in A$, $\mathbf{P}_a$*, a probability transition matrix over* S*;*
- π_0*, the initial distribution over states and* $F \subseteq Q$ *the final states.*

As in Markov chains, when the initial distribution is a Dirac distribution concentrated on q_0, one says that q_0 is the initial state.

Example 5.2 (A simple PA) *Figure 5.1 depicts a* PA *with an initial state* q_0 *and a final state* q_1*. In order to get a compact view of the automaton, an edge from a state to another one is labelled by a vector of transition probabilities indexed by* A*. In order to make the state indices explicit, the vector is denoted by a formal sum. For instance, the transition from* q_0 *to itself is labelled by* $1a + 0.5b$ *meaning that when a (resp. b) is chosen in state* q_0*, the probability that the next state is* q_0*,* $\mathbf{P}_a[q_0, q_0]$ *(resp.* $\mathbf{P}_b[q_0, q_0]$*), is equal to* 1 *(resp.* 0.5*).*

When some finite word $w \stackrel{\text{def}}{=} a_1 \dots a_n$ is selected, we are interested in the probability to be in a final state using w as a policy. The following definition is straightforwardly justified by the analysis of MDP's.

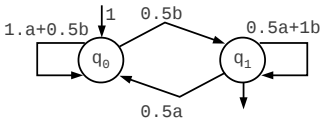

Figure 5.1: A simple PA

Definition 5.3 *Given $\mathcal{A}$ a* PA *and $w \stackrel{def}{=} a_1 \dots a_n \in A^*$ a word, the* acceptance probability *of w by $\mathcal{A}$ is defined by:*

$$\mathbf{Pr}_{\mathcal{A}}(w) \stackrel{def}{=} \sum_{q \in Q} \pi_{\mathbf{0}}[q] \sum_{q' \in F} \left(\prod_{i=1}^{n} \mathbf{P}_{a_i} \right) [q, q']$$

Notation. Given a word $w \stackrel{\text{def}}{=} a_1 \dots a_n$, the probability matrix $\mathbf{P}_w$ is defined by $\mathbf{P}_w \stackrel{\text{def}}{=} \prod_{i=1}^{n} \mathbf{P}_{a_i}$. In particular $\mathbf{P}_\varepsilon = \mathbf{Id}$. Thus $\mathbf{Pr}_{\mathcal{A}}(w) = \pi_0 \mathbf{P}_w \mathbf{1}_F^T$. where $\mathbf{1}_F$ is the indicator vector of subset F.

Let us compute in example 5.2, $\mathbf{Pr}_{\mathcal{A}}(abba)$. Since there are only two states, it is enough to keep the acceptance probability of the prefixes of *abba*. One starts with the empty word ε, $\mathbf{Pr}_{\mathcal{A}}(\varepsilon) = 0$. Then:

- $\mathbf{Pr}_{\mathcal{A}}(a) = \frac{1}{2}\mathbf{Pr}_{\mathcal{A}}(\varepsilon) = 0$,
- $\mathbf{Pr}_{\mathcal{A}}(ab) = \mathbf{Pr}_{\mathcal{A}}(a) + \frac{1}{2}(1 - \mathbf{Pr}_{\mathcal{A}}(a)) = \frac{1}{2}$
- $\mathbf{Pr}_{\mathcal{A}}(abb) = \mathbf{Pr}_{\mathcal{A}}(ab) + \frac{1}{2}(1 - \mathbf{Pr}_{\mathcal{A}}(ab)) = \frac{3}{4}$
- $\mathbf{Pr}_{\mathcal{A}}(abba) = \frac{1}{2}\mathbf{Pr}_{\mathcal{A}}(abb) = \frac{3}{8}$

More generally, the following recursive equations hold:

$$\mathbf{Pr}_{\mathcal{A}}(wa) = \frac{1}{2}\mathbf{Pr}_{\mathcal{A}}(w) \text{ and } \mathbf{Pr}_{\mathcal{A}}(wb) = \frac{1}{2}(1 + \mathbf{Pr}_{\mathcal{A}}(w))$$

from which one can derive an explicit expression of the acceptance probability:

$$\mathbf{Pr}_{\mathcal{A}}(a_1 \dots a_n) = \sum_{i=1}^{n} 2^{i-1-n} \cdot \mathbf{1}_{a_i = b}$$

Observe that $\sup(\mathbf{Pr}_{\mathcal{A}}(w) \mid w \in A^*) = 1$ and this value is not reached by any finite word.

We are interested in useful policies which directly leads to the introduction of *stochastic languages*.

Definition 5.4 *Let $\mathcal{A}$ be a probabilistic automaton, $\theta \in [0,1]$ a* threshold *also called a* cut point *and $\bowtie \in \{<, \leq, >, \geq, =, \neq\}$ be a comparison operator. Then $L_{\bowtie\theta}(\mathcal{A})$ is defined by:*

$$L_{\bowtie\theta}(\mathcal{A}) = \{w \in A^* \mid \mathbf{Pr}_{\mathcal{A}}(w) \bowtie \theta\}$$

The stochastic languages of example 5.2 have a simple interpretation. Let us introduce $v_a \stackrel{\text{def}}{=} 0$ and $v_b \stackrel{\text{def}}{=} 1$. The acceptance probability of a word $w_1 \dots w_n$ is now the binary number $0.v_{w_n} \dots v_{w_1}$. So for instance $\mathcal{L}_{\geq 0.5}(\mathcal{A})$ is the set of representations of binary numbers greater or equal than 0.5. Of course the representation of a number is not unique due to trailing zeros.

Example 5.5 (Counting with PA) *Figure 5.2 depicts a* PA *over alphabet $\{a, b\}$. This is a succinct representation where we have omitted an absorbing rejecting state and all transitions towards*

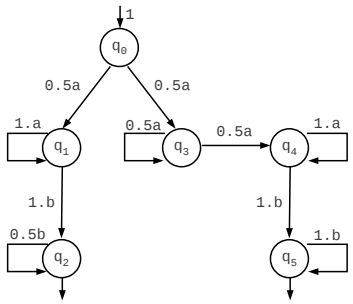

Figure 5.2: $\mathcal{L}_{=0.5}(\mathcal{A}) = \{a^n b^n \mid n > 0\}$

this state. For instance, in q_0 reading b one goes to this state with probability 1 and in q_2 reading b one goes to this state with probability 0.5.

Any word z different from $a^m b^n$ with $m > 0, n > 0$ cannot be accepted and so $\mathbf{Pr}_{\mathcal{A}}(z) = 0$. Let $w \stackrel{def}{=} a^m b^n$ with $m > 0, n > 0$. w can be accepted by a path q_0, q_1^m, q_2^n or by a family of paths q_0, q_3^r, q_4^s, q_5^n with $0 < r, s$ and $r + s = m$. The probability of the former path is $\frac{1}{2^n}$ while the cumulated probability of the latter paths is $\frac{1}{2} - \frac{1}{2^m}$, summing one obtains $\frac{1}{2} + \frac{1}{2^n} - \frac{1}{2^m}$. Thus $\mathcal{L}_{=0.5}(\mathcal{A}) = \{a^n b^n \mid n > 0\}$.

One wants to use a PA as input for algorithms and without restrictions on its representation this cannot be done. So we introduce a subclass of PA.

Definition 5.6 *A* rational PA *is a* PA *with probability distributions over $\mathbb{Q}^Q$. A* rational stochastic language *is a stochastic language specified by a rational* PA *and a rational threshold.*

Example 5.2 shows that $\{a^n b^n \mid n > 0\}$ is a rational stochastic language.

5.2 Properties of stochastic languages

5.2.1 Expressiveness

The first issue of expressiveness is related to the way we define stochastic languages: can we limit the threshold and the comparison operators while preserving the same expressive power? We first show that we can always use threshold $\frac{1}{2}$.

Proposition 5.7 *Let $\mathcal{A}$ be a probabilistic automaton, $\theta \in [0,1]$ a threshold and $\bowtie \in \{<, \leq, >, \geq, =, \neq\}$ be a comparison operator. Then there exists a probabilistic automaton $\mathcal{A}'$ such that:*

$$L_{\bowtie \frac{1}{2}}(\mathcal{A}') = L_{\bowtie \theta}(\mathcal{A})$$

Furthermore if $\mathcal{A}$ is a rational probabilistic automaton and θ is rational then $\mathcal{A}'$ is a rational probabilistic automaton.

Proof

We now show that equality and disequality operators can be omitted.

Proposition 5.8 *Let $\mathcal{A}$ be a probabilistic automaton. Then there exists a probabilistic automaton $\mathcal{A}'$ such that:*

$$L_{=\frac{1}{4}}(\mathcal{A}') = L_{\geq \frac{1}{4}}(\mathcal{A}') = L_{=\frac{1}{2}}(\mathcal{A}) \text{ and so } L_{<\frac{1}{4}}(\mathcal{A}') = L_{\neq \frac{1}{2}}(\mathcal{A})$$

Furthermore if $\mathcal{A}$ is a rational probabilistic automaton then $\mathcal{A}'$ is a rational probabilistic automaton.

Proof

This proposition has an important corollary.

Corollary 5.9 *The family of languages $\{L_{=\theta}(\mathcal{A})\}_{\mathcal{A},\theta}$ is closed under intersection. This closure property also holds for rational stochastic languages.*

Proof

Finally by complementing the final states, the comparison operator $\leq$ (resp. $<$) can be simulated by operator $>$ (resp. $\geq$).

Proposition 5.10 *Let $\mathcal{A}$ be a probabilistic automaton and $\mathcal{A}'$ be like $\mathcal{A}$ except that $F' \stackrel{def}{=} Q \setminus F$. Then:*

$$L_{\geq\theta}(\mathcal{A}') = L_{<\theta}(\mathcal{A}) \text{ and } L_{>\theta}(\mathcal{A}') = L_{\leq\theta}(\mathcal{A})$$

The second issue is related to the situation of the class of stochastic languages w.r.t. the standard classes of languages. Consider the languages $L_{>\theta}(\mathcal{A})$ where $\mathcal{A}$ is the PA presented in example 5.2. Given $\theta < \theta'$ there exists a binary number b such that $\theta < b < \theta'$ so that $L_{>\theta'}(\mathcal{A}) \subsetneq L_{>\theta}(\mathcal{A})$. Thus the family of languages $\{L_{>\theta}(\mathcal{A})\}_{0\leq\theta\leq 1}$ is uncountable. Since the family of recursively enumerable languages is countable, one gets the following proposition.

Proposition 5.11 *There exists a non recursively enumerable stochastic language.*

This proposition is somewhat unsatisfactory since it uses the uncountability of the reals and as we will see soon does not hold for rational stochastic langugages.

A deterministic complete finite automaton is particular case of probabilistic automaton where the distribution associated with s, a have always an atom of probability 1: the destination state $\delta(s, a)$. So the class of regular languages is included in the class of rational stochastic languages. There is a lot of ways to obtain strict inclusion: for instance the language of example 5.5 is not regular. In order to show how one can use irrational stochastic languages, we carry on studying example 5.2.

Lemma 5.12 *Let $\mathcal{A}$ be the* PA *of example 5.2. Then $L_{>\theta}(\mathcal{A})$ is regular iff θ is rational.*

Proof

While the automaton of figure 5.1 uses only values in $\{0, \frac{1}{2}, 1\}$, the threshold must be irrational to obtain non regularity. Nevertheless this result is used in the proof of next proposition.

Proposition 5.13 *The class of regular languages is strictly included in the class of rational stochastic languages.*

Proof

The two following propositions show that context-free languages and (rational) stochastic languages are incomparable.

Proposition 5.14 *There exists a context-free language that is not a stochastic language. More precisely $L \stackrel{def}{=} \{a^{n_1}ba^{n_2}b\ldots a^{n_k}ba^* \mid k \geq 2 \wedge \exists i > 1\ n_i = n_1\}$ is not stochastic.*

Proof

Proposition 5.15 *There exists a rational stochastic language that is not context-free. More precisely $L \stackrel{def}{=} \{a^n b^n c^n \mid n > 0\}$ is a rational stochastic language.*

Proof

Observe that the membership problem is decidable for rational stochastic languages. Elaborating on it, one gets an interesting inclusion for this class of languages.

Proposition 5.16 *The class of rational stochastic languages is strictly included in the class of context-sensitive languages.*

Proof

We end this section with an interesting result showing that allowing "weights" instead of probabilities does not extend the expressiveness of such automata. First we introduce generalized PA and their corresponding languages.

Definition 5.17 *A* generalized PA $\mathcal{A} = (Q, A, \{\mathbf{P}_a\}_{a\in A}, \pi_0, \pi_f)$ *is defined by:*

- Q*, the finite set of states;*
- A*, the finite alphabet;*
- *For all* $a \in A$*,* $\mathbf{P}_a$*, a real matrix over* $Q \times Q$*;*
- π_0*, the initial real vector over* Q *and* π_f*, the final real vector over* Q*.*

Definition 5.18 *Given* $\mathcal{A}$ *a generalized* PA *and* $w \in A^*$ *a word, the* acceptance weight *of* w *by* $\mathcal{A}$ *is defined by:*

$$\mathbf{Pr}_{\mathcal{A}}(w) \stackrel{def}{=} \pi_0 \mathbf{P}_w \pi_f^T$$

where as before $\mathbf{P}_w \stackrel{def}{=} \prod_{i=1}^{n} \mathbf{P}_{a_i}$ *for* $w \stackrel{def}{=} a_1 \dots a_n$*.*

We define the family of *generalized stochastic languages.*

Definition 5.19 *Let* $\mathcal{A}$ *be a generalized* PA*,* $\theta \in \mathbb{R}$ *a threshold and* $\bowtie \in \{<, \leq, >, \geq, =, \neq\}$ *be a comparison operator. Then* $L_{\rhd\bowtie\theta}(\mathcal{A})$ *is defined by:*

$$L_{\rhd\bowtie\theta}(\mathcal{A}) = \{w \in A^* \mid \mathbf{Pr}_{\mathcal{A}}(w) \bowtie \theta\}$$

In order to keep the proof readable, we proceed by a sequence of lemmas.

Lemma 5.20 *Let* $\mathcal{A}$ *be a generalized* PA*. Then there exists* $\mathcal{A}'$ *a generalized* PA *such that for all* $a \in A$*, the sum of items of any row or column of* $\mathbf{P}'_a$ *is null and for all* θ *and all* $\bowtie$*:*

$$L_{\rhd\bowtie\theta}(\mathcal{A}') = L_{\rhd\bowtie\theta}(\mathcal{A})$$

Proof

Lemma 5.21 *Let* $\mathcal{A}$ *be a generalized* PA *such that for all* $a \in A$*, the sum of items of any row or column of* $\mathbf{P}_a$ *is null. Then there exists* $\mathcal{A}'$ *a generalized* PA *such that for all* $a \in A$*, any items of* $\mathbf{P}'_a$ *is non negative and for all* θ *and all* $\bowtie$*:*

$$L_{\rhd\bowtie\theta}(\mathcal{A}') = L_{\rhd\bowtie\theta}(\mathcal{A})$$

Proof

Lemma 5.22 *Let* $\mathcal{A}$ *be a generalized* PA *such that for all* $a \in A$*, any item of* $\mathbf{P}_a$ *is non negative. Then there exists* $\mathcal{A}'$ *a generalized* PA *such that for all* $a \in A$*,* $\mathbf{P}'_a$ *is a probability transition matrix and for all* θ *and all* $\bowtie$*:*

$$L_{\rhd\bowtie 0}(\mathcal{A}') = L_{\rhd\bowtie\theta}(\mathcal{A})$$

Proof

Lemma 5.23 *Let* $\mathcal{A}$ *be a generalized* PA *such that for all* $a \in A$*,* $\mathbf{P}_a$ *is a probability transition matrix. Then there exists* $\mathcal{A}'$ *a generalized* PA *such that* π'_0 *is a distribution,* $\pi'_f \cdot \mathbf{1} = 0$*, for all* $a \in A$*,* $\mathbf{P}'_a$ *is a probability transition matrix and for all* $\bowtie$*:*

$$L_{\rhd\bowtie 0}(\mathcal{A}') = L_{\rhd\bowtie 0}(\mathcal{A})$$

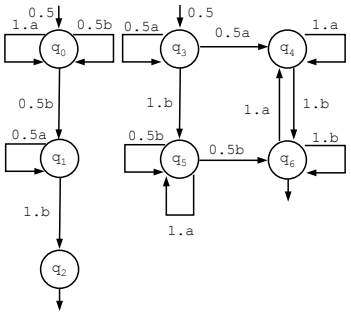

Figure 5.3: A PA for $\{a^{m_1}b \dots ba^{m_k}b \mid 1 < k \wedge m_1 = m_k\}$

Proof

Lemma 5.24 *Let $\mathcal{A}$ be a generalized* PA *such that π_0 is a distribution, $\pi_f \cdot \mathbf{1} = 0$ and for all $a \in A$, $\mathbf{P}_a$ is a probability transition matrix. Then there exists $\theta' > 0$ and $\mathcal{A}'$ a generalized* PA *such that π'_0 is a distribution, π'_f is positive, $\pi'_f \cdot \mathbf{1} = |Q'|\theta'$, for all $a \in A$, $\mathbf{P}'_a$ is a probability transition matrix, and for all $\bowtie$:*

$$L_{\rhd\lhd\theta'}(\mathcal{A}') = L_{\rhd\lhd 0}(\mathcal{A})$$

Proof

Lemma 5.25 *Let $\theta > 0$ and $\mathcal{A}$ be a generalized* PA *such that π_0 is a distribution, π_f is positive, $\pi_f \cdot \mathbf{1} = |Q|\theta$ and for all $a \in A$, $\mathbf{P}_a$ is a probability transition matrix. Then there exists $\mathcal{A}'$ a* PA *such that for all $\bowtie$:*

$$L_{\bowtie \frac{1}{|Q|}}(\mathcal{A}') = L_{\bowtie\theta}(\mathcal{A})$$

Proof

Combining the previous lemmas we obtain the theorem we are looking for. In addition, observe that: (1) starting from a rational generalized stochastic language one obtains a rational stochastic language and (2) that all transformations are performed in polynomial time.

Theorem 5.26 *The families of (rational) generalized stochastic languages and (rational) stochastic languages are identical.*

5.2.2 Closure

The next proposition shows that as for most of the standard families of languages, stochastic languages are closed by intersection and union with regular languages. What is interesting here is that the probabilistic automaton used in the proof has a size linear w.r.t. the size of the two input automata (contrary to standard synchronized product used in other constructions).

Proposition 5.27 *The family of (rational) stochastic languages is closed under intersection and union with regular languages.*

Proof

In order to prove non closure results, we exhibit a particular stochastic language.

Lemma 5.28 *Let $\mathcal{A}$ be the automaton of figure 5.3. Then:*

$$L_{=\frac{1}{2}}(\mathcal{A}) = \{a^{m_1}b \dots ba^{m_k}b \mid 1 < k \wedge m_1 = m_k\}$$

Proof

Proposition 5.29 *The family of (rational) stochastic languages is not closed under concatenation with a regular language.*

Proof

Proposition 5.30 *The family of (rational) stochastic languages is not closed under Kleene star.*

Proof

Proposition 5.31 *The family of (rational) stochastic languages is not closed under homomorphism.*

Proof

We end this section with a theorem whose involved proof is beyond the scope of these notes. Observe that the result does not hold for rational stochastic languages for which closure under intersection and union remains an open issue.

Theorem 5.32 ([FLI 74]) *The family of stochastic languages is not closed under intersection and union even for a one-letter alphabet.*

5.3 Decidability results

In this section we illustrate the tight frontier between decidability and undecidability studying two close problems: the equivalence of probabilistic automata and the equality of stochastic languages.

Definition 5.33 *Let $\mathcal{A}$ and $\mathcal{A}'$ be two probabilistic automata over the same alphabet A. $\mathcal{A}$ and $\mathcal{A}'$ are* equivalent *if for all word $w \in A^*$:*

$$\mathbf{Pr}_{\mathcal{A}}(w) = \mathbf{Pr}_{\mathcal{A}'}(w)$$

W.l.o.g. we assume that $F \cup F' \neq \emptyset$ (why?). Let us describe algorithm 6. It tries to establish non equivalence by finding a counter-example whose length is increasing starting with word ε. If it does not succeed then it manages a stack of words w from which it tries to find counter-examples aw. In order to avoid redundant computations, it also keeps in the stack the pair of vectors $(\mathbf{P}_w\mathbf{1}_F, \mathbf{P}'_w\mathbf{1}_{F'})$.

Without "pruning", the algorithm would be a semi-algorithm that only terminates when $\mathcal{A}$ and $\mathcal{A}'$ are not equivalent. So it manages *Gen* a set of independent orthogonal vectors of $\mathbb{R}^{Q\cup Q'}$. When a word w is not a counter-example, the algorithm checks that the vector $(\mathbf{P}_w\mathbf{1}_F, \mathbf{P}'_w\mathbf{1}_{F'})$ is not in the vector space generated by *Gen*. It performs this test by producing the orthogonal projection of the vector on this subspace and then comparing it to the original vector. If the vector is independent, then the word w is added to the stack, the difference between the vector and its orthogonal projection is added to *Gen* thus preserving the property of orthogonality.

By construction, when the algorithm finds a counter-example it has established the non equivalence. More subtle is the proof of equivalence when the algorithm has not found a counter-example.

Proposition 5.34 *Algorithm 6 operates in $O(|A|n^3)$ where $n = |Q| + |Q'|$ and decides whether $\mathcal{A}$ and $\mathcal{A}'$ are equivalent.*

Proof

Consider the problem of equality of languages $L_{\rhd\!\lhd\theta}(\mathcal{A})$ and $L_{\rhd\!\lhd'\theta'}(\mathcal{A}')$. Of course if $\mathcal{A}$ and $\mathcal{A}'$ are equivalent and $\rhd\!\lhd\ \theta$ equals $\rhd\!\lhd'\ \theta'$, the languages are equal. Unfortunately this is only a sufficient condition and the *a priori* simpler problem of language emptiness is already undecidable.

Let us recall the Post correspondence problem (PCP). Given an alphabet A and two morphisms φ_1, φ_2 from A to $\{0,1\}^+$ does there exist a word $w \in A^+$ such that $\varphi_1(w) = \varphi_2(w)$? This problem

Algorithm 6: Checking equivalence of two probabilistic automata

`Equivalence`$(\mathcal{A}, \mathcal{A}')$
Input: $\mathcal{A}, \mathcal{A}'$, two PA over alphabet A such that $F \cup F' \neq \emptyset$

Output: the status of equivalence with a witness for non equivalence

Data: $\mathbf{v}, \mathbf{x}, \mathbf{y}, \mathbf{z}$ vectors of $\mathbb{R}^Q$ and $\mathbf{v}', \mathbf{x}', \mathbf{y}', \mathbf{z}'$ vectors of $\mathbb{R}^{Q'}$
Data: *Stack* whose items are pairs of a vector of $\mathbb{R}^{Q \cup Q'}$ and a word
Data: *Gen*, a set of (non null) orthogonal vectors of $\mathbb{R}^{Q \cup Q'}$, a a letter

if $\pi_0 \cdot \mathbf{1}_F \neq \pi'_0 \cdot \mathbf{1}_{F'}$ **then return**(**false**, ε)
$Gen \leftarrow \{(\mathbf{1}_F, \mathbf{1}_{F'})\}$; `Push`$(Stack, ((\mathbf{1}_F, \mathbf{1}_{F'}), \varepsilon))$
repeat
  $((\mathbf{v}, \mathbf{v}'), w) \leftarrow$ `Pop`$(Stack)$
  for $a \in A$ **do**
    $\mathbf{z} \leftarrow \mathbf{P}_a \mathbf{v}$; $\mathbf{z}' \leftarrow \mathbf{P}'_a \mathbf{v}$
    if $\pi_0 \cdot \mathbf{z} \neq \pi'_0 \cdot \mathbf{z}'$ **then return**(**false**, aw)
    $\mathbf{y} \leftarrow \mathbf{0}$; $\mathbf{y}' \leftarrow \mathbf{0}$
    for $(\mathbf{x}, \mathbf{x}') \in Gen$ **do**
      $\mathbf{y} \leftarrow \mathbf{y} + \frac{\mathbf{z} \cdot \mathbf{x}}{\mathbf{x} \cdot \mathbf{x}} \mathbf{x}$
      $\mathbf{y}' \leftarrow \mathbf{y}' + \frac{\mathbf{z}' \cdot \mathbf{x}'}{\mathbf{x}' \cdot \mathbf{x}'} \mathbf{x}'$
    end
    if $(\mathbf{z}, \mathbf{z}') \neq (\mathbf{y}, \mathbf{y}')$ **then**
      `Push`$(Stack, ((\mathbf{z}, \mathbf{z}'), aw))$
      $Gen \leftarrow Gen \cup \{(\mathbf{z} - \mathbf{y}, \mathbf{z}' - \mathbf{y}')\}$
    end
  end
until `IsEmpty`$(Stack)$
return(**true**)

is undecidable. We consider here an a priori restricted variation where the images of letters lie in $(10+11)^+$. In fact this is not a restriction since inserting a 1 before each letter of the image reduces the former problem to the latter. A word $w \stackrel{\text{def}}{=} a_1 \dots a_n \in (10+11)^+$ defines a value $val(w) \in [0,1]$ by: $val(w) \stackrel{\text{def}}{=} \sum_{i=1}^{n} \frac{a_i}{2^{n-i}}$. Furthermore since every word starts with a 1, $val(w) = val(w')$ implies $w = w'$. We now elaborate on example 5.2 in order to establish the proof of the next proposition.

Proposition 5.35 *Given $\mathcal{A}$ a rational stochastic automaton, the question $L_{=\frac{1}{2}}(\mathcal{A}) = \emptyset$? is undecidable.*

Proof

Using proposition 5.8 we immediately obtain the following corollary.

Corollary 5.36 *Given $\mathcal{A}$ a rational stochastic automaton and θ a rational number, the question $L_{\geq\theta}(\mathcal{A}) = \emptyset$? is undecidable.*

In fact with small work, we also obtain the next corollary.

Corollary 5.37 *Given $\mathcal{A}$ a rational stochastic automaton and θ a rational number, the question $L_{>\theta}(\mathcal{A}) = \emptyset$? is undecidable.*

Proof

5.4 Proofs

5.4.1 Proofs of section 5.2

Proof of proposition 5.7
Given $\mathcal{A}$ and $\theta \neq \frac{1}{2}$, one builds $\mathcal{A}'$ as described in figure 5.4. One adds a state q_0. The initial distribution is modified as follows: $\pi'_0[q_0] \stackrel{\text{def}}{=} 1 - \alpha$ and for all $q \in Q$, $\pi'_0[q] \stackrel{\text{def}}{=} \alpha\pi_0[q]$ where $0 < \alpha < 1$. State q_0 is an absorbing state whatever the letter chosen.
The value α depends on θ in the following way:

- If $\theta > \frac{1}{2}$ then $q_0 \notin F$ and $\alpha \stackrel{\text{def}}{=} \frac{1}{2\theta}$ so that for all $w \in A^*$, $\mathbf{Pr}_{\mathcal{A}'}(w) = \frac{1}{2\theta}\mathbf{Pr}_{\mathcal{A}'}(w)$.
  Thus $w \in L_{\bowtie\frac{1}{2}}(\mathcal{A}')$ iff $w \in L_{\bowtie\theta}(\mathcal{A})$.

- If $\theta < \frac{1}{2}$ then $q_0 \in F$ and $\alpha \stackrel{\text{def}}{=} \frac{1}{2(1-\theta)}$ so that for all $w \in A^+$, $\mathbf{Pr}_{\mathcal{A}'}(w) = \frac{1-2\theta+\mathbf{Pr}_{\mathcal{A}'}(w)}{2(1-\theta)}$.
  Thus $w \in L_{\bowtie\frac{1}{2}}(\mathcal{A}')$ iff $w \in L_{\bowtie\theta}(\mathcal{A})$.

q.e.d. (proposition 5.7) ◊◊◊

Proof of proposition 5.8
We build $\mathcal{A}'$ as follows.

- The set of states $Q' \stackrel{\text{def}}{=} Q \times Q$;

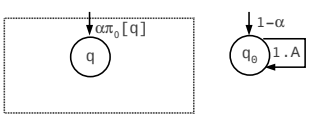

Figure 5.4: A PA for threshold $\frac{1}{2}$

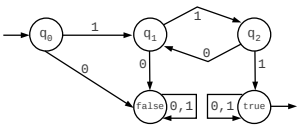

Figure 5.5: A finite automaton for threshold $\frac{5}{6}$

- $\mathbf{P}'_a[(q_1, q_2), (q'_1, q'_2)] \stackrel{\text{def}}{=} \mathbf{P}_a[q_1, q'_1]\mathbf{P}_a[q_2, q'_2]$;
- $\pi'_0[q_1, q_2] \stackrel{\text{def}}{=} \pi_0[q_1]\pi_0[q_2]$ and $F' \stackrel{\text{def}}{=} F \times (Q \setminus F)$.

Once a word w is selected, the two components of the DES behave independently and so:
$\mathbf{Pr}_{\mathcal{A}'}(w) = \mathbf{Pr}_{\mathcal{A}}(w)(1 - \mathbf{Pr}_{\mathcal{A}}(w))$
Consequently $\mathbf{Pr}_{\mathcal{A}'}(w) \leq \frac{1}{4}$ with equality iff $\mathbf{Pr}_{\mathcal{A}}(w) = \frac{1}{2}$.

q.e.d. (proposition 5.8) ◊◊◊

Proof of corollary 5.9
Due to proposition 5.7, we pick two arbitrary languages $L_{=\frac{1}{2}}(\mathcal{A}_1)$ and $L_{=\frac{1}{2}}(\mathcal{A}_2)$. Let $\mathcal{A}'_1$ and $\mathcal{A}'_2$ be the automata corresponding to proposition 5.8. One builds $\mathcal{A}$ as follows:

- The set of states $Q \stackrel{\text{def}}{=} Q'_1 \times Q'_2$;
- $\mathbf{P}_a[(q_1, q_2), (q'_1, q'_2)] \stackrel{\text{def}}{=} (\mathbf{P}'_1)_a[q_1, q'_1](\mathbf{P}'_2)_a[q_2, q'_2]$;
- $\pi'_0[q_1, q_2] \stackrel{\text{def}}{=} \pi_{1,0}[q_1]\pi_{2,0}[q_2]$ and $F \stackrel{\text{def}}{=} F'_1 \times F'_2$.

By construction, $\mathbf{Pr}_{\mathcal{A}}(w) = \mathbf{Pr}_{\mathcal{A}'_1}(w)\mathbf{Pr}_{\mathcal{A}'_2}(w)$. So for all word w, $\mathbf{Pr}_{\mathcal{A}}(w) \leq \frac{1}{16}$ and $\mathbf{Pr}_{\mathcal{A}}(w) = \frac{1}{16}$ iff $\mathbf{Pr}_{\mathcal{A}'_1}(w) = \mathbf{Pr}_{\mathcal{A}'_2}(w) = \frac{1}{4}$. Consequently, $L_{=\frac{1}{16}}(\mathcal{A}) = L_{=\frac{1}{2}}(\mathcal{A}_1) \cap L_{=\frac{1}{2}}(\mathcal{A}_2)$.

q.e.d. (corollary 5.9) ◊◊◊

Proof of lemma 5.12
Let $u_\theta \stackrel{\text{def}}{=} u_1 u_2 \ldots$ be the binary development of θ (when θ is binary we choose the development that ends with 0^ω). A finite word w belongs to $L_{>\theta}(\mathcal{A})$ if $\widetilde{w} > u_\theta$ for lexicographic order where $\widetilde{w}$ is the mirror of w. Since the mirror image of a language is regular iff the original language is regular, we study the language $L_\theta \stackrel{\text{def}}{=} \{w \mid w > u_\theta\}$.
Assume that θ is rational. Then its binary development is ultimately periodic, i.e. $u_\theta = u_1 \ldots u_k (u_{k+1} \ldots u_l)^\omega$. For instance, $u_{\frac{5}{6}} = 1(10)^\omega$. The deterministic automaton that accepts L_θ is defined as follows.

- The set of states $Q = \{q_0, \ldots, q_l, \mathit{true}, \mathit{false}\}$ with q_0 the initial state and *true* the single final state.
- The set of transitions is defined by four subsets.
  - for all $0 \leq i < l$, $\delta(q_i, u_{i+1}) = q_{i+1}$ and $\delta(q_l, u_{l+1}) = q_{k+1}$.
  - for all $0 \leq i \leq l$ such that $u_{i+1} = 0$, $\delta(q_i, 1) = \mathit{true}$.
  - for all $0 \leq i \leq l$ such that $u_{i+1} = 1$, $\delta(q_i, 0) = \mathit{false}$.
  - for all $q \in \{\mathit{true}, \mathit{false}\}$, $\delta(q, 0) = \delta(q, 1) = q$.

We have represented in figure 5.5 the automaton corresponding to $\theta = \frac{5}{6}$. We let the reader check the correctness of this automaton.

Assume that θ is irrational and that L_θ is regular. Let $\mathcal{A}$ be a deterministic automaton whose language is L_θ. Then there is a (single) infinite path $q_0 q_1 \dots$ in $\mathcal{A}$ corresponding to the infinite word u_θ. This path never meets a final state (why?) and there is at least one state q that is infinitely met by the path so that $u_\theta = w_0 w_1 \dots$ where the w_i's are words and forall $i > 0$, $\delta(q, w_i) = q$. Since u_θ is not ultimately periodic, there exist $0 < i < j$ with $w_i \neq w_j$. Assume that $w_i < w_j$, then $w \stackrel{\text{def}}{=} w_0 \dots w_{i-1} w_j$ should be accepted but $\delta(q_0, w) = q$. Assume that $w_i > w_j$, then $w' \stackrel{\text{def}}{=} w_0 \dots w_{j-1} w_i$ should be accepted but $\delta(q_0, w') = q$. So we conclude that L_θ is not regular.

q.e.d. (lemma 5.12) ◊◊◊

Proof of proposition 5.13
Let $L_{>\frac{1}{\sqrt{2}}}(\mathcal{A})$ with $\mathcal{A}$ the automaton of example 5.2. This language is non regular. Now build $\mathcal{A}'$ with a construction similar to proof of proposition 5.8.

- The set of states $Q' \stackrel{\text{def}}{=} Q \times Q$;
- $\mathbf{P}'_a[(q_1, q_2), (q'_1, q'_2)] \stackrel{\text{def}}{=} \mathbf{P}_a[q_1, q'_1]\mathbf{P}_a[q_2, q'_2]$;
- $\pi'_0[(q_0, q_0)] \stackrel{\text{def}}{=} 1$ and $F' \stackrel{\text{def}}{=} F \times F$.

Once a word w is selected, the two components of the DES behave independently and so:
$\mathbf{Pr}_{\mathcal{A}'}(w) = \mathbf{Pr}_{\mathcal{A}}(w)^2$. So $L_{>\frac{1}{2}}(\mathcal{A}') = L_{>\frac{1}{\sqrt{2}}}(\mathcal{A})$ is non regular.

q.e.d. (proposition 5.13) ◊◊◊

Proof of proposition 5.14
Let $L \stackrel{\text{def}}{=} \{a^{n_1} b a^{n_2} b \dots a^{n_k} b a^* \mid \exists i > 1 \; n_i = n_1\}$. L is context-free. Indeed with a counter (i.e. a stack over one letter), one counts n_1 the number of a's until the first occurrence of b. Then one guesses an occurrence of b and decrements the counter by the occurrences of a until the next occurrence of b. If the counter is zero the word is accepted.

Assume that (1) $L = L_{>\theta}(\mathcal{A})$ or (2) $L = L_{\geq\theta}(\mathcal{A})$ for some probabilistic automaton.
Let $\sum_{i=0}^n c_i x^i$ be the minimal polynomial of $\mathbf{P}_a$.
Since 1 is an eigenvalue of $\mathbf{P}_a$, one gets $\sum_{i=0}^n c_i = 0$ and there are positive and negative coefficients.
By definition, $\sum_{i=0}^n c_i \mathbf{P}_{a^i} = 0$ and so for any word w, $\sum_{i=0}^n c_i \mathbf{P}_{a^i w} = 0$.
Let $c_{i_1}, \dots, c_{i_k}$ be the positive coefficients of this polynomial.
Choose $w \stackrel{\text{def}}{=} b a^{i_1} b \dots b a^{i_k} b$.

Case $L = L_{>\theta}(\mathcal{A})$**.** Let $0 \leq i \leq n$, by definition of L, $\pi_0 \mathbf{P}_{a^i w} \mathbf{1}_F^T > \theta$ iff $i \in \{i_1, \dots, i_k\}$.
So: $0 = \sum_{i=0}^n c_i \pi_0 \mathbf{P}_{a^i w} \mathbf{1}_F^T > (\sum_{i=0}^n c_i)\theta = 0$
leading to a contradiction.

Case $L = L_{\geq\theta}(\mathcal{A})$**.** Let $0 \leq i \leq n$, by definition of L, $\pi_0 \mathbf{P}_{a^i w} \mathbf{1}_F^T \geq \theta$ iff $i \in \{i_1, \dots, i_k\}$.
So: $0 = \sum_{i=0}^n c_i \pi_0 \mathbf{P}_{a^i w} \mathbf{1}_F^T > (\sum_{i=0}^n c_i)\theta = 0$
leading to a contradiction.

q.e.d. (proposition 5.14) ◊◊◊

Proof of proposition 5.15

Using Ogden's lemma it can be easily proved that $L \stackrel{\text{def}}{=} \{a^n b^n c^n \mid n > 0\}$ is not context-free (see for instance [HMU 06]).

We observe that $L = L_1 \cap L_2$ with $L_1 \stackrel{\text{def}}{=} \{a^n b^n c^+ \mid n > 0\}$ and $L_2 \stackrel{\text{def}}{=} \{a^+ b^n c^n \mid n > 0\}$. Using corollary 5.9, it is sufficient to prove that $L_i = L_{=\frac{1}{2}}(\mathcal{A}_i)$ for some rational probabilistic automaton

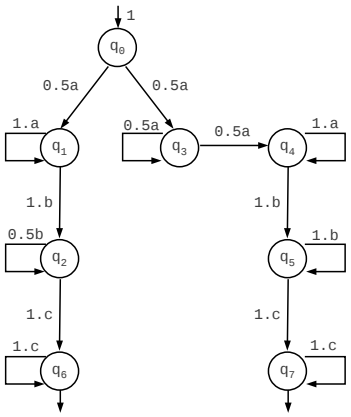

Figure 5.6: $\mathcal{L}_{=0.5}(\mathcal{A}) = \{a^n b^n c^+ \mid n > 0\}$

$\mathcal{A}_i$. Both automata are straightforward variations of example 5.5. We describe in figure 5.6 $\mathcal{A}_1$ and let the reader design $\mathcal{A}_2$.

q.e.d. (proposition 5.15) $\Diamond\Diamond\Diamond$

Proof of proposition 5.16
Context-sensitive languages are exactly the languages for which membership checking can be performed by a non deterministic procedure in linear space (see for instance [HMU 06]).

In fact we show that we can perform membership checking by a deterministic procedure in linear space (which is far from being optimal). First one computes the least common multiple, say b, of denominators of (1) θ, (2) items of matrices $\{\mathbf{P}_a\}_{a\in A}$ and, (3) items of vector π_0. This is done in constant space (i.e. independent of n, the size of the word $w \stackrel{\text{def}}{=} a_1 \dots a_n$ to be checked). Then one builds the integer matrices $\mathbf{P}'_a \stackrel{\text{def}}{=} b\mathbf{P}_a$ and integer vector $\pi'_0 \stackrel{\text{def}}{=} b\pi_0$ again in constant space. The membership problem becomes $\pi'_0(\prod_{i=1}^n \mathbf{P}'_{a_i})\mathbf{1}_F^T \bowtie \theta b^{n+1}$? Observe that the space needed to compute θb^{n+1} is $O(n)$. One also computes $\mathbf{v} \stackrel{\text{def}}{=} \pi'_0(\prod_{i=1}^n \mathbf{P}'_{a_i})$ by initializing $\mathbf{v}$ to π'_0 and then iteratively multiply it by $\mathbf{P}'_{a_i}$. Observe that at the ith iteration the sum of the coefficients of $\mathbf{v}$ is exactly b^{i+1}. So again this can be performed in space $O(n)$. Finally the comparison only requires indices for bits to be compared again in space $O(n)$.

q.e.d. (proposition 5.16) $\Diamond\Diamond\Diamond$

Proof of lemma 5.20
Let $Q' \stackrel{\text{def}}{=} Q \uplus \{q_0, q_1\}$. Then:

- for all $q \in Q$, $\pi'_0[q] \stackrel{\text{def}}{=} \pi_0[q]$ and $\pi'_0[q_0] \stackrel{\text{def}}{=} \pi'_0[q_1] \stackrel{\text{def}}{=} 0$;
- for all $a \in A$, for all $q \in Q'$, $\mathbf{P}'_a[q_0, q] \stackrel{\text{def}}{=} 0$ and $\mathbf{P}'_a[q, q_1] \stackrel{\text{def}}{=} 0$;
- for all $a \in A$, for all $q, q' \in Q$, $\mathbf{P}'_a[q, q'] \stackrel{\text{def}}{=} \mathbf{P}_a[q, q']$ and $\mathbf{P}'_a[q, q_0] \stackrel{\text{def}}{=} -\sum_{q'\in Q} \mathbf{P}'_a[q, q']$;
- for all $a \in A$, for all $q \in Q'$, $\mathbf{P}'_a[q_1, q] \stackrel{\text{def}}{=} -\sum_{q'\in Q} \mathbf{P}'_a[q', q]$;
- for all $q \in Q$, $\pi'_f[q] \stackrel{\text{def}}{=} \pi_f[q]$ and $\pi'_f[q_0] \stackrel{\text{def}}{=} \pi'_f[q_1] \stackrel{\text{def}}{=} 0$.

Let $w \stackrel{\text{def}}{=} a_1 \dots a_n \in A^*$, then:
$\mathbf{Pr}_{\mathcal{A}'}(w) = \sum_{q \in Q'} \sum_{q' \in Q'} \pi'_0[q]\pi'_f[q'] \left(\prod_{i=1}^{n} \mathbf{P}'_{a_i}\right)[q, q'] = \sum_{q \in Q} \sum_{q' \in Q} \pi'_0[q]\pi'_f[q'] \left(\prod_{i=1}^{n} \mathbf{P}'_{a_i}\right)[q, q']$
Observe now that: for all $q \in Q$, and all $a \in A$, $\mathbf{P}_a[q, q_1] = \mathbf{P}_a[q_0, q] = 0$.
Thus $\mathbf{Pr}_{\mathcal{A}'}(w) = \mathbf{Pr}_{\mathcal{A}}(w)$.

By construction, the sum of items of any row and column of all $\mathbf{P}'_a$ is null.

q.e.d. (lemma 5.20) $\Diamond\Diamond\Diamond$

Proof of lemma 5.21
Let $\delta \stackrel{\text{def}}{=} \max(|\mathbf{P}_a[q, q']| \mid a \in A, q, q' \in Q)$.
Let $Q^+ \stackrel{\text{def}}{=} \{q^+ \mid q \in Q\}$, $Q^- \stackrel{\text{def}}{=} \{q^- \mid q \in Q\}$ and $Q' \stackrel{\text{def}}{=} Q^+ \uplus Q^- \uplus \{q_0, q_1\}$. Then:

- for all $q \in Q$, $\pi'_0[q^+] \stackrel{\text{def}}{=} \pi'_0[q^-] \stackrel{\text{def}}{=} \pi_0[q]$ and $\pi'_0[q_0] \stackrel{\text{def}}{=} \pi_0 \cdot \pi_f$ and $\pi'_0[q_1] \stackrel{\text{def}}{=} 0$;
- for all $a \in A$, for all $q \neq q_1$, $\mathbf{P}'_a[q_0, q] \stackrel{\text{def}}{=} \mathbf{P}'_a[q_1, q] \stackrel{\text{def}}{=} 0$ and $\mathbf{P}'_a[q_0, q_1] \stackrel{\text{def}}{=} \mathbf{P}'_a[q_1, q_1] \stackrel{\text{def}}{=} 1$;
- for all $a \in A$, for all $q, q' \in Q$, $\mathbf{P}'_a[q^+, {q'}^+] \stackrel{\text{def}}{=} \mathbf{P}_a[q, q'] + \delta$ and $\mathbf{P}'_a[q^-, {q'}^-] \stackrel{\text{def}}{=} \delta$;
- for all $a \in A$, for all $q \in Q$, $q' \notin Q^+$ and $q'' \notin Q^-$, $\mathbf{P}'_a[q^+, q'] \stackrel{\text{def}}{=} \mathbf{P}'_a[q^-, q''] \stackrel{\text{def}}{=} 0$;
- for all $q \in Q$, $\pi'_f[q^+] \stackrel{\text{def}}{=} \pi_f[q]$, $\pi'_f[q^-] \stackrel{\text{def}}{=} -\pi_f[q]$, $\pi'_f[q_0] \stackrel{\text{def}}{=} 1$ and $\pi'_f[q_1] \stackrel{\text{def}}{=} 0$.

Observe that for all a, $\mathbf{P}'_a$ is a block-diagonal matrix where the block corresponding to Q^+ is $\mathbf{P}_a + \delta\mathbf{1}$ ($\mathbf{1}$ is a matrix with all items equal to 1), the block corresponding to Q^- is $\delta\mathbf{1}$ and the block corresponding to $\{q_0, q_1\}$ is $\begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}$, an idempotent matrix.

Observe also that for all a, $\mathbf{P}_a\mathbf{1} = \mathbf{1}\mathbf{P}_a = \mathbf{0}$ due to the property of matrices $\mathbf{P}_a$.

Thus by induction on $n \geq 1$, given $w = a_1 \dots a_n$, $\mathbf{P}'_w$ is a block-diagonal matrix where the block corresponding to Q^+ is $\mathbf{P}_w + \delta^n\mathbf{1}$, the block corresponding to Q^- is $\delta^n\mathbf{1}$ and the block corresponding to $\{q_0, q_1\}$ is $\begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}$.

So $\mathbf{Pr}_{\mathcal{A}'}(w) = \pi_0(\mathbf{P}_w + \delta^n\mathbf{1})\pi_f - \pi_0(\delta^n\mathbf{1})\pi_f = \mathbf{Pr}_{\mathcal{A}}(w)$
Finally $\mathbf{Pr}_{\mathcal{A}'}(\varepsilon) = \pi_0\pi_f - \pi_0\pi_f + \pi_0\pi_f = \mathbf{Pr}_{\mathcal{A}}(\varepsilon)$

q.e.d. (lemma 5.21) $\Diamond\Diamond\Diamond$

Proof of lemma 5.22
Let $\delta \stackrel{\text{def}}{=} 1 + \max(\sum_{q' \in Q} \mathbf{P}_a[q, q'] \mid a \in A, q \in Q)$.
Let $Q' \stackrel{\text{def}}{=} Q \uplus \{q_0, q_1\}$. Then:

- for all $q \in Q$, $\pi'_0[q] \stackrel{\text{def}}{=} \pi_0[q]$, $\pi'_0[q_1] \stackrel{\text{def}}{=} \theta$, and $\pi'_0[q_0] \stackrel{\text{def}}{=} 0$;
- for all $a \in A$, for all $q \neq q_0$, $\mathbf{P}'_a[q_0, q] \stackrel{\text{def}}{=} 0$
  and $\mathbf{P}'_a[q_0, q_0] \stackrel{\text{def}}{=} 1$ (q_0 is an absorbing state);
- for all $a \in A$, for all $q, q' \in Q$, $\mathbf{P}'_a[q, q'] \stackrel{\text{def}}{=} \delta^{-1}\mathbf{P}_a[q, q']$, $\mathbf{P}'_a[q, q_0] \stackrel{\text{def}}{=} 1 - \sum_{q' \in Q} \mathbf{P}'_a[q, q']$ and $\mathbf{P}'_a[q, q_1] \stackrel{\text{def}}{=} 0$;
- for all $a \in A$, for all $q \notin \{q_1, q_0\}$, $\mathbf{P}'_a[q_1, q] \stackrel{\text{def}}{=} 0$, $\mathbf{P}'_a[q_1, q_1] \stackrel{\text{def}}{=} \delta^{-1}$ and $\mathbf{P}'_a[q_1, q_0] \stackrel{\text{def}}{=} 1 - \delta^{-1}$;
- for all $q \in Q$, $\pi'_f[q] \stackrel{\text{def}}{=} \pi_f[q]$, $\pi'_f[q_1] \stackrel{\text{def}}{=} -1$, and $\pi'_f[q_0] \stackrel{\text{def}}{=} 0$.

Let $w = a_1 \dots a_n \in A^*$. Observe that states contributing to $\mathbf{Pr}_{\mathcal{A}'}(w)$ are those of Q and q_1. By construction, for all $q, q' \in Q$,
$\mathbf{P}'_w[q, q'] = \delta^{-n}\mathbf{P}_w[q, q']$, $\mathbf{P}'_w[q_1, q_1] = \delta^{-n}$ and $\mathbf{P}'_w[q, q_1] = \mathbf{P}'_w[q_1, q] = 0$.
So $\mathbf{Pr}_{\mathcal{A}'}(w) = \delta^{-n}\mathbf{Pr}_{\mathcal{A}}(w) - \delta^{-n}\theta$.

q.e.d. (lemma 5.22) ◊◊◊

Proof of lemma 5.23
Let $\delta \stackrel{\text{def}}{=} 1 + \max(|\pi_0[q]| \mid q \in Q)$ and $\rho \stackrel{\text{def}}{=} 2|Q|\delta + \sum_{q \in Q} \pi_0[q]$.
Let $Q^+ \stackrel{\text{def}}{=} \{q^+ \mid q \in Q\}$, $Q^- \stackrel{\text{def}}{=} \{q^- \mid q \in Q\}$ and $Q' \stackrel{\text{def}}{=} Q^+ \uplus Q^-$. Then:

- for all $q \in Q$, $\pi'_0[q^+] \stackrel{\text{def}}{=} \rho^{-1}(\pi_0[q] + \delta)$ and $\pi'_0[q^-] \stackrel{\text{def}}{=} \rho^{-1}\delta$;
- for all $a \in A$, for all $q, q' \in Q$, $\mathbf{P}'_a[q^+, {q'}^+] \stackrel{\text{def}}{=} \mathbf{P}'_a[q^-, {q'}^-] \stackrel{\text{def}}{=} \mathbf{P}_a[q, q']$ and $\mathbf{P}'_a[q^-, {q'}^+] \stackrel{\text{def}}{=} \mathbf{P}'_a[q^+, {q'}^-] \stackrel{\text{def}}{=} 0$;
- for all $q \in Q$, $\pi'_f[q^+] \stackrel{\text{def}}{=} \pi_f[q]$, $\pi'_f[q^-] \stackrel{\text{def}}{=} -\pi_f[q]$.

Observe that for all a, $\mathbf{P}'_a$ is a block-diagonal matrix where the blocks corresponding to Q^+ and Q^- are $\mathbf{P}_a$. So:
$\mathbf{Pr}_{\mathcal{A}'}(w) = (\rho^{-1}\pi_0 + \rho^{-1}\delta\mathbf{1})\mathbf{P}_w\pi_f - \rho^{-1}\delta\mathbf{1}\mathbf{P}_w\pi_f = \rho^{-1}\mathbf{Pr}_{\mathcal{A}}(w)$

q.e.d. (lemma 5.23) ◊◊◊

Proof of lemma 5.24
Let $\theta' \stackrel{\text{def}}{=} 1 + \max(|\pi_f[q]| \mid q \in Q)$.
Let $Q' \stackrel{\text{def}}{=} Q$, $\pi'_0 \stackrel{\text{def}}{=} \pi_0$ and for all $a \in A$, $\mathbf{P}'_a = \mathbf{P}_a$.
For all $q \in Q$, $\pi'_f[q] \stackrel{\text{def}}{=} \pi_f[q] + \theta'$. Then:
$\mathbf{Pr}_{\mathcal{A}'}(w) = \pi_0\mathbf{P}_w(\pi_f + \theta'\mathbf{1})^T = \mathbf{Pr}_{\mathcal{A}'}(w) + \theta'$

q.e.d. (lemma 5.24) ◊◊◊

Proof of lemma 5.25
Let $Q' \stackrel{\text{def}}{=} Q^2$. Then:

- for all $(q_1, q_2) \in Q^2$, $\pi'_0[q_1, q_2] \stackrel{\text{def}}{=} \frac{\pi_0[q_2]}{|Q|}$;
- for all $a \in A$, for all $(q_1, q_2), (q'_1, q'_2) \in Q^2$, $\mathbf{P}'_a[(q_1, q_2), (q'_1, q'_2)] \stackrel{\text{def}}{=} \frac{\pi_f[q'_1]}{\pi_f \cdot \mathbf{1}}\mathbf{P}_a[q_2, q'_2]$;
- $F \stackrel{\text{def}}{=} \{(q, q) \mid q \in Q\}$.

By induction on $n \geq 1$, given $w = a_1 \dots a_n$, $\mathbf{P}'_w[(q_1, q_2), (q'_1, q'_2)] = \frac{\pi_f[q'_1]}{\pi_f \cdot \mathbf{1}}\mathbf{P}_w[q_2, q'_2]$.
So: $(\pi'_0\mathbf{P}'_w)[(q'_1, q'_2)] = \sum_{(q_1,q_2) \in Q^2} \frac{\pi_0[q_2]}{|Q|}\frac{\pi_f[q'_1]}{\pi_f \cdot \mathbf{1}}\mathbf{P}_w[q_2, q'_2] = \sum_{q_2 \in Q} \frac{\pi_0[q_2]\pi_f[q'_1]}{\pi_f \cdot \mathbf{1}}\mathbf{P}_w[q_2, q'_2]$.
Thus: $\mathbf{Pr}_{\mathcal{A}'}(w) = \sum_{q_2 \in Q}\sum_{q'_2 \in Q} \frac{\pi_0[q_2]\pi_f[q'_2]}{\pi_f \cdot \mathbf{1}}\mathbf{P}_w[q_2, q'_2] = \frac{\mathbf{Pr}_{\mathcal{A}}(w)}{\pi_f \cdot \mathbf{1}}$.
Let us recall that $\frac{\theta}{\pi_f \cdot \mathbf{1}} = \frac{1}{|Q|}$. Thus the two languages are identical except possibly w.r.t. the empty word.

However given a stochastic language $L_{\bowtie\theta}(\mathcal{A})$, one can easily add or remove the empty word by the following construction:

- $Q_0 \stackrel{\text{def}}{=} \{(q, 0) \mid q \in Q\}$ and $Q' \stackrel{\text{def}}{=} Q \uplus Q_0$;
- $\pi'_0[(q, 0)] \stackrel{\text{def}}{=} \pi_0[q]$ and $\pi'_0[q] \stackrel{\text{def}}{=} 0$;

- for all $a \in A$, for all $q, q' \in Q$, $\mathbf{P}'_a[(q,0),q'] \stackrel{\text{def}}{=} \mathbf{P}'_a[q,q'] \stackrel{\text{def}}{=} \mathbf{P}_a[q,q']$ and $\mathbf{P}'_a[(q,0),(q',0)] \stackrel{\text{def}}{=} \mathbf{P}'_a[q,(q',0)] \stackrel{\text{def}}{=} 0$;
- $F' \stackrel{\text{def}}{=} F \cup G$ where $G \subseteq Q_0$ is appropriately chosen (depending on $\bowtie \theta$) in order to add or remove ε.

We let the details to the reader.

q.e.d. (lemma 5.25) $\Diamond\Diamond\Diamond$

Proof of proposition 5.27
Let $L_{\bowtie\theta}(\mathcal{A}_1)$ be a stochastic regular language (with $\bowtie \in \{>, \geq\}$) and $L_{=1}(\mathcal{A}_2)$ be a regular language (where $\mathcal{A}_2$ is a probabilistic automaton with Dirac distributions). W.l.o.g we assume that $\bowtie \theta$ is different from > 1. Then $\mathcal{A}$ is defined by:

- $Q \stackrel{\text{def}}{=} Q_1 \uplus Q_2$;
- For all $i \in \{1,2\}$ and $q \in Q_i$, $\pi_0(q) \stackrel{\text{def}}{=} \frac{1}{2}\pi_{i,0}(q)$;
- For all $a \in A$, $q_1, q'_1 \in Q_1$, $q_2, q'_2 \in Q_2$,
  $\mathbf{P}_a[q_1,q'_1] \stackrel{\text{def}}{=} \mathbf{P}_{1,a}[q_1,q'_1]$, $\mathbf{P}_a[q_2,q'_2] \stackrel{\text{def}}{=} \mathbf{P}_{2,a}[q_2,q'_2]$ and $\mathbf{P}_a[q_1,q'_2] \stackrel{\text{def}}{=} \mathbf{P}_a[q_2,q'_1] \stackrel{\text{def}}{=} 0$;
- $F \stackrel{\text{def}}{=} F_1 \uplus F_2$.

We let the reader check that $L_{\bowtie\frac{\theta}{2}}(\mathcal{A}) = L_{\bowtie\theta}(\mathcal{A}_1) \cup L_{=1}(\mathcal{A}_2)$ and $L_{\bowtie\frac{1+\theta}{2}}(\mathcal{A}) = L_{\bowtie\theta}(\mathcal{A}_1) \cap L_{=1}(\mathcal{A}_2)$.

q.e.d. (proposition 5.27) $\Diamond\Diamond\Diamond$

Proof of lemma 5.28
Let $\mathcal{A}$ be the automaton of figure 5.3. For any word $w \in A^* \setminus \{a^{m_1}b \dots ba^{m_k}b \mid 1 < k\}$, one has $\mathbf{Pr}_{\mathcal{A}}(w) = 0$.

Let $w \stackrel{\text{def}}{=} a^{m_1}b \dots ba^{m_k}b$ with $1 < k$. It can be accepted either by a path starting from q_0 or by a path starting from q_3.

- When the path starts from q_0, in order to be accepted it must stay in q_0 for all b's except for the one that preceeds a^{m_k}. Then it must stay in q_1 for all a's. This leads to acceptance probability of $\frac{1}{2^{k+m_k}}$.
- When the path starts from q_3, in order to be rejected it must stay in q_3 for all a's that preceeds the first b and then must stay in q_5 when reading the remaining b's. This leads to a rejection probability of $\frac{1}{2^{k+m_1}}$.

So $\mathbf{Pr}_{\mathcal{A}}(w) = \frac{1}{2} - \frac{1}{2^{k+m_1}} + \frac{1}{2^{k+m_k}}$. Thus w is only accepted when $m_1 = m_k$.

q.e.d. (lemma 5.28) $\Diamond\Diamond\Diamond$

Proof of proposition 5.29
Let $L \stackrel{\text{def}}{=} \{a^{m_1}b \dots ba^{m_k}b \mid 1 < k \wedge m_1 = m_k\}$ the stochastic language of lemma 5.28. Then $LA^* = \{a^{m_1}ba^{m_2}b \dots a^{m_k}ba^* \mid \exists i > 1 \; m_i = m_1\}$ which is not a stochastic language as established by proposition 5.14.

q.e.d. (proposition 5.29) $\Diamond\Diamond\Diamond$

Proof of proposition 5.30
Let $L \stackrel{\text{def}}{=} \{a^{m_1}b \dots ba^{m_k}b \mid 1 < k \wedge m_1 = m_k\}$ the stochastic language of lemma 5.28. Assume that $L^* = L_{\bowtie\theta}(\mathcal{A})$ with $\bowtie \in \{>, \geq\}$.

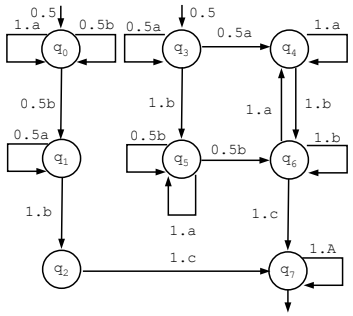

Figure 5.7: A PA for $\{a^{m_1}b\dots ba^{m_k}bcA^* \mid 1 < k \wedge m_1 = m_k\}$

Let $\sum_{i=0}^{n} c_i x^i$ be the minimal polynomial of $\mathbf{P}_a$.
Since 1 is an eigenvalue of $\mathbf{P}_a$, one gets $\sum_{i=0}^{n} c_i = 0$ and there are positive and negative coefficients.
By definition, $\sum_{i=0}^{n} c_i \mathbf{P}_{a^i} = 0$ and so for any word w, $\sum_{i=0}^{n} c_i \mathbf{P}_{a^i w} = 0$.
Let $c_{i_1}, \dots, c_{i_k}$ be the positive coefficients of this polynomial.
Choose $w \stackrel{\text{def}}{=} ba^{i_1}b(a^{i_2}b)^2 \dots (a^{i_k}b)^2$.
Observe that $a^i w \in L^*$ iff $i \in \{i_1, \dots, i_k\}$.

Case $L^* = L_{>\theta}(\mathcal{A})$**.** Let $0 \leq i \leq n$, using the observation $\pi_0 \mathbf{P}_{a^i w} \mathbf{1}_F^T > \theta$ iff $i \in \{i_1, \dots, i_k\}$.
So: $0 = \sum_{i=0}^{n} c_i \pi_0 \mathbf{P}_{a^i w} \mathbf{1}_F^T > (\sum_{i=0}^{n} c_i)\theta = 0$
leading to a contradiction.

Case $L^* = L_{\geq\theta}(\mathcal{A})$**.** Let $0 \leq i \leq n$, using the observation $\pi_0 \mathbf{P}_{a^i w} \mathbf{1}_F^T \geq \theta$ iff $i \in \{i_1, \dots, i_k\}$.
So: $0 = \sum_{i=0}^{n} c_i \pi_0 \mathbf{P}_{a^i w} \mathbf{1}_F^T > (\sum_{i=0}^{n} c_i)\theta = 0$
leading to a contradiction.

q.e.d. (proposition 5.30) ◊◊◊

Proof of proposition 5.31
Let $L \stackrel{\text{def}}{=} \{a^{m_1}b\dots ba^{m_k}bcA^* \mid 1 < k \wedge m_1 = m_k\}$ where $A \stackrel{\text{def}}{=} \{a, b, c\}$. We let the reader check that $L = L_{=\frac{1}{2}}(\mathcal{A})$ where $\mathcal{A}$ is the automaton of figure 5.7.

Define the homomorphism h from A to $A' \stackrel{\text{def}}{=} \{a, b\}$ by:

$$h(a) \stackrel{\text{def}}{=} a \qquad h(b) \stackrel{\text{def}}{=} b \qquad h(c) \stackrel{\text{def}}{=} \varepsilon$$

Then $h(L) = \{a^{m_1}ba^{m_2}b\dots a^{m_k}ba^* \mid \exists i > 1 \; m_i = m_1\}$ which is not a stochastic language as established by proposition 5.14.

q.e.d. (proposition 5.31) ◊◊◊

5.4.2 Proofs of section 5.3

Proof of proposition 5.34
As the dimension of the vector space generated by Gen is at most n, there are at most n iterations of main loop. The index of the first inner loop ranges over A while the index of the most inner loop ranges over Gen^2. This leads to a time complexity of $O(n^3|A|)$.

Assume now that the automata are not equivalent and that the algorithm has returned true. Let u be a word such that $\mathbf{Pr}_{\mathcal{A}}(u) \neq \mathbf{Pr}_{\mathcal{A}'}(u)$. Thus u has not be examined by the algorithm. Let $u \stackrel{\text{def}}{=} w'w$ with w the greatest suffix examined by the algorithm. Among such words u, pick one

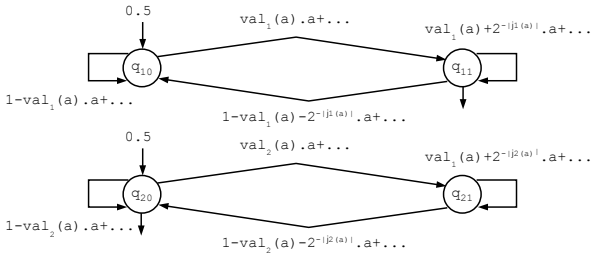

Figure 5.8: a PA for reduction of PCP

word such that $|w'|$ is minimal. We claim that there exists a word w'' that has been inserted in the stack before w such that $\mathbf{Pr}_{\mathcal{A}}(w'w'') \neq \mathbf{Pr}_{\mathcal{A}'}(w'w'')$.

Indeed since w has not been inserted in the stack, considering $w_1, \ldots, w_k$ that have been previously inserted in the stack, there exist $\lambda_1, \ldots, \lambda_k$ such that:

$$\mathbf{P}_w \mathbf{1}_F = \sum_{i=1}^{k} \lambda_i \mathbf{P}_{w_i} \mathbf{1}_F \text{ and } \mathbf{P}'_w \mathbf{1}_{F'} = \sum_{i=1}^{k} \lambda_i \mathbf{P}'_{w_i} \mathbf{1}_{F'}$$

So:

$$\mathbf{Pr}_{\mathcal{A}}(w'w) \stackrel{\text{def}}{=} \pi_0 \mathbf{P}_{w'} \mathbf{P}_w \mathbf{1}_F = \sum_{i=1}^{k} \lambda_i \pi_0 \mathbf{P}_{w'} \mathbf{P}_{w_i} \mathbf{1}_F = \sum_{i=1}^{k} \lambda_i \mathbf{Pr}_{\mathcal{A}}(w'w_i)$$

Similarly:

$$\mathbf{Pr}_{\mathcal{A}'}(w'w) = \sum_{i=1}^{k} \lambda_i \mathbf{Pr}_{\mathcal{A}'}(w'w_i)$$

Thus the claim follows.

Let us rewrite $w' \stackrel{\text{def}}{=} w'''a$. Since w_i has been inserted in the stack, aw_i is examined by the algorithm. So the word $u' \stackrel{\text{def}}{=} w'w_i$ has a decomposition $u' \stackrel{\text{def}}{=} z'z$ where z the greatest suffix examined by the algorithm has for suffix aw_i. So $|z'| < |w'|$ yielding a contradiction.

q.e.d. (proposition 5.34) $\Diamond\Diamond\Diamond$

Proof of proposition 5.35

Given a PCP, one builds a PA $\mathcal{A}$ such that $L_{=\frac{1}{2}}(\mathcal{A}) = \{\varepsilon\}$ iff the PCP does not have a solution.

For $w \in A^+$ and $i \in \{1, 2\}$, we define $val_i(w) \stackrel{\text{def}}{=} val(\varphi_i(a))$. Then $\mathcal{A}$ is defined by:

- $Q \stackrel{\text{def}}{=} \{q_{10}, q_{11}, q_{20}, q_{21}\}$;
- $\pi_0[q_{10}] \stackrel{\text{def}}{=} \pi_0[q_{20}] \stackrel{\text{def}}{=} \frac{1}{2}$ and $\pi_0[q_{11}] \stackrel{\text{def}}{=} \pi_0[q_{21}] \stackrel{\text{def}}{=} 0$;
- For all $a \in A$ and $i \in \{1, 2\}$,
  $\mathbf{P}_a[q_{i0}, q_{i1}] \stackrel{\text{def}}{=} 1 - \mathbf{P}_a[q_{i0}, q_{i0}] \stackrel{\text{def}}{=} val_i(a)$,
  $\mathbf{P}_a[q_{i1}, q_{i1}] \stackrel{\text{def}}{=} 1 - \mathbf{P}_a[q_{i1}, q_{i0}] \stackrel{\text{def}}{=} val_i(a) + 2^{-|\varphi_i(a)|}$,
  and all other items of transition matrices are null;
- $F \stackrel{\text{def}}{=} \{q_{11}, q_{20}\}$.

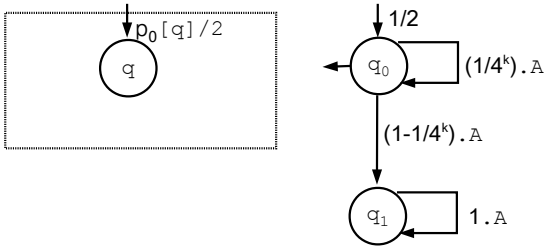

Figure 5.9: a PA for reduction of a large inequality to a strict inequality

So for all $w \in A^*$ and $a \in A$

$$\mathbf{1}_{q_{i0}}\mathbf{P}_{wa}\mathbf{1}_{q_{i1}}^T = \mathbf{1}_{q_{i0}}\mathbf{P}_w\mathbf{1}_{q_{i1}}^T(val_i(a)+2^{-|\varphi_i(a)|})+(1-\mathbf{1}_{q_{i0}}\mathbf{P}_w\mathbf{1}_{q_{i1}}^T)val_i(a) = val_i(a)+2^{-|\varphi_i(a)|}\mathbf{1}_{q_{i0}}\mathbf{P}_w\mathbf{1}_{q_{i1}}^T$$

By induction we obtain that for all $w \stackrel{\text{def}}{=} a_1 \dots a_n$:

$$\mathbf{1}_{q_{i0}}\mathbf{P}_w\mathbf{1}_{q_{i1}}^T = \sum_{j=1}^{n} val_i(a_j)2^{-\sum_{j<k\leq n}|\varphi_i(a_k)|} = val_i(w)$$

So for $w \in A^+$: $\mathbf{Pr}_{\mathcal{A}}(w) = \frac{1}{2}(val_1(w)+1-val_2(w))$. Thus $w \in L_{=\frac{1}{2}}(\mathcal{A})$ iff $val(\varphi_1(w)) = val(\varphi_2(w))$ implying (due to our assumption on images) that $\varphi_1(w) = \varphi_2(w)$ which means that w is a solution of the PCP.

In the proof of lemma 5.25, we have built $\mathcal{A}'$ with twice the number of states of $\mathcal{A}'$ such that $L_{=\frac{1}{2}}(\mathcal{A}') = L_{=\frac{1}{2}}(\mathcal{A}) \setminus \{\varepsilon\}$. So $L_{=\frac{1}{2}}(\mathcal{A}') = \emptyset$ iff the PCP does not have a solution.

q.e.d. (proposition 5.35) $\Diamond\Diamond\Diamond$

Proof of corollary 5.37

All the probabilities of the automaton corresponding to the reduction of PCP in the proof of proposition 5.35 are multiples of 2^k for some k depending on the PCP. The transformation of proposition 5.8 produces an automaton, say $\mathcal{A}$ whose probabilities are product of the initial probabilities, so they are multiples of 4^k and such that $L_{\geq\frac{1}{4}}(\mathcal{A}) = \emptyset$ iff the corresponding PCP does not have a solution.

Due to the transition probabilities, for all word $w \in A^+$, $\mathbf{Pr}_{\mathcal{A}}(w) = \frac{d}{4^{k|w|}}$ where d is an integer depending on w. So $\mathbf{Pr}_{\mathcal{A}}(w) \geq \frac{1}{4}$ iff $\mathbf{Pr}_{\mathcal{A}}(w) > \frac{1}{4} - \frac{1}{4^{k|w|}}$.

Let $\mathcal{A}'$ be defined by:

- $Q' \stackrel{\text{def}}{=} Q \cup \{q_0, q_1\}$;
- foral $q \in Q$, $\pi'_0[q] \stackrel{\text{def}}{=} \frac{\pi_[q]}{2}$, $\pi'_0[q_0] \stackrel{\text{def}}{=} \frac{1}{2}$ and $\pi'_0[q_1] \stackrel{\text{def}}{=} 0$;
- For all $a \in A$ and $q, q' \in Q$,
  $\mathbf{P}'_a[q, q'] \stackrel{\text{def}}{=} \mathbf{P}'_a[q, q']$,
  $\mathbf{P}'_a[q_0, q_0] \stackrel{\text{def}}{=} 1 - \mathbf{P}'_a[q_0, q_1] \stackrel{\text{def}}{=} \frac{1}{4^k}$,
  $\mathbf{P}'_a[q_1, q_1] \stackrel{\text{def}}{=} 1$ and all other items of transition matrices are null;
- $F' \stackrel{\text{def}}{=} F \cup \{q_0\}$.

$\mathcal{A}'$ is represented in figure 5.9.
Thus for all $w \in A^+$,

$$\mathbf{Pr}_{\mathcal{A}'}(w) = \frac{\mathbf{Pr}_{\mathcal{A}}(w)}{2} + \frac{1}{2 \times 4^{k|w|}}$$

which can be rewritten as:

$$\mathbf{Pr}_{\mathcal{A}}(w) = 2\mathbf{Pr}_{\mathcal{A}'}(w) - \frac{1}{4^{k|w|}}$$

So $\mathbf{Pr}_{\mathcal{A}}(w) > \frac{1}{4} - \frac{1}{4^{k|w|}}$ iff $\mathbf{Pr}_{\mathcal{A}'}(w) > \frac{1}{8}$.
So $L_{>\frac{1}{8}}(\mathcal{A}') = L_{\geq\frac{1}{4}}(\mathcal{A}) \cup \{\varepsilon\}$.
As previously done, we eliminate ε by doubling the number of states.

q.e.d. (corollary 5.37) $\Diamond\Diamond\Diamond$

Bibliography

[AHU 74] Aho A. V., Hopcroft J. E. , Ullman J. D. The Design and Analysis of Computer Algorithms. Addison-Wesley Publishing Company, 1974

[BRE 98] Brémaud P. Markov Chains. Gibbs Fields, Monte Carlo Simulation, and Queues. Springer-Verlag, 1998

[CHV 83] Chvatal V. Linear Programming Series of Books in the Mathematical Sciences W. H. Freeman, 1983

[CIN 75] Çinlar E. Introduction to Stochastic Processes. Prentice Hall, 1975,

[FEL 68] Feller W. An Introduction to Probability Theory and its Applications. Volume I. John Wiley & Sons, 1968, (third edition).

[FEL 71] Feller W. An Introduction to Probability Theory and its Applications. Volume II. John Wiley & Sons, 1971, (second edition).

[FLI 74] Fliess M. Propriétés booléennes des langages stochastiques. *Mathematical Systems Theory*, 7(4): 353-359, 1974

[FOA 98] Foata D., Fuchs A. Calcul des probabilités. Dunod, 1998, Seconde édition.

[FOA 02] Foata D., Fuchs A. Processus stochastiques. Processus de Poisson, chaînes de Markov et martingales. Dunod, 2002.

[FOX 88] Fox B. L., Glynn P. W. Computing Poisson probabilities *Commun. ACM*, vol. 31, no. 4, pp. 440-445, 1988.

[HMU 06] Hopcroft J.E., Motwani R. and Ullman J.D. Introduction to Automata Theory, Languages, and Computation. Addison-Wesley, Third edition 2006.

[JEN 53] Jensen A. Markov chains as an aid in the study of Markov processes. *Skand. Aktuarietidskrift*, vol. 3, pages 87-91, 1953.

[KS 60] Kemeny J.G., Snell J.L. Finite Markov Chains. D. Van Nostrand-Reinhold, New York, NY, 1960.

[KSK 76] Kemeny J.G., Snell J.L., Knapp A.W. Denumerable Markov Chains. Springer-Verlag, 1976

[PUT 94] Puterman M. Markov Decision Processes: Discrete Stochastic Dynamic Programming. John Wiley & Sons inc., 1994

[RAB 63] Rabin, M. O. Probabilistic Automata. *Information and Control* 6, pp. 230-245, 1963

[RTV 97] Roos C. , Terlaky T. and Vial J. P. Theory and Algorithms for Linear Optimization: An Interior Point Approach. John Wiley and Sons inc., 1997

[STE 94] Stewart W. J., Introduction to the numerical solution of Markov chains. Princeton University Press, USA, 1994.