Perturbed Markov chains

Eilon Solan[*] and Nicolas Vieille[†‡]

February 24, 2002

Abstract

We obtain results on the sensitivity of the invariant measure and other statistical quantities of a Markov chain with respect to perturbations of the transition matrix. We use graph-theoretic techniques, in contrast with the matrix analysis techniques previously used.

keywords: Perturbed Markov chains, conductance.

Primary subject classification: 60J10

Secondary subject classification: 60F10

[*]MEDS Department, Kellogg School of Management, Northwestern University, *and* the School of Mathematical Sciences, Tel Aviv University, Tel Aviv 69978, Israel. e-mail: eilons@post.tau.ac.il

[†]Département Finance et Economie, HEC, 1, rue de la Libération, 78 351 Jouy-en-Josas, France. e-mail: vieille@hec.fr

[‡]We acknowledge the financial support of the Arc-en-Ciel/Keshet program for 2001/2002. The research of the second author was supported by the Israel Science Foundation (grant No. 03620191).

1 Introduction

Consider an irreducible Markov chain (S, q) with finite state space, and denote its invariant measure by $\mu = (\mu_s)_{s \in S}$. The main purpose of the paper is to quantify the change in μ when q slightly changes to $\widehat{q}$. This question has a long history, starting with Schweizer [10]. There, the problem is analyzed using matrix analysis. It is shown that the modulus of continuity of $q \mapsto \mu$ is related to the fundamental matrix of the Markov chain (S, q). This formalism has later been exploited in numerous studies extending the results of [10], see e.g. [6], [11], [12]. There has also been some literature dealing with the numerical analysis aspects of the computation of the fundamental matrix, see e.g. [4]. There is also a wide literature that quantifies changes in the eigenvalues of the fundamental matrix as the matrix itself changes, see e.g. [2].

Our work differs both in the type of results we obtain and in the techniques we use. First, rather than matrix analysis, we use the graph techniques developed by Freidlin and Wenzell [3] and extensively used in the analysis of Markov chains with rare transitions (see, e.g., Catoni [1]). Next, we obtain results of the following nature: provided the ratio $\frac{q(t|s)}{\widehat{q}(t|s)}$ is close to one for the most frequent transitions $s \to t$, $|\mu_s - \widehat{\mu}_s|$ is small compared to μ_s. In addition, we provide information on other quantities of interest.

Our notion of closeness between q and $\widehat{q}$ is the following. Given $\varepsilon, \beta > 0$, we say that $\widehat{q}$ is (ε, β)-close to q if for every two states $s, t \in S$, $\left|1 - \frac{\widehat{q}(t|s)}{q(t|s)}\right| \leq \beta$ whenever (a) $\mu_s q(t|s) \geq \varepsilon$ or (b) $\mu_s \widehat{q}(t|s) \geq \varepsilon$. Condition (a) holds whenever the transition from s to t occurs frequently. Condition (b) is not analogous to (a), since it involves the invariant measure of q, and the transition function $\widehat{q}$.

Provided ε and β are small enough, we show that $\widehat{q}$ is irreducible, and $\frac{\mu_s}{\widehat{\mu}_s}$ close to 1 for each s. In addition, given any proper subset C of S, we obtain estimates on the exit distribution from C, and on the average length of visits to C under q and $\widehat{q}$ respectively.

The main motivation for this question is a problem of a statistical nature, that arises in the analysis of a class of stochastic games, see Rosenberg et *al,* [9]. There, the transition function q is unknown, and an outside observer wishes to estimate q or its invariant measure, on the basis of the N first visited states $s_0, s_1, \ldots, s_N$. The observer can calculate the *empirical* transition function $\widehat{q}$, defined by

$$\widehat{q}(t \mid s) = \#\{n < N \mid s_n = s, s_{n+1} = t\} / \#\{n < N \mid s_n = s\},$$

and compute the invariant measure $\widehat{\mu}$ of μ. If N is large enough, then with high probability $\widehat{q}$ is (ε, β)-close to q, therefore $\widehat{\mu}$ is close to μ.

The paper is organized as follows. Section 2 contains the statements of the main results. Section 3 briefly recalls standard formulas, and states few elementary properties. Section 4 is devoted to the proof of the main result. The last section deals with the variation that is used in [9].

2 Notations and results

Let S be a finite set, fixed through the paper, with at least two elements. For every subset $C \subseteq S$, $\overline{C} = S \setminus C$ is the complement of C in S, $|C|$ is its cardinality, and $\Delta(C)$ is the set of probability distributions over C. For $s \in C \subseteq S$, we denote $C \setminus s$ instead of the more cumbersome $C \setminus \{s\}$.

2.1 Main results

Let q be an irreducible transition function over S, with invariant measure $\mu = (\mu_s)_{s \in S}$. For every $C \subseteq S$ we denote $\mu_C = \sum_{s \in C} \mu_s$. Let $\widehat{q}$ be another transition function over S. Assuming $\widehat{q}$ is irreducible, we wish to bound the distance between μ and $\widehat{\mu}$.

Our notion of closeness of $\widehat{q}$ to q involves a measure of how mixing q is. Our measure involves the quantity

$$\zeta_q = \min_{\emptyset \subset C \subset S} \sum_{s \in C} \mu_s q(\overline{C} \mid s), \tag{1}$$

which is a variant of the *conductance*, see e.g. [5], [7], [8]. Given $C \subset S$, the quantity $\sum_{s \in C} \mu_s q(\overline{C} \mid s)$ measures the average frequency of transitions out of C. Hence, ζ_q , being the lowest such frequency, is a measure of how isolated a subset C may be. Formally,

Definition 1 *Let $\varepsilon, \beta > 0$. We say that a transition function $\widehat{q}$ is (ε, β)-*close to q *if for every two states $s, t \in S$, $\left|1 - \frac{\widehat{q}(t|s)}{q(t|s)}\right| \le \beta$ whenever $\mu_s q(t|s) \ge \varepsilon\zeta_q$ or $\mu_s \widehat{q}(t|s) \ge \varepsilon\zeta_q$.*

Note that this closeness notion is not symmetric, since we use only the invariant distribution of q.

Denote $L = \sum_{n=1}^{|S|-1} \binom{|S|}{n} n^{|S|}$. We now state our main result.

Theorem 2 *Let $\beta \in (0, 1/2^{|S|})$ and let $\varepsilon \in (0, \frac{\beta(1-\beta)}{L \times |S|^4})$. For every irreducible transition function q on S and every transition function $\widehat{q}$ that is (ε, β)-close to q:*

1. *$\widehat{q}$ is irreducible.*
2. *Its invariant distribution $\widehat{\mu}$ satisfies $\left|1 - \frac{\widehat{\mu}_s}{\mu_s}\right| \le 18\beta L$ for each $s \in S$.*

In Theorem 2, the transition functions are required to be close over the whole state space S. For the analysis in [9] we need a variation of Theorem 2, where the two transition functions are close in a subset S_1 of S, and are identical outside S_1.

Let S_1 be a subset of S, with $|S_1| > 1$. Define

$$\zeta_q^1 = \min_{\emptyset \ne C \subset S_1} \sum_{s \in C} \mu_s q(\overline{C} \mid s).$$

Let $(\mathbf{s}_n)$ be a Markov chain with transition function q. We denote by $\mathbf{P}_{s,q}$ the law of $(\mathbf{s}_n)$ when the initial state is s, and by $\mathbf{E}_{s,q}$ the corresponding expectation.

For every proper subset C of S we let $T_C = \min\{n \ge 0, \mathbf{s}_n \in C\}$ denote the first hitting time of C and $T_C^+ = \min\{n \ge 1, \mathbf{s}_n \in C\}$ the first return to C . By convention, the minimum over an empty set is $+\infty$.

Definition 3 *Let $\varepsilon, \beta > 0$. We say that a transition function $\widehat{q}$ is (ε, β)-*close to q on S_1 *if for every two states $s, t \in S$, $\left|1 - \frac{\widehat{q}(t|s)}{q(t|s)}\right| \le \beta$ whenever $\mu_s q(t|s) \ge \varepsilon\zeta_q^1$ or $\mu_s \widehat{q}(t|s) \ge \varepsilon\zeta_q^1$.*

We now state the Theorem that corresponds to Theorem 2.

Theorem 4 *Let $\beta \in (0, 1/2^{|S|})$, $a > 0$ and $\varepsilon \in (0, \frac{1}{2}\left(\frac{a}{L}\right)^{|S|} \times \frac{\beta(1-\beta)}{L\times|S|^4})$. Let q be an irreducible transition function such that $\mathbf{P}_{s,q}(T^+_{S_1\cup\{t\}} = T^+_{\{t\}}) \geq a$ for every $s, t \in S_1$. Then, for every transition function $\widehat{q}$ that is (ε, β)-close to q on S_1 and that coincides with q on $S\backslash S_1$, we have*

1. *All states of S_1 belong to the same recurrent set R for $\widehat{q}$.*

2. *The invariant distribution $\widehat{\mu}$ of $\widehat{q}$ on R satisfies*

$$\left|1 - \frac{\widehat{\mu}(s|S_1)}{\mu(s|S_1)}\right| \leq 18\beta L, \text{ for each } s \in S_1, \tag{2}$$

   where $\mu(s \mid S_1) = \mu_s/\mu_{S_1}$.

Note that the claims in Theorem 4 differ from those in Theorem 2. It is no longer claimed that $\widehat{q}$ is irreducible, nor that the unconditional invariant measures μ and $\widehat{\mu}$ are close. The statements in Theorem 4 are optimal in this respect. This is due to the fact that the quantity ζ^1_q contains no information on the frequency of transitions out of S_1. To emphasize this point, consider the following example.

Assume that $S = \{a, b, c\}$ and $S_1 = \{a, b\}$. Let $\varepsilon, \beta \in (0, 1/2)$ be given. Let two additional parameters λ and η be given in $(0, 1)$, and define q as follows. From state a (resp. b) a chain with transition function q moves to c with probability η, and otherwise to b (resp. to a). From state c, the chain remains in c with probability $1-\lambda$, and otherwise moves to a or b with equal probability $\frac{1}{2}\lambda$.

Plainly, q is irreducible, and the value of $\mu_a = \mu_b$ depends on the ratio λ/η: this common value may be arbitrary close to 0 (resp. to 1/2) provided λ/η is close enough to 0 (resp. to $+\infty$). Note that $\zeta^1_q = \mu_a q(\{b, c\}\,|a) = \mu_a$. Let now $\widehat{q}$ be defined exactly as q, except that the parameter η is replaced by another parameter $\widehat{\eta} \in [0, 1]$. As soon as $\eta, \widehat{\eta} < \min(\varepsilon, \beta)$, $\widehat{q}$ is (ε, β)-close to q. This is in particular the case if $\widehat{\eta} = 0$, in which case $\widehat{q}$ fails to be irreducible. On the other hand, even if $\widehat{\eta} > 0$, the values of $\eta, \widehat{\eta}$ and λ can be chosen in such a way that the inequalities $\eta, \widehat{\eta} < \min\{\varepsilon, \beta\}$ are satisfied, and $\eta \ll \lambda \ll \widehat{\eta}$. Hence, even if $\widehat{q}$ is irreducible, its unconditional invariant measure $\widehat{\mu}$ may be arbitrarily far from μ.

2.2 Other results

Our graph-theoretic approach allows us to obtain information on other quantities of interest. We here present the statements of the corresponding results. Additional extensions will also be suggested.

We let $\mathbf{Q}_{s,q}(\cdot|C)$ denote the law of the exit state from C: $\mathbf{Q}_{s,q}(t|C) = \mathbf{P}_{s,q}(T_{\overline{C}} = T_t)$ for $t \notin C$. Next, we set

$$\nu_C(s) := \frac{\sum_{t \in \overline{C}} \mu_t q(s|t)}{\sum_{t \in \overline{C}} \mu_t q(C|t)} \text{ for } C \subset S \text{ and } s \in C, \text{ and} \tag{3}$$

$$K_C := \sum_{s \in C} \nu_C(s) \mathbf{E}_{s,q}[e_C] \text{ for } C \subset S.$$

The numerator (resp. the denominator) in (3) is the long run frequency of transitions from $\overline{C}$ to s (resp. from $\overline{C}$ to C). Thus, $\nu_C(s)$ is the probability that the first stage in C the process visits is s, while K_C is the average length of a visit to C.

Assuming $\widehat{q}$ is irreducible, the corresponding quantities for $\widehat{q}$ will be denoted by $\mathbf{Q}_{s,\widehat{q}}$, $\widehat{\nu}_C(s)$ and $\widehat{K}_C$. We now state the results on $\mathbf{Q}_{s,\widehat{q}}$ and $\widehat{K}_C$ that hold in the framework of Theorems 2 and 4 respectively.

Theorem 5 *Set $c = 2|S|^2$. Under the assumptions of Theorem 2, the following holds: for each $C \subset S$,*

1. $\left\|\mathbf{Q}_{s,q}(\cdot \mid C) - \mathbf{Q}_{s,\widehat{q}}(\cdot \mid C)\right\| < 12\beta L$ *for every* $s \in C$
2. $\frac{1}{c} K_C \leq \widehat{K}_C \leq c K_C$.

Theorem 6 *Set $c = 2|S|^2$. Under the assumptions of Theorem 4, the following holds: for each $C \subset S_1$,*

1. $\left\|\mathbf{Q}_{s,q}(\cdot \mid C) - \mathbf{Q}_{s,\widehat{q}}(\cdot \mid C)\right\| < 12\beta L$ *for every* $s \in C$

2. $\frac{1}{c}K_C \le \widehat{K}_C \le cK_C$

3. $\frac{1}{c}K_{S_1} \le \widehat{K}_{S_1} \le cK_{S_1}$ *or* $K_{S_1}, \widehat{K}_{S_1} \ge \frac{1}{2\varepsilon|S|} \times \frac{\mu_{S_1}}{\zeta_q^1}$.

We let q be an irreducible transition function over S. It is fixed throughout the paper.

3 Preliminaries

Our computations are based on formulas due to Freidlin and Wenzell [3], that express invariant measure, exit distributions and expected hitting times in graph-theoretic terms. For a discussion of some applications, we refer to Catoni [1]. These tools have also been used in the context of stochastic games in [14] and [13].

The weight of a graph is obtained from the transition probabilities corresponding to the different edges of the graph. We recall these formulas in section 3.1. Next, we compare the weights of a given graph under a transition function $\widehat{q}$ that is close to q.

3.1 Reminder

Given $C \subset S$, a *C-graph* is a directed graph without cycle g over S such that:[1]

- For $s \in C$, there is exactly one edge starting at s, denoted by $(s, g(s))$.
- For $s \in \overline{C}$, there is no edge starting at s.

Thus, given $s \in C$, there is a unique path starting at s and ending at some $t \in \overline{C}$. We say that *s leads to t along g*. We denote by $G(C)$ the set of C-graphs; for $s \in C, t \in \overline{C}$, $G_{s,t}(C)$ is the subset of graphs $g \in G(C)$ such that s leads to t along g. Note that $G(C)$ depends only on C, and not on the transition function. Note also that L bounds the number of graphs: $L \ge \sum_{\emptyset \subset C \subset S} |G(C)|$.

[1] Our C-graphs correspond to $\overline{C}$-graphs in [3], [1].

We identify each C-graph g with the collection of its edges: $g = \cup_{s \in C}\{(s, g(s))\}$.

Given $D \subseteq C$, and $g \in G(C)$, the *restriction of* g *to* D is defined to be the subgraph of g that contains exactly those edges of g that start in D. Thus, it is the D-graph $g' = \cup_{s \in D} \{(s, g(s))\}$.

For every $g \in G(C)$, we define the *weight* of g under q by

$$p(g) := \prod_{(s,t) \in g} q(t|s).$$

Proposition 7 (Freidlin-Wenzell, 1984) *Let* (S, q) *be a Markov chain.*

- *If* q *is irreducible then for every* $s \in S$

$$\mu_s = \frac{\sum_{G(S \setminus \{s\})} p(g)}{\sum_{y \in S} \sum_{G(S \setminus \{y\})} p(g)}. \tag{4}$$

- *For every proper subset* C *of* S *and every* $s \in C$,

$$\mathbf{E}_{s,q}\left[T_{\overline{C}}\right] = \frac{\sum_{G(C \setminus \{s\})} p(g) + \sum_{t \in C, t \neq s} \sum_{G_{s,t}(C \setminus \{t\})} p(g)}{\sum_{G(C)} p(g)}, \tag{5}$$

and

$$\mathbf{Q}_{s,q}(t|C) = \frac{\sum_{G(C)} p(g)}{\sum_{G_{s,t}(C)} p(g)} \text{ for each } t \notin C. \tag{6}$$

3.2 Basic properties

In this section we provide basic properties of weights of graphs. The transition function q is here arbitrary.

Definition 8 *Let* C *be a proper subset of* S, *and let* $\eta > 0$. *A graph* $g \in G(C)$ *is* η-maximal *if*

$$p(g) \geq \eta \max_{g' \in G(C)} p(g').$$

We denote by $G^\eta(C)$ the set of η-maximal C-graphs. For simplicity of notations, we do not emphasize the dependency of $G^\eta(C)$ on the transition function. Clearly, $G^\eta(C)$ is non-empty, for every $\eta \leq 1$ and $C \subset S$. It is worth listing a few basic properties of graphs that we use repeatedly.

Proposition 9 ***P0*** *Let $C_1 \cap C_2 = \emptyset$, and $g_i \in G(C_i)$, for $i = 1, 2$. If all paths of g_1 lead to $\overline{C_1 \cup C_2}$, then $g_1 \cup g_2$ is a $C_1 \cup C_2$-graph.*

P1 *Let $C_1 \cap C_2 = \emptyset$, $g \in G^\eta(C_1 \cup C_2)$, and g_i the restriction of g to C_i. If all paths of g_2 lead to $\overline{C_1 \cup C_2}$, then $g_1 \in G^\eta(C_1)$.*

P2 *Let $C_1 \cap C_2 = \emptyset$, and $g_i \in G^{\eta_i}(C_i)$ for $i = 1, 2$. If $g_1 \cup g_2$ is a $C_1 \cup C_2$-graph, then it is $\eta_1\eta_2$-maximal.*

Proof. **P0** and **P2** follow from the definitions. We now show that **P1** holds. Otherwise, there is $g'_1 \in G(C_1)$ such that $p(g_1) < \eta p(g'_1)$. By **P0**, $g' = g'_1 \cup g_2$ is in $G(C_1 \cup C_2)$, but $p(g) < \eta p(g')$, a contradiction. ∎

Note that **P1** needs not hold without the condition that all paths of g_2 lead to $\overline{C_1 \cup C_2}$. Indeed, take $S = \{1, 2, 3, 4\}$, $C_1 = \{1\}$, $C_2 = \{2\}$, and $q(2 \mid 1) = q(1 \mid 2) = 1 - q(3 \mid 1) = 1 - q(4 \mid 2) = 2/3$. The C_2-graph $g_1 = (2 \to 4)$ is 1/2-maximal, and the $C_1 \cup C_2$-graph $(1 \to 2, 2 \to 4)$ is 1-maximal.

Lemma 10 *Let C be a proper subset of S, let $\eta > 0$, and let H be a set of graphs such that $G^\eta(C) \subseteq H \subseteq G(C)$. Then*

$$0 \leq \frac{\sum_{g \in G(C)} p(g)}{\sum_{g \in H} p(g)} - 1 < \eta L.$$

In particular,

$$0 \leq 1 - \frac{\sum_{g \in H} p(g)}{\sum_{g \in G(C)} p(g)} < \eta L.$$

Proof. Since $H \subseteq G(C)$, and by the definition of $G^\eta(C)$,

$$0 \leq \frac{\sum_{g \in G(C)} p(g)}{\sum_{g \in H} p(g)} - 1 = \frac{\sum_{g \in G(C) \setminus H} p(g)}{\sum_{g \in H} p(g)} \leq \frac{\sum_{g \in G(C) \setminus G^\eta(C)} p(g)}{\sum_{g \in G^1(C)} p(g)} < \eta L,$$

as desired. ∎

4 Proof of the main results

We here prove Theorems 2 and 5. We let $\varepsilon, \beta \in (0, 1)$ satisfy the assumptions of Theorem 2, and $\widehat{q}$ be another transition function over S. We assume that $\widehat{q}$ is (ε, β)-close to q.

4.1 On graphs

For every proper subset C of S and every $\eta > 0$, we denote by $\widehat{G}^\eta(C)$ the set of η-maximal graphs under $\widehat{q}$. For every C-graph g, $\widehat{p}(g) = \prod_{s\in C} \widehat{q}(g(s) \mid s)$ is the weight of g under $\widehat{q}$.

Lemma 11 *For every proper subset C of S,*

$$\frac{1-\beta}{|S|^2}\sum_{s\in C}\mu_s q(\overline{C}|s) \le \sum_{s\in C}\mu_s \widehat{q}(\overline{C}|s) \le (1+\beta)|S|^2 \sum_{s\in C}\mu_s q(\overline{C}|s). \tag{7}$$

Proof. Let $s_0 \in C$ and $t_0 \in \overline{C}$ maximize the quantity $\mu_s q(t \mid s)$ amongst $s \in C$ and $t \in \overline{C}$. Then $\mu_{s_0} q(t_0 \mid s_0) \ge \sum_{s\in C}\mu_s q(\overline{C}|s)/|S|^2 \ge \zeta_q/|S|^2 > \varepsilon\zeta_q$. Since $\widehat{q}$ is (ε,β)-close to q, $\widehat{q}(t_0 \mid s_0) \ge (1-\beta)q(t_0 \mid s_0)$. In particular,

$$\sum_{s\in C}\mu_s \widehat{q}(\overline{C} \mid s) \ge \mu_{s_0}\widehat{q}(t_0 \mid s_0) \ge (1-\beta)\mu_{s_0} q(t_0 \mid s_0) \ge \frac{1-\beta}{|S|^2}\sum_{s\in C}\mu_s q(\overline{C}|s), \tag{8}$$

and the left hand side inequality in (7) holds.

Let $s_1 \in C$ and $t_1 \in \overline{C}$ maximize the quantity $\mu_s \widehat{q}(t \mid s)$ amongst $s \in C$ and $t \in \overline{C}$. By (8), $\mu_{s_1}\widehat{q}(t_1 \mid s_1) \ge \sum_{s\in C}\mu_s \widehat{q}(\overline{C} \mid s)/|S|^2 \ge (1-\beta)\zeta_q/|S|^4 > \varepsilon$. Since $\widehat{q}$ is (ε,β)-close to q, $q(t_1 \mid s_1) \ge \widehat{q}(t_1 \mid s_1)/(1+\beta)$. Therefore

$$\sum_{s\in C}\mu_s q(\overline{C} \mid s) \ge \mu_{s_1} q(t_1 \mid s_1) \ge \frac{1}{1+\beta}\mu_{s_1}\widehat{q}(t_1 \mid s_1) \ge \frac{1}{(1+\beta)|S|^2}\sum_{s\in C}\mu_s \widehat{q}(\overline{C} \mid s),$$

and the right hand side inequality holds as well. ■

Lemma 12 *Let $C \subset S$ and $s \in C$ be given. For every $g \in G^\beta(C)$ (resp. $g \in \widehat{G}^\beta(C)$) $\mu_s q(g(s) \mid s) \ge \varepsilon\zeta_q$ (resp. $\mu_s \widehat{q}(g(s) \mid s) \ge \varepsilon\zeta_q$).*

Note that the second claim is not symmetric to the first, since in both we use the invariant distribution of q.

Proof. The proof is quite similar for $g \in G^\beta(C)$ and $g \in \widehat{G}^\beta(C)$. We prove the lemma for the former, and mention where the proof for the latter differs.

Let $g \in G^\beta(C)$ be arbitrary. The proof is by induction over the number of states in C.

If $|C| = 1$, then $C = \{s\}$ for some $s \in S$. Since g is β-maximal, $\mu_s q(g(s) \mid s) \geq \beta/|S|\mu_s q(\overline{C}|s) \geq \frac{\zeta_q}{|S|}\beta$ (for $g \in \widehat{G}^\beta(C)$, by Lemma 11, $\mu_s \widehat{q}(g(s) \mid s) \geq \frac{\beta}{|S|}\mu_s \widehat{q}(\overline{C}|s) \geq \beta\frac{1-\beta}{|S|^3}\mu_s q(\overline{C}|s) \geq \beta\frac{1-\beta}{|S|^3}\zeta_q$).

Consider now the case $|C| > 1$.

We first assume that there are at least two edges of g whose endpoints do not belong to C. Let $s_1 \neq s_2 \in C$ Let g_i be the restriction of g to $C \setminus \{s_i\}$, $i = 1, 2$. By **P1**, $g_i \in G^\beta(C \setminus \{s_i\})$. Since any edge of g is an edge of g_1 or g_2 (or both), the induction hypothesis applied to $C \setminus \{s_i\}$ and g_i, $i = 1, 2$, implies that the claim holds for g.

Assume now that there is a unique state $s_1 \in C$ such that $g(s_1) \notin C$. Let g_1 be the restriction of g to $C \setminus \{s_1\}$. By **P1**, $g_1 \in G^\beta(C \setminus \{s_1\})$. By the induction hypothesis applied to $C \setminus \{s_1\}$ and g_1, $\mu_s q(g(s) \mid s) \geq \varepsilon\zeta_q$ for every $s \in C \setminus \{s_1\}$. Thus, it remains to show that $\mu_{s_1} q(g(s_1) \mid s_1) \geq \varepsilon\zeta_q$.

Let $s_2 \in C$ maximize the quantity $\mu_s q(\overline{C} \mid s)$ amongst $s \in C$ (for $g \in \widehat{G}^\beta(C)$, it is chosen to maximize $\mu_s \widehat{q}(\overline{C} \mid s)$). By the definition of ζ_q, $\mu_{s_2} q(\overline{C} \mid s_2) \geq \zeta_q/|S|$ (for $g \in \widehat{G}^\beta(C)$, by Lemma 11, $\mu_{s_2}\widehat{q}(\overline{C} \mid s_2) \geq (1-\beta)\zeta_q/|S|^3$). Let $\widehat{g} \in G^1(S \setminus C)$ (for $g \in \widehat{G}^\beta(C)$, one also chooses $\widehat{g} \in G^1(S \setminus C)$) . By **P0** and **P2**, $\widehat{g} \cup g_1 \in G(S \setminus \{s_1\})$.

Let $\overline{g} \in G^1(S \setminus \{s_2\})$. Since $\overline{g}_{|S\setminus C}$ is a $S \setminus C$-graph, we have $p(\widehat{g}) \geq p(\overline{g}_{|S\setminus C})$. Since for every $t \in \overline{C}$, $\overline{g}_{|C\setminus\{s_2\}} \cup (s_2, t)$ is a C-graph, $p(g) \geq p(\overline{g}_{|C\setminus\{s_2\}})q(t \mid s_2)$. In particular, $p(\widehat{g})p(g) \geq \beta p(\overline{g})q(t \mid s_2)$ for every $t \in \overline{C}$, and therefore $p(\widehat{g})p(g) \geq \frac{\beta}{|S|}p(\overline{g})q(\overline{C} \mid s_2)$.

Denote $\sum = \sum_{y \in S} \sum_{g \in G(S \setminus \{y\})} p(g)$. By (4), $\mu_s = \frac{1}{\sum} \sum_{g \in G(S \setminus \{s\})} p(g)$. In particular,

$$\begin{aligned}
\frac{\zeta_q}{|S|} &\le \mu_{s_2} q(\overline{C} \mid s_2) \le \frac{\sum_{g \in G(S \setminus \{s_2\})} p(g)}{\sum} \times q(\overline{C} \mid s_2) \\
&\le \frac{L p(\overline{g}) q(\overline{C} \mid s_2)}{\sum} \le \frac{L \times |S|}{\sum \beta} p(\widehat{g}) p(g) \\
&= \frac{L \times |S|}{\sum \beta} p(\widehat{g} \cup g \setminus (s_1, g(s_1))) q(g(s_1) \mid s_1) \\
&\le \frac{L \times |S|}{\sum \beta} \sum_{g \in G(S \setminus \{s_1\})} p(g) \times q(g(s_1) \mid s_1) \\
&= \frac{L \times |S|}{\beta} \mu_{s_1} q(g(s_1) \mid s_1).
\end{aligned}$$

But then $\mu_{s_1} q(g(s_1) \mid s_1) \ge \varepsilon \zeta_q$, as desired. The calculation for $g \in \widehat{G}^\beta(C)$ is analogous. ■

Corollary 13 *For every proper subset C of S,*

$$\left| 1 - \frac{\widehat{p}(g)}{p(g)} \right| \le (|S| + 1)\beta, \text{ for every } g \in G^\beta(C) \cup \widehat{G}^\beta(C) \tag{9}$$

and

$$\left| \frac{\sum_{g \in H} \widehat{p}(g)}{\sum_{g \in H} p(g)} - 1 \right| < (|S| + 1)\beta, \text{ where } H = G^\beta(C) \cup \widehat{G}^\beta(C).$$

Thus, the weights of β-maximal graphs under q and $\widehat{q}$ are close.

Proof. Note first that the second inequality follows immediately from the first one. Let us prove (9). Let $g \in G^\beta(C)$. By Lemma 12, $\mu_s q(g(s) \mid s) \ge \varepsilon \zeta_q$ for every $s \in C$. Since $\widehat{q}$ is (ε, β)-close to q, $(1 - \beta) q(g(s) \mid s) \le \widehat{q}(g(s) \mid s) \le (1 + \beta) q(g(s) \mid s)$. Multiplying this inequality over $s \in C$ yields $(1 - \beta)^{|C|} p(g) \le \widehat{p}(g) \le (1 + \beta)^{|C|} p(g)$, and (9) follows.

The proof for $g \in \widehat{G}^\beta(C)$ is similar. ■

4.2 Proof of Theorem 2

Proposition 14 *The transition function $\widehat{q}$ is irreducible.*

Proof. It is enough to prove that for every non-empty subset $C \subset S$, there exists $s \in C$, and $t \notin C$ such that $\widehat{q}(t \mid s) > 0$.

Let $s_1 \in C$ and $t_1 \notin C$ be such that $\mu_{s_1} q(t_1 \mid s_1) \geq \zeta_q/|S|^2 > \varepsilon\zeta_q$. Since $\widehat{q}$ is (ε, β)-close to q, $\widehat{q}(t_1 \mid s_1) > (1-\beta)q(t \mid s) > 0$. ■

We need the following technical Lemma.

Lemma 15 *1. Let $(a_i)_{i=1}^I$ and $(b_i)_{i=1}^I$ be positive numbers, and let $\varepsilon > 0$. If $\left|\frac{a_i}{b_i} - 1\right| < \varepsilon$ for every $i = 1, \ldots, I$ then $\left|\frac{\sum_{i=1}^I a_i}{\sum_{i=1}^I b_i} - 1\right| < \varepsilon$ and $\left|\frac{\min\{a_1,a_2,\ldots,a_I\}}{\min\{b_1,b_2,\ldots,b_I\}} - 1\right| < \varepsilon$.*

2. Let $\varepsilon \in (0, 1/3)$, and let $a, A, b, B > 0$. If $\left|\frac{a}{b} - 1\right| < \varepsilon$ and $\left|\frac{b}{B} - 1\right| < \varepsilon$ then $\left|\frac{a/b}{A/B} - 1\right| < 3\varepsilon$.

Proof. The proof of the first part is left to the reader. For the second part, note that $1/(1+\varepsilon) < B/b < 1/(1-\varepsilon)$, which implies that $B/b - 1 < \varepsilon/(1-\varepsilon)$. In particular,

$$\left|\frac{a/b}{A/B} - 1\right| \leq \left(\left|\frac{a}{A} - 1\right| + 1\right)\left|\frac{B}{b} - 1\right| + \left|\frac{a}{A} - 1\right| < (1+\varepsilon)\frac{\varepsilon}{1-\varepsilon} + \varepsilon < 3\varepsilon.$$

■

Proposition 16 *For each $s \in S$,*

$$|1 - \frac{\widehat{\mu}_s}{\mu_s}| < 18\beta L.$$

Proof. Fix $s \in S$. By (4),

$$\mu_s = \frac{\sum_{G(S\setminus\{s\})} p(g)}{\sum_{y\in S}\sum_{G(S\setminus\{y\})} p(g)} \text{ and } \widehat{\mu}_s = \frac{\sum_{G(S\setminus\{s\})} \widehat{p}(g)}{\sum_{y\in S}\sum_{G(S\setminus\{y\})} \widehat{p}(g)}.$$

For every $y \in S$, define $H_y = G^\beta(S\setminus\{y\}) \cup \widehat{G}^\beta(S\setminus\{y\})$. Define

$$\mu'_s = \frac{\sum_{H_s} p(g)}{\sum_{y\in S}\sum_{H_y} p(g)} \text{ and } \widehat{\mu}'_s = \frac{\sum_{H_s} \widehat{p}(g)}{\sum_{y\in S}\sum_{H_y} \widehat{p}(g)}.$$

By Lemma 10 and Lemma 15, $\left|\frac{\mu_s}{\mu'_s} - 1\right| < 3\beta L$ and $\left|\frac{\widehat{\mu}_s}{\widehat{\mu}'_s} - 1\right| < 3\beta L$. By Lemmas 10 and 15, $\left|\frac{\widehat{\mu}'_s}{\mu'_s} - 1\right| < 3(|S|+1)\beta$. Since $L \geq |S| \geq 2$, the result follows by Lemma 15(2). ■

4.3 Proof of Theorem 5

Proposition 17 *For every proper subset C of S, every $s \in C$ and $t \notin C$,*

$$\left|\mathbf{Q}_{s,q}(t \mid C) - \mathbf{Q}_{s,\widehat{q}}(t \mid C)\right| < 12\beta L.$$

Proof. Denote $H = G^{\beta}(C) \cup \widehat{G}^{\beta}(C)$, and $H_{s,t} = H \cap G_{s,t}(C)$.

Assume first that $H_{s,t} \neq \emptyset$. By (6), one has

$$\left|\mathbf{Q}_{s,q}(t|C) - \frac{\sum_{H \cap G_{s,t}(C)} p(g)}{\sum_{G(C)} p(g)}\right| \leq \beta L.$$

Since $\left|\frac{\sum_H p(g)}{\sum_{G(C)} p(g)} - 1\right| \leq \beta L$, this yields, by Lemma 15(2),

$$\left|\mathbf{Q}_{s,q}(t|C) - \frac{\sum_{H \cap G_{s,t}(C)} p(g)}{\sum_{H} p(g)}\right| \leq \beta L + 3\beta L \leq 4\beta L, \tag{10}$$

and a similar inequality holds with q replaced by $\widehat{q}$.

By Corollary 13 and Lemma 15(1),

$$\left|\frac{\sum_{H \cap G_{s,t}(C)} \widehat{p}(g)}{\sum_{H \cap G_{s,t}(C)} p(g)} - 1\right| \leq \beta(|S|+1) \text{ and } \left|\frac{\sum_H \widehat{p}(g)}{\sum_H p(g)}\right| \leq \beta(|S|+1). \tag{11}$$

By Lemma 15(2), (11) implies

$$\left|\frac{\sum_{H \cap G_{s,t}(C)} p(g)}{\sum_{H} p(g)} - \frac{\sum_{H \cap G_{s,t}(C)} \widehat{p}(g)}{\sum_{H} \widehat{p}(g)}\right| \leq 3\beta(|S|+1),$$

which implies, using (10),

$$\left|\mathbf{Q}_{s,q}(t|C) - \mathbf{Q}_{s,\widehat{q}}(t|C)\right| \leq \beta(8L + 3(|S|+1)).$$

If, on the other hand, $H_{s,t} = \emptyset$, then by (6) and the definition of $H_{s,t}$, $\mathbf{Q}_{s,q}(t \mid C), \mathbf{Q}_{s,\widehat{q}}(t \mid C) \leq \beta L$. ■

Proposition 18 *For every proper subset C of S,*

$$\frac{1}{2\,|S|^2} K_C \leq \widehat{K}_C \leq 2\,|S|^2\, K_C.$$

Proof. We first argue that

$$K_C = \frac{\sum_{s\in C}\mu_s}{\sum_{s\in C}\mu_s q(\overline{C}\mid s)}. \tag{12}$$

Indeed, define the r.v. ρ_n as the average length of visits to C that end before stage $n+1$:

$$\rho_n = \frac{\sum_{p=1}^n 1_{s_p\in C}}{\sum_{p=1}^n 1_{s_p\in C}1_{s_{p+1}\notin C}}. \tag{13}$$

By the ergodic theorem, the sequence (ρ_n) converges, $\mathbf{P}_{s_1,q}$-a.s. to K_C, while the right hand side in (13) converges $\mathbf{P}_{s_1,q}$-a.s. to $\frac{\sum_{s\in C}\mu_s}{\sum_{s\in C}\mu_s q(\overline{C}|s)}$. The identity (12) follows.

By Proposition 16, for every $s\in C$,

$$(1-18\beta L)\mu_s < \widehat{\mu}_s < (1+18\beta L)\mu_s. \tag{14}$$

By Lemma 11,

$$\frac{1-\beta}{|S|^2}\sum_{s\in C}\mu_s q(\overline{C}|s) \le \sum_{s\in C}\mu_s\widehat{q}(\overline{C}|s) \le (1+\beta)|S|^2\sum_{s\in C}\mu_s q(\overline{C}|s). \tag{15}$$

Eqs. (14) and (15) yield

$$\frac{(1-18\beta L)(1-\beta)}{|S|^2}\sum_{s\in C}\mu_s q(\overline{C}|s) \le \sum_{s\in C}\widehat{\mu}_s\widehat{q}(\overline{C}|s) \le (1+\beta)(1+18\beta L)|S|^2\sum_{s\in C}\mu_s q(\overline{C}|s). \tag{16}$$

Summing up Eq. (14) over $s\in C$ gives

$$(1-18\beta L)\sum_{s\in C}\mu_s < \sum_{s\in C}\widehat{\mu}_s < (1+18\beta L)\sum_{s\in C}\mu_s. \tag{17}$$

The Proposition follows by dividing (17) by (16). ■

5 Proof of the variations

We here prove Theorems 4 and 6. We shall follow the previous proofs, and will point out which changes are needed. We let a,ε,β be given, that satisfy the assumptions of Theorem 4. The result of Section 4.1 still hold for every proper subset C of S_1, namely Lemmas 11, 12 and Corollary 13 are still valid, provided the assumption $C\subset S$ is replaced by the assumption $C\subset S_1$.

5.1 Proof of Theorem 4

We need the following observation.

Lemma 19 *For every $y \in S_1$, there exists a $(a/L)^{|S|}$-maximal graph $\overline{g} \in G(S_1 \backslash y)$ such that all paths of $\overline{g}$ lead to y.*

Proof. By **P2**, for every $s \in S_1 \backslash y$ there is a $\frac{a}{L}$-maximal $S_1 \backslash y$-graph g_s in which s leads to a state in y. Let h_s be the path in g_s that connects s to y (this is a set of edges). Let $\overline{g}$ be a $S_1 \backslash y$-graph that is contained in $\cup_{s \in S_1 \backslash y} h_s$. Then $\overline{g}$ satisfies the conditions. ■

We next prove the two assertions of Theorem 4.

Lemma 20 *All states of S_1 belong to the same recurrent set for $\widehat{q}$.*

Proof. It is enough to prove that for each $C \subset S_1$, there exists $s \in C$ and $t \in \overline{C}$ such that $\widehat{q}(t|s) > 0$. The proof of Proposition 14 still applies, provided ζ_q is replaced by ζ_q^1. ■

Lemma 21 *For each $s \in S_1$,*

$$\left| 1 - \frac{\widehat{\mu}(s|S_1)}{\mu(s|S_1)} \right| \leq 18\beta L.$$

Proof. The proof goes essentially as in Proposition 16. Set $\eta = \beta/(a/L)^{|S|} < (a/L)^S$, and fix $s \in S_1$. By (4),

$$\mu(s|S_1) = \frac{\sum_{G(S\backslash\{s\})} p(g)}{\sum_{y \in S_1} \sum_{G(S\backslash\{y\})} p(g)} \text{ and } \widehat{\mu}(s|S_1) = \frac{\sum_{G(S\backslash\{s\})} \widehat{p}(g)}{\sum_{y \in S_1} \sum_{G(S\backslash\{y\})} \widehat{p}(g)}.$$

For every $y \in S_1$, define $H_y = G^\eta(S\backslash\{y\}) \cup \widehat{G}^\eta(S\backslash\{y\})$. Define

$$\mu'(s|S_1) = \frac{\sum_{H_s} p(g)}{\sum_{y \in S_1} \sum_{H_y} p(g)} \text{ and } \widehat{\mu}'(s|S_1) = \frac{\sum_{H_s} \widehat{p}(g)}{\sum_{y \in S_1} \sum_{H_y} \widehat{p}(g)}.$$

Fix for a moment $y \in S_1$. By Lemma 19 there is a $(a/L)^{|S|}$-maximal $S_1\backslash\{y\}$-graph $\overline{g}$ such that all its paths lead to y. Let $g \in G^\eta(S\backslash\{y\})$, and $g_{S_1\backslash\{y\}}$, $g_{S\backslash S_1}$ its restrictions to $S_1\backslash\{y\}$ and $S\backslash S_1$. Using the above remark, the graph $\overline{g} \cup g_{S \setminus S_1}$ is a $S \setminus \{y\}$-graph. Therefore, $g_{S_1 \setminus \{y\}}$ is $\eta(a/L)^{|S|}$-maximal ($= \beta$-maximal). By Corollary 13 one has $\left|1 - \frac{\widehat{p}(g_{S_1 \setminus \{y\}})}{p(g_{S_1 \setminus \{y\}})}\right| < (|S| + 1)\beta$. Since q and $\widehat{q}$ coincide outside S_1, $p(g_{S \setminus S_1}) = \widehat{p}(g_{S \setminus S_1})$. Thus, $\left|1 - \frac{\widehat{p}(g)}{p(g)}\right| < (|S| + 1)\beta$. Lemma 10 and Lemma 15 implies that $\left|\frac{\mu(s|S_1)}{\mu'(s|S_1)} - 1\right| < 3\beta L$ and $\left|\frac{\widehat{\mu}(s|S_1)}{\widehat{\mu}'(s|S_1)} - 1\right| < 3\beta L$. By Lemmas 10 and 15, $\left|\frac{\widehat{\mu}'(s|S_1)}{\mu'(s|S_1)} - 1\right| < 3(|S| + 1)\beta$. Since $L \geq |S| \geq 2$, the Lemma follows by Lemma 15(2). ■

5.2 Proof of Theorem 6

The proof of the first two assertions in Theorem 6 is identical to the proof of the two assertions in Theorem 5 (see Propositions 17 and 18). We omit it. We now prove the last assertion.

Proposition 22 *One has*

$$\frac{1}{c} K_{S_1} \leq \widehat{K}_{S_1} \leq c K_{S_1} \text{ or } K_{S_1}, \widehat{K}_{S_1} \geq \frac{1}{2\varepsilon\,|S|} \times \frac{\mu_{S_1}}{\zeta_q^1}.$$

Proof. Fix $s \in S_1$. By (12),

$$K_{S_1} = \frac{1}{\sum_{t \in S_1} \mu(t|S_1) q(\overline{S_1} \mid t)},$$

and a similar equality holds for $\widehat{K}_{S_1}$, involving $\widehat{\mu}$ and $\widehat{q}$. By Theorem 4(2) and Lemma 15, the ratio between $\widehat{K}_{S_1}$ and $\frac{1}{\sum_{t \in S_1} \mu(t|S_1)\widehat{q}(\overline{S_1}|t)}$ is between $1 - 54\beta L$ and $1 + 54\beta L$.

If for every $t \in S_1$ and $u \notin S_1$, $\mu_t q(u \mid t) < \varepsilon\zeta_q^1$ and $\mu_t \widehat{q}(u \mid t) < \varepsilon\zeta_q^1$, then $K_{S_1} \geq \frac{\mu_{S_1}}{|S|^2 \varepsilon \zeta_q^1}$ and $\widehat{K}_{S_1} \geq (1 - 54\beta L) \times \frac{\mu_{S_1}}{|S|^2 \varepsilon \zeta_q^1}$, as desired.

If, on the other hand, there exist $t \in S_1$ and $u \notin S_1$ such that $\mu_t q(u \mid t) \geq \varepsilon\zeta_q^1$ or $\mu_t \widehat{q}(u \mid t) \geq \varepsilon\zeta_q^1$ then, since q and $\widehat{q}$ are (ε, β)-close, $|1 - \frac{\widehat{q}(u|t)}{q(u|t)}| \leq \beta$, and therefore $\mu_t q(u \mid t) \geq (1 - \beta)\varepsilon\zeta_q^1$ and $\mu_t \widehat{q}(u \mid t) \geq (1 - \beta)\varepsilon\zeta_q^1$. For every $t \in S_1$ and $u \notin S_1$ such that $\mu_t q(u \mid t) < \varepsilon\zeta_q^1$ and $\mu_t \widehat{q}(u \mid t) < \varepsilon\zeta_q^1$ we have $\mu_t q(u \mid t) \leq \sum_{t \in S_1} \mu_t q(\overline{S}_1 \mid t)$ and $\mu_t \widehat{q}(u \mid t) \leq \sum_{t \in S_1} \mu_t \widehat{q}(\overline{S}_1 \mid t)$. It follows that the ratio between $\sum_{t \in S_1} \mu_t q(\overline{S}_1 \mid t)$ and $\sum_{t \in S_1} \mu_t \widehat{q}(\overline{S}_1 \mid t)$ is at most $|S|^2$. The result follows. ■

References

[1] O. Catoni. Simulated annealing algorithms and Markov chains with rare transitions. In M. Ledoux, M. Yor, J. Azéma, M. Emery (eds), *Séminaire de Probabilités XXXIII*, pages 69–119. Springer, 1999.

[2] P. Diaconis and L. Saloff-Coste. Comparison theorems for reversible Markov chains. *The Annals of Applied Probability*, 3:696–730, 1993.

[3] M. Freidlin and A. Wentzell. *Random Perturbations of Dynamical Systems.* Springer, Berlin, 1984.

[4] J.J. Hunter. Stationary distributions of perturbed Markov chains. *Linear Algebra and its applications*, 82:201–214, 1986.

[5] M. Jerrum and A. Sinclair. Approximating the permanent. *SIAM Journal on Computing*, 18:1149–1178, 1989.

[6] J.B. Lasserre. Formula for singular perturbations of Markov chains. *Journal of Applied Probability*, 31:829–833, 1994.

[7] L. Lovasz and R. Kannan. Faster mixing via average conductance. *Annual ACM Symposium on Theory of Computing*, 282–287, 1999.

[8] L. Lovasz and M. Simonovits. The mixing rate of Markov chains, an isoperimetric inequality, and computing the volume. *31st Annual Symposium on Foundations of Computer Science, Vol. I, II*, 346–354.

[9] D. Rosenberg, E. Solan and N. Vieille. Stochastic games with imperfect monitoring. mimeo, 2002.

[10] P.J. Schweizer. Perturbation theory and finite Markov chains. *Journal of Applied Probability*, 5:401–413, 1968.

[11] E. Seneta. Perturbation of the stationary distribution measured by ergodicity coefficients. *Advances in Applied Probability*, 20:228–230, 1988.

[12] E. Seneta. Sensitivity of finite Markov chains under perturbations. *Statistics and Probability Letters*, 17:163–168, 1993.

[13] E. Solan and N. Vieille. Correlated equilibrium in stochastic games. *Games and Economic Behavior*, 38:362–399, 2002.

[14] N. Vieille. Small perturbations and stochastic games. *Israel Journal of Mathematics*, 119:127–142, 2000.