**ASYMPTOTICALLY EFFICIENT ADAPTIVE STRATEGIES
IN REPEATED GAMES,
PART I: CERTAINTY EQUIVALENCE STRATEGIES**

By

Nahum Shimkin

IMA Preprint Series # 1092

December 1992

Asymptotically Efficient Adaptive Strategies in Repeated Games, Part I: Certainty Equivalence Strategies

Nahum Shimkin¹

Institute for Mathematics and its Applications
University of Minnesota
Minneapolis, MN 55455

Abstract

We consider the problem of dynamic decision making in an uncertain environment while facing an opponent. Using the simplified model of a repeated matrix game with incomplete information, we concentrate on a worst-case analysis for an uninformed decision maker (DM1), and rely on ideas from non-Bayesian stochastic adaptive control to formulate our problem. Rewards are random, and perfect observations of both actions and rewards are assumed. Performance is measured in terms of the *total* reward, and accordingly the purpose of DM1 is to guarantee for himself a total reward which is as close as possible to what he could guarantee under complete information. This paper considers adaptive decision strategies of the Certainty Equivalence type, based on a Maximum Likelihood estimator, and studies their asymptotic performance. A sequel paper will be devoted to “asymptotically optimal” strategies.

Key words: repeated matrix games, incomplete information, adaptive control, total reward criterion.

¹Part of this research was performed while the author was with the department of Electrical Engineering at the Technion – Israel Institute of Technology.

1 Introduction

This paper studies certain aspects of dynamic decision making under uncertainty and competition, and extends some ideas from stochastic adaptive control to this setting. Basically, we place ourselves in the position of a decision maker (DM1) facing a dynamic system, on which he has incomplete initial information, and which is also being affected by others (DM2). Taking a worst-case view, we assume that DM2 is doing his best to obstruct us. We also assume that DM1 is non-Bayesian, in the sense that no prior probabilities are prescribed on his initial uncertainty. Learning is however facilitated by observing the system response over time.

We consider the simplest system dynamics, consisting of a fixed one-stage decision problem which is repeated in time. More specifically, a two-person matrix game with random rewards (payoffs) is played repeatedly, so the probability distribution which determines the reward (to DM1²) at a given stage depends only on the actions of DM1 and DM2 at this stage. However, DM1's *prior* information regarding this game is incomplete: While he knows that the game matrix (whose entries are the probability distributions of the rewards) is one of a finite specified set, he does not know which one. We shall parameterize this set by a parameter θ , and refer interchangeably to the "true matrix game" or the "true parameter" θ_0 . Note that the choices of DM1 and DM2 in each matrix game are termed actions, while 'strategy' is reserved for the overall decision rule in the repeated game.

This model belongs in the general class of repeated games with incomplete information. These games offer a convenient model for the isolation and examination of various aspects of learning and information manipulation in dynamic conflict situations. They have been extensively studied within the classical Bayesian framework for incomplete-information games, typically under the limiting expected average payoff criterion. See, e.g., [20,23,11] for surveys of this field.

In this paper we focus on the (long-run) *total* reward, which is a refinement of the average criterion. This, together with the non-Bayesian setting, distinguishes the present work from the mainstream of existing research on repeated games with incomplete information. Furthermore, we are concerned only with the special case where the system history (actions and rewards) is perfectly observed. The approach taken here is closer in spirit to that of non-Bayesian adaptive control ([12,13,15]), and is indeed based on certain ideas and methods from that field.

Let us next introduce the performance criterion for DM1. Generally speaking, DM1 wishes to secure a large total reward, for each and every possible value of the unknown parameter. To highlight the dependence on the unknown parameter and the role of learning, it will be convenient to express this goal in terms of the *relative loss* with respect to a complete-information performance level. Let the number of stages n and the true parameter θ_0 be fixed. Note that if DM1 knew θ_0 , then after n stages he could guarantee an expected total reward of $nv(\theta_0)$, where $v(\theta_0)$ is the (minimax) value of the true matrix game. Obviously, the best he can hope for in the incomplete information game is to come near to this level. Define then the relative loss as the dif-

²Since we concentrate on a worst-case analysis for DM1, the reward to DM2 is immaterial. It will however be convenient to think of the game as being zero-sum.

ference between $nv(\theta_0)$ and the expected total reward, and let the worst-case relative loss (WCRL) be the maximal relative loss over all strategies of DM2. The long run performance of DM1 may now be measured in terms of the asymptotic rate of increase of the WCRL (e.g. linear, sub-linear, logarithmic, bounded, etc.). An efficient strategy for DM1 must guarantee a low rate of increase for *all* possible values of the true game parameter.

Note that the WCRL criterion is a refinement of the expected *average* reward criterion, and in fact supplied rates of convergence for the later. Indeed, an “ideal performance” with respect to the average criterion requires only that the WCRL be $o(n)$; better results, namely $O(\log n)$, will be obtained here. We note that if rewards are not observed, then a convergence rate of $O(n^{-1/2})$ for the average reward to the value of the game (in a Bayesian setting) cannot be generally improved upon ([25]). Other relevant results concerning the average criterion may be found in of [5] and [19], which consider a completely uninformed non-Bayesian player and non-random rewards, and also in [17]. For a related problem of (incomplete information) statistical games against nature see [24] and references therein.

What are the strategic problems that confront DM1? Obviously, he may use the observed game history to estimate, in a statistical sense, the true game parameter. We are then confronted with the dual role of control: actions which are good for (statistical) information acquisition may be inefficient in terms of rewards, and vice-versa. On top of this, the effect of DM2 on both aspects should be considered. In particular, the information content of the observations depends also on DM2’s actions, and DM1 may not be able to guarantee consistent estimation of the true game parameter. This implies that DM1 cannot isolate the problem of information acquisition (probing), as is possible in comparable adaptive control problems ([13,15,16,1]). Instead, he should rely on inter-relations between information and reward, in trying to make sure that, no matter what DM2 may do, low information will be compensated by high enough reward as to make this information immaterial. This theme will be dominant in the following analysis.

Our formulation of the total reward criterion is closely related to that of [16], where a theory of asymptotic total reward optimality is presented for the statistical Bandit Problem. These results have been extended to various adaptive control models, e.g. [1,2,3]. In particular, our game model reduces to that of [1] if DM2 is removed. Asymptotic optimality results in that vein will be presented for the game model in a sequel paper [22].

In the present paper we shall focus on some relatively simple strategies, which are intuitively appealing and easy to implement, study their performance, and identify conditions under which they perform well. These strategies are based on the Certainty Equivalence principle, where a point estimator of the true parameter is substituted in a complete-information optimal control law. First, we consider a strategy which is based on the Maximum Likelihood Estimator (MLE). It turns out that performance is poor in general, in that the WCRL might increase at a linear rate. However, certain conditions are provided which ensure a *bounded* WCRL (which is of course the best conceivable “rate of increase”). These results are related to the “closed loop identification problem” which arises in the adaptive control of Markov chains ([13,15]). Motivated by an idea of [14], we then consider a Certainty Equivalence strategy with a modified estimator, the value-biased

MLE. Considerable improvement in performance is obtained: here the WCRL is $O(\log n)$ at the most and, moreover, a bounded WCRL is guaranteed under weaker conditions.

The paper is organized as follows. The next section describes the model. Section 3 develops some preliminary results required in the analysis. Sections 4 and 5 consider Certainty Equivalence strategies which are based on the MLE and value-biased MLE, respectively, and Section 6 is devoted to the proof of Theorem 5.2. The paper ends with some concluding remarks in Section 7.

Notations: Let $M = \{M(i, j)\}$ be an $\mathcal{I} \times \mathcal{J}$ matrix with real or vector entries, where $\mathcal{I}, \mathcal{J}$ are finite sets of actions. $\mathcal{P}(\mathcal{I})$ denotes the set of probability vectors over $\mathcal{I}$. An element $x = (x_i)$ in $\mathcal{P}(\mathcal{I})$ is referred to as a “mixed action”, and similarly for $y \in \mathcal{P}(\mathcal{J})$. Let

$$M[x, y] = \sum_{i,j} x_i M(i, j) y_j \quad (1.1)$$

denote the mean (expected) value of M with respect to the mixed actions x and y . The notation $M(i, y)$ and $M(x, j)$ will also be used, with similar interpretation. For positive sequences $\{f_n\}$ and $\{g_n\}$, we write $f_n = o(g_n)$ if $\limsup_n (f_n/g_n) = 0$, and $f_n \leq O(g_n)$ if $\limsup_n (f_n/g_n) < \infty$. Finally, $1\{\cdot\}$ denotes the indicator function.

2 The Model

The game model involves the following objects: (i) Two decision makers, DM1 and DM2. (ii) Finite sets $\mathcal{I}$ and $\mathcal{J}$ of actions for DM1 and DM2, respectively. (iii) A real, finite set $\mathcal{A}$ of possible rewards. (iv) Finite parameter set Θ . (v) A set $\{G(\theta) : \theta \in \Theta\}$ of $\mathcal{I} \times \mathcal{J}$ matrix games with random rewards. Each game $G(\theta)$ is specified by a collection $\{p_{\theta,i,j}(\cdot) : i \in \mathcal{I}, j \in \mathcal{J}\}$ of probability distributions over the reward set $\mathcal{A}$.

The repeated game Γ_∞ proceeds as follows. At stage 0, Nature chooses an element $\theta_0 \in \Theta$ (“the true parameter”). This choice is told to DM2, but not to DM1. Then, at each stage $n = 1, 2, \dots$: DM1 chooses an action $i_n \in \mathcal{I}$ and DM2 simultaneously chooses an action $j_n \in \mathcal{J}$. A reward $a_n \in \mathcal{A}$ is then randomly chosen according to the probability distribution p_{θ_0, i_n, j_n} . At the end of each stage, both decision makers observe the actions (i_n, j_n) and the reward a_n . Perfect recall of past information is assumed. Rewards accumulate to form the total n -stage reward $s_n = \sum_{t=1}^n a_t$.

A strategy for each decision maker is a (possibly randomized) rule for choosing his actions at each stage. Since perfect recall is assumed, it follows by the Kuhn-Aumann theorem ([4]) that one can restrict attention (at least as far as the reward sequence distribution is concerned) to *behavioral* strategies, where randomizations are performed independently at each stage. Formally, a (behavioral) strategy for DM1 is a collection of maps $\sigma_n : H_n \rightarrow \mathcal{P}(\mathcal{I})$, $n = 1, 2, \dots$, where H_n denotes the set of all possible observed histories $h_n = \{i_m, j_m, a_m\}_{m=1}^{n-1}$ up to stage n . Thus $x_n := \sigma_n(h_n)$ is the randomized action of DM1 at stage n . Strategies of DM2 are defined similarly except that they are allowed to depend explicitly on the true parameter θ_0 . A strategy of DM2 is denoted by τ , his strategy set by $\mathcal{T}$, and his randomized action at stage n is denoted y_n . Let $P_{\theta_0}^{\sigma, \tau}$

and $E_{\theta_0}^{\sigma, \tau}$ denote the probability measure and the expectation induced by the a triplet (θ_0, σ, τ) on the sample space H_∞ .

For each matrix game $G(\theta)$, let $\bar{A}_\theta(i, j) = \sum_{a \in \mathcal{A}} a p_{\theta, i, j}(a)$ denote the expected reward given actions i and j . The (minimax) value of $G(\theta)$ is then given by

$$v(\theta) = \max_{x \in \mathcal{P}(\mathcal{I})} \min_{y \in \mathcal{P}(\mathcal{J})} \bar{A}_\theta[x, y], \quad (2.1)$$

where the notation (1.1) is used. We define the *one-stage relative loss* as

$$d_\theta(i, j) = v(\theta) - \bar{A}_\theta(i, j). \quad (2.2)$$

Note that $d_\theta(i, j)$ may be negative.

Let $X^*(\theta)$ denote the (closed convex) set of optimal (maximin) randomized actions of DM1 in $G(\theta)$, and let $Y^*(\theta)$ denote the set of optimal (minimax) randomized actions of DM2. Let $\mathcal{I}^*(\theta) \subset \mathcal{I}$ be the set of relevant actions for DM1 in $G(\theta)$, i.e. the set of actions which are given strictly positive probability by some $x^* \in X^*(\theta)$. Similarly define $\mathcal{J}^*(\theta)$ as the set of relevant actions for DM2. Let $x^*(\theta)$ be an arbitrary point in the relative interior of $X^*(\theta)$, which we fix for the rest of the paper. Note that $i \in \mathcal{I}^*(\theta)$ if and only if $x^*(\theta)_i > 0$. It follows from Theorems 3.1.2 and 3.1.16 in [21] that $\mathcal{J}^*(\theta)$ is exactly the set of actions which minimize the expected reward against $x^*(\theta)$, that is

$$d_\theta[x^*(\theta), j] := v(\theta) - \bar{A}_\theta[x^*(\theta), j] = 0 \quad \text{for } j \in \mathcal{J}^*(\theta), \quad (2.3)$$

$$d_\theta[x^*(\theta), j] < 0 \quad \text{for } j \notin \mathcal{J}^*(\theta). \quad (2.4)$$

An important quantity to be used in the sequel is the *information divergence* between $p_{\theta, i, j}$ and $p_{\theta', i, j}$, defined by

$$\hat{I}_{\theta, \theta'}(i, j) = \sum_{a \in \mathcal{A}} p_{\theta, i, j}(a) \log \frac{p_{\theta, i, j}(a)}{p_{\theta', i, j}(a)} \quad (2.5)$$

(where $0 \log 0 \triangleq 0$). The information divergence, also known as the cross-entropy or Kullback-Leibler distance, is a well known measure of statistical distinguishability between probability distributions, and arises naturally as the expected log-likelihood ratio. It is easily verified ([9]) that $\hat{I}_{\theta, \theta'}(i, j) \geq 0$, and $\hat{I}_{\theta, \theta'}(i, j) = 0$ if and only if $p_{\theta, i, j} = p_{\theta', i, j}$. Since $\hat{I}$ may be infinite, it will be convenient to define a “truncated” version:

$$I_{\theta, \theta'}(i, j) = \sum_{a \in \mathcal{A}} p_{\theta, i, j}(a) \min\{M_o, \log \frac{p_{\theta, i, j}(a)}{p_{\theta', i, j}(a)}\}. \quad (2.6)$$

Here $M_o > 0$ is a large enough constant so that $\hat{I}_{\theta, \theta'}(i, j) > 0$ implies $I_{\theta, \theta'}(i, j) > 0$ (such a constant obviously exists since the sets $\Theta, \mathcal{I}, \mathcal{J}, \mathcal{A}$ are finite). It follows that $0 \leq I_{\theta, \theta'}(i, j) \leq M_o$, and

$$I_{\theta, \theta'}(i, j) = 0 \quad \text{if and only if } p_{\theta, i, j} = p_{\theta', i, j}. \quad (2.7)$$

We now define the performance measure for DM1. For each strategy pair (σ, τ) , $\theta_0 \in \Theta$ and $n \geq 1$ define the *relative loss*:

$$L_n^{\sigma, \tau}(\theta_0) = E_{\theta_0}^{\sigma, \tau} \left(n v(\theta_0) - \sum_{t=1}^n a_t \right). \quad (2.8)$$

Note that $nv(\theta_0)$ is the highest n -stage expected reward which DM1 could guarantee if he knew the true parameter θ_0 . Thus, $L_n^{\sigma, \tau}(\theta_0)$ is the deficiency in the expected total reward as compared with this “ideal” performance. Maximizing over all strategies of DM2 yields the *worst case relative loss* (WCRL):

$$L_n^{\sigma}(\theta_0) = \max_{\tau} L_n^{\sigma, \tau}(\theta_0). \quad (2.9)$$

As motivated in the introduction, the WCRL will be used as a performance measure for DM1.

We note that the relative loss may also be written as:

$$L_n^{\sigma, \tau} = E_{\theta_0}^{\sigma, \tau} \sum_{t=1}^n d_{\theta_0}(i_t, j_t) = E_{\theta_0}^{\sigma, \tau} \sum_{t=1}^n d_{\theta_0}[x_t, j_t]. \quad (2.10)$$

These relations follow by using the definition (2.2) of d_{θ_0} and applying appropriate conditioning on each term separately.

3 Preliminaries: Controlled I.I.D. Processes

This section develops some results which will be used, in essence, to bound the deviation of the the log-likelihood ratio from its (conditional) mean. These results are derived within a general “controlled i.i.d. process”, whose exact relation to the repeated game model is indicated in Lemma 3.1 below.

We consider the following controlled process, which is similar to a one-player version of the repeated game. Let U denote the action space, and Z the state space. For each $u \in U$, let $q(\cdot|u) \in \mathcal{P}(Z)$ be a probability distribution over Z , and let $r : Z \rightarrow \mathbb{R}$ be the reward function. At each stage $n \geq 1$, the controller chooses an action $u_n \in U$, the state $z_n \in U$ is randomly chosen according to $q(\cdot|u_n)$, and a one-stage reward $r_n = r(z_n)$ is collected. A control policy π is a sequence of mappings $\pi_n : H_n \rightarrow \mathcal{P}(U)$, which associate with each observed history sequence $h_n = (u_1, z_1, \dots, u_{n-1}, z_{n-1})$ a probability distribution over U . (Since the sets U and Z are not necessarily finite, some measurability conditions are required for the above description to induce a well defined stochastic process. Thus, assume that U, Z are measurable sets, $q : U \rightarrow \mathcal{P}(Z)$ and $\pi_n : H_n \rightarrow \mathcal{P}(U)$ are transition probabilities, where H_n is endowed with the product σ -algebra, and r is a Borel-measurable function.) Let Π denote the set of control policies.

Let R_u denote a random variable distributed as the one-stage reward under action $u \in U$ (i.e. $R_u = r(Z_u)$ where $Z_u \sim q(\cdot|u)$), and let $\bar{R}(u)$ denote its expected value: $\bar{R}(u) = E(R_u) = \int_Z r(z)q(dz|u)$. Define also the total reward $S_n = \sum_{t=1}^n r_t$. The following conditions are imposed throughout:

Assumption 3.1

- (i) The reward function r is bounded: $|r(z)| \leq \hat{R}$, $z \in Z$.
- (ii) $\bar{R}(u) \geq 0$, $u \in U$.
- (iii) There exists a constant $C_o > 0$ such that $E(R_u^2) \leq C_o \bar{R}(u)$ for every $u \in U$.

Here item (ii) requires that the expected reward is non-negative for each action u . The essence of (iii) is that if $\bar{R}(u)$ is small, then so is (the second moment of) R_u .

The next lemma and remark summarize the correspondence which will be used between this model and the game model.

Lemma 3.1 *Let $\theta_0, \theta \in \Theta$ be fixed parameters. Consider the following definition:*

$$u_n = (x_n, j_n), \quad z_n = (i_n, j_n, a_n),$$

$$r(z) = \min\{M_o, \log \frac{p_{\theta_0, i, j}(a)}{p_{\theta, i, j}(a)}\} \quad \forall z \equiv (i, j, a)$$

where M_o is the same as in (2.6), and we arbitrarily define $r(z) = 0$ if $p_{\theta_0, i, j}(a) = 0$. Then Assumption 3.1 is satisfied.

Remark: Under this definition, we have

$$S_n = \sum_{t=1}^n \min\{M_o, \log \frac{p_{\theta_0, i, j}(a)}{p_{\theta, i, j}(a)}\} \triangleq \tilde{\Lambda}_n(\theta_0, \theta),$$

(the “truncated log-likelihood ratio”), and $\bar{R}(u) = \sum_i x_i I_{\theta, \theta'}(i, j) = I_{\theta, \theta'}[x, j]$ for every $u \equiv (x, j)$. We note that the choice $U = \mathcal{P}(\mathcal{I}) \times \mathcal{J}$, instead of just $U = \mathcal{I} \times \mathcal{J}$, is essential since we shall need to bound expressions which depend directly on $I_{\theta_0, \theta}[x_n, j_n]$.

Proof: Since $Z = \mathcal{I} \times \mathcal{J} \times \mathcal{A}$ is a finite set, then $r(z)$ is obviously bounded. Also, $\bar{R}(u) = I_{\theta_0, \theta}[x, j] \geq 0$. It therefore remains only to establish part (iii) of Assumption 3.1. Consider first the finite set $U_o = \{(\delta_i, j) : i \in \mathcal{I}, j \in \mathcal{J}\}$, where $\delta_i \in \mathcal{P}(\mathcal{I})$ is the randomized action which always chooses i . We claim that for every $u \in U_o$ there exists a constant $C(u) > 0$ such that

$$E(R_u^2) \leq C(u) \bar{R}(u). \quad (3.1)$$

Indeed, let $u = (\delta_i, j) \in U_o$, and note that $\bar{R}(u) = I_{\theta_0, \theta}(i, j) \geq 0$. If $\bar{R}(u) = 0$, then (2.7) implies that $R_u = 0$ w.p. 1, so that (3.1) hold trivially for any $C(u) > 0$. Otherwise, if $\bar{R}(u) > 0$, then (3.1) obviously holds for $C(u)$ large enough, (note that $E(R_u^2) \leq \hat{R}^2 < \infty$). Hence, since U_o is finite, there exists a constant $C_o > 0$ such that $E(R_u^2) \leq C_o \bar{R}(u)$ for every $u \in U_o$. But then, for every $u = (x, j) \in U$,

$$E(R_u^2) = \sum_i x_i E(R_{(\delta_i, j)}^2) \leq C_o \sum_i x_i \bar{R}((\delta_i, j)) = C_o \bar{R}(u). \quad \square$$

The next lemma is required in the proof of Lemma 3.3, which is the main result of this section.

Lemma 3.2 *There exist constants $\lambda > 0, \mu > 0$ such that:*

$$(i) \quad E e^{-\lambda R_u} \leq 1 - \mu \bar{R}(u), \quad u \in U, \quad (3.2)$$

$$(ii) \quad \sup_{\pi \in \Pi} P_\pi\{S_n \leq -K\} \leq e^{-\lambda K}, \quad \forall K \geq 0. \quad (3.3)$$

Proof:

(i) For every $u \in U$ we have $E(e^{-\lambda R_u})|_{\lambda=0} = 1$,

$$\frac{d}{d\lambda} E(e^{-\lambda R_u})|_{\lambda=0} = -E(R_u) = -\bar{R}(u) < 0 \quad (3.4)$$

and, by Assumption 3.1(i, iii),

$$\begin{aligned} \frac{d^2}{d\lambda^2} E(e^{-\lambda R_u}) &= E(R_u^2 e^{-\lambda R_u}) \leq E(R_u^2 e^{|R_u|}) \\ &\leq e^{\hat{R}} E(R_u^2) \leq (C_o e^{\hat{R}}) \bar{R}(u), \quad \forall \lambda \in [0, 1]. \end{aligned} \quad (3.5)$$

Therefore, a second-order power series expansion of $E(e^{-\lambda R_u})$ around $\lambda = 0$ yields

$$E(e^{-\lambda R_u}) \leq 1 - \lambda \bar{R}(u) + \frac{\lambda^2}{2} (C_o e^{\hat{R}}) \bar{R}(u), \quad 0 \leq \lambda \leq 1, \quad (3.6)$$

and the result follows with $\lambda = (C_o e^{\hat{R}})^{-1}$, $\mu = \lambda/2$.

(ii) Let $\lambda > 0$ be the constant for which (3.2) holds. Since $\mathbf{1}\{\beta \leq 0\} \leq e^{-\lambda\beta}$ for every $\beta \in \mathbb{R}$,

$$P_\pi\{S_n \leq -K\} \leq E_\pi e^{-\lambda(S_n+K)} = e^{-\lambda K} E_\pi\left(\prod_{t=1}^n e^{-\lambda r_t}\right). \quad (3.7)$$

Now, by standard Dynamic Programming arguments applied to multiplicative cost functionals (e.g. [6, p.66]), it follows that

$$E_\pi\left(\prod_{t=1}^n e^{-\lambda r_t}\right) \leq [\sup_u E(e^{-\lambda R_u})]^n \leq 1 \quad (3.8)$$

where the last inequality follows from (3.2). $\square$

Lemma 3.3 *Let $\{\beta_n\}$ be an $o(n)$ positive non-decreasing sequence such that $\beta_n \rightarrow \infty$. Let $\bar{S}_n = \sum_{t=1}^n \bar{R}(u_t)$. Then*

(i) *There exists a constant $Q < \infty$ such that: $\sup_\pi E_\pi \left(\sum_{t=1}^\infty \bar{R}(u_t) \mathbf{1}\{S_{t-1} \leq 0\} \right) \leq Q \leq \infty$.*

(ii) $\limsup_{n \rightarrow \infty} \frac{1}{\beta_n} \sup_\pi E_\pi \left(\sum_{t=1}^n \bar{R}(u_t) \mathbf{1}\{S_{t-1} \leq \beta_n\} \right) \leq 1$.

(iii) $\sum_{n=1}^\infty \sup_\pi P_\pi\{\bar{S}_n \geq \eta n, S_n \leq \beta_n\} < \infty, \quad \forall \eta > 0$.

(iv) $\sum_{n=1}^\infty \sup_\pi P_\pi \left\{ \sum_{t=1}^n \bar{R}(u_t) \mathbf{1}\{S_{t-1} \leq \beta_n\} \geq \eta n \right\} < \infty, \quad \forall \eta > 0$.

(v) For any $\alpha > 0$, $\epsilon > 0$, there exists a constant $Q = Q(\alpha, \epsilon) < \infty$ such that

$$E_\pi \sum_{t=1}^{\infty} \bar{R}(u_t) \mathbf{1}\{\bar{S}_{t-1} \geq (1 + \epsilon)\beta_t - \alpha, S_{t-1} \leq \beta_t\} \leq Q, \quad \forall \pi \in \Pi.$$

Proof:

(i) Let $\lambda > 0$, $\mu > 0$ be as in Lemma 3.2(i). Note that $\mathbf{1}\{S_{t-1} < 0\} \leq e^{-\lambda S_{t-1}}$, so it suffices to prove that the bound

$$J_1^n \triangleq \sup_{\pi} E \left\{ \sum_{t=1}^n \bar{R}(u_t) e^{-\lambda S_{t-1}} \right\} \leq Q \quad (3.9)$$

holds for some $Q < \infty$ and every n . Fixing $n > 1$, define for every $1 < m \leq n$:

$$J_m^n = \sup_{\pi} E_{\pi} \left\{ \sum_{t=m}^n \bar{R}(u_t) \exp(-\lambda S_m^{t-1}) \right\} \quad (3.10)$$

where $S_m^{t-1} = \sum_{k=m}^{t-1} r_k$ for $t > m$, and $S_m^{m-1} = 0$. Further define $J_{n+1}^n = 0$. Then, for every $1 \leq m \leq n$,

$$\begin{aligned} J_m^n &= \sup_{\pi} E_{\pi} \left\{ \bar{R}(u_m) + e^{-\lambda r_m} \sum_{t=m+1}^n \bar{R}(u_t) \exp(-\lambda S_{m+1}^{t-1}) \right\} \\ &= \sup_{\pi, u} E_{\pi} \left\{ \bar{R}(u) + e^{-\lambda R_u} \sum_{t=m+1}^n \bar{R}(u_t) \exp(-\lambda S_{m+1}^{t-1}) \right\} \\ &= \sup_u \{ \bar{R}(u) + E(e^{-\lambda R_u}) J_{m+1}^n \} \end{aligned} \quad (3.11)$$

(which is in essence just the optimality principle of Dynamic Programming). Hence, by Lemma 3.2(i),

$$J_m^n \leq \sup_u \{ \bar{R}(u) + (1 - \mu \bar{R}(u)) J_{m+1}^n \} = J_{m+1}^n + \sup_u \{ \bar{R}(u)(1 - \mu J_{m+1}^n) \}. \quad (3.12)$$

Since $0 \leq \bar{R}(u) \leq \hat{R}$, it follows that $J_m^n \leq J_{m+1}^n + \hat{R}$, and also that $J_m^n = J_{m+1}^n$ if $J_{m+1}^n \geq \mu^{-1}$. Since $J_{n+1}^n = 0$, this clearly implies that $J_m^n \leq \hat{R} + \mu^{-1}$, and in particular $J_1^n \leq \hat{R} + \mu^{-1}$.

(ii) Let $0 < \epsilon < 1$ be fixed. Then, recalling that $0 \leq \bar{R}(u) \leq \hat{R}$,

$$\begin{aligned} J_n &\triangleq \sum_{t=1}^n \bar{R}(u_t) \mathbf{1}\{S_{t-1} \leq \beta_n\} \\ &= \sum_{t=1}^n \bar{R}(u_t) \mathbf{1}\{S_{t-1} \leq \beta_n, S_{t-1} > (1 - \epsilon)\bar{S}_{t-1}\} \\ &\quad + \sum_{t=1}^n \bar{R}(u_t) \mathbf{1}\{S_{t-1} \leq \beta_n, S_{t-1} \leq (1 - \epsilon)\bar{S}_{t-1}\} \\ &\leq \sum_{t=1}^n \bar{R}(u_t) \mathbf{1}\{\bar{S}_{t-1} < \frac{\beta_n}{1 - \epsilon}\} + \sum_{t=1}^n \bar{R}(u_t) \mathbf{1}\{S_{t-1} \leq (1 - \epsilon)\bar{S}_{t-1}\} \\ &\leq \frac{\beta_n}{1 - \epsilon} + \hat{R} + \sum_{t=1}^n \bar{R}(u_t) \mathbf{1}\{S_{t-1} \leq (1 - \epsilon)\bar{S}_{t-1}\}. \end{aligned} \quad (3.13)$$

In order to bound the last term, define a modified reward function: $r'(z) = r(z) - (1 - \epsilon)\bar{R}(u)$. Denoting all quantities related to r' by a prime, it follows that $\bar{R}'(u) = \epsilon\bar{R}(u)$, and $S'_t = S_t - (1 - \epsilon)\bar{S}_t$. Note further that the model assumptions are satisfied for this modified reward function. In particular, r' is bounded, and

$$E(R'_u)^2 \leq E(R_u^2) \leq C_o \bar{R}(u) = (\epsilon^{-1}C_o)\bar{R}'(u). \quad (3.14)$$

Therefore, by item (i) of the present lemma,

$$E_\pi \sum_{t=1}^n \bar{R}(u_t) \mathbf{1}\{S_{t-1} \leq (1 - \epsilon)\bar{S}_{t-1}\} = \epsilon^{-1} E_\pi \sum_{t=1}^n \bar{R}'(u_t) \mathbf{1}\{S'_{t-1} \leq 0\} \leq \epsilon^{-1} Q(\epsilon). \quad (3.15)$$

Combining (3.13) and (3.15) gives

$$\limsup_{n \rightarrow \infty} \frac{1}{\beta_n} \sup_{\pi} E_\pi(J_n) \leq \limsup_{n \rightarrow \infty} \frac{1}{\beta_n} \left(\frac{\beta_n}{1 - \epsilon} + \hat{R} + \epsilon^{-1} Q(\epsilon) \right) = \frac{1}{1 - \epsilon}. \quad (3.16)$$

The required result then follows by letting $\epsilon \rightarrow 0$.

(iii) Define, as in the proof of (ii), a modified reward function $r'(z) = r(z) - \frac{1}{2}\bar{R}(u)$. Then, for every $\pi \in \Pi$,

$$\begin{aligned} q_n(\pi) &\triangleq P_\pi \left\{ \bar{S}_n \geq \delta n, S_n \leq \beta_n \right\} \leq P_\pi \left\{ S_n - \frac{1}{2}\bar{S}_n \leq \beta_n - \frac{1}{2}\delta n \right\} \\ &\leq P_\pi \left\{ S'_n \leq \left(\beta_n - \frac{1}{2}\delta n \right) \right\} \leq \exp \left(-\lambda' \left(\frac{1}{2}\delta n - \beta_n \right) \right) \triangleq \alpha_n \end{aligned} \quad (3.17)$$

holds for some $\lambda' > 0$, where the last step follows by Lemma 3.2(ii). Since $\beta_n = o(n)$, then $\{\alpha_n\}$ is summable and (iii) follows.

(iv) For every $\pi \in \Pi$,

$$\begin{aligned} q'_n(\pi) &\triangleq P_\pi \left\{ \sum_{t=1}^n \bar{R}(u_t) \mathbf{1}\{S_{t-1} \leq \beta_n\} \geq \delta n \right\} \\ &\leq P_\pi \left\{ \exists m, 1 \leq m \leq n, \text{ s.t. } \sum_{t=1}^m \bar{R}(u_t) \geq \delta n, S_{m-1} \leq \beta_n \right\} \\ &\leq \sum_{m=1}^n P \left\{ \sum_{t=1}^m \bar{R}(u_t) \geq \delta n, S_{m-1} \leq \beta_n \right\} \leq n \alpha_n, \end{aligned} \quad (3.18)$$

where the last inequality follows exactly as in (3.17), with α_n as defined there. Since the sequence $\{n\alpha_n\}$ is summable, the result follows.

(v) Define, as in the proof of (ii), a modified reward function: $r'(z) = (1 + \epsilon)r(z) - \bar{R}(u)$. Then

$$\begin{aligned} J_n &:= \sum_{t=1}^n \bar{R}(u_t) \mathbf{1}\{\bar{S}_{t-1} \geq (1 + \epsilon)\beta_t - \alpha, S_{t-1} \leq \beta_t\} \\ &\leq \sum_{t=1}^n \bar{R}(u_t) \mathbf{1}\{(1 + \epsilon)S_{t-1} - \bar{S}_{t-1} \leq \alpha\} \\ &= \epsilon^{-1} \sum_{t=1}^n \bar{R}'(u_t) \mathbf{1}\{S'_{t-1} \leq \alpha\}. \end{aligned} \quad (3.19)$$

Now $E_\pi J_n$ can be bounded by applying the proof of (i) to the process with this modified reward function. Indeed note that for any $\lambda > 0$,

$$E_\pi J_n \leq \epsilon^{-1} e^{\lambda \alpha} E_\pi \sum_{t=1}^n \overline{R}'(u_t) e^{-\lambda S'_{t-1}}.$$

Comparing this expression with (3.9), it follows that for some $\lambda' > 0$,

$$E_\pi J_n \leq \epsilon^{-1} e^{\lambda' \alpha} Q' := Q < \infty \quad \forall \pi, n.$$

□

4 Certainty Equivalence: MLE

The Maximum Likelihood Estimator (MLE) of θ_0 just prior to stage $n \geq 1$ is given by:

$$\hat{\theta}_n = \arg \max \{ \lambda_{n-1}(\theta) : \theta \in \Theta \}, \quad (4.1)$$

where

$$\lambda_{n-1}(\theta) = \prod_{t=1}^{n-1} p_{\theta, i_t, j_t}(a_t) \quad (4.2)$$

is the likelihood function. To define $\hat{\theta}_n$ uniquely, we assume that ties in (4.1) are decided according to some fixed ordering of Θ ; also, let $\lambda_o(\theta) \triangleq 1$. For every $\theta_0, \theta \in \Theta$, define the log-likelihood ratio:

$$\Lambda_n(\theta_0, \theta) = \log \frac{\lambda_n(\theta_0)}{\lambda_n(\theta)} = \sum_{t=1}^n \log \frac{p_{\theta_0, i_t, j_t}(a_t)}{p_{\theta, i_t, j_t}(a_t)}, \quad (4.3)$$

and the “truncated” log-likelihood ratio:

$$\tilde{\Lambda}_n(\theta_0, \theta) = \sum_{t=1}^n \min \{ M_o, \log \frac{p_{\theta_0, i_t, j_t}(a_t)}{p_{\theta, i_t, j_t}(a_t)} \}, \quad (4.4)$$

where $M_o > 0$ is the same constant as in (2.6). Note that, by definition of the MLE,

$$\hat{\theta}_n = \theta \implies \Lambda_n(\theta_0, \theta) \leq 0. \quad (4.5)$$

The following strategy $\sigma 1$ of DM1 will be considered in this section:

Strategy $\sigma 1$: $x_n = x^*(\hat{\theta}_n)$, where $\hat{\theta}_n$ is the MLE (4.1).

Thus, DM1 is using at each stage its optimal randomized action in the matrix game which corresponds to the estimated parameter. Control policies of this type, namely Certainty Equivalence with the MLE, have been well studied in the context of stochastic adaptive control (e.g. [18,7,8,13,15]). Performance of these schemes is often hampered by the *closed-loop identification problem*: the prescribed control signals may not be adequate to ensure sufficient excitation of the system

for identification purposes (even though consistent estimation of the unknown parameters may be possible by a different choice of control signals), and poor performance might result. These observations have led to two research directions. The first is to specify appropriate identifiability conditions on the model which ensure “optimal” performance, see [Ibid.]. The second is to consider modifications of the basic policy which alleviate the need for such conditions; this will be further discussed in the next section.

Returning to the strategy $\sigma 1$, we proceed here to formulate an identifiability-type condition which will be shown to guarantee a *bounded* WCRL for this strategy. The results are first stated and proved, and then discussed at the end of the section.

For each $\theta^\circ \in \Theta$, define the following conditions:

Condition $C_1(\theta^\circ)$: For every θ and j , $\bar{A}_{\theta^\circ}[x^*(\theta), j] < v(\theta^\circ)$ implies $I_{\theta^\circ, \theta}[x^*(\theta), j] > 0$.

Condition C_1 : $C_1(\theta^\circ)$ is satisfied for every $\theta^\circ \in \Theta$.

We then have the following result.

Theorem 4.1 *Assume that DM1 is using the strategy $\sigma 1$.*

- (i) *If $C_1(\theta_0)$ is satisfied, then $\limsup_{n \rightarrow \infty} L_n^{\sigma 1}(\theta_0) < \infty$.*
- (ii) *Thus, if C_1 is satisfied, then the WCRL is bounded for every $\theta_0 \in \Theta$.*

Proof: The idea of the proof is to upper-bound the relative loss by the information content of the data (quantified by the information divergence) over the times when the MLE is different from the true parameter (see (4.7), (4.8) below). To bound the latter, we rely on the observation that large information steers the estimator towards the true parameter.

Recalling (2.10), and noting that $x_t = x^*(\hat{\theta}_t)$ under $\sigma 1$, it follows that for every θ_0 and every strategy τ of DM2:

$$L_n^{\sigma 1, \tau}(\theta_0) = E_{\theta_0}^{\sigma 1, \tau} \sum_{t=1}^n d_{\theta_0}[x^*(\hat{\theta}_t), j_t]. \quad (4.6)$$

Consider a fixed θ_0 such that $C_1(\theta_0)$ is satisfied. Recalling the definition of d_θ in (2.2), $C_1(\theta_0)$ reads: For all θ and j , $d_{\theta_0}[x^*(\theta), j] > 0$ implies $I_{\theta_0, \theta}[x^*(\theta), j] > 0$. Since $I_{\theta_0, \theta}$ is non-negative and Θ as well as $\mathcal{J}$ are finite sets, it follows that for some $M < \infty$:

$$d_{\theta_0}[x^*(\theta), j] \leq M I_{\theta_0, \theta}[x^*(\theta), j], \quad \forall \theta, j. \quad (4.7)$$

Also, since $x^*(\theta_0)$ is optimal in the matrix game $G(\theta_0)$, then $d_{\theta_0}[x^*(\theta_0), j] \leq 0$ for every j . Using the last two bounds in (4.6) gives (with E standing for $E_{\theta_0}^{\sigma 1, \tau}$):

$$\begin{aligned} L_n^{\sigma 1, \tau}(\theta_0) &= E \sum_{t=1}^n \sum_{\theta \in \Theta} d_{\theta_0}[x^*(\theta), j_t] \mathbf{1}\{\hat{\theta}_t = \theta\} \\ &\leq E \sum_{t=1}^n \sum_{\theta \neq \theta_0} d_{\theta_0}[x^*(\theta), j_t] \mathbf{1}\{\hat{\theta}_t = \theta\} \end{aligned}$$

$$\begin{aligned}
&\leq M \sum_{\theta \neq \theta_0} E \sum_{t=1}^n I_{\theta_0, \theta}[x^*(\theta), j_t] \mathbf{1}\{\hat{\theta}_t = \theta\} \\
&= M \sum_{\theta \neq \theta_0} E \sum_{t=1}^n I_{\theta_0, \theta}[x_t, j_t] \mathbf{1}\{\hat{\theta}_t = \theta\}.
\end{aligned} \tag{4.8}$$

Recall that $\Lambda_n(\theta_0, \theta)$ is the log-likelihood ratio (4.3). By (4.5) we have

$$\mathbf{1}\{\hat{\theta}_t = \theta\} \leq \mathbf{1}\{\Lambda_{t-1}(\theta_0, \theta) \leq 0\}, \tag{4.9}$$

so that, noting that $I_{\theta_0, \theta} \geq 0$,

$$L_n^{\sigma_1, \tau}(\theta_0) \leq M \sum_{\theta \neq \theta_0} E \sum_{t=1}^{\infty} I_{\theta_0, \theta}[x_t, j_t] \mathbf{1}\{\Lambda_{t-1}(\theta_0, \theta) \leq 0\}. \tag{4.10}$$

We now proceed to upper-bound the expressions

$$F^\tau(\theta) \triangleq E_{\theta_0}^{\sigma_1, \tau} \sum_{t=1}^{\infty} I_{\theta_0, \theta}[x_t, j_t] \mathbf{1}\{\Lambda_{t-1}(\theta_0, \theta) \leq 0\} \tag{4.11}$$

for each $\theta \neq \theta_0$ separately by employing Lemma 3.3(i) and the correspondence indicated in Lemma 3.1. For that purpose two adjustments are required. First, we have to “replace” $\Lambda_{t-1}(\theta_0, \theta)$ with its truncated version $\tilde{\Lambda}_{t-1}(\theta_0, \theta)$ (see (4.5)), in order to comply with the required correspondence. Second, we shall have to extend the strategy set of DM2 in order to comply with the “controlled i.i.d. model” of section 3.

Note first that $\tilde{\Lambda}_{t-1}(\theta_0, \theta) \leq \Lambda_{t-1}(\theta_0, \theta)$ by its definition, so that

$$F^\tau(\theta) \leq \tilde{F}^\tau(\theta) \triangleq E_{\theta_0}^{\sigma_1, \tau} \sum_{t=1}^{\infty} I_{\theta_0, \theta}[x_t, j_t] \mathbf{1}\{\tilde{\Lambda}_{t-1}(\theta_0, \theta) \leq 0\}, \tag{4.12}$$

and it suffices to upper-bound $\tilde{F}^\tau(\theta)$. Now, since the strategy of DM1 is fixed, DM2 can be regarded as a single controller (maximizer) in (4.12). However, since $x_t = x^*(\hat{\theta}_t)$ depends on the process history, then DM2 is *not* facing a “controlled i.i.d. process”. Let us therefore extend the original set $\mathcal{T}$ of strategies available to DM2 by letting him choose at each stage n both j_n , as before, and also $x_n \in \mathcal{P}(\mathcal{I})$. Denote this extended set by Π . Lemma 3.3(i) can now be applied, which gives

$$F^\tau(\theta) \leq \tilde{F}^\tau(\theta) \leq \sup_{\tau \in \mathcal{T}} \tilde{F}^\tau(\theta) \leq \sup_{\pi \in \Pi} \tilde{F}^\pi(\theta) \leq Q(\theta) < \infty, \quad \forall \tau \in \mathcal{T}. \tag{4.13}$$

Combining (4.10), (4.11) and (4.13) yields

$$L_n^{\sigma_1, \tau}(\theta_0) \leq M \sum_{\theta \neq \theta_0} Q(\theta) < \infty, \quad \forall \tau, n,$$

and (i) is proved. Since (ii) follows immediately from (i), the proof is complete. $\square$

Discussion: We conclude this section with a few remarks concerning the strategy σ_1 and condition C_1 .

1. Condition $C_1(\theta_0)$ imposes a certain relation between one-stage rewards and information. Essentially, it requires the following: if DM1 uses the randomized action $x^*(\theta)$, for any $\theta \in \Theta$, then either the expected reward is at least $v(\theta_0)$, or otherwise the observed data carries some information for discriminating θ_0 and θ .
2. Condition $C_1(\theta_0)$ does *not* guarantee that the true parameter θ_0 may be consistently estimated by (any strategy of) DM1. For example, DM2 may still have a “non-revealing” action j_o under which all the possible games are indistinguishable, i.e. $I_{\theta_0, \theta}[x, j_o] = 0$ for all x and θ . However, in that case $C_1(\theta_0)$ implies that $\bar{A}_{\theta_0}[x^*(\theta), j_o] \geq v(\theta_0)$ for every θ , so that knowledge of the true parameter is not required in order to guarantee a one-stage reward of $v(\theta_0)$ or more against j_o .
3. If condition $C_1(\theta)$ is not satisfied, then the proposed strategy σ_1 may perform poorly. Indeed, let θ be a parameter which violates $C_1(\theta_0)$, i.e. for some action j' we have

$$(i) \quad \bar{A}_{\theta_0}[x^*(\theta), j'] < v(\theta_0), \quad \text{and} \quad (ii) \quad I_{\theta_0, \theta}[x^*(\theta), j'] = 0. \quad (4.14)$$

Assume that at some stage n the MLE is $\hat{\theta}_n = \theta$, so that $x_n = x^*(\theta)$ by definition of σ_1 . If DM2 chooses $j_n = j'$, then (i) implies that a (positive) relative loss is incurred, while (ii) implies that the likelihood ratio $\Lambda_n(\theta_0, \theta)$ will not change. This situation may therefore persist, thus leading to an average reward which is lower than $v(\theta_0)$ (or, equivalently, to a relative loss which increases as $O(n)$).

4. An important property of condition $C_1(\theta_0)$ is the following: any parameter θ which violates $C_1(\theta_0)$ must have a value lower than that of θ_0 , i.e. $v(\theta) < v(\theta_0)$. Indeed, let θ be such that (4.14) is satisfied for some $j' \in J$. Then,

$$v(\theta) \leq \bar{A}_{\theta}[x^*(\theta), j'] = \bar{A}_{\theta_0}[x^*(\theta), j'] < v(\theta_0), \quad (4.15)$$

where the first inequality follows by optimality of $x^*(\theta)$ in $G(\theta)$, and the equality follows from (4.14)(ii).

5. Noting the last two remarks, we can summarize the basic shortcoming of the strategy σ_1 : if condition C_1 is not satisfied, then the estimator may get “stuck” on the wrong parameter θ' , while the relative loss increases. However, this is only possible if $v(\theta') < v(\theta_0)$. This latter observation is the key to the improved strategy of the next section.

5 Certainty Equivalence: Value-Biased MLE

In this section we base our strategy on a modified estimator, which now depends also on the reward structure of the model. While keeping the simple Certainty Equivalence structure, considerable improvement in performance is attained as compared with the MLE-based strategy. Here the worst-case relative loss (WCRL) is always kept down to $O(\log n)$, and, moreover, finite WCRL is guaranteed under weaker conditions.

As indicated at the end of the previous section, a basic problem of the MLE-based strategy is that the estimator may adhere to parameters with a lower value than that of the true parameter. To prevent that, a certain bias will be introduced in the estimator in favor of parameters with high value. Naturally, this bias has to be delicate enough so that the identification capability of the estimator will not be destroyed.

The biasing method proposed here relies on the introduction of confidence levels in the estimator. Instead of just the MLE, which is the single parameter that maximizes the likelihood function, consider the set of parameters which *nearly* maximize the likelihood function (to within a prescribed time-varying threshold, or confidence level). We shall refer to this set as the *likely parameters set*. The estimator is then chosen as the member of this set which has the highest value.

The value-biased scheme is closely related to the cost-biased MLE algorithm, introduced in [14] in the context of adaptive control of Markov chains with the average cost criterion. There the bias is introduced by adding a cost-dependent term to the likelihood function; we shall comment on this biasing method at the end of the section. Several adaptive control algorithms have been proposed which incorporate confidence levels in the estimation scheme (e.g. [10,8,16,1]), but not for the purpose of cost related biasing which is crucial here.

Let $\{K_n, n \geq 1\}$ be a sequence of positive numbers, such that

$$\begin{aligned} (i) \quad & K_n \uparrow \infty, \quad K_n \geq 1 \\ (ii) \quad & \log K_n = o(n) \\ (iii) \quad & \sum_{n=1}^{\infty} K_n^{-1} < \infty \end{aligned} \tag{5.1}$$

Two particular examples are:

$$K_n = n^{1+\epsilon}, \quad \epsilon > 0 \tag{5.2}$$

$$K_n = n(\log n)^P + 1, \quad P > 1. \tag{5.3}$$

Let $\hat{\theta}_n$ be the MLE (4.1), and further define the *likely parameters set*:

$$\hat{\Theta}_n = \left\{ \theta \in \Theta : \Lambda_{n-1}(\hat{\theta}_n, \theta) \triangleq \log \frac{\lambda_{n-1}(\hat{\theta}_n)}{\lambda_{n-1}(\theta)} \leq \log K_n \right\}, \tag{5.4}$$

and the *value-biased maximum likelihood estimator*:

$$\bar{\theta}_n = \arg \max \left\{ v(\theta) : \theta \in \hat{\Theta}_n \right\}. \tag{5.5}$$

The following strategy will be considered in the greater part of this section.

Strategy σ_2 : $x_n = x^*(\bar{\theta}_n)$, where $\bar{\theta}_n$ is the value-biased MLE (5.5).

Before presenting the main results, we state some basic properties of the proposed estimator. By definition of $\bar{\theta}_n$, the following implications hold (for every $\theta_0 \in \Theta$):

$$v(\bar{\theta}_n) < v(\theta_0) \implies \theta_0 \notin \hat{\Theta}_n \implies \frac{\lambda_{n-1}(\hat{\theta}_n)}{\lambda_{n-1}(\theta_0)} > K_n. \tag{5.6}$$

$$v(\bar{\theta}_n) = v(\theta_0) \implies \Lambda_{n-1}(\theta_0, \bar{\theta}_n) \leq 0. \quad (5.7)$$

$$v(\bar{\theta}_n) \geq v(\theta_0) \implies \Lambda_{n-1}(\theta_0, \bar{\theta}_n) \leq \log K_n. \quad (5.8)$$

The following lemma, a consequence of (5.6), indicates that the biasing scheme achieves the purpose for which it was introduced.

Lemma 5.1

$$(i) \quad E_{\theta_0}^{\sigma, \tau} \sum_{t=1}^{\infty} \mathbf{1}\{\theta_0 \notin \hat{\Theta}_t\} \leq Q_1 \quad \text{for some finite constant } Q_1 \text{ and all } \sigma \text{ and } \tau.$$

$$(ii) \quad \text{Consequently, } E_{\theta_0}^{\sigma, \tau} \sum_{t=1}^{\infty} \mathbf{1}\{v(\bar{\theta}_t) < v(\theta_0)\} \leq Q_1 \quad \text{for all } \sigma \text{ and } \tau.$$

Proof:

(i) Let σ, τ be arbitrary strategies. Then, defining $\lambda(\theta, \theta_0) \triangleq \lambda_n(\theta)/\lambda_n(\theta_0)$,

$$\begin{aligned} E \sum_{t=1}^{\infty} \mathbf{1}\{\theta_0 \notin \hat{\Theta}_t\} &= \sum_{t=1}^{\infty} P\{\theta_0 \notin \hat{\Theta}_t\} \\ &= \sum_{t=1}^{\infty} P\{\lambda_{t-1}(\hat{\theta}_t, \theta_0) > K_t\} \\ &\leq \sum_{\theta \neq \theta_0} \sum_{t=1}^{\infty} P\{\lambda_{t-1}(\theta, \theta_0) > K_t\}. \end{aligned}$$

Now, as is well known the likelihood ratio $\{\lambda_n(\theta, \theta_0)\}_{n \geq 0}$ is a positive martingale with an expected value of 1. It then follows by Markov's inequality that:

$$E \sum_{t=1}^{\infty} \mathbf{1}\{\theta_0 \notin \hat{\Theta}_t\} \leq \sum_{\theta \neq \theta_0} \sum_{t=1}^{\infty} \frac{1}{K_t}, \quad (5.9)$$

which is finite by the choice (5.1) of $\{K_n\}$.

(ii) From (5.6) it follows that $\mathbf{1}\{v(\bar{\theta}_t) < v(\theta_0)\} \leq \mathbf{1}\{\theta_0 \in \hat{\Theta}_t\}$, and (ii) is therefore implied by (i). $\square$

Lemma 5.1 implies roughly that the the biased estimator $\bar{\theta}_n$ will equal a parameter with a value lower than that of the true parameter not more than a finite number of times. Thus, the effect of such parameters of lower value is no longer significant. Note however that the biasing scheme introduces a new potential problem: Since the estimator is biased towards parameters with high value, it may favor those over the true parameter, even if the unbiased likelihood function is maximized by the true parameter. The main effort in following analysis will be directed to bound the loss associated with this effect.

We are now in a position to present the first main result of this section.

Theorem 5.1 *For every $\theta_0 \in \Theta$ there exists a constant $\beta(\theta_0)$ such that*

$$\limsup_{n \rightarrow \infty} \frac{1}{\log K_n} L_n^{\sigma 2}(\theta_0) \leq \beta(\theta_0). \quad (5.10)$$

Thus, under strategy $\sigma 2$ the WCRL is $O(\log K_n)$ at most. In particular, if K_n is chosen as in (5.2) or (5.3), then the WCRL is at most $O(\log n)$.

The proof of Theorem 5.1 proceeds through some lemmas. We assume in the following that θ_0 is fixed. Define for future reference the constant

$$\hat{D} = \max_{i,j} d_{\theta_0}(i, j), \quad (5.11)$$

and the following three sets which are, respectively, the set of parameters with higher value than θ_0 , same value as θ_0 , and the union of the first two:

$$H(\theta_0) = \{\theta \in \Theta : v(\theta) > v(\theta_0)\} \quad (5.12)$$

$$S(\theta_0) = \{\theta^\circ \in \Theta : \theta^\circ \neq \theta_0, v(\theta^\circ) = v(\theta_0)\} \quad (5.13)$$

$$H_o(\theta_0) = \{\theta \in \Theta : \theta \neq \theta_0, v(\theta) \geq v(\theta_0)\}. \quad (5.14)$$

Lemma 5.2 *There exists a constant $M < \infty$ such that $d_{\theta_0}[x^*(\theta), j] \leq M \cdot I_{\theta_0, \theta}[x^*(\theta), j]$ holds for every j and every $\theta \in \Theta$ which satisfies $v(\theta) \geq v(\theta_0)$.*

Proof: Since Θ and J are finite sets and $I_{\theta_0, \theta}$ is non-negative, it is enough to show that $v(\theta) \geq v(\theta_0)$ and $I_{\theta_0, \theta}[x^*(\theta), j] = 0$ together imply $d_{\theta_0}[x^*(\theta), j] \leq 0$. Indeed, $I_{\theta_0, \theta}[x^*(\theta), j] = 0$ implies that $\bar{A}_{\theta_0}[x^*(\theta), j] = \bar{A}_\theta[x^*(\theta), j]$ (cf. (2.7)), and since $x^*(\theta)$ is optimal in the game matrix $\bar{A}_\theta$, we get

$$d_{\theta_0}[x^*(\theta), j] \triangleq v(\theta_0) - \bar{A}_{\theta_0}[x^*(\theta), j] = v(\theta_0) - \bar{A}_\theta[x^*(\theta), j] \leq v(\theta_0) - v(\theta) \leq 0. \quad (5.15)$$

□

Lemma 5.3 *For every θ with $v(\theta) > v(\theta_0)$:*

$$\limsup_{n \rightarrow \infty} \frac{1}{\log K_n} \max_{\tau \in T} E_{\theta_0}^{\sigma 2, \tau} \sum_{t=1}^n I_{\theta_0, \theta}[x_t, j_t] \mathbf{1}\{\Lambda_{t-1}(\theta_0, \theta) \leq \log K_n\} \leq 1.$$

Proof: Follows from Lemma 3.3(ii), by using exactly the same considerations that were used to bound (4.11). □

Proof of Theorem 5.1: By (2.10) and the definition of $\sigma 2$:

$$L_n^{\sigma 2, \tau}(\theta_0) = E_{\theta_0}^{\sigma 2, \tau} \sum_{t=1}^n d_{\theta_0}[x_t, j_t] = E_{\theta_0}^{\sigma 2, \tau} \sum_{t=1}^n d_{\theta_0}[x^*(\bar{\theta}_t), j_t]. \quad (5.16)$$

Now,

$$\begin{aligned} \sum_{t=1}^n d_{\theta_0}[x^*(\bar{\theta}_t), j_t] &= \sum_{t=1}^n d_{\theta_0} \left[x^*(\bar{\theta}_t), j_t \cdot [\mathbf{1}\{v(\bar{\theta}_t) < v(\theta_0)\} + \mathbf{1}\{v(\bar{\theta}_t) \geq v(\theta_0)\}] \right] \\ &\leq \hat{D} \sum_{t=1}^n \mathbf{1}\{v(\bar{\theta}_t) < v(\theta_0)\} + \sum_{\theta \in H_o(\theta_0)} \sum_{t=1}^n d_{\theta_0}[x^*(\theta), j_t] \mathbf{1}\{\bar{\theta}_t = \theta\}, \end{aligned} \quad (5.17)$$

where $\hat{D}$ and $H_o(\theta_0)$ are defined in (5.11), (5.14). We next bound the last term in (5.17). Using Lemma 5.2, the control law $x_t = x^*(\bar{\theta}_t)$ and (5.8), it follows that for every $\theta \in H_o(\theta_0)$,

$$\begin{aligned} d_{\theta_0}[x^*(\theta), j_t] \mathbf{1}\{\bar{\theta}_t = \theta\} &\leq M I_{\theta_0, \theta}[x^*(\theta), j_t] \mathbf{1}\{\bar{\theta}_t = \theta\} \\ &= M I_{\theta_0, \theta}[x_t, j_t] \mathbf{1}\{\bar{\theta}_t = \theta\} \\ &\leq M I_{\theta_0, \theta}[x_t, j_t] \mathbf{1}\{\Lambda_{t-1}(\theta_0, \theta) \leq \log K_t\}. \end{aligned} \quad (5.18)$$

Therefore

$$\begin{aligned} \sum_{t=1}^n d_{\theta_0}[x^*(\bar{\theta}_t), j_t] &\leq \hat{D} \sum_{t=1}^n \mathbf{1}\{v(\bar{\theta}_t) < v(\theta_0)\} \\ &\quad + M \sum_{\theta \in G(\theta_0)} \sum_{t=1}^n I_{\theta_0, \theta}[x_t, j_t] \cdot \mathbf{1}\{\Lambda_{t-1}(\theta_0, \theta) \leq \log K_t\}. \end{aligned} \quad (5.19)$$

The proof now follows by taking the expectation and applying Lemma 5.2(ii) and Lemma 5.3. $\square$

The proof above supports the following heuristic explanation for Theorem 5.1. As already noted, Lemma 5.1 implies that the effect of parameters with lower value than $v(\theta_0)$ on the loss may be ignored. Consider then the effect of parameters with a higher value than $v(\theta_0)$. The biased estimator would prefer such parameters over θ_0 – unless the observations provide sufficient statistical information to overcome the bias. Let θ be a parameter with $v(\theta) > v(\theta_0)$, i.e. $\theta \in H(\theta_0)$, and assume that $\bar{\theta}_n = \theta$ at some stage, so that $x_n = x^*(\theta)$. Now, the basic information–loss relation of Lemma 5.2 implies that any (positive) relative loss incurred at that stage is accompanied by a proportional information for discriminating θ_0 and θ . Thus, if the relative loss associated over the times when $\theta_t = \theta$ builds up to $O(\log K_n)$, so does the information for discriminating θ_0 and θ , and this information is just sufficient to overcome the bias so that θ is ruled out by the estimator. Consequently, this loss cannot exceed $O(\log K_n)$.

We turn now to formulate conditions that guarantee a *bounded* WCRL under the strategy σ_2 . Essentially, a certain information–loss relation is required to hold under the optimal action $x^*(\theta_0)$; this provides an additional “source of information” for discrimination between θ_0 and parameters with higher value.

For each $\theta^\circ \in \Theta$, recall that $H(\theta^\circ) \triangleq \{\theta : v(\theta) > v(\theta^\circ)\}$, and define

Condition $C_2(\theta^\circ)$: For each $j \in J$, at least one of the following holds:

(i) $\bar{A}_{\theta^\circ}[x^*(\theta^\circ), j] > v(\theta^\circ)$, or (ii) $\min_{\theta \in H(\theta^\circ)} I_{\theta^\circ, \theta}[x^*(\theta^\circ), j] > 0$.

Condition C_2 : Condition $C_2(\theta^\circ)$ holds for every $\theta^\circ \in \Theta$.

Theorem 5.2 Assume that DM1 employs strategy $\sigma 2$.

(i) Given $\theta_0 \in \Theta$, if $C_1(\theta^\circ)$ is satisfied for every $\theta^\circ \in S(\theta_0)$ (i.e. $v(\theta^\circ) = v(\theta_0)$), then

$$\limsup_{n \rightarrow \infty} L_n^{\sigma 2}(\theta_0) < \infty. \quad (5.20)$$

(ii) Consequently, if C_2 is satisfied, then the WCRL is bounded for every $\theta_0 \in \Theta$.

The proof of this result is presented in the next Section. Here we compare condition C_2 with condition C_1 of the previous section. It should be noted that for a given θ° , conditions $C_1(\theta^\circ)$ and $C_2(\theta^\circ)$ are not comparable, since the first pertains to parameters with lower value than θ° , while the second pertains to parameters with higher value. However, as the following lemma shows the global condition C_2 is weaker than C_1 , thus implying that the modified strategy $\sigma 2$ guarantees a bounded WCRL under weaker conditions than those required for $\sigma 1$.

Lemma 5.4 C_1 implies C_2 .

Proof: We shall prove that if C_2 is not satisfied, then C_1 is not satisfied; more specifically if $C_1(\theta^\circ)$ is not satisfied for some θ° , then $C_2(\theta)$ is not satisfied for some $\theta \in H(\theta^\circ) \triangleq \{\theta' : v(\theta') > v(\theta^\circ)\}$. Assume then that $C_2(\theta^\circ)$ is not satisfied. This means that for some j and $\theta \in H(\theta^\circ)$, we have (i') $\bar{A}_{\theta^\circ}[x^*(\theta^\circ), j] = v(\theta^\circ)$, and (ii') $I_{\theta^\circ, \theta}[x^*(\theta^\circ), j] = 0$, which imply that

$$\bar{A}_\theta[x^*(\theta^\circ), j] = \bar{A}_{\theta^\circ}[x^*(\theta^\circ), j] = v(\theta_0) < v(\theta). \quad (5.21)$$

But (5.21) and (ii') together imply that $C_1(\theta)$ is not satisfied. $\square$

An alternative biasing method:

We now consider briefly an alternative biasing method, where the estimator is the maximizer of a biased likelihood function. A similar method was employed in [14].

Let $\{w(\theta) : \theta \in \Theta\}$ be a set of real numbers which are increasing in $v(\theta)$, i.e. $w(\theta) > w(\theta')$ if and only if $v(\theta) > v(\theta')$. Let $\delta = \min\{|w(\theta) - w(\theta')| : w(\theta) \neq w(\theta')\}$ denote the minimal separation between this numbers, and let $\{k_n\}$ be a positive sequence such that (compare (5.1)): (i) $k_n \uparrow \infty$, (ii) $\log k_n = o(n)$, (iii) $\sum_{n=1}^{\infty} k_n^{-1} < \infty$. The value-biased MLE is then defined by

$$\hat{\theta}_n^b = \arg \max_{\theta \in \Theta} (k_n)^{w(\theta)} \lambda_{n-1}(\theta) \quad (5.22)$$

where $\lambda_{n-1}(\theta)$ is the likelihood function (4.2).

The main results of this section, namely Theorems 5.1 and 5.2 (with $\log K_n$ replaced by $\log k_n$), remain valid under the strategy $x_n = x^*(\hat{\theta}_n^b)$, namely the Certainty Equivalence strategy based on this new estimator. This is easily verified by noting that this estimator satisfies properties similar to (5.6)–(5.8), which were the key properties of the estimator $\bar{\theta}_n$.

We note, however, that a certain difference exists between the two biasing schemes. The estimator $\bar{\theta}_n$ effectively provides a uniform bias to all the parameters in $H(\theta_0)$, while in the estimator $\hat{\theta}_n^b$ the bias necessarily increases with $v(\theta)$. The uniform biasing property will prove to be of critical importance in [22], where an “asymptotically optimal” strategy is constructed based on the estimator $\bar{\theta}_n$.

6 Proof of Theorem 5.2

This Section contains the proof of Theorem 5.2. First we summarize the basic information–loss relations required here, in addition to those of Lemma 5.2.

Lemma 6.1 *There exist positive constants M, δ such that, for every $\theta_0 \in \Theta$ and $j \in J$:*

- (i) $d_{\theta_0}[x^*(\theta), j] \leq -\delta + M I_{\theta_0, \theta}[x^*(\theta), j]$ for every $\theta \in H(\theta_0) := \{\theta : v(\theta) > v(\theta_0)\}$.
- (ii) If $C_2(\theta_0)$ holds then $d_{\theta_0}[x^*(\theta_0), j] \leq -\delta + M \min_{\theta \in H(\theta_0)} I_{\theta_0, \theta}[x^*(\theta_0), j]$.
- (iii) For every θ° , if $v(\theta^\circ) = v(\theta_0)$ and $C_2(\theta^\circ)$ are satisfied then

$$d_{\theta_0}[x^*(\theta^\circ), j] \leq -\delta + M I_{\theta_0, \theta^\circ}[x^*(\theta^\circ), j] + M \min_{\theta \in H(\theta_0)} I_{\theta_0, \theta}[x^*(\theta^\circ), j].$$

Proof: Since $I_{\theta_0, \theta_0}[x, j] \equiv 0$, then (ii) is a special case of (iii) for $\theta^\circ = \theta_0$. It is therefore only required to prove (i) and (iii). Since Θ and J are finite sets and $I_{\theta_0, \theta}[x, j] \geq 0$ for all θ, x, j , it is sufficient to establish the following claims (i') and (iii'):

- (i') For every $\theta \in H(\theta_0)$, $I_{\theta_0, \theta}[x^*(\theta), j] = 0$ implies $d_{\theta_0}[x^*(\theta), j] < 0$.
- (iii') If $v(\theta^\circ) = v(\theta_0)$ and $C_2(\theta^\circ)$ hold, then $I_{\theta_0, \theta^\circ}[x^*(\theta^\circ), j] = \min_{\theta \in H(\theta_0)} I_{\theta_0, \theta}[x^*(\theta), j] = 0$ implies $d_{\theta_0}[x^*(\theta^\circ), j] < 0$.

Claim (i') follows exactly as in the proof of Lemma 5.2, where in the last line the strict inequality $v(\theta_0) - v(\theta) < 0$ may now be used. To establish (iii'), we assume that the assertions there are satisfied, and show that $d_{\theta_0}[x^*(\theta^\circ), j] < 0$. Noting (2.7), $I_{\theta_0, \theta^\circ}[x^*(\theta^\circ), j] = 0$ implies that $I_{\theta_0, \theta}[x^*(\theta^\circ), j] = I_{\theta^\circ, \theta}[x^*(\theta^\circ), j]$ for all θ , and in particular for all $\theta \in H(\theta_0)$. Note that $H(\theta^\circ) = H(\theta_0)$ since $v(\theta^\circ) = v(\theta_0)$. Therefore

$$\min_{\theta \in H(\theta^\circ)} I_{\theta^\circ, \theta}[x^*(\theta^\circ), j] = \min_{\theta \in H(\theta_0)} I_{\theta_0, \theta}[x^*(\theta^\circ), j] = 0. \quad (6.1)$$

By $C_2(\theta^\circ)$ this implies that $\bar{A}_{\theta^\circ}[x^*(\theta^\circ), j] > v(\theta_0)$, so that

$$d_{\theta_0}[x^*(\theta^\circ), j] \triangleq v(\theta_0) - \bar{A}_{\theta^\circ}[x^*(\theta^\circ), j] < 0. \quad (6.2)$$

□

Remark: Lemma 6.1 reflects the following relations between one-stage relative loss and information. Item (i) implies that for $x_n = x^*(\theta)$ with $\theta \in H(\theta_0)$, DM1 obtains either some (strictly

positive) $I_{\theta_0, \theta}$ -information (i.e., information for discriminating θ_0 and θ) or a strictly negative relative loss (i.e. average reward higher than $v(\theta_0)$). Item (ii) means that for $x_n = x^*(\theta_0)$, DM1 obtains either some positive $I_{\theta_0, \theta}$ -information for *every* $\theta \in H(\theta_0)$, or a strictly negative relative loss. Item (iii) may be interpreted similarly.

Consider henceforth a fixed $\theta_0 \in \Theta$ and a fixed strategy τ of DM2. Let $\hat{D}, H(\theta_0)$ and $H_o(\theta_0)$ be defined as in (5.11)–(5.14). Recall from (5.16) and (5.17) that

$$L_n^{\sigma^2, \tau}(\theta_0) \leq \hat{D} E_{\theta_0}^{\sigma^2, \tau} \sum_{t=1}^n \mathbf{1}\{v(\bar{\theta}_t) < v(\theta_0)\} + E_{\theta_0}^{\sigma^2, \tau} \sum_{t=1}^n d_{\theta_0}[x^*(\bar{\theta}_t), j_t] \mathbf{1}\{v(\bar{\theta}_t) \geq v(\theta_0)\}. \quad (6.3)$$

Since the first term on the right-hand-side is bounded by Lemma 5.1(ii), it remains to bound the last term. To this end, define for every $n \geq 1$

$$\ell_n = \sum_{t=1}^n d_{\theta_0}[x^*(\bar{\theta}_t), j_t] \mathbf{1}\{v(\bar{\theta}_t) \geq v(\theta_0)\}, \quad (6.4)$$

$$\Delta \ell_n = \ell_n - \ell_{n-1} = d_{\theta_0}[x^*(\bar{\theta}_n), j_n] \mathbf{1}\{v(\bar{\theta}_n) \geq v(\theta_0)\} \quad (6.5)$$

(where $\ell_o \triangleq 0$). The required upper-bound is established in the following lemmas, where the basic idea is that (the expected value of) ℓ_n cannot increase “too much” over the time instants when ℓ_n is positive.

Lemma 6.2

- (i) $\ell_n \leq \hat{D} + \hat{D} \sum_{t=1}^{n-1} (\Delta \ell_{t+1})^+ \mathbf{1}\{\ell_t \geq 0\}$ for every $n \geq 1$.
- (ii) Consequently, there exists a finite constant Q_3 such that

$$E_{\theta_0}^{\sigma^2, \tau} \ell_n \leq Q_3 + \hat{D} \sum_{t=1}^{\infty} P_{\theta_0}^{\sigma^2, \tau} \{\ell_t \geq 0, \bar{\theta}_{t+1} \in H(\theta_0)\}.$$

Proof:

- (i) Fix $n \geq 1$, and let $m_o = \max\{0 \leq t \leq n-1 : \ell_t \leq 0\}$. Note that $\Delta \ell_t \leq \hat{D}$ for every t . Then

$$\begin{aligned} \ell_n &\leq \ell_n - \ell_{m_o} \leq \sum_{t=m_o}^{n-1} (\Delta \ell_{t+1})^+ \\ &\leq \hat{D} + \sum_{t=m_o+1}^{n-1} (\Delta \ell_{t+1})^+ = \hat{D} + \sum_{t=m_o+1}^{n-1} (\Delta \ell_{t+1})^+ \mathbf{1}\{\ell_t > 0\} \\ &\leq \hat{D} + \sum_{t=1}^{n-1} (\Delta \ell_{t+1})^+ \mathbf{1}\{\ell_t \geq 0\}. \end{aligned} \quad (6.6)$$

- (ii) Recalling the definitions in (5.12) and (5.14) of $H(\theta_0)$ and $S(\theta_0)$, (6.5) may be rewritten as

$$\Delta \ell_t = d_{\theta_0}[x^*(\bar{\theta}_t), j_t] \mathbf{1}\{\bar{\theta}_t \in \{\theta_0\} \cup H(\theta_0) \cup S(\theta_0)\}. \quad (6.7)$$

Noting that $d_{\theta_0}[x^*(\theta_0), j] \leq 0$ and $d_{\theta_0}[x^*(\theta), j] \leq \hat{D}$ for $\theta \in H(\theta_0)$, and using Lemma 5.2 for $\theta^\circ \in S(\theta_0)$, we get

$$(\Delta \ell_t)^+ \leq \hat{D} \mathbf{1}\{\bar{\theta}_t \in H(\theta_0)\} + M \sum_{\theta^\circ \in S(\theta_0)} I_{\theta_0, \theta^\circ}[x^*(\theta^\circ), j_t] \mathbf{1}\{\bar{\theta}_t = \theta^\circ\}. \quad (6.8)$$

Therefore, by (i)

$$\ell_n \leq \hat{D} + \hat{D} \sum_{t=1}^{n-1} \mathbf{1}\{\ell_t \geq 0, \bar{\theta}_{t+1} \in H(\theta_0)\} + M \sum_{\theta \in S(\theta_0)} \sum_{t=1}^{n-1} I_{\theta_0, \theta}[x^*(\theta), j_{t+1}] \mathbf{1}\{\bar{\theta}_{t+1} = \theta\}. \quad (6.9)$$

Extending the summations to ∞ and taking expectation then gives

$$\begin{aligned} E_{\theta_0}^{\sigma_2, \tau} \ell_n &\leq \hat{D} + \hat{D} \sum_{t=1}^{n-1} P\{\ell_t \geq 0, \bar{\theta}_{t+1} \in H(\theta_0)\} \\ &\quad + M \sum_{\theta \in S(\theta_0)} E_{\theta_0}^{\sigma_2, \tau} \sum_{t=1}^{\infty} I_{\theta_0, \theta}[x^*(\theta), j_{t+1}] \mathbf{1}\{\bar{\theta}_{t+1} = \theta\}. \end{aligned} \quad (6.10)$$

It remains only to bound the last term. Let $\theta \in S(\theta_0)$. Since $v(\theta) = v(\theta_0)$, it follows from (5.7) that

$$E_{\theta_0}^{\sigma_2, \tau} \sum_{t=1}^{\infty} I_{\theta_0, \theta}[x^*(\theta^\circ), j_t] \mathbf{1}\{\bar{\theta}_t = \theta\} \leq E_{\theta_0}^{\sigma_2, \tau} \sum_{t=1}^{\infty} I_{\theta_0, \theta}[x_t, j_t] \mathbf{1}\{\Lambda_{t-1}(\theta_0, \theta) \leq 0\}. \quad (6.11)$$

This last term can now be bounded exactly in the same way that $F^\tau(\theta)$ of (4.11) was bounded. $\square$

Lemma 6.3 *Assume that $C_2(\theta^\circ)$ is satisfied for every $\theta^\circ \in \Theta$ such that $v(\theta^\circ) = v(\theta_0)$. Assume that DM1 employs strategy σ_2 , and DM2 is using any strategy τ . Then there exists a constant $\eta > 0$ such that, for every $n \geq 1$: if $\ell_n > 0$, then at least one of the following events $\Omega_1(n) - \Omega_3(\theta)$ holds:*

$$\begin{aligned} \Omega_1(n) : & \quad \sum_{t=1}^n \mathbf{1}\{v(\bar{\theta}_t) < v(\theta_0)\} \geq \eta n. \\ \Omega_2(n) : & \quad \sum_{t=1}^n I_{\theta_0, \theta}[x_t, j_t] \mathbf{1}\{\bar{\theta}_t = \theta\} \geq \eta n \text{ for some } \theta \in H_o(\theta_0). \\ \Omega_3(n) : & \quad \min_{\theta \in H(\theta_0)} \sum_{t=1}^n I_{\theta_0, \theta}[x_t, j_t] \geq \eta n. \end{aligned}$$

Proof: Similarly to (6.7), we have

$$\ell_n = \sum_{t=1}^n \sum_{\theta' \in \Theta} d_{\theta_0}[x^*(\theta'), j_t] \mathbf{1}\{\bar{\theta}_t = \theta' \in \{\theta_0\} \cup S(\theta_0) \cup H(\theta_0)\}. \quad (6.12)$$

Now, using the appropriate bounds from Lemma 6.1 for each θ' in the above sum gives

$$\begin{aligned}
\ell_n &\leq \sum_{t=1}^n \{ [-\delta + M \min_{\theta \in H(\theta_0)} I_{\theta_0, \theta}[x^*(\theta_0), j_t]] \mathbf{1}\{\bar{\theta}_t = \theta_0\} \\
&\quad + \sum_{\theta^\circ \in S(\theta_0)} [-\delta + M I_{\theta_0, \theta^\circ}[x^*(\theta^\circ), j_t] + M \min_{\theta \in H(\theta_0)} I_{\theta_0, \theta}[x^*(\theta^\circ), j_t]] \mathbf{1}\{\bar{\theta}_t = \theta^\circ\} \\
&\quad + \sum_{\theta \in H(\theta_0)} [-\delta + M I_{\theta_0, \theta}[x^*(\theta), j_t]] \mathbf{1}\{\bar{\theta}_t = \theta\} \\
&= -\delta \sum_{t=1}^n \mathbf{1}\{v(\bar{\theta}_t) \geq v(\theta_0)\} + M \sum_{t=1}^n \sum_{\theta \in H_o(\theta_0)} I_{\theta_0, \theta}[x^*(\theta), j_t] \mathbf{1}\{\bar{\theta}_t = \theta\} \\
&\quad + M \sum_{t=1}^n \left[\min_{\theta \in H_o(\theta_0)} I_{\theta_0, \theta}[x_t, j_t] \mathbf{1}\{\bar{\theta}_t \in \{\theta_0\} \cup S(\theta_0)\} \right] \\
&\leq -\delta n + \delta \sum_{t=1}^n \mathbf{1}\{v(\bar{\theta}_t) < v(\theta_0)\} + M \sum_{\theta \in H_o(\theta_0)} \sum_{t=1}^n I_{\theta_0, \theta}[x_t, j_t] \mathbf{1}\{\bar{\theta}_t = \theta\} \\
&\quad + M \min_{\theta \in H(\theta_0)} \sum_{t=1}^n I_{\theta_0, \theta}[x_t, j_t] \tag{6.13}
\end{aligned}$$

where in the last step we have used the facts that $x_t = x^*(\bar{\theta}_t)$ under σ_2 , and that $I_{\theta_0, \theta} \geq 0$.

Defining $\eta = \delta/(\delta + M|\Theta|)$, it follows from the last inequality that ℓ_n will be negative unless one of $\Omega_1(n) - \Omega_3(n)$ is satisfied. $\square$

Lemma 6.4 *Let $\Omega_1(n) - \Omega_3(n)$ be defined as in the previous lemma. Then, for some $Q_2 < \infty$ and every $\tau \in \mathcal{T}$,*

$$\begin{aligned}
(i) \quad &\sum_{n=1}^{\infty} P_{\theta_0}^{\sigma_2, \tau} \{\Omega_1(n)\} \leq Q_2, \\
(ii) \quad &\sum_{n=1}^{\infty} P_{\theta_0}^{\sigma_2, \tau} \{\Omega_2(n)\} \leq Q_2, \\
(iii) \quad &\sum_{n=1}^{\infty} P_{\theta_0}^{\sigma_2, \tau} \{\Omega_3(n), \bar{\theta}_n \in H(\theta_0)\} \leq Q_2.
\end{aligned}$$

Proof:

(i) Recall from (5.6) that $v(\bar{\theta}_t) < v(\theta_0)$ implies $\lambda_{t-1}(\hat{\theta}_t, \theta_0) := \lambda_{t-1}(\hat{\theta}_t)/\lambda_{t-1}(\theta_0) > K_t$. Therefore (denoting $P := P_{\theta_0}^{\sigma_2, \tau}$),

$$\begin{aligned}
P\{\Omega_1(n)\} &\leq P\left\{ \sum_{t=1}^n \mathbf{1}\{\lambda_{t-1}(\hat{\theta}_t, \theta_0) > K_t\} \geq \eta n \right\} \\
&\leq P\{\lambda_{t-1}(\hat{\theta}_t, \theta_0) > K_t \text{ for some } t \geq \eta n\} \tag{6.14}
\end{aligned}$$

$$\begin{aligned}
&\leq P \left\{ \sup_{t \geq 1} \lambda_{t-1}(\hat{\theta}_t, \theta_0) > K_{[\eta n]} \right\} \\
&\leq \sum_{\theta \in \Theta} P \left\{ \sup_{t \geq 1} \lambda_{t-1}(\theta, \theta_0) > K_{[\eta n]} \right\},
\end{aligned}$$

where $[\eta n]$ is the integer part of ηn . Now, since the likelihood ratio $\{\lambda_t(\theta, \theta_0)\}$ is a positive Martingale with expected value 1, it follows by Doob's inequality that $P\{\Omega_1(n)\} \leq |\Theta| (K_{[\eta n]})^{-1}$. Since $\{K_n^{-1}\}$ is summable by (5.1), this implies

$$\sum_{n=1}^{\infty} P\{\Omega_1(n)\} \leq |\Theta| \sum_{n=1}^{\infty} K_{[\eta n]}^{-1} < \infty. \quad (6.15)$$

(ii) Using the union bound and (5.8),

$$\begin{aligned}
P\{\Omega_2(n)\} &\leq \sum_{\theta \in H_o(\theta_0)} P \left\{ \sum_{t=1}^n I_{\theta_0, \theta}[x_t, j_t] \mathbf{1}\{\bar{\theta}_t = \theta\} \geq \eta n \right\} \\
&\leq \sum_{\theta \in H_o(\theta_0)} P \left\{ \sum_{t=1}^n I_{\theta_0, \theta}[x_t, j_t] \mathbf{1}\{\Lambda_{t-1}(\theta_0, \theta) \leq \log K_t\} \geq \eta n \right\}.
\end{aligned} \quad (6.16)$$

Now, using the same procedure as in the proof of Theorem 4.1 (i.e., DM2 is allowed to choose (x_t, j_t) , and Λ_{t-1} is replaced by a “truncated” $\tilde{\Lambda}_{t-1}$), it follows by Lemmas 3.1 and 3.3(iv) (with $\beta_n = \log K_n$) that, for every $\tau \in \mathcal{T}$,

$$\sum_{n=1}^{\infty} P\{\Omega_2(n)\} \leq \sum_{\theta \in H_o(\theta_0)} Q(\theta) < \infty. \quad (6.17)$$

(iii) Similarly to the proof of (ii), it follows that

$$\begin{aligned}
P\{\Omega_3(n), \bar{\theta}_{n+1} \in H(\theta_0)\} &\leq \sum_{\theta \in H(\theta_0)} P \left\{ \sum_{t=1}^n I_{\theta_0, \theta}[x_t, j_t] > \eta n, \bar{\theta}_t = \theta \right\} \\
&\leq \sum_{\theta \in H(\theta_0)} P \left\{ \sum_{t=1}^n I_{\theta_0, \theta}[x_t, j_t] \geq \eta n, \Lambda_n(\theta_0, \theta) \leq \log K_{n+1} \right\}.
\end{aligned} \quad (6.18)$$

The bound now follows exactly as in (ii), except that Lemma 3.3(iii) is used in place of Lemma 3.3(iv). $\square$

We are now ready to conclude the proof of Theorem 5.2. By (6.3), (6.4), Lemma 5.1(ii) and Lemma 5.6(ii), it follows that for every $\tau \in \mathcal{T}$ and $n \geq 1$:

$$L_n^{\sigma^2, \tau}(\theta_0) \leq \hat{D}Q_1 + E_{\theta_0}^{\sigma^2, \tau} \ell_n \leq \hat{D}Q_1 + Q_3 + \hat{D} \sum_{t=1}^{\infty} P\{\ell_t \geq 0, \bar{\theta}_{t+1} \in H(\theta_0)\}. \quad (6.19)$$

Moreover, by Lemmas 5.7 and 5.8,

$$\begin{aligned} \sum_{t=1}^{\infty} P\{\ell_t \geq 0, \bar{\theta}_{t+1} \in H(\theta_0)\} &\leq \sum_{t=1}^{\infty} P\{C_1(n)\} + \sum_{t=1}^{\infty} P\{C_2(n)\} + \sum_{t=1}^{\infty} P\{C_3(n), \bar{\theta}_t > \theta_0\} \\ &\leq 3Q_2 < \infty, \end{aligned} \tag{6.20}$$

so that $L_n^{\sigma^2}(\theta_0) \leq \hat{D}Q_1 + Q_3 3Q_2 < \infty$ for every $n \geq 1$. $\square$

7 Concluding Remarks

We have examined in this paper the long-term performance of Certainty Equivalence strategies, for a decision maker faced with an uncertain dynamic game situation. It was shown that these schemes potentially suffer from closed-loop identification problems, similar to those found in comparable adaptive control models, and that these can be essentially eliminated by properly modifying the estimator. In particular, Theorem 5.1 established that the worst-case relative loss can be kept down to $O(\log n)$ by using the value-biased Maximum Likelihood Estimator.

While this latter result seems quite satisfactory, it is still natural to ask whether this is the best that can be attained (without imposing further conditions on the model). In the sequel paper [22] it will be shown that an increase rate of $O(\log n)$ is in fact the best that can be guaranteed by any strategy; furthermore, the optimal coefficient associated with this increase rate (i.e., the smallest possible coefficient $\beta(n)$ in Theorem 5.1) will be characterized, and a strategy which attains this “asymptotically optimal” performance will be constructed.

The basic model of this paper may be extended in several directions. Here we studied the case of finite parameter and actions sets; more general sets may be of interest. It should also be of interest to consider system with non-trivial dynamics, e.g. controlled Markov processes (leading to stochastic game models). More ideas and methods from the field of adaptive control may prove applicable to such models.

References

- [1] R. Agrawal, D. Teneketzis, and V. Anantharam, "Asymptotically efficient adaptive allocation schemes for controlled i.i.d. processes: Finite parameter space," *IEEE Trans. Automat. Control* **34**, pp. 258–267, 1989.
- [2] R. Agrawal, D. Teneketzis, and V. Anantharam, "Asymptotically efficient adaptive allocation schemes for controlled Markov chains: Finite parameter space," *IEEE Trans. Automat. Control* **34**, pp. 1249–1259, 1989.
- [3] V. Anantharam, P. Varaiya, and J. Walrand, "Asymptotically efficient allocation rules for the multiarmed bandit problem with multiple plays. Part I: iid rewards, Part II: Markovian rewards," *IEEE Trans. Automat. Control* **32**, pp. 968–982, 1987.
- [4] R. J. Aumann, "Mixed and behavior strategies in infinite extensive games," in: *Advances in Game Theory: Ann. of Math. Studies* **52**, M. Dresher et al. (eds.), Princeton University Press, NJ, pp. 627–650, 1964.
- [5] A. Baños, "On pseudo-games," *Ann. Math. Statist.* **39**, pp. 1932–1945, 1968.
- [6] D.P. Bertsekas, *Dynamic Programming and Stochastic Control*, Academic Press, NY, 1976.
- [7] V. Borkar and P. Varaiya, "Adaptive control of Markov chains, I: Finite parameter set," *IEEE Trans. Automat. Control* **24**, pp. 953–958, 1979.
- [8] V. Borkar and R. Varaiya, "Identification and adaptive control of Markov chains," *SIAM J. Control Optim.* **20**, pp. 470–489, 1982.
- [9] T.M. Cover and J.A. Thomas, *Elements of Information Theory*, Wiley, NY, 1991.
- [10] B. Doshi and S.E. Shreve, "Strong consistency of a modified maximum likelihood estimator for controlled Markov chains," *J. Appl. Prob.* **17**, pp. 726–734, 1980.
- [11] D. Fudenberg and J. Tirole, *Game Theory*, MIT Press, Cambridge, MA, 1991.
- [12] G.C. Goodwin and K.S. Sin, *Adaptive Filtering, Prediction and Control*, Prentice-Hall, NJ, 1984.
- [13] P. R. Kumar, "A survey of some results in stochastic adaptive control," *SIAM J. Control Optim.* **23**, pp. 329–380, 1985.
- [14] P.R. Kumar and A. Becker, "A new family of optimal adaptive controllers for Markov chains," *IEEE Trans. Automat. Control* **27**, pp. 137–146, 1982.
- [15] P.R. Kumar and P. Varaiya, *Stochastic Systems: Estimation, Identification and Adaptive Control*, Prentice-Hall, NJ, 1986.
- [16] T.L. Lai and H. Robbins, "Asymptotically efficient adaptive allocation rules," *Adv. in Appl. Math.* **6**, pp. 4–22, 1985.
- [17] S. Lakshmivarahan, *Learning Algorithms Theory and Applications*, Springer-Verlag, NY, 1981.
- [18] P. Mandl, "Estimation and control in Markov chains," *Advances in Appl. Prob.* **6**, pp. 40–60, 1974.

- [19] N. Megiddo, "On repeated games with incomplete information played by non-Bayesian players", *Internat. J. Game Theory* **9**, pp. 157–167, 1980.
- [20] J.-F. Mertens, *Repeated Games*, CORE reprint No. 8624, Centre for Operations Research and Economics, Universite Catholique de Louvain, Belgium 1986.
- [21] T. Parthasarathy and T.E.S. Ragahavan, *Some Topics in Two-Person Games*, American Elsevier Publishing Co., NY, 1971.
- [22] N. Shimkin, "Asymptotically efficient adaptive strategies in repeated games. Part 2: Asymptotic optimality," in preparation, 1992.
- [23] S. Sorin, "An introduction to two-person zero-sum repeated games with incomplete information," IMSS-Economics, Technical Report No. 312, Stanford University, 1980.
- [24] J. Van Ryzin, "Repetitive play in finite statistical games with unknown distributions," *Ann. Math. Statist.* **37**, pp. 976–974, 1966.
- [25] S. Zamir, "On the relation between finitely and infinitely repeated games with incomplete information," *Internat. J. Game Theory* **1**, pp. 215–229, 1972.

#	Author/s	Title
1011	E.G. Kalnins, Willard Miller, Jr. and Sanchita Mukherjee,	Models of q -algebra representations: Matrix Elements of $U_q(su_2)$
1012	Zhangxin Chen and Bernardo Cockburn,	Error estimates for a finite element method for the drift-diffusion semiconductor device equations
1013	Chaocheng Huang,	Drying of gelatin asymptotically in photographic film
1014	Richard E. Ewing and Hong Wang,	Eulerian-Lagrangian localized adjoint methods for reactive transport in groundwater
1015	Bing-Yu Zhang,	Taylor series expansion for solutions of the Korteweg-de Vries equation with respect to their initial values
1016	Kenneth R. Driessel,	Some remarks on the geometry of some surfaces of matrices associated with Toeplitz eigenproblems
1017	C.J. Van Duijn and Peter Knabner,	Flow and reactive transport in porous media induced by well injection: Similarity solution
1018	Wasin So,	Rank one perturbation and its application to the Laplacian spectrum of a graph
1019	G. Baccarani, F. Odeh, A. Gnudi and D. Ventura,	A critical review of the fundamental semiconductor equations
1020	T.R. Hoffend Jr.,	Magnetostatic interactions for certain types of stacked, cylindrically symmetric magnetic particles
1021	IMA Summer Program for Graduate Students,	Mathematical Modeling
1022	Wayne Barrett, Charles R. Johnson, and Pablo Tarazaga,	The real positive definite completion problem for a simple cycle
1023	Charles A. McCarthy,	Fourth order accuracy for a cubic spline collocation method
1024	Martin Hanke, James Nagy, and Robert Plemmons,	Preconditioned iterative regularization for ill-posed problems
1025	John R. Gilbert, Esmond G. Ng, and Barry W. Peyton,	An efficient algorithm to compute row and column counts for sparse Cholesky factorization
1026	Xinfu Chen,	Existence and regularity of solutions of a nonlinear nonuniformly elliptic system arising from a thermistor problem
1027	Xinfu Chen and Weiqing Xie,	Discontinuous solutions of steady state, viscous compressible Navier-Stokes equations
1028	E.G. Kalnins, Willard Miller, Jr., and Sanchita Mukherjee,	Models of q -algebra representations: Matrix elements of the q -oscillator algebra
1029	W. Miller, Jr. and Lee A. Rubel,	Functional separation of variables for Laplace equations in two dimensions
1030	I. Gohberg and I. Koltracht,	Structured condition numbers for linear matrix structures
1031	Xinfu Chen,	Hele-Shaw problem and area preserved curve shortening motion
1032	Zhangxin Chen and Jim Douglas, Jr.	Modelling of compositional flow in naturally fractured reservoirs
1033	Harald K. Wimmer,	On the existence of a least and negative-semidefinite solution of the discrete-time algebraic Riccati equation
1034	Harald K. Wimmer,	Monotonicity and parametrization results for continuous-time algebraic Riccati equations and Riccati inequalities
1035	Bart De Moor, Peter Van Overschee, and Geert Schelfhout,	H_2 model reduction for SISO systems
1036	Bart De Moor,	Structured total least squares and L_2 approximation problems
1037	Chjan Lim,	Nonexistence of Lyapunov functions and the instability of the Von Karman vortex streets
1038	David C. Dobson and Fadil Santosa,	Resolution and stability analysis of an inverse problem in electrical impedance tomography – dependence on the input current patterns
1039	C.N. Dawson, C.J. van Duijn, and M.F. Wheeler,	Characteristic-Galerkin methods for contaminant transport with non-equilibrium adsorption kinetics
1040	Bing-Yu Zhang,	Analyticity of solutions of the generalized Korteweg-de Vries equation with respect to their initial values
1041	Neerchal K. Nagaraj and Wayne A. Fuller,	Least squares estimation of the linear model with autoregressive errors
1042	H.J. Sussmann & W. Liu,	A characterization of continuous dependence of trajectories with respect to the input for control-affine systems
1043	Karen Rudie & W. Murray Wonham,	Protocol verification using discrete-event systems
1044	Rohan Abeyaratne & James K. Knowles,	Nucleation, kinetics and admissibility criteria for propagating phase boundaries
1045	Gang Bao & William W. Symes,	Computation of pseudo-differential operators
1046	Srdjan Stojanovic,	Nonsmooth analysis and shape optimization in flow problem
1047	Miroslav Tuma,	Row ordering in sparse QR decomposition

- 1048 **Onur Toker & Hitay Özbay**, On the computation of suboptimal H^∞ controllers for unstable infinite dimensional systems
- 1049 **Hitay Özbay**, H^∞ optimal controller design for a class of distributed parameter systems
- 1050 **J.E. Dunn & Roger Fosdick**, The Weierstrass condition for a special class of elastic materials
- 1051 **Bei Hu & Jianhua Zhang**, A free boundary problem arising in the modeling of internal oxidation of binary alloys
- 1052 **Eduard Feireisl & Enrique Zuazua**, Global attractors for semilinear wave equations with locally distributed nonlinear damping and critical exponent
- 1053 **I-Heng McComb & Chjan C. Lim**, Stability of equilibria for a class of time-reversible, $D_n \times O(2)$ -symmetric homogeneous vector fields
- 1054 **Ruben D. Spies**, A state-space approach to a one-dimensional mathematical model for the dynamics of phase transitions in pseudoelastic materials
- 1055 **H.S. Dumas, F. Golse, and P. Lochak**, Multiphase averaging for generalized flows on manifolds
- 1056 **Bei Hu & Hong-Ming Yin**, Global solutions and quenching to a class of quasilinear parabolic equations
- 1057 **Zhangxin Chen**, Projection finite element methods for semiconductor device equations
- 1058 **Peter Guttorp**, Statistical analysis of biological monitoring data
- 1059 **Wensheng Liu & Héctor J. Sussmann**, Abnormal sub-Riemannian minimizers
- 1060 **Chjan C. Lim**, A combinatorial perturbation method and Arnold's whiskered Tori in vortex dynamics
- 1061 **Yong Liu**, Axially symmetric jet flows arising from high speed fiber coating
- 1062 **Li Qiu & Tongwen Chen**, $\mathcal{H}_2$ and $\mathcal{H}_\infty$ designs of multirate sampled-data systems
- 1063 **Eduardo Casas & Jiongmin Yong**, Maximum principle for state-constrained optimal control problems covered by quasilinear elliptic equations
- 1064 **Suzanne M. Lenhart & Jiongmin Yong**, Optimal control for degenerate parabolic equations with logistic growth
- 1065 **Suzanne Lenhart**, Optimal control of a convective-diffusive fluid problem
- 1066 **Enrique Zuazua**, Weakly nonlinear large time behavior in scalar convection-diffusion equations
- 1067 **Caroline Fabre, Jean-Pierre Puel & Enrike Zuazua**, Approximate controllability of the semilinear heat equation
- 1068 **M. Escobedo, J.L. Vazquez & Enrike Zuazua**, Entropy solutions for diffusion-convection equations with partial diffusivity
- 1069 **M. Escobedo, J.L. Vazquez & Enrike Zuazua**, A diffusion-convection equation in several space dimensions
- 1070 **F. Fagnani & J.C. Willems**, Symmetries of differential systems
- 1071 **Zhangxin Chen, Bernardo Cockburn, Joseph W. Jerome & Chi-Wang Shu**, Mixed-RKDG finite element methods for the 2-D hydrodynamic model for semiconductor device simulation
- 1072 **M.E. Bradley & Suzanne Lenhart**, Bilinear optimal control of a Kirchhoff plate
- 1073 **Héctor J. Sussmann**, A cornucopia of abnormal subriemannian minimizers. Part I: The four-dimensional case
- 1074 **Marek Rakowski**, Transfer function approach to disturbance decoupling problem
- 1075 **Yuncheng You**, Optimal control of Ginzburg-Landau equation for superconductivity
- 1076 **Yuncheng You**, Global dynamics of dissipative modified Korteweg-de Vries equations
- 1077 **Mario Taboada & Yuncheng You**, Nonuniformly attracting inertial manifolds and stabilization of beam equations with structural and Balakrishnan-Taylor damping
- 1078 **Michael Böhm & Mario Taboada**, Global existence and regularity of solutions of the nonlinear string equation
- 1079 **Zhangxin Chen**, BDM mixed methods for a nonlinear elliptic problem
- 1080 **J.J.L. Velázquez**, On the dynamics of a closed thermosyphon
- 1081 **Frédéric Bonnans & Eduardo Casas**, Some stability concepts and their applications in optimal control problems
- 1082 **Hong-Ming Yin**, $\mathcal{L}^{2,\mu}(Q)$ -estimates for parabolic equations and applications
- 1083 **David L. Russell & Bing-Yu Zhang**, Smoothing and decay properties of solutions of the Korteweg-de Vries equation on a periodic domain with point dissipation
- 1084 **J.E. Dunn & K.R. Rajagopal**, Fluids of differential type: Critical review and thermodynamic analysis
- 1085 **Mary Elizabeth Bradley & Mary Ann Horn**, Global stabilization of the von Kármán plate with boundary feedback acting via bending moments only
- 1086 **Mary Ann Horn & Irena Lasiecka**, Global stabilization of a dynamic von Kármán plate with nonlinear boundary feedback
- 1087 **Vilmos Komornik**, Decay estimates for a petrovski system with a nonlinear distributed feedback
- 1088 **Jesse L. Barlow**, Perturbation results for nearly uncoupled Markov chains with applications to iterative methods
- 1089 **Jong-Shenq Guo**, Large time behavior of solutions of a fast diffusion equation with source
- 1090 **Tongwen Chen & Li Qiu**, $\mathcal{H}_\infty$ design of general multirate sampled-data control systems
- 1091 **Satyanad Kichenassamy & Walter Littman**, Blow-up surfaces for nonlinear wave equations, I
- 1092 **Nahum Shimkin**, Asymptotically efficient adaptive strategies in repeated games, Part I: certainty equivalence strategies
- 1093 **Caroline Fabre, Jean-Pierre Puel & Enrique Zuazua**, On the density of the range of the semigroup for semilinear heat equations