ASYMPTOTICALLY EFFICIENT ADAPTIVE STRATEGIES IN REPEATED GAMES, PART II: ASYMPTOTIC OPTIMALITY

By

Nahum Shimkin

and

Adam Shwartz

IMA Preprint Series # 1121

March 1993

Asymptotically Efficient Adaptive Strategies in Repeated Games, Part II: Asymptotic Optimality

NAHUM SHIMKIN* and ADAM SHWARTZ†

March 1993

Abstract

This paper continues the analysis of a repeated game model with incomplete information on one side, in which rewards are random, perfect observations are assumed, and the emphasis is on strategies of DM1 (the uninformed decision maker) which maximize his worst-case total reward in a non-Bayesian sense, namely for all possible states of nature. An asymptotic bound on performance is first established, followed by the construction of strategies which achieve this bound. The analysis highlights the efficient acquisition of (statistical) information under conflict conditions, and especially the relations between information and payoff which are inherent in this problem.

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

*Institute for Mathematics and its Applications, University of Minnesota, Minneapolis, MN 55455, USA.
†Department of Electrical Engineering, Technion – Israel Institute of Technology, Haifa 32000, Israel.

1 Introduction

This paper continues the study of asymptotically efficient strategies for the model considered in [7]. For completeness we summarize briefly the model and relevant notation. Further details and background may be found in [7].

The game model involves two decision makers, DM1 (the maximizer) and DM2, which repeatedly play a matrix game $G(\theta_0)$, known to be a member of a finite set $\{G(\theta),\, \theta \in \Theta\}$. Each $G(\theta)$ is a zero-sum matrix game with random rewards, and with finite action sets $\mathcal{I}$ for for DM1 and $\mathcal{J}$ for DM2. The reward structure is thus specified by the probability distributions $\{p_{\theta,i,j}(\,\cdot\,) : i \in \mathcal{I},\ j \in \mathcal{J}\}$ on a finite reward set $\mathcal{A}$. Perfect monitoring is assumed, namely, at the end of each stage n both decision makers observe and remember the actions (i_n, j_n) and the reward a_n. Rewards accumulate to form the total n-stage reward $\sum_{t=1}^{n} a_t$.

We denote by Σ and $\mathcal{T}$ the sets of (behavioral) strategies for DM1 and DM2, respectively. Thus, decisions at stage n may depend on the history $h_{n-1} = \{i_t, j_t, a_t\}_{t=1}^{n-1}$, and randomized actions may be used. DM1 does not know the value of the true parameter θ_0 (except that it belongs in Θ), so that his strategies cannot depend on θ_0. Such dependence is allowed for DM2 (although it does not appear explicitly in the problem formulation below). For each triplet (θ_0, σ, τ) in $\Theta \times \Sigma \times \mathcal{T}$, let $P_{\theta_0}^{\sigma,\tau}$ and $E_{\theta_0}^{\sigma,\tau}$ denote the induced probability measure and expectation on the actions–rewards process. Further notations include x_n and y_n – the randomized actions of DM1 and DM2 at stage n, $v(\theta)$ – the minimax value of the matrix game $G(\theta)$, and $\overline{A}_\theta(i,j) = \sum_{a\in\mathcal{A}} a \cdot p_{\theta,i,j}(a)$ – the expected reward in $G(\theta)$ given actions (i,j).

The performance measure for DM1 will be defined in terms of the relative loss. For fixed σ, τ, θ_0 and $n \geq 1$ define the *relative loss* $L_n^{\sigma,\tau}(\theta_0)$ and the *worst-case relative loss* $L_n^{\sigma}(\theta_0)$ (WCRL) by

$$L_n^{\sigma}(\theta_0) \stackrel{\Delta}{=} \max_{\tau} L_n^{\sigma,\tau}(\theta_0) \stackrel{\Delta}{=} \max_{\tau} E_{\theta_0}^{\sigma,\tau}(nv(\theta_0) - \sum_{t=1}^{n} a_t)\,. \tag{1.1}$$

For each strategy σ, the WCRL represents the deficiency in the worst-case (over all strategies of DM2) expected total reward as compared with the complete-information minimax value of the n-stage game. It is important to note that L_n^{σ} depends on the (unknown) parameter θ_0. It therefore supplies, for each fixed n and σ, a *vector* of performance measures, which contains one entry for each possible parameter, and which DM1 would ideally like to minimize (reduce to zero) simultaneously.

While the previous paper [7] focused on performance of strategies of relatively simple structure, the present paper is concerned with asymptotic (long-term) optimality. In defining a meaningful sense of optimality, we shall follow the asymptotic theory introduced by Lai and Robbins [5] in relation with the statistical multiarmed bandit problem, and its extension in [1] to controlled i.i.d. processes. The idea is that the *rate of increase* of $L_n^{\sigma}(\theta_0)$ may be simultaneously minimized for all θ_0. First, a lower bound on the rate of increase of $L_n^{\sigma}(\theta_0)$ will be established, which is logarithmic in n. More precisely, the bound holds for any strategy σ which is *uniformly good*, i.e., achieves a "satisfactory" (and achievable) rate of increase for *every* possible value of the true parameter θ_0 (cf. Definition 3.1). It establishes that the rate of increase of $L_n^{\sigma}(\theta_0)$ is at least $b(\theta_0)\log n$, with $b(\theta_0)$ a

non-negative constant which is explicitly specified. We then proceed to construct strategies which are asymptotically optimal, in the sense that they satisfy the lower bound. These strategies are based on the value-biased Certainty Equivalence strategies which were analyzed in [7], but with various modifications which will be required to achieve asymptotic optimality.

In the adaptive control problem [1], which corresponds to the present model without DM2, it was possible to achieve asymptotic optimality by using a standard parameter estimation scheme, modified by adding a special "probing" phase. Probing is performed whenever it seems that insufficient statistical information was obtained, and is done by choosing actions which are solely dedicated to the efficient acquisition of information. In the present game model, such clear separation of an information acquisition phase is no longer possible. Indeed, DM2 may be able to deny all information regarding the true parameter by using, e.g., "non-revealing" actions, so that information acquisition may not be guaranteed by any action of DM1. Instead, a delicate balance must be maintained between information acquisition and immediate rewards. Certain (sub-) strategies which are related to Blackwell's approachability theory [2] will be constructed for that purpose.

The remainder of the paper is organized as follows. In Section 2 we introduce some simplifying assumptions on the model, as well as additional notation. Section 3 contains the asymptotic bound on the WCRL, and the definition of an asymptotically optimal strategy. Construction of such a strategy commences in Section 4, where two classes of sub-strategies are developed. These form the basis for the overall strategy, which is presented in Section 5.

2 Assumptions and Further Notation

In this paper we shall introduce some additional technical assumptions on the model.

Assumption A1: For each $i \in I$ and $j \in J$, the distributions $\{p_{\theta,i,j}(\cdot)\}_{\theta\in\Theta}$ are mutually absolutely continuous. That is, for every θ, θ' and a, $p_{\theta,i,j}(a) = 0$ if and only if $p_{\theta',i,j}(a) = 0$.

Assumption A2: In each matrix game $G(\theta)$, the optimal strategies of DM1 and DM2, denoted x^*_θ and y^*_θ, are unique.

Assumption A3: The values of the matrix games $\{G(\theta)\}$ are distinct, namely $v(\theta) \neq v(\theta')$ for $\theta \neq \theta'$.

These assumptions simplify the presentation of the results and the construction of optimal strategies. The basic methodology of this paper should be applicable without these assumptions; however, the extension is technically not trivial and involves additional technical effort which might obscure the main ideas.

Some additional definitions and basic relations from [7] are next recalled. $\mathcal{P}(\mathcal{I})$ denotes the set of probability vectors over the (finite) set $\mathcal{I}$. For any matrix $M = \{M(i,j)\}$, the following notation denotes averaging over rows or columns: $M[x,j) \stackrel{\Delta}{=} \sum_i x_i M(i,j)$, $M[x,y] \stackrel{\Delta}{=} \sum_{i,j} x_i y_j M(i,j)$, and similarly for $M(i,y]$.

The *one-stage relative loss* is defined by $d_\theta(i,j) = v(\theta) - \overline{A}_\theta(i,j)$. In this notation the (total)

relative loss may be written as

$$L_n^{\sigma,\tau}(\theta_0) \;=\; E_{\theta_0}^{\sigma,\tau} \sum_{t=1}^{n} d_{\theta_0}(i_t,\, j_t)\,, \tag{2.1}$$

where $d_{\theta_0}(i_t, j_t)$ may be replaced (by appropriate conditioning) by $d_{\theta_0}[x_t, j_t)$, $d_{\theta_0}(i_t, y_t]$ or $d_{\theta_0}[x_t, y_t]$.

The *log-likelihood ratio* equals

$$\Lambda_n(\theta, \theta') = \sum_{t=1}^{n} \log \frac{p_{\theta,i_t,j_t}(a_t)}{p_{\theta',i_t,j_t}(a_t)}\,. \tag{2.2}$$

The corresponding *information divergence* (or *Kullback Leibler information*) is defined by

$$I_{\theta,\theta'}(i,j) = \sum_{a\in\mathcal{A}} p_{\theta,i,j}(a) \log \frac{p_{\theta,i,j}(a)}{p_{\theta',i,j}(a)} \tag{2.3}$$

It is always true that $I_{\theta,\theta'} \geq 0$, and under assumption A1 $I_{\theta,\theta'}$ is finite. Thus we need not introduce a truncated version, as done in [7].

The parameters in Θ are assumed ordered according to the values $v(\theta)$, so that $\theta > \theta'$ stands for $v(\theta) > v(\theta')$, and $\theta \geq \theta'$ for $v(\theta) \geq v(\theta')$. Finally, $\|\cdot\|$ denotes the Euclidean norm, and $\|\cdot\|_\infty$ the sup-norm.

3 A Lower Bound on the Loss

In this section we derive an asymptotic lower bound on the WCRL. This will be used to define a meaningful non-Bayesian sense of optimal performance for DM1.

The stated objective of DM1 is to minimize (the rate of increase of) the WCRL. However, in general the WCRL cannot be minimized simultaneously for every possible θ_0. For example, if DM1 plays at every stage his optimal (maximin) strategy in $G(\theta)$ for some fixed $\theta \in \Theta$, then he guarantees zero loss if θ happens to be the true parameter. But if the true parameter is different, his WCRL may grow *linearly* in n.

To exclude such non-adaptive strategies, we shall restrict attention to strategies which perform "reasonably well" for every parameter, as specified in the following definition (compare [5], [1]).

Definition 3.1 *A strategy σ of DM1 is said to be* uniformly good *if for every $\theta_0 \in \Theta$:*

$$L_n^{\sigma}(\theta_0) = o(n^{\alpha}) \quad \text{for every } \alpha > 0\,. \tag{3.1}$$

From [7] we know that the set of uniformly good strategies is non-empty, and in fact there exist strategies which guarantee that the WCRL is $O(\log n)$. Thus, strategies outside this set need not be considered.

For each parameter θ, define an associated set of "bad" parameters (see the end of the section for interpretation of this set and discussion of the lower bound; note that θ is not included in $B(\theta)$):

$$B(\theta) = \{\theta' \in \Theta : \ v(\theta') > v(\theta), \ I_{\theta,\theta'}[x^*_\theta, y^*_\theta] = 0\} \,. \tag{3.2}$$

Since $I_{\theta,\theta'}$ is non-negative, the requirement $I_{\theta,\theta'}[x^*_\theta, y^*_\theta] = 0$ in the last definition is equivalent to: $I_{\theta,\theta'}(i,j) = 0$ for every pair of *relevant actions* in $G(\theta)$, namely $i \in \mathcal{I}^*_\theta \stackrel{\triangle}{=} \{i : (x^*_\theta)_i\} > 0\}$ and $j \in \mathcal{J}^*_\theta \stackrel{\triangle}{=} \{j : (y^*_\theta)_j\} > 0\}$.

Theorem 3.1 *Let $\theta \in \Theta$ be such that $B(\theta) \neq \emptyset$. Then, for every uniformly good strategy σ of DM1,*

$$\liminf_{n\to\infty} \frac{L^\sigma_n(\theta)}{\log n} \geq b(\theta) \,, \tag{3.3}$$

where (defining $0/0 \stackrel{\triangle}{=} \infty$)

$$b(\theta) = \min_{x \in \mathcal{P}(\mathcal{I})} \frac{d_\theta[x, y^*_\theta]}{\min_{\theta' \in B(\theta)} I_{\theta,\theta'}[x, y^*_\theta]} > 0 \,. \tag{3.4}$$

Proof: Consider a fixed $\theta \in \Theta$ such that $B(\theta) \neq \emptyset$. Let $\tau^\theta = \{y^*_\theta\}$ denote the stationary strategy of DM2, in which $y_n = y^*_\theta$ at each stage (note that this strategy does not depend on the true parameter of the game). It will be proved below that for any uniformly good strategy σ

$$\liminf_{n\to\infty} \frac{L^{\sigma,\tau^\theta}_n(\theta)}{\log n} \geq b(\theta) \,, \tag{3.5}$$

which clearly establishes the required bound.

Given that DM2 uses τ^θ, DM1 in effect is facing a "controlled i.i.d. process" of the type considered in [1] with "state" $X_n = (a_n, j_n)$. However, the lower bound from [1] does not apply here. The reason is that our definition of the loss, which is appropriate for the game situation, does not coincide with the one used in [1]. Indeed, in the single-controller problem treated there the relative loss is naturally defined with respect to the maximal one-stage reward (i.e. $\max_i d_{\theta_0}(i, y^*_\theta]$), whereas our definition uses $v(\theta_0)$ in that role. Still, it will be possible to follow the proof of [1] after establishing the next two lemmas. We note that the following one relies on σ being uniformly good against any strategy of DM2 (and not just against τ^θ).

Lemma 3.1 (compare [1, Lemma 3.2]) *Assume that σ is a uniformly good strategy of DM1. Then, for any $\theta' \in B(\theta)$,*

$$P^{\sigma,\tau^\theta}_{\theta'} \left\{ \sum_{t=1}^{n} d_\theta(i_t, y^*_\theta] < K \log n \right\} = o(n^{\alpha-1}) \quad \textit{for every } \alpha > 0 \textit{ and } K > 0 \,. \tag{3.6}$$

Proof: Fix $\theta' \in B(\theta)$. It is first shown that a small loss under θ' implies a large loss under θ (see (3.10) below). Let $\mathcal{I}^*_\theta = \{i \in \mathcal{I} : (x^*_\theta)_i > 0\}$ be the set of relevant actions of DM1 in $G(\theta)$. It is

well known that this is exactly the set of actions which maximize the (expected) reward against y^*_θ ([6], Theorems 3.1.2 and 3.1.16); that is, $\overline{A}_\theta(i,\, y^*_\theta] = v(\theta)$ for $i \in \mathcal{I}^*_\theta$, and $\overline{A}_\theta(i,\, y^*_\theta] < v(\theta)$ for $i \notin \mathcal{I}^*_\theta$. Consequently,

$$d_\theta(i,\, y^*_\theta] \equiv v(\theta) - \overline{A}_\theta(i, y^*_\theta] = 0 \quad \text{for} \quad i \in \mathcal{I}^*_\theta\,, \tag{3.7}$$

and, since the action set is finite, there exists a positive constant δ_1 such that

$$d_\theta(i,\, y^*_\theta] \geq \delta_1 > 0 \quad \text{for} \quad i \notin \mathcal{I}^*_\theta\,. \tag{3.8}$$

Consider now the game $G(\theta')$. By definition, $\theta' \in B(\theta)$ implies $I_{\theta,\theta'}[x^*_\theta, y^*_\theta] = 0$ and $v(\theta') > v(\theta)$. Noting (3.7),

$$d_{\theta'}(i, y^*_\theta] \;=\; v(\theta') - \overline{A}_{\theta'}(i, y^*_\theta] \;=\; v(\theta') - \overline{A}_\theta(i, y^*_\theta] \;=\; v(\theta') - v(\theta) \;\stackrel{\triangle}{=}\; \delta_o \;>\; 0\,, \qquad i \in \mathcal{I}^*_\theta\,. \tag{3.9}$$

Denoting $D = \max_i d_{\theta'}(i, y^*_\theta]$, it follows from (3.7)–(3.9) that for any $m \geq 1$,

$$\begin{aligned} \sum_{t=1}^{m} d_{\theta'}(i_t, y^*_\theta] \;&\geq\; \delta_o m - (D + \delta_o) \sum_{t=1}^{m} 1\{i_t \notin \mathcal{I}^*_\theta\} \\ &\geq\; \delta_o m - \delta_2 \sum_{t=1}^{m} d_\theta(i_t, y^*_\theta]\,, \end{aligned} \tag{3.10}$$

where $\delta_2 \stackrel{\triangle}{=} (D + \delta_o)/\delta_1 > 0$. Thus, using again the fact that $d_\theta(i, y^*_\theta] \geq 0$ by optimality of y^*_θ (cf. (3.7) and (3.8)),

$$\begin{aligned} P^{\sigma,\tau^\theta}_{\theta'} \left\{ \sum_{t=1}^{n} d_\theta(i_t, y^*_\theta] < K \log n \right\} \;&=\; P^{\sigma,\tau^\theta}_{\theta'} \left\{ \sum_{t=1}^{m} d_\theta(i_t, y^*_\theta] < K \log n\,,\;\; \forall m \leq n \right\} \\ &\leq\; P^{\sigma,\tau^\theta}_{\theta'} \left\{ \sum_{t=1}^{m} d_{\theta'}(i_t, y^*_\theta] \geq \delta_o m - \delta_2 K \log n\,,\;\; \forall m \leq n \right\} \\ &\stackrel{\triangle}{=}\; P^{\sigma,\tau^\theta}_{\theta'}\{A_n\}\,, \end{aligned} \tag{3.11}$$

where A_n denotes the corresponding event. To establish the lemma, it remains to show that the last probabilities decay as $o(n^{\alpha-1})$. (Note that application of Chebycheff's inequality, as in the proof of Lemma 3.2 in [1], is impossible here since the loss $\sum_1^n d_{\theta'}(i_t, y^*_\theta]$ may be negative.) Fix $n \geq 1$, and consider a the following strategy τ' of DM2. First define a stopping time T by

$$T = \min\{1 \leq m \leq n \;:\; \sum_{t=1}^{m} d_{\theta'}(i_t, y^*_\theta] < \delta_o m - \delta_2 K \,\log n\,\}\,,$$

and $T = n$ if the minimized set is empty. Define τ' as the strategy which chooses $y_t = y^*_\theta$ for $t \leq T$, and $y_t = y^*_{\theta'}$ thereafter. Since τ^θ and τ' coincide on the event A_n,

$$P^{\sigma,\tau^\theta}_{\theta'}\{A_n\} = P^{\sigma,\tau'}_{\theta'}\{A_n\}\,. \tag{3.12}$$

Also, noting that $d_{\theta'}(i, y^*_{\theta'}] \geq 0$ (by optimality of $y^*_{\theta'}$ in $G(\theta')$) and by definitions of T and A_n (note in particular that $T = n$ on A_n), we obtain under τ':

$$\begin{aligned}
\sum_{t=1}^{n} d_{\theta'}(i_t, y_t] &= \sum_{t=1}^{T} d_{\theta'}(i_t, y^*_\theta] + \sum_{t=T+1}^{n} d_{\theta'}(i_t, y^*_{\theta'}] \\
&\geq \sum_{t=1}^{T-1} d_{\theta'}(i_t, y^*_\theta] + d_{\theta'}(i_T, y^*_\theta] \\
&\geq -\delta_2 K \log n - D' + \delta_o n \, \mathbf{1}\{A_n\}\,, \qquad P^{\sigma,\tau'}_{\theta'}\text{-a.s.}\,, \qquad (3.13)
\end{aligned}$$

where $D' = \max_i |d_{\theta'}(i, y^*_\theta]|$. Now, since σ is uniformly good, it follows by (3.1), (1.1), (2.1) and (3.13) that for every $\alpha > 0$:

$$\begin{aligned}
o(n^\alpha) &= L^\sigma_n(\theta') \geq L^{\sigma,\tau'}_n(\theta') = E^{\sigma,\tau'}_{\theta'}\left(\sum_{t=1}^{n} d_{\theta'}(i_t, y_t]\right) \\
&\geq -\delta_2 K \log n - D' + \delta_o n \, P^{\sigma,\tau'}_{\theta'}\{A_n\}\,, \qquad (3.14)
\end{aligned}$$

so that:

$$P^{\sigma,\tau'}_{\theta'}\{A_n\} \leq \frac{o(n^\alpha) + \delta_2 K \log n + D'}{\delta_o n} = o(n^{\alpha-1})\,. \qquad (3.15)$$

The lemma follows from (3.11), (3.12) and (3.15). □

Lemma 3.2 *Assume $B(\theta) \neq \emptyset$. Then*

(*i*) $0 < b(\theta) < \infty$.

(*ii*) *The minimization in the definition (3.4) of $b(\theta)$ can alternatively be taken over the set* $X(\theta) = \{x \in \mathcal{P}(\mathcal{I}) : x_i = 0 \text{ if } (x^*_\theta)_i > 0\}$.

Proof: The inequalities $b(\theta) > 0$ and $b(\theta) < \infty$ follow from the following facts (a) and (b), respectively:

(a) For every $x \in \mathcal{P}(\mathcal{I})$, $d_\theta[x, y^*_\theta] = 0$ implies $I_{\theta,\theta'}[x, y^*_\theta] = 0$ for every $\theta' \in B(\theta)$. Indeed, fixing x, take any i for which $x_i > 0$. Recall that $d_\theta[x, y^*_\theta] = \sum_i x_i d_\theta(i, y^*_\theta]$, and $d_\theta(i, y^*_\theta] \geq 0$ by optimality of y^*_θ in $G(\theta)$. Thus $d_\theta[x, y^*_\theta] = 0$ implies that $d_\theta(i, y^*_\theta] = 0$. By (3.8) it follows that $i \in \mathcal{I}^*_\theta$, i.e. $(x^*_\theta)_i > 0$. But, by definition of $B(\theta)$, this implies that $I_{\theta,\theta'}(i, y^*_\theta] = 0$ for every $\theta' \in B(\theta)$.

(b) $\min_{\theta' \in B(\theta)} I_{\theta,\theta'}[x, y^*_\theta] > 0$ for some $x \in \mathcal{P}(\mathcal{I})$. To show that, note that for every $\theta' \in B(\theta)$,

$$\overline{A}_{\theta'}[x^*_{\theta'}, y^*_\theta] \geq v(\theta') > v(\theta) \geq \overline{A}_\theta[x^*_{\theta'}, y^*_\theta]\,,$$

i.e. rewards under θ are different from those under θ', which implies $I_{\theta,\theta'}[x^*_{\theta'}, y^*_\theta] \neq 0$. Thus fact (b) is satisfied by choosing x as a convex combination of $\{x^*_{\theta'} : \theta' \in B(\theta)\}$.

Item (ii) of the lemma follows since $i \in \mathcal{I}_\theta^*$ implies that $d_\theta(i, y_\theta^*] = 0$ (see (3.7)), and that $I_{\theta,\theta'}(i, y_\theta^*] = 0$ for all $\theta' \in B(\theta)$ (by definition of $B(\theta)$). $\square$

Based on these lemmas, the proof of Theorem 3.1 may be concluded exactly as the proof of the lower bound in [1]. For the reader's convenience the main steps will be outlined here using the present notations. Fixing a uniformly good strategy σ, our objective is to establish (3.5). Fix $\rho > 0$, and for each $n \geq 1$ define the event

$$A_n \;=\; \left\{ \sum_{t=1}^{n} d_\theta(i_t, y_\theta^*] < \frac{b(\theta)}{1+2\rho} \log n \right\} .$$

Recalling that $d_\theta(i, y_\theta^*] \geq 0$, we obtain

$$L_n^{\sigma,\tau^\theta}(\theta) \;=\; E_\theta^{\sigma,\tau^\theta} \sum_{t=1}^{n} d_\theta(i_t, y_\theta^*] \;\geq\; (1 - P_\theta^{\sigma,\tau^\theta}\{A_n\}) \frac{b(\theta)}{1+2\rho} \log n \,.$$

Thus, since $\rho > 0$ is arbitrary, to establish (3.5) it is sufficient to prove that $P_\theta^{\sigma,\tau^\theta}\{A_n\} \to 0$, to which we proceed.

Let B denote the number of elements in $B(\theta)$. Denote $P_{\theta'} = P_{\theta'}^{\sigma,\tau^\theta}$, and $P_B = B^{-1} \sum_{\theta' \in B(\theta)} P_{\theta'}$. Consider the following change of measure, for any event D_n measurable on the sigma algebra H_n generated by $\{i_t, j_t, a_t\}_{t=1}^n$:

$$P_\theta\{D_n\} = \int_{D_n} \frac{dP_\theta}{dP_B} dP_B \leq \int_{D_n} B \min_{\theta' \in B(\theta)} \frac{dP_\theta}{dP_{\theta'}} dP_B = B \int_{D_n} \min_{\theta' \in B(\theta)} \exp\{\Lambda_n(\theta, \theta')\} dP_B \,,$$

where Λ_n is the log-likelihood ratio (2.2). Note that $\Lambda_n(\theta, \theta')$ is the sum of the (controlled i.i.d.) random variables $X_t \stackrel{\Delta}{=} \log(p_{\theta,i_t,j_t}/p_{\theta',i_t,j_t})$, with conditional expectation $E_\theta^{\sigma,\tau^\theta}(X_t | i_t = i) = I_{\theta,\theta'}(i, y_\theta^*]$. It follows from, e.g., Lemma 3.1 in [1] that for every $\epsilon > 0$ and $\rho > 0$ there exist a constant $K(\epsilon, \rho)$ and an event $A(\epsilon, \rho)$ with $P_\theta\{A(\epsilon, \rho)\} > 1 - \epsilon$, such that on this event

$$\Lambda_n(\theta, \theta') \;\leq\; (1+\rho)\, n\, I_{\theta,\theta'}[\hat{x}_n, y_\theta^*] + K(\epsilon, \rho) \,, \quad \forall\, n \geq 1 \,,\; \theta' \in B(\theta) \,,$$

where $\hat{x}_n(i) = n^{-1} \sum_{t=1}^n \mathbf{1}\{i_t = i\}$ denotes the empirical distribution of the actions $\{i_t\}$; note that $I_{\theta,\theta'}[\hat{x}_n, y_\theta^*] = \sum_i \hat{x}_n(i) I_{\theta,\theta'}(i, y_\theta^*]$. Since $\hat{x}_n \in \mathcal{P}(\mathcal{I})$, it follows from the definition of $b(\theta)$ in (3.4) that

$$\min_{\theta' \in B(\theta)} I_{\theta,\theta'}[\hat{x}_n, y_\theta^*] \;=\; d_\theta[\hat{x}_n, y_\theta^*] \frac{\min_{\theta' \in B(\theta)} I_{\theta,\theta'}[\hat{x}_n, y_\theta^*]}{d_\theta[\hat{x}_n, y_\theta^*]} \;\leq\; d_\theta[\hat{x}_n, y_\theta^*]\; b(\theta)^{-1} \;=\; \frac{1}{n} \sum_{t=1}^{n} d_\theta(i_t, y_\theta^*]\; b(\theta)^{-1} \,.$$

Thus, noting the definitions of A_n and $A(\rho, \epsilon)$,

$$P_\theta\{A_n \cap A(\rho, \epsilon)\} \leq B \exp\{(1+\rho) \frac{\log n}{1+2\rho} + K(\epsilon, \rho)\} P_B\{A_n\} = B e^{K(\epsilon,\rho)} n^{\frac{1+\rho}{1+2\rho}} P_B\{A_n\} \,,$$

which converges to 0 as a consequence of Lemma 3.1. Finally, letting $\epsilon \to 0$ establishes $P_\theta\{A_n\} \to 0$. The proof of Theorem 3.1 is thus complete. $\square$

Theorem 3.1 provides a lower bound on the asymptotic WCRL for those parameters which satisfy $B(\theta) \neq \emptyset$. Therefore, the best performance that DM1 can hope for (in terms of the asymptotic increase rate of the WCRL) is to achieve the lower bound for those parameters for which it applies, while keeping the WCRL *finite* for the rest. This leads to the following definition of asymptotic optimality.

Definition 3.2: A strategy σ of DM1 is said to be *asymptotically optimal* if

(i) $\limsup_{n\to\infty} L_n^\sigma(\theta_0) < \infty$ for every $\theta_0 \in \Theta$ s.t. $B(\theta_0) = \emptyset$,

(ii) $\limsup_{n\to\infty} \frac{L_n^\sigma(\theta_0)}{\log n} = b(\theta_0)$ for every $\theta_0 \in \Theta$ s.t. $B(\theta_0) \neq \emptyset$.

Discussion: The lower bound of Theorem 3.1 can be rendered a simplified but useful heuristic interpretation, in accordance with [5]. Suppose that DM2 uses the strategy $\tau^\theta = \{y_\theta^*\}$ for some θ with $B(\theta) \neq \emptyset$. Suppose that DM1 has (statistical) indications that θ is the true parameter. If this is indeed the case, to achieve zero relative loss he must choose his actions in the relevant set $\mathcal{I}_\theta^*$. Unfortunately, this may lead to undesired consequences if in fact some $\theta' \in B(\theta)$ is the true parameter. Since $I_{\theta,\theta'}(i, x_\theta^*] = 0$ for every $i \in \mathcal{I}_\theta^*$ and $\theta' \in B(\theta)$ (by definition of the latter), these actions do not yield any statistical information for discriminating θ from θ'. Furthermore, under θ' positive relative loss will be incurred at each stage (cf. (3.9)), leading to $O(n)$ WCRL.

Therefore, in a uniformly good strategy (against τ^θ), DM1 must "probe" the system by playing outside $\mathcal{I}_\theta^*$. To minimize the associated relative loss, he should choose a probing strategy which gives the best "loss to information ratio". This is the essence of the constant $b(\theta)$, where information is quantified by the Kullback-Leibler information.

The lower bound (3.3) may now be interpreted as follows. For a strategy of DM1 to be uniformly good (against τ^θ), if θ is the true parameter he must maintain his total information (i.e., a measure of statistical value of the data for discriminating θ and $B(\theta)$, related to the Kullback-Leibler information) at a level of $\log n$ at least. By performing the required probing optimally, he can keep the probing loss down to $b(\theta)\log n$.

4 Optimal Strategies: Preliminary Results

4.1 Discussion and Results

This section is an intermediate step in the construction of an asymptotically optimal strategy. This strategy will essentially be based on the *Certainty Equivalence strategy with biased MLE* which was introduced in [7]. To indicate the required modifications in this basic strategy, we start by recalling its definition and performance. This will expose its deficiencies as compared with the required optimal performance. Two families of (sub-) strategies will be introduced to overcome these deficiencies. These strategies are not in themselves adaptive, i.e., each is designed with a specific parameter θ in mind. They will however be used as building blocks for the overall (adaptive) optimal strategy, to be presented in the next section.

Recall the following definitions from [7]. The maximum likelihood estimator (MLE) $\hat{\theta}_n$ is the maximizer of the likelihood function $\lambda_{n-1}(\theta) = \prod_{t=1}^{n-1} p_{\theta,i_t,j_t}(a_t)$. For some fixed $Q > 1$, define the sequence

$$K_n = n(\log n)^Q + 1\,. \tag{4.1}$$

(this is the "smallest" sequence which satisfies requirements (5.1) in [7]). Further define the *likely parameters set*:

$$\hat{\Theta}_n = \{\theta \in \Theta : \Lambda_{n-1}(\hat{\theta}_n, \theta) \leq \log K_n\}\,, \tag{4.2}$$

and the *value-biased MLE*:

$$\overline{\theta}_n \;=\; \arg\max\{v(\theta) : \theta \in \hat{\Theta}_n\}\,. \tag{4.3}$$

The Certainty Equivalence strategy with biased MLE, denoted $\sigma 2$, is simply specified by $x_n = x^*(\overline{\theta}_n)$. The following results have been established for this strategy ([7], Theorems 5.1 and 5.2):

Theorem 4.1 *For every $\theta_0 \in \Theta$,*

(*i*) $L_n^{\sigma 2}(\theta_0) \leq O(\log n)$.

(*ii*) *Assume that $B_2(\theta_0) = \emptyset$, where $B_2(\theta_0) \stackrel{\Delta}{=} \{\theta' \in \Theta : I_{\theta_0,\theta'}[x^*_\theta, j) = 0$ for some $j \in \mathcal{J}^*_{\theta_0}\}$. Then $L_n^{\sigma 2}(\theta_0)$ is bounded.*

Note that the requirement $B_2(\theta) = \emptyset$ is equivalent, under Assumption A3, to condition C_2 of [7]. It may be readily verifies that $B_2(\theta) \supset B(\theta)$, hence $B_2(\theta_0) = \emptyset$ implies $B(\theta) = \emptyset$, so that the last result is compatible with the lower bound of Section 3.

Compared with the definition of asymptotic optimality, the performance guaranteed by σ_2 falls short in the following two cases:

I. $B(\theta_0) = \emptyset$, but $B_2(\theta_0) \neq \emptyset$. Asymptotic optimality demands that the WCRL be bounded. However, $\sigma 2$ guarantees only $O(\log n)$ WCRL in this case.

II. $B(\theta_0) \neq \emptyset$. Then $B_2(\theta_0) \neq \emptyset$, and again $\sigma 2$ guarantees an $O(\log n)$ WCRL. However, it does not guarantee that the optimal coefficient $b(\theta_0)$ of the lower bound is achieved.

Consider case I. We shall identify the key properties which enabled to bound the WCRL for the strategy $\sigma 2$ when $B_2(\theta_0) = \emptyset$, and then attempt to guarantee similar properties (by appropriate strategies) under the weaker condition $B(\theta_0) = \emptyset$. For any parameter θ, consider those times when the estimator $\bar{\theta}_n = \theta$. According to $\sigma 2$, x^*_θ is played at these times. Now the key properties which were used in the proof of Theorem 4.1(*ii*) are the following relations between loss (or reward) and information:

(*a*) $d_\theta[x^*_\theta, j) \leq 0$ for all j.

(*b*) $B_2(\theta) = \emptyset$ implies $d_\theta[x^*_\theta, j) \leq -\delta + M \min_{\theta' > \theta} I_{\theta,\theta'}[x^*_\theta, j)$.

(*c*) $d_{\theta'}[x^*_\theta, j) \leq -\delta + M\, I_{\theta',\theta}[x^*_\theta, j)$ for every $\theta' < \theta$.

(The first property is of course a consequence of optimality of x^*_θ in $G(\theta)$, and the other two were established in [7], Lemma 6.1.) The interpretation in the context of $\sigma 2$ is as follows. Assume that

x^*_θ is played (which occurs at the times when $\bar{\theta}_n = \theta$). If θ happens to be the true parameter, then (a) guarantees non-positive loss, and moreover, by (b), if no information is attained with respect to some $\theta' > \theta$ (i.e. $I_{\theta,\theta'}$ is low), this will be compensated by strictly negative loss. Also, if some $\theta' < \theta$ happens to be the true parameter, then (c) low $I_{\theta',\theta}$-information is compensated by strictly negative loss.

Unfortunately, property (b) does not hold if $B_2(\theta) \neq \emptyset$, which may be easily seen from the definition. Nonetheless, as long as the (smaller) set $B(\theta_0)$ is empty, a generalized version of these properties may still be achieved. This requires to deviate from playing x^*_θ whenever θ is the estimated parameter, and instead use a modified (non-stationary, history-dependent) strategy over these times. The precise formulation is the content of the following proposition.

Proposition 4.1 *There exist strategies $\{\sigma^*(\theta) \in \Sigma : \theta \in \Theta\}$ and positive constants M_1 and δ_1 such that, for every strategy τ of DM2 and every $n \geq 1$, the following hold:*

(i) $\sum_{t=m}^{n} d_\theta[x_t, j_t] \leq M_1, \quad \forall\, 1 \leq m \leq n\,.$

(ii) $\sum_{t=1}^{n} d_\theta[x_t, j_t] \leq -\delta_1 n + M_1 + M_1 \min_{\theta' \in G_o(\theta)} \sum_{t=1}^{n} I_{\theta,\theta'}[x_t, j_t]\,, \quad$ *where* $G_o(\theta) = \{\theta' :\, \theta' > \theta\} - B(\theta).$

(iii) $d_{\theta'}[x_t, j_t] \leq -\delta_1 + M_1 I_{\theta',\theta}[x_t, j_t]$ *for every* $\theta' < \theta$.

Remark 4.1: The relations in Proposition 4.1, as well as in the rest of this section, hold in a sample-path sense, and are independent of the true parameter θ_0. Indeed, all quantities (d_θ etc.) are deterministic functions of the actions, and θ_0 does not appear. Note, moreover, that explicit dependence on DM1's actions is only through x_n (and not i_n). Accordingly, all (non-stationary) strategies of DM1 which appear in this section may be expressed as functions of DM2's actions only; that is, $\sigma_n = \sigma_n(j_1, \ldots, j_{n-1})$.

Remark 4.2: Properties (i)–(iii) are a generalization of properties (a)–(c) which were pointed out above. In fact, when $B_2(\theta) = \emptyset$ then $\sigma^*(\theta)$ may simply be taken as the stationary strategy $\{x^*_\theta\}_{n \geq 1}$.

Remark 4.3: Note that (i) bounds the relative loss over any time interval $[m, n]$, and not just $[1, n]$. This will be essential for the results of the next section (cf. the proof of Lemma 5.1).

The proof, as well as definition of the strategy $\sigma^*(\theta)$, are presented in the second part of this section. The main idea in constructing this strategy is as follows. The negation of property (ii) above (or of $B_2(\theta) = \emptyset$) may be written as:

$$I_{\theta,\theta'}[x^*_\theta, y] = 0 \;\text{ and }\; d_\theta[x^*_\theta, y] = 0 \quad \text{for some } \theta' > \theta \text{ and } y \in \mathcal{P}(\mathcal{J})\,. \tag{4.4}$$

However, when $B(\theta_0) = \emptyset$, (4.4) cannot hold for $y = y^*_\theta$. In other words, (4.4) is then satisfied only if DM2 plays a strategy y which not his optimal in the matrix game $G(\theta)$. Thus, *if* that y was known in advance to DM1, he could achieve an expected reward greater than $v(\theta)$ (i.e., strictly negative relative loss) in the matrix game $G(\theta)$. When the game is repeated, a similar effect can be achieved (in the long run) by DM1 even if the y_n's are unknown in advance; see Proposition 4.3 below.

Let us turn to the second deficiency noted above, namely case II. As discussed at the end of the last section, to achieve the optimal coefficient $b(\theta_0)$, DM1's strategy should include an "optimal probing" phase. This phase is intended to accumulate statistical information (quantified by $I_{\theta_0,\theta}$ for $\theta \in B(\theta)$) at a minimal loss-per-information ratio. Also, some safeguards should be activated if (due to DM2's actions) insufficient information is revealed.

If DM2 is playing $y_n = y^*_{\theta_0}$ at every stage, then such "optimal probing" may be achieved on a single-stage basis by any $x^o \in \mathcal{P}(\mathcal{I})$ which is a minimizer in (3.4). (This is trivially satisfied in the single-controller case; cf. [1].) However, since DM2 may play differently, then a stationary strategy $x_n \equiv x^o$ might not yield the desired result: The loss-per-information ratio may then be larger than $b(\theta_0)$, or possibly no information will be obtained.

Again, the problem will be resolved by "punishing" DM2 for playing off $y^*_{\theta_0}$. This can in principle be accomplished by superimposing the one-stage probing strategy x^o on a strategy similar to $\sigma^*(\theta_0)$ of Proposition 4.1. The following result may be thus obtained:

Proposition 4.2 *Let $\theta \in \Theta$ be such that $B(\theta) \neq \emptyset$. Then there exists a strategy $\sigma^o(\theta)$ of DM1 and positive constants M_2, δ_2 such that, for every $\tau \in \mathcal{T}$ and $n \geq 1$,*

(i) For every $\epsilon > 0$ and $1 \leq m \leq n$,

$$\sum_{t=m}^{n} d_\theta[x_t, j_t] \leq (1+\epsilon)\, b(\theta) \min_{\theta' \in B(\theta)} \sum_{t=1}^{n} I_{\theta,\theta'}[x_t, j_t] + M(\epsilon)\,,$$

where $b(\theta)$ is defined in (3.4), and $M(\epsilon) > 0$ is a constant which depends only on ϵ.

(ii) $$\sum_{t=1}^{n} d_\theta[x_t, j_t] \leq -\delta_2 n + M_2 + M_2 \min_{\theta' > \theta} \sum_{t=1}^{n} I_{\theta,\theta'}[x_t, j_t]\,.$$

(iii) $$d_{\theta'}[x_t, j_t] \leq -\delta_2 + M_2 I_{\theta',\theta}[x_t, j_t] \quad \text{for every } \theta' < \theta.$$

The bounds in Proposition 4.2 may be roughly interpreted as follows. (i) implies that the information-per-loss ratio (with respect to $B(\theta)$) is close to optimal, provided that information is indeed accumulated (say, at an $O(n)$ rate). Item (ii) implies that if the information rate (with respect to any $\theta' > \theta$, and in particular for $\theta' \in B(\theta)$) is smaller than some critical linear rate, then a strictly negative loss results; compare with (ii) of Proposition 4.1. Finally, (iii) is analogous to Proposition 4.1(iii) or property (c) above.

4.2 Proofs

The proofs of Propositions 4.1 and 4.2 depend on a basic result for repeated matrix games, established by different methods in [4] and [3], which essentially states the following. In a (complete information) repeated matrix game, each player can asymptotically guarantee for himself an average reward which is no less than what he could guarantee if he knew in advance the empirical frequencies of his opponent's actions. The following (somewhat non-standard) version of this result will be required here:

Proposition 4.3

(i) For every $\theta \in \Theta$, there exists a strategy $\overline{\sigma}(\theta)$ of DM1 such that

$$\frac{1}{n}\sum_{t=1}^{n} \overline{A}_\theta[x_t, j_t] \geq \max_{x \in \mathcal{P}(\mathcal{I})} \overline{A}_\theta[x, \overline{y}_n] - \frac{B}{\sqrt{n}}, \quad \forall \tau \in \mathcal{T},\ n \geq 1,$$

where B is a positive constant, and $\overline{y}_n$ is the empirical distribution of DM2's actions, namely $\overline{y}_n = n^{-1}\sum_{t=1}^{n} e_{j_t}$ with e_j the unit vector with 1 in the j's entry.

(ii) The strategy $\overline{\sigma}(\theta)$ may be defined as follows. Let

$$Q = \{(a, y) \in \mathbb{R} \times \mathcal{P}(\mathcal{J}) :\ a \geq \max_{x \in \mathcal{P}(\mathcal{I})} \overline{A}_\theta[x, y]\},$$

and consider a point $(a, y) \in \mathbb{R} \times \mathcal{P}(\mathcal{J})$ such that $(a, y) \notin Q$. Let c denote the closest point in Q to (a, y), and $(\alpha, \xi) = c - (a, y)$. Finally, let $x^*(a, y)$ be an optimal (maximin) strategy of DM1 in the matrix game with augmented payoff matrix: $A^{(\alpha,\xi)} \stackrel{\Delta}{=} \overline{A}_\theta + \underline{1}'\xi \equiv (\overline{A}_\theta(i, j) + \xi_j)$. Then

$$\overline{\sigma}(\theta)_n(h_{n-1}) = \begin{cases} x^*(\overline{a}_{n-1}, \overline{y}_{n-1}) & \text{if } n \geq 2 \text{ and } (\overline{a}_{n-1}, \overline{y}_{n-1}) \notin Q \\ \text{arbitrary} & \text{otherwise} \end{cases}$$

where $\overline{a}_n = n^{-1}\sum_{t=1}^{n} \overline{A}_\theta[x_t, j_t)$, and $\overline{y}_n$ as defined in (i).

Proof: As observed in [3], the proof follows by applying general approachability results ([2]) to the set Q. Although the approachability result required here is not standard (in that the one-stage payoff depends directly on x_t instead of i_t, and a.s. relations are required), it may be easily inferred from the version which appears, e.g., in [8]. A direct proof is supplied in the Appendix. □

We note that the strategy $\overline{\sigma}(\theta)$ as defined in (ii) depends on the history only through DM2's actions, since $\overline{A}_\theta$ is deterministic, and x_t may be recursively eliminated from the equations. That is, $\overline{\sigma}(\theta)_n(h_{n-1}) = f_n(j_1, \cdots, j_{n-1})$.

The following lemma will also be required:

Lemma 4.1 *Let A be an $|\mathcal{I}| \times |\mathcal{J}|$ zero-sum game matrix with value $v(A)$. Assume that y^* is a unique optimal (minimax) strategy for DM2. Then for some $\delta > 0$ and every $y \in \mathcal{P}(\mathcal{J})$:*

$$\max_{x \in \mathcal{P}(\mathcal{I})} A[x, y] \geq v(A) + \delta \|y - y^*\| \tag{4.5}$$

Proof: Consider $f(y) \stackrel{\Delta}{=} \max_x A[x, y]$, $y \in \mathcal{P}(\mathcal{J})$. Since y^* is a unique optimal strategy, it follows that $f(y^*) = v(A)$ and $f(y) > v(A)$ for $y \neq y^*$, so that y^* is the unique minimizer of f. Note further that $f(y) = \max_i A(i, y]$, so that f is the maximum of a finite number of linear functions. The inequality (4.5) is an easy consequence of these facts. □

The next lemma will be useful in establishing property (iii) in Propositions 4.1 and 4.2:

Lemma 4.2 *There exist a (small enough) constant* $0 < \mu \leq 1/2$ *and positive constants* δ_3, M_3 *such that* $\|x - x^*_\theta\|_\infty \leq \mu$ *implies*

$$d_{\theta'}[x,j] \leq -\delta_3 + M_3 I_{\theta',\theta}[x,j], \quad \forall\, j, \theta, \theta' < \theta\,.$$

Proof: By Lemma 6.1(i) in [7], there exist positive δ and M such that for every j and $\theta' < \theta$: $d_{\theta'}[x^*_\theta, j) \leq -\delta + M I_{\theta',\theta}[x^*_\theta, j)$. (This is exactly property (a) discussed at the beginning of this section.) The lemma follows by continuity of $d_{\theta'}$ and $I_{\theta,\theta'}$ in x. □

We proceed now to the proof of Proposition 4.1. It will convenient to use in the remainder of this section the abbreviated notation:

$$d_\theta\{m:n\} \stackrel{\triangle}{=} \sum_{t=m}^{n} d_\theta[x_t, j_t)\,, \quad I_{\theta,\theta'}\{m:n\} \stackrel{\triangle}{=} \sum_{t=m}^{n} I_{\theta,\theta'}[x_t, j_t)\,.$$

Also recall that $G_o(\theta) = \{\theta' > \theta\} - B(\theta)$. As an intermediate step, the following type of strategies is required:

Lemma 4.3 *For each* $\theta \in \Theta$, *there exists a strategy* $\sigma^1(\theta)$ *for DM1 and positive constants* M_4, δ_4, ϵ_4 *such that the following hold for every* $\tau \in \mathcal{T}$ *and* $n \geq 1$.

(i) $d_\theta\{1:n\} \leq M_4$.

(ii) $\min_{\theta' \in G_o(\theta)} I_{\theta,\theta'}\{1:n\} \leq \epsilon_4 n$ *implies* $d_\theta\{1:n\} \leq -\delta_4 \sqrt[4]{n} + M_4$.

(iii) $\|x_n - x^*_\theta\|_\infty \leq \mu$, *where* μ *is as in Lemma 4.2.*

Remark: The strategy $\sigma^1(\theta)$ will be based on the strategy $\overline{\sigma}(\theta)$ of Proposition 4.3. However, an essential improvement is property (i), i.e. bounded WCRL. In contrast, in Proposition 4.3(i) the total relative loss may be as high as $B\sqrt{n}$ if $\overline{y}_n = y^*_\theta$.

Proof: Let θ be fixed, and let $\overline{\sigma}(\theta)$ be the strategy of Proposition 4.3. For each $0 < \xi < 1$, define the strategy:

$$\sigma(\xi):\ \sigma(\xi)_n = \xi \overline{\sigma}(\theta)_n + (1-\xi) x^*_\theta\,. \tag{4.6}$$

The required strategy $\sigma^1(\theta)$ will be defined by restarting $\sigma(\xi)$ at prespecified times, with ξ diminishing to zero. This scheme makes it possible to guarantee property (i), namely bounded loss.

Let $0 < \mu \leq 1/2$ be as defined in Lemma 4.2. Choose a real sequence $\{\xi_k\}_{k\geq 0}$ and a sequence of integers $0 = T_o < T_1 < \cdots$, such that: $0 < \xi_k \leq \mu$, $\xi_k \downarrow 0$, and for some finite constants C_1, C_2,:

$$\sum_{k=0}^{\infty} \xi_k \sqrt{T_{k+1} - T_k} \leq C_1\,, \tag{4.7}$$

$$\xi_k\, T_k \geq C_2 \sqrt[4]{T_{k+1}} \quad \forall k \geq 1\,. \tag{4.8}$$

Two specific examples are (with $0 < \epsilon < \frac{1}{4}$): (a) $T_k = 2^k - 1$, $\xi_k = \mu 2^{-k(\frac{1}{2}+\epsilon)}$. (b) $T_k = k^3$, $\xi_k = \mu(1+k)^{-(2+\epsilon)}$.

Finally, define $\sigma^1(\theta)$ as follows:

Strategy $\sigma^1(\theta)$ *:* At stages $n = 1, 2, \cdots, T_1$, play according to $\sigma(\xi_o)$.
At stage $T_k, k \geq 1$, reset the history counter to 0, and then play for $n = T_k + 1, \cdots, T_{k+1}$ according to $\sigma(\xi_k)$. More precisely, for $T_k < n \leq T_{k+1}$, $x_n = \xi_k \overline{\sigma}(\theta)_{n-T_k}(j_{T_k+1}, \cdots, j_{n-1}) + (1 - \xi_k)x^*_\theta$.

We proceed to upper-bound the loss and lower-bound the information under $\sigma^1(\theta)$. Both bounds will be in terms of $\|\overline{y}_n - y^*_\theta\|$, DM2's average deviation from his optimal strategy in $G(\theta)$. It is assumed in the following that DM2 is using any strategy $\tau \in \mathcal{T}$.

Consider first the strategy $\sigma(\xi)$ defined above, with ξ fixed. Suppose for the moment that this strategy is used throughout by DM1. Then $x_n = \xi\tilde{x}_n + (1 - \xi)x^*_\theta$, where $\tilde{x}_n = \bar{\sigma}(\theta)_n(j_1, \cdots j_{n-1})$. Therefore, by optimality of x^*_θ, Proposition 4.3 and Lemma 4.1:

$$\begin{aligned}
\frac{1}{n}\sum_{t=1}^{n} \overline{A}_\theta[x_t, j_t) &= \frac{1}{n}\sum_{t=1}^{n} \left\{\xi \overline{A}_\theta[\tilde{x}_t, j_t) + (1-\xi)\overline{A}_\theta[x^*_\theta, j_t)\right\} \\
&\geq \xi \frac{1}{n}\sum_{t=1}^{n} \overline{A}_\theta[\tilde{x}_t, j_t) + (1-\xi)v(\theta) \\
&\geq \xi(\max_x \overline{A}_\theta[x, \overline{y}_n] - B/\sqrt{n}) + (1-\xi)v(\theta) \\
&\geq \xi\left(v(\theta) + \delta_\theta\|\overline{y}_n - y^*_\theta\| - B/\sqrt{n}\right) + (1-\xi)v(\theta) \\
&= v(\theta) + \xi\delta_\theta\|\overline{y}_n - y^*_\theta\| - \xi B/\sqrt{n}\,, \qquad (4.9)
\end{aligned}$$

where B and δ_θ are positive constants. This can be written equivalently as:

$$d_\theta\{1:n\} \equiv nv(\theta) - \sum_{t=1}^{n} \overline{A}_\theta[x_t, j_t) \leq -\xi\delta_\theta n\|\overline{y}_n - y^*_\theta\| + \xi B\sqrt{n}\,. \qquad (4.10)$$

Returning to the strategy $\sigma^1(\theta)$, assume henceforth that this strategy is used by DM1. Let $T_k < m \leq T_{k+1}$ for some $k \geq 0$. Observe that $\sigma(\xi_k)$ is started at $t = T_k + 1$. Therefore, (4.10) implies:

$$d_\theta\{T_k + 1 : m\} \leq -\xi_k\delta_\theta(m - T_k)\|\Delta y(m, T_k)\| + \xi_k B\sqrt{m - T_k}\,,$$

where

$$\Delta y(m, T_k) = \frac{1}{m - T_k}\sum_{t=T_k+1}^{m} e_{j_t} - y^*_\theta\,.$$

Therefore, for any $T_K < n \leq T_{K+1}$, $K \geq 0$:

$$\begin{aligned}
d_\theta\{1:n\} \leq &\sum_{k=0}^{K-1}\left(-\xi_k\delta_\theta\Delta T_k\|\Delta y(T_{k+1}, T_k)\| + \xi_k B\sqrt{\Delta T_k}\right) \\
&+ \left(-\xi_K\delta_\theta(n - T_K)\|\Delta y(n, T_K)\| + \xi_K B\sqrt{n - T_K}\right)\,,
\end{aligned}$$

where $\Delta T_k = T_{k+1} - T_k$. Now, using the fact that $\{\xi_k\}$ is decreasing, the triangle inequality, and (4.7):

$$\begin{aligned}
d_\theta\{1:n\} &\leq -\xi_K\delta_\theta\left(\sum_{k=0}^{K-1}\Delta T_k\|\Delta y(T_{k+1}, T_k)\| + (n - T_K)\|\Delta y(n, T_K)\|\right) + B\sum_{k=0}^{\infty}\xi_k\sqrt{\Delta T_k} \\
&\leq -\xi_K\delta_\theta n\|\overline{y}_n - y^*_\theta\| + BC_1\,.
\end{aligned}$$

Moreover, since $T_K < n \le T_{K+1}$, it follows by (4.8) that:

$$\xi_K n \ge \xi_K T_K \ge C_2 \sqrt[4]{T_{K+1}} \ge C_2 \sqrt[4]{n}\,,$$

so that, finally, we obtain the upper bound

$$d_\theta\{1:n\} \;\le\; -C_2\delta_\theta \sqrt[4]{n}\|\overline{y}_n - y^*_\theta\| + BC_1\,. \tag{4.11}$$

Next, the information will be bounded. Let $\theta' \in G_o(\theta)$. Since $x_t = (1-\xi_k)x^*_\theta + (\ldots)$ with $1-\xi_k \ge 1-\mu \ge 1/2$, and noting that $I_{\theta,\theta'} \ge 0$,

$$\begin{aligned}
\sum_{t=1}^{n} I_{\theta,\theta'}[x_t,\, j_t) \;&\ge\; \frac{1}{2}\sum_{t=1}^{n} I_{\theta,\theta'}[x^*_\theta,\, j_t) \;=\; \frac{1}{2}\, n\, I_{\theta,\theta'}[x^*_\theta, \overline{y}_n] \\
&=\; \frac{1}{2}\, n\, I_{\theta,\theta'}[x^*_\theta, y^*_\theta] + \frac{1}{2}\, n\, I_{\theta,\theta'}[x^*_\theta, \overline{y}_n - y^*_\theta] \\
&\ge\; \frac{1}{2}\,\beta_1 n - \frac{1}{2}\,\beta_2 n\|\overline{y}_n - y^*_\theta\|, \qquad\qquad \theta' \in G_o(\theta)
\end{aligned} \tag{4.12}$$

where

$$\beta_1 = \min_{\theta' \in G_o(\theta)} I_{\theta,\theta'}[x^*_\theta, y^*_\theta] > 0\,, \qquad \beta_2 = \max_{\theta' \in G_o(\theta)} \max_j I_{\theta,\theta'}[x^*_\theta, j)\,.$$

Note that β_1 is positive by the definition of $G_o(\theta)$ and $B(\theta)$ in Proposition 4.1 and (3.2).

We may now proceed to establish (i)–(iii) of the lemma.

(i) Follows immediately from (4.11) (since $C_2\delta_\theta > 0$), for any $M_4 \ge BC_1$.

(ii) Assume that for some $\theta' \in G_o(\theta)$, $\sum_{t=1}^{n} I_{\theta,\theta'}[x_t, j_t) \le \epsilon n$ where $\epsilon > 0$ will be specified shortly. Then, from (4.12), $\|\overline{y}_n - y^*_\theta\| \ge (\beta_1 - 2\epsilon)/\beta_2$. Therefore, for $\epsilon = \beta_1/4$, (4.11) gives:

$$\sum_{t=1}^{n} d_\theta[x_t, j_t) \le -C_2\delta_\theta \frac{\beta_1}{2\beta_2}\sqrt[4]{n} + BC_1$$

which clearly implies (ii) for any $\delta_4 \le C_2\delta_\theta\beta_1/2\,\beta_2$ and $M_4 \ge BC_1$.

(iii) Recall that, for $T_K < n \le T_{K+1}$, $x_n = \xi_K\tilde{x}_n + (1-\xi_K)x^*_\theta$ for some $\tilde{x}_n \in \mathcal{P}(\mathcal{I})$, and $\xi_K \le \mu$. Therefore, $\|x_n - x^*_\theta\|_\infty \le \mu\|\tilde{x}_n - x^*_\theta\|_\infty \le \mu$. □

Proof of Proposition 4.1:

To motivate the definition of $\sigma^*(\theta)$ below, note that property (i) in the proposition requires the relative loss to be bounded on any time interval $[m,n]$. However, the strategy $\sigma^1(\theta)$ of the previous lemma guarantees that only on $[1,n]$. Thus, if the loss is negative on $[0, m-1]$, say, it might be large on $[m,\, n]$.

To rectify this problem, we define $\sigma^*(\theta)$ as the strategy which follows $\sigma^1(\theta)$ as long as the loss is above a certain (negative) threshold. However, as soon as it goes below this threshold, the clock is reset and $\sigma^1(\theta)$ is restarted with a new history. The precise definition follows.

Strategy $\sigma^*(\theta)$*:* Let C_1 be a positive constant. Let $\{m_k\}_{k\geq 0}$ be the sequence of stopping times (possibly infinite) defined recursively by: $m_o = 0$, $m_{k+1} = \inf\{m \geq m_k+1 : d_\theta\{m_k+1 : m\} \leq -C_1\}$. Then, for $m_k+1 \leq n \leq m_{k+1}$, $\sigma^*(\theta)_n(h_{n-1}) \stackrel{\triangle}{=} \sigma^1(\theta)_{n-m_k}(h_{n-1}^{(k)})$, where $h_{n-1}^{(k)} \stackrel{\triangle}{=} (j_{m_k+1}, \ldots, j_{n-1})$.

Assume that DM1 uses this strategy $\sigma^*(\theta)$. Let $\tau \in \mathcal{T}$ and $n \geq 1$ be fixed, and let $K \geq 0$ be such that $m_K < n \leq m_{K+1}$. Define:

$V_k = \{m_k + 1, \cdots, m_{k+1}\}$, $0 \leq k \leq K-1$: the k'th (terminated) interval.

$V_K = \{m_K + 1, \cdots, n\}$: the last (K'th) interval.

By definition of $\{m_k\}$, it follows that on each interval:

$$d_\theta\{m_k + 1 : m\} \geq -C_1 - \hat{D} \qquad \forall m \in V_k,\ 0 \leq k \leq K\,, \tag{4.13}$$

where $\hat{D} = \max_{i,j} |d_\theta(i,j)|$. On the other hand, on each terminated interval:

$$d_\theta\{V_k\} \stackrel{\triangle}{=} \sum_{t \in V_k} d_\theta[x_t, j_t) \leq -C_1, \quad 0 \leq k \leq K-1\,. \tag{4.14}$$

Finally, since $\sigma^1(\theta)$ is used on each interval, and in particular on the last interval, it follows from Lemma 4.3(i) that:

$$d_\theta\{V_K\} \leq M_4\,. \tag{4.15}$$

We proceed now to prove assertions (i)–(iii) of the proposition. Note that it is enough to prove each assertion with different constants (M_1, δ_1), since then the maximal M_1 and minimal δ_1 satisfy the assertions simultaneously.

(i) Fix $1 < m \leq n$. Then $m-1 \in V_k$ for some $0 \leq k \leq K$, so that by (4.13),

$$d_\theta\{m_k + 1 : m-1\} \geq -C_1 - \hat{D}\,. \tag{4.16}$$

Also, by (4.14) and (4.15) it follows that $d_\theta\{m_k+1 : n\} \leq M_4$. Subtracting the last two inequalities gives $d_\theta\{m : n\} \leq M_4 + C_1 + \hat{D}$, so that ($i$) holds for $M_1 = K_4 + C_1 + \hat{D}$.

(ii) Since $\sigma^1(\theta)$ is used on each interval, it follows by Lemma 4.3(ii) that, for some positive constants ϵ_4, δ_4 and every $0 \leq k \leq K$, $d_\theta\{V_k\} > -\delta_4 \sqrt[4]{|V_k|} + M_4$ implies

$$\min_{\theta' \in G_o(\theta)} I_{\theta,\theta'}\{V_k\}[x_t, j_t) \geq \epsilon_4 |V_k|\,. \tag{4.17}$$

Let $L > 0$ be some large enough constant so that $-\delta_4 \sqrt[4]{L} + M_4 < -C_1 - \hat{D}$. It follows then from (4.14) that on each interval V_k, $0 \leq k \leq K$, for which $|V_k| \geq L$:

$$d_\theta\{V_k\} \ \geq\ -C_1 - \hat{D} \ >\ -\delta_4 \sqrt[4]{L} + M_4 \ \geq\ -\delta_4 \sqrt[4]{|V_k|} + M_4\,,$$

so that (4.17) holds on the that interval. Therefore,

$$\begin{aligned} I_{\theta,\theta'}\{1 : n\} \ &\geq\ \epsilon_4 \sum_{k=0}^{K} |V_k|\, 1\{|V_k| \geq L\} \\ &=\ \epsilon_4 (\, n - \sum_{k=0}^{K} |V_k|\, 1\{|V_k| < L\}\,) \\ &\geq\ \epsilon_4 [n - L(K+1)], \qquad \forall \theta' \in G_o(\theta)\,. \end{aligned} \tag{4.18}$$

On the other hand, by (4.14) and (4.15) it follows that $d_\theta\{1:n\} \leq -C_1 K + M_4$. Using (4.18) to eliminate K from this equation, we finally obtain

$$d_\theta\{1:n\} \leq -\frac{C_1}{L}n + (M_4 + C_1) + \frac{C_1}{L\epsilon_4} I_{\theta,\theta'}\{1:n\} \qquad \forall \theta' \in G_o(\theta)\,,$$

which implies (ii) with $M_1 = \max\{M_4 + C_1, C_1/(L\epsilon_4)\}$ and $\delta_1 = C_1/L$.

(iii) Recall from Lemma 4.3(i) that, under $\sigma^1(\theta)$, $\|x_n - x^*_\theta\|_\infty \leq \mu$ for every $n \geq 1$. By definition of $\sigma^*(\theta)$, this is valid under $\sigma^*(\theta)$ as well. Therefore, by Lemma 4.2,

$$d_{\theta'}[x_n, j_n) \leq -\delta_3 + M_3 I_{\theta',\theta}[x_n, j_n), \quad \forall \theta' < \theta\,,$$

which implies (iii) for any $\delta_1 \leq \delta_3, M_1 \geq M_3$. Thus, the proof of Proposition 4.1 is complete. □

We turn now to the proof of Proposition 4.2, which proceeds through the following lemmas.

Lemma 4.4 *Let $\theta \in \Theta$ be such that $B(\theta) \neq \emptyset$. Then there exists a strategy $\sigma^2(\theta)$ for DM1 and positive constants B_5, M_5, δ_5 such that, for every $\tau \in \mathcal{T}$ and $n \geq 1$, the following hold:*

(i) $\quad d_\theta\{1:n\} \leq b(\theta) \min_{\theta' \in B(\theta)} I_{\theta,\theta'}\{1:n\} + B_5\sqrt{n}\,.$

(ii) $\quad d_\theta\{1:n\} \leq -\delta_5 n + M_5 + M_5 \min_{\theta' > \theta} I_{\theta,\theta'}\{1:n\}\,.$

(iii) $\quad \|x_t - x^*_\theta\|_\infty \leq \mu$, *with μ as in Lemma 4.2.*

Proof: Consider some fixed θ such that $B(\theta) \neq \emptyset$. Let $x^o = x^o(\theta) \in \mathcal{P}(\mathcal{I})$ be a minimizer in (3.4), and let $\overline{\sigma}(\theta)$ be the strategy of Proposition 4.3. For any $0 < \epsilon < 1$, define a strategy σ^ϵ by:

$$\sigma^\epsilon:\ \sigma^\epsilon_n = (1-\mu)x^*_\theta + \mu[\epsilon x^o + (1-\epsilon)\overline{\sigma}(\theta)_n]\,.$$

It will be proved that, for ϵ small enough, σ^ϵ satisfies the lemma.

Denote $d_o = d_\theta[x^o, y^*_\theta]$. Note that, by definition of x^o:

$$d_o = b(\theta) \min_{\theta' \in B(\theta)} I_{\theta,\theta'}[x^o, y^*_\theta] > 0\,. \tag{4.19}$$

Assume now that DM1 uses the strategy σ^ϵ, and DM2 any strategy $\tau \in \mathcal{T}$. Proceeding similarly to (4.9),(4.10), the loss may then be bounded by:

$$d_\theta\{1:n\} \ \leq\ \mu\epsilon n\, d_\theta[x^o, \overline{y}_n] - \mu(1-\epsilon)\left[\delta_\theta\, n\|\overline{y}_n - y^*_\theta\| - B\sqrt{n}\right]\,,$$

where B and δ_θ are positive constants. Furthermore, note that

$$d_\theta[x^o, \overline{y}_n] \ \leq\ d_\theta[x^o, y^*_\theta] + \beta_1\|\overline{y}_n - y^*_\theta\|\,,$$

where $\beta_1 \stackrel{\triangle}{=} \max_j |d_\theta[x^o, j)|$. Therefore,

$$d_\theta\{1:n\} \ \leq\ \mu\,\epsilon\, d_o n - \mu[\delta_\theta - \epsilon(\delta_\theta + \beta_1)]\, n\, \|\overline{y}_n - y^*_\theta\| + \mu B\sqrt{n}\,. \tag{4.20}$$

Next, the information will be lower-bounded. Consider first any $\theta' \in B(\theta)$. Then, by definition of σ^ϵ and (4.19):

$$\begin{aligned} I_{\theta,\theta'}\{1:n\} &\equiv \sum_{t=1}^{n} I_{\theta,\theta}[x_t, j_t) \geq \sum_{t=1}^{n} \mu\,\epsilon\, I_{\theta,\theta'}[x^o, j_t) = \mu\,\epsilon\, n\, I_{\theta,\theta'}[x^o, \overline{y}_n] \\ &= \mu\,\epsilon\, n\,(I_{\theta,\theta'}[x^o, y^*_\theta] + I_{\theta,\theta'}[x^o, \overline{y}_n - y^*_\theta]) \\ &\geq \mu\,\epsilon\, n\,(b(\theta)^{-1} d_o - \beta_2 \|\overline{y}_n - y^*_\theta\|), \qquad \theta' \in B(\theta)\,, \end{aligned} \tag{4.21}$$

where $\beta_2 \stackrel{\Delta}{=} \max_{j,\,\theta'>\theta} I_{\theta,\theta'}[x^o, j)$.

Consider now $\theta' \in G_o(\theta) = \{\theta' > \theta\} - B(\theta)$. Then, similarly,

$$\begin{aligned} I_{\theta,\theta'}\{1:n\} &\geq (1-\mu)\, n\, I_{\theta,\theta'}[x^*_\theta, \overline{y}_n] \\ &\geq (1-\mu)\, n\,(I_{\theta,\theta'}[x^*_\theta, y^*_\theta] - \beta_3 \|\overline{y}_n - y^*_\theta\|) \\ &\geq (1-\mu) n(\beta_4 - \beta_3 \|\overline{y}_n - y^*_\theta\|)\,, \qquad \theta' \in G_o(\theta)\,, \end{aligned} \tag{4.22}$$

where $\beta_3 \stackrel{\Delta}{=} \max_{j,\,\theta' \in G_o(\theta)} I_{\theta,\theta'}[x^*_\theta, j)$, $\beta_4 \stackrel{\Delta}{=} \min_{\theta' \in G_o(\theta)} I_{\theta,\theta'}[x^*_\theta, y^*_\theta]$. Note that $\beta_4 > 0$ by definition of $G_o(\theta)$ and $B(\theta)$.

Choose $\epsilon > 0$ small enough so that:

$$\epsilon(\delta_\theta + \beta_1 + b(\theta)\beta_2) \leq \frac{1}{2}\,\delta_\theta\,, \quad \epsilon(\delta_\theta + \beta_1 + d_o\beta_3/\beta_4) \leq \frac{1}{2}\,\delta_\theta\,. \tag{4.23}$$

Assertions (i)–(iii) of the lemma now follow from the bounds derived above by simple algebra:

(i) Multiplying (4.21) by $b(\theta)$, subtracting from (4.20) and rearranging yields:

$$d_\theta\{1:n\} \;\leq\; b(\theta) I_{\theta,\theta'}\{1:n\} - \beta_5 n \|\overline{y}_n - y^*_\theta\| + \mu B\sqrt{n}, \quad \theta' \in B(\theta)$$

where $\beta_5 = \mu[\delta_\theta - \epsilon(\delta_\theta + \beta_1 + b(\theta)\beta_2)]$. Since $\beta_5 \geq 0$ by the choice (4.23) of ϵ, (i) follows with $B_5 = \mu B$.

(ii) Using (4.21) to eliminate $\|\overline{y}_n - y^*_\theta\|$ from (4.20) gives:

$$d_\theta\{1:n\} \leq -\beta_7 n + \beta_6 I_{\theta,\theta'}\{1:n\} + \mu B\sqrt{n}\,, \quad \theta' \in B(\theta)\,, \tag{4.24}$$

where $\beta_6 = \delta_\theta/(\epsilon\beta_2)$, $\beta_7 = [\delta_\theta - \epsilon(\delta_\theta + \beta_1 + b(\theta)\beta_2)]/(b(\theta)\beta_2)$. Note that $\beta_7 > 0$ by the choice (4.23) of ϵ. Now, a similar calculation with (4.22) used instead of (4.21) gives:

$$d_\theta\{1:n\} \leq -\beta_9 n + \beta_8 I_{\theta,\theta'}\{1:n\} + \mu B\sqrt{n}\,, \quad \theta' \in G_o(\theta)\,, \tag{4.25}$$

where $\beta_8 = \delta_\theta/(1-\mu)\beta_3$, $\beta_9 = \mu\beta_4\beta_3^{-1}[\delta_\theta - \epsilon(\delta_\theta + \beta_1 + d_0\beta_3/\beta_4)]$; note that $\beta_9 > 0$ by choice of ϵ. Combining (4.24) and (4.25) establishes (ii) with, e.g., $\delta_5 = \frac{1}{2}\min\{\beta_7, \beta_9\}$ and $M_5 = \max\{\beta_8, \beta_9, \max_{n\geq 1}(\mu B\sqrt{n} - \delta_5 n)\}$.

(iii) Follows directly from the definition of σ^ϵ. □

Lemma 4.5 *For the strategy $\sigma^2(\theta)$ of the previous lemma, and under the same conditions,*

(*i*) *For every* $\epsilon > 0$ *there exists* $M'(\epsilon) > 0$ *such that:*

$$d_\theta\{1:n\} \;\le\; b(\theta)(1+\epsilon)I_{\theta,\theta'}\{1:n\} + M'(\epsilon)\,, \quad \forall\, n \ge 1,\ \theta' \in B(\theta)\,.$$

(*ii*) *Let* $\epsilon_5 = \delta_5/(2M_5)$, $\delta_6 = \delta_5/2$. *Then* $\min_{\theta'>\theta} I_{\theta,\theta'}\{1:n\} \le \epsilon_5 n$ *implies* $d_\theta\{1:n\} \le -\delta_6 n + M_5$.

Proof: Since (*ii*) follows directly from Lemma 4.4(*ii*), it remains to prove (*i*). Assume first that $S_n \stackrel{\triangle}{=} \min_{\theta' \in B(\theta)} I_{\theta,\theta'}\{1:n\} \le \epsilon_5 n$. Noting that $B(\theta) \subset \{\theta' :\ \theta' > \theta\}$, it follows by item (*ii*) of the present lemma that

$$d_\theta\{1:n\} \le M_5\,. \tag{4.26}$$

Assume now that $S_n > \epsilon_5 n$. Then, by Lemma 4.4(*i*):

$$d_\theta\{1:n\} \;\le\; b(\theta)S_n + B_5\sqrt{n} \;\le\; b(\theta)(1+\epsilon)S_n + (B_5\sqrt{n} - \epsilon\, b(\theta)\,\epsilon_5\, n)\,. \tag{4.27}$$

Let $M'(\epsilon) \stackrel{\triangle}{=} \max\{M_5,\, \max_{n\ge 1}(B_5\sqrt{n} - \epsilon b(\theta)\epsilon_5 n)\}$; assertion (*i*) follows now from (4.26) and (4.27). □

Proof of Proposition 4.2:

Similarly to the proof of Proposition 4.1, it is required to modify the strategy $\sigma^2(\theta)$ so that (*i*) will hold on any interval $[m,n]$. Thus, define

Strategy $\sigma^o(\theta)$*:* Defined similarly to $\sigma^*(\theta)$ in the proof of Proposition 4.1, except that $\sigma^1(\theta)$ in that definition is replaced by $\sigma^2(\theta)$ of Lemma 4.4.

Let $\tau \in \mathcal{T}$ and $n \ge 1$ be fixed. Retain the notations in the proof of Proposition 4.1 (i.e. C_1, m_k, V_k and K). We proceed to prove (*i*)–(*iii*) of Proposition 4.2.

(*i*) Consider a fixed $1 \le m \le n$. Let $0 \le k \le K$ be such that $m-1 \in V_k$, and note that (4.14) and (4.16) hold true. Moreover, since $\sigma^2(\theta)$ is restarted at $t = m_K + 1$, it follows by Lemma 4.5(*ii*) that for every $\epsilon > 0$.

$$\begin{aligned} d_\theta\{m_k+1:n\} &= \sum_{k'=k}^{K} d_\theta\{V_{k'}\} \;\le\; d_\theta\{V_K\} \\ &\le (1+\epsilon)b(\theta) \min_{\theta' \in B(\theta)} I_{\theta,\theta'}\{V_K\} + M'(\epsilon) \\ &\le (1+\epsilon)b(\theta) \min_{\theta' \in B(\theta)} I_{\theta,\theta'}\{1:n\} + M'(\epsilon)\,. \end{aligned} \tag{4.28}$$

Then (*i*) follows by subtracting (4.16) from (4.28), with $M(\epsilon) = M'(\epsilon) + C_1 + \hat{D}$.

(*ii*) By Lemma 4.5(*ii*), it follows that for every $0 \le k \le K$, $d_\theta\{V_k\} > -\delta_6|V_k| + M_5$ implies

$$\min_{\theta'>\theta} I_{\theta,\theta'}\{V_k\} > \epsilon_5|V_k|\,. \tag{4.29}$$

Let $L > 0$ be a large enough constant so that $-\delta_6 L + M_5 < -C_1 - \hat{D}$. Then on each interval V_k such that $|V_k| \geq L$,

$$d_\theta\{V_k\} \;\geq\; -C_1 - \hat{D} \;>\; -\delta_6 L + M_5 \;\geq\; -\delta_6 |V_k| + M_5\,,$$

so that (4.29) holds on that interval. Therefore, for every $\theta' > \theta$,

$$\sum_{k=0}^{K-1} I_{\theta,\theta'}\{V_k\} \;\geq\; \epsilon_5 \sum_{k=0}^{K-1} |V_k|\, 1\{|V_k| \geq L\} \;\geq\; \epsilon_5 (n - LK - |V_K|)\,. \tag{4.30}$$

On the other hand, by (4.14) and Lemma 4.4(ii) (applied to the last interval),

$$\sum_{t=1}^{n} d_\theta[x_t,\, j_t) \leq -C_1 K - \delta_5 |V_K| + M_5 + M_5 \min_{\theta' > \theta} I_{\theta,\theta'}\{V_K\}\,. \tag{4.31}$$

Using (4.30) to eliminate K from (4.31) and noting that $C_1/L < \delta_5$ by choice of L, it follows that for every $\theta' > \theta$,

$$\begin{aligned} d_\theta\{1:n\} \;&\leq\; -\frac{C_1}{L}\left(-\frac{1}{\epsilon_5}\sum_{k=0}^{K-1} I_{\theta,\theta'}\{V_k\} + n - |V_K|\right) - \delta_5 |V_K| + M_5 I_{\theta,\theta'}\{V_K\} \\ &\leq\; -\frac{C_1}{L} n + M_2 I_{\theta,\theta'}\{1:n\} + M_2\,, \end{aligned}$$

where $M_2 \stackrel{\triangle}{=} \max\{M_5,\, C_1/(L\epsilon_5)\}$. Thus, defining $\delta_2 \stackrel{\triangle}{=} C_1/L_2 > 0$, ($ii$) is established.

(iii) Follows by Lemma 4.4(iii) and Lemma 4.2, exactly as in the proof of Proposition 4.1(iii). □

5 The Optimal Strategy

We are now in a position to present a strategy σ^* which is asymptotically optimal. The following definitions will be required.

Definition 5.1 *Let $\{m_k\}_{k\geq 1}$ be a strictly increasing sequence of stopping times with respect to the history σ-algebras $\{H_n\}_{n\geq 0}$. Let σ be a given (behavioral) strategy of DM1. By* the strategy σ restricted to the times $\{m_k\}$ *we refer to the following selection rule at the times m_k, $k \geq 1$: $x_{m_k} = \sigma_k(\tilde{h}_{k-1})$, where $\tilde{h}_{k-1} \stackrel{\triangle}{=} \{i_{m_\ell}, j_{m_\ell}, a_{m_\ell}\}_{\ell=1}^{k-1}$.*

Let $\hat{\theta}_n$, $\hat{\Theta}_n$, $\overline{\theta}_n$ be defined as in Section 4.1. For every $\theta \in \Theta$ and $n \geq 1$, define the following conditions $C_n^*(\theta)$, $C_n^o(\theta)$:

$C_n^*(\theta)$: $\overline{\theta}_n = \theta$, and either $B(\theta) = \emptyset$ or else

$$\min_{\theta' \in B(\theta)} \sum_{t=1}^{n-1} I_{\theta,\theta'}[x_t,\, j_t)\, \mathbf{1}\{\overline{\theta}_t = \theta\} > \log K_n\,. \tag{5.1}$$

$C_n^o(\theta)$: $\overline{\theta}_n = \theta, B(\theta) \neq \emptyset$, and

$$\min_{\theta' \in B(\theta)} \sum_{t=1}^{n-1} I_{\theta,\theta'}[x_t, j_t)\, \mathbf{1}\{\overline{\theta}_t = \theta\} \leq \log K_n \,. \tag{5.2}$$

Note that exactly one of $C_n^*(\theta)$ and $C_n^o(\theta)$ is satisfied when $\overline{\theta}_n = \theta$. To introduce the following strategy, observe that for each fixed θ, the times t at which condition $C_t^*(\theta)$ [or $C_t^o(\theta)$] is satisfied form a sequence of increasing stopping times. We shall consider each such sequence separately, and apply on it a restricted version (according to Definition 5.1) of an appropriate sub-strategy.

Strategy σ^*. *For $n = 1, 2, \ldots$: Denote $\overline{\theta} = \overline{\theta}_n$.*
If $C_n^*(\overline{\theta})$ is satisfied, then play according to the strategy $\sigma^*(\overline{\theta})$ of Proposition 4.1, restricted to the times when $C_t^*(\overline{\theta})$ is satisfied.
If $C_n^o(\overline{\theta})$ is satisfied, then play according to the strategy $\sigma^o(\overline{\theta})$ of Proposition 4.2, restricted to the times when $C_t^o(\overline{\theta})$ is satisfied.

The strategy σ^* just defined may be interpreted as follows. At each stage n, the value-biased MLE $\overline{\theta} \stackrel{\Delta}{=} \overline{\theta}_n$ is computed. Then the level of information for discriminating $\overline{\theta}$ from $B(\overline{\theta})$ (quantified as in (5.1) or (5.2)) is evaluated, and compared with the critical level $\log K_n$ (which is slightly larger than $\log n$). If below that level, then the probing strategy $\sigma^o(\overline{\theta})$ is followed. This strategy ensures that, if indeed $\overline{\theta} = \theta_0$, additional information will be obtained at a loss-to-information ratio close to $b(\theta_0)$, or else *negative* relative loss is guaranteed to accumulate.

If the information level is below the critical level (or if $B(\overline{\theta})$ is empty), then the strategy $\sigma^*(\overline{\theta})$ is used. As discussed in Section 4.1, this strategy replaces the stationary strategy $\{x_{\overline{\theta}}\}$, and its stronger properties guarantee that the loss associated with the parameters in $(B_2(\theta_0) - B(\theta_0))$ is finite.

Observe that the "information level" in (5.1) and (5.2) is evaluated only over the times when the estimator was identical to the current one. This turns out to be important for the proof of the following theorem, which is the main result of this paper.

Theorem 5.1 *The strategy σ^* defined above is asymptotically optimal, in the sense of Definition 3.2.*

The remainder of this section will be devoted to the proof of this theorem. This proof extends the proof of Theorem 5.2 in [7], and some of the results established there will be used here.

Assume henceforth that DM1 uses the strategy σ^*, and let $\theta_0 \in \Theta$, $\tau \in \mathcal{T}$, $n \geq 1$ be fixed. In what follows, all relations between random variables hold $P_{\theta_0}^{\sigma^*,\tau}$–a.s. Also, all constants (M, δ, Q, n_o etc.) are independent of τ and n, unless otherwise stated.

It will be convenient to use the abbreviated notation: $d_t = d_{\theta_0}[x_t, j_t)$, $(d_t)^+ = \max\{d_t, 0\}$, $\hat{D} = \max_{i,j} d_{\theta_0}(i,j)$, $I_{\theta_0,\theta}(t) = I_{\theta_0,\theta}[x_t, j_t)$, $E = E_{\theta_0}^{\sigma^*,\tau}$, and finally $\ell_n = \sum_{t=1}^n d_t\, \mathbf{1}\{\overline{\theta}_t \geq \theta_0\}$.

By (2.1),

$$L_n^{\sigma^*,\tau}(\theta_o) \;=\; E\sum_{t=1}^n d_t \;=\; E\sum_{t=1}^n d_t\, \mathbf{1}\{\overline{\theta}_t < \theta_o\} + E\sum_{t=1}^n d_t \mathbf{1}\{\overline{\theta}_t \geq \theta_0\}$$

$$\leq \hat{D}E\sum_{t=1}^{n} \mathbf{1}\{\bar{\theta}_t < \theta_0\} + E\ell_n \leq \hat{D}Q_1 + E\ell_n\,, \tag{5.3}$$

where the last inequality is a basic property of the value-biased estimator $\bar{\theta}_n$, as established (for some $Q_1 < \infty$) in [7, Lemma 5.1(ii)]. We proceed then to bound $E\ell_n$.

Lemma 5.1 *For every $\epsilon > 0$, there exists a constant $M_o(\epsilon)$ such that,*

$$\ell_n \leq (1+\epsilon)b(\theta_0)\log K_n + M_o(\epsilon) + \sum_{t=1}^{n}(d_t)^+\mathbf{1}\{\ell_t > 0,\, \bar{\theta}_t > \theta_0\}\,, \tag{5.4}$$

where $b(\theta_0)$ is defined by (3.4) if $B(\theta_0) \neq \emptyset$, and $b(\theta_0) \stackrel{\triangle}{=} 0$ otherwise.

Proof: Noting Assumption A3, one has $\ell_n = \ell_n^a + \ell_n^b$, where

$$\ell_n^a = \sum_{t=1}^{n} d_t\,\mathbf{1}\{\bar{\theta}_t = \theta_0\}\,, \quad \ell_n^b = \sum_{t=1}^{n} d_t\,\mathbf{1}\{\bar{\theta}_t > \theta_0\}\,.$$

Define the stopping time $m = \max\{0 \leq t \leq n :\, \ell_t \leq 0\}$, where $\ell_o \stackrel{\triangle}{=} 0$. Then

$$\ell_n \leq \ell_n - \ell_m = (\ell_n^a - \ell_m^a) + (\ell_n^b - \ell_m^b)\,. \tag{5.5}$$

Now,

$$\begin{aligned}\ell_n^b - \ell_m^b &\leq \sum_{t=m+1}^{n}(d_t)^+\,\mathbf{1}\{\bar{\theta}_t > \theta_0\} = \sum_{t=m+1}^{n}(d_t)^+\,\mathbf{1}\{\ell_t > 0,\, \bar{\theta}_t > \theta_0\} \\ &\leq \sum_{t=1}^{n}(d_t)^+\,\mathbf{1}\{\ell_t > 0,\, \bar{\theta}_t > \theta_o\}\,,\end{aligned} \tag{5.6}$$

where the last equality follows by definition of m.

It remains to upper-bound the term:

$$\begin{aligned}\ell_n^a - \ell_m^a &= \sum_{t=m+1}^{n} d_t\,\mathbf{1}\{\bar{\theta}_t = \theta_0\} \\ &= \sum_{t=m+1}^{n} d_t\,\mathbf{1}\{C_t^*(\theta_0)\} + \sum_{t=m+1}^{n} d_t\mathbf{1}\{C_t^o(\theta_0)\}\,,\end{aligned} \tag{5.7}$$

where $C_t^*(\theta_0)$ stands for "$C_t^*(\theta_0)$ is satisfied", and similarly for $C_t^o(\theta_0)$. Note that, by definition of σ^*, DM1's strategy on the times in which $C_t^*(\theta_0)$ is satisfied is the restriction of $\sigma^*(\theta_0)$ to these times. Therefore, by Proposition 4.1(i),

$$\sum_{t=m+1}^{n} d_t\,\mathbf{1}\{C_t^*(\theta_0)\} \leq M_1\,. \tag{5.8}$$

To bound the term involving $\{C_t^o(\theta_0)\}$, note first that if $C_t^o(\theta_0)$ is not satisfied for any $m+1 \le t \le n$, then that term vanishes (this is trivially the case if $B(\theta_0) = \emptyset$). Otherwise, note that DM1's strategy on the times when $C_t^o(\theta_0)$ is satisfied is the restriction of $\sigma^o(\theta_0)$ to these times. Thus, by Proposition 4.2(i) it follows that for every $\epsilon > 0$,

$$\sum_{t=m+1}^{n} d_t \, \mathbf{1}\{C_t^o(\theta_0)\} \;\le\; (1+\epsilon) b(\theta_0) \min_{\theta \in B(\theta_0)} \sum_{t=m+1}^{n} I_{\theta_0,\theta}(t) \, \mathbf{1}\{C_t^o(\theta_0)\} \;+\; M(\epsilon) \,.$$

Define $m' = \max\{m+1 \le t \le n : \; C_t^o(\theta_0) \text{ is satisfied}\}$. Then,

$$\begin{aligned} \min_{\theta \in B(\theta_0)} \sum_{t=m+1}^{n} I_{\theta_0,\theta}(t) \, \mathbf{1}\{C_t^o(\theta_0)\} \;&\le\; \min_{\theta \in B(\theta_0)} \sum_{t=1}^{m'} I_{\theta_0,\theta}(t) \, \mathbf{1}\{C_t^o(\theta_0)\} \\ &\le\; \hat{I} + \log K_{m'} \;\le\; \hat{I} + \log K_n \,, \end{aligned}$$

where $\hat{I} = \max_{i,j,\theta} I_{\theta_0,\theta}(i,j)$, and the next to last inequality follows by definition of condition $C_t^o(\theta_0)$ (which is satisfied at $t = m'$). Thus,

$$\sum_{t=m+1}^{n} d_t \, \mathbf{1}\{C_t^*(\theta_0)\} \;\le\; (1+\epsilon) b(\theta_0) \log K_n + [(1+\epsilon) b(\theta_0) \hat{I} + M(\epsilon)] \,, \tag{5.9}$$

(which holds trivially if $B(\theta_0) = \emptyset$, with $b(\theta_0) = 0$).

The lemma now follows from (5.5)–(5.9), with $M_o(\epsilon) = M_1 + (1+\epsilon) b(\theta_0) \hat{I} + M(\epsilon)$. □

To upper-bound the (expected value of) the last term in (5.4), the following lemmas will be required.

Lemma 5.2 *There exist positive constants η and n_o such that, for any $n \ge n_o$, $\ell_n > 0$ implies that at least one of the following conditions $\Omega_1(n) - \Omega_4(n)$ is satisfied:*

$\Omega_1(n)$: $\sum_{t=1}^{n} \mathbf{1}\{\bar{\theta}_t < \theta_0\} \ge \eta n$.

$\Omega_2(n)$: $\sum_{t=1}^{n} I_{\theta_0,\theta}(t) \mathbf{1}\{\bar{\theta}_t = \theta\} \ge \eta n$ *for some* $\theta > \theta_0$.

$\Omega_3(n)$: $\min_{\theta > \theta_0} \sum_{t=1}^{n} I_{\theta_0,\theta}(t) \ge \eta n$.

$\Omega_4(n)$: *both* $\Omega_{4a}(n)$ *and* $\Omega_{4b}(n)$ *below are satisfied;*

$\Omega_{4a}(n)$: $\min_{\theta \in G_o(\theta_0)} \sum_{t=1}^{n} I_{\theta_0,\theta}(t) \ge \eta n$, *where* $G_o(\theta_0) = \{\theta > \theta_0\} - B(\theta_0)$.

$\Omega_{4b}(n)$: $N_n^*(\theta_0) \stackrel{\Delta}{=} \sum_{t=1}^{n} \mathbf{1}\{C_t^*(\theta_0)\} \ge \frac{1}{2} n$.

Proof: Let us first translate the relevant relations in Propositions 4.1 and 4.2 to the setting of the present strategy σ^*. For that purpose, define for each $\theta \in \Theta$:

$$N_n^*(\theta) = \sum_{t=1}^{n} \mathbf{1}\{C_t^*(\theta)\} \,, \quad N_n^o(\theta) = \sum_{t=1}^{n} \mathbf{1}\{C_t^o(\theta)\} \,,$$

$$N_n(\theta) = N_n^*(\theta) + N_n^o(\theta) = \sum_{t=1}^{n} \mathbf{1}\{\bar{\theta}_t = \theta\} \,.$$

Let $M_1, M_2, \delta_1, \delta_2$ be the constants for which Propositions 4.1 and 4.2 hold, and define $M = \max\{M_1, M_2\}$, $\delta = \min\{\delta_1, \delta_2\} > 0$. It then follows from items (i) and (ii) of Proposition 4.1 (upon substituting $\theta \to \theta_0$ and $\theta' \to \theta$) that

$$\sum_{t=1}^{n} d_t \, \mathbf{1}\{C_t^*(\theta_0)\} \le M \,, \tag{5.10}$$

$$\sum_{t=1}^{n} d_t \, \mathbf{1}\{C_t^*(\theta_0)\} \le -\delta N_n^*(\theta_0) + M + M \min_{\theta \in G_o(\theta_0)} \sum_{t=1}^{n} I_{\theta_0,\theta}(t) \, \mathbf{1}\{C_t^*(\theta_0)\} \,. \tag{5.11}$$

Similarly, by Proposition 4.2(ii),

$$\sum_{t=1}^{n} d_t \, \mathbf{1}\{C_t^o(\theta_0)\} \le -\delta N_n^o(\theta_0) + M + M \min_{\theta > \theta_0} \sum_{t=1}^{n} I_{\theta_0,\theta}(t) \, \mathbf{1}\{C_t^o(\theta_0)\} \,. \tag{5.12}$$

Finally, combining Propositions 4.1(iii) and 4.2(iii) (with $\theta' \to \theta_0$) gives

$$\sum_{t=1}^{n} d_t \, \mathbf{1}\{\overline{\theta}_t = \theta\} \le -\delta N_n(\theta) + M \sum_{t=1}^{n} I_{\theta_0,\theta}(t) \, \mathbf{1}\{\overline{\theta}_t = \theta\}, \quad \forall \theta > \theta_0 \,. \tag{5.13}$$

Assume now, in contradiction, that $\Omega_1(n) - \Omega_4(n)$ are false. It is required to show that $\ell_n \le 0$. Write $\overline{\Omega}_i(n)$ for '$\Omega_i(n)$ is false'. Then by $\overline{\Omega}_3(n)$ and (5.12):

$$\sum_{t=1}^{n} d_t \, \mathbf{1}\{C_t^o(\theta_0)\} \le -\delta N_n^o(\theta_0) + M + M\eta n \,. \tag{5.14}$$

By (5.13) and $\overline{\Omega}_2(n)$,

$$\sum_{t=1}^{n} d_t \, \mathbf{1}\{\overline{\theta} = \theta\} \le -\delta N_n(\theta) + M\eta n \,, \quad \forall \theta > \theta_0 \,. \tag{5.15}$$

Note also that $\overline{\Omega}_4(n)$ implies that at least one of the following holds:

(a) $\overline{\Omega}_{4b}(n)$, i.e. $N_n^*(\theta_0) < \frac{1}{2} n$.

(b) $\Omega_{4b}(n)$ (i.e. $N_n^*(\theta_0) \ge \frac{1}{2} n$), and $\overline{\Omega}_{4a}(n)$.

We consider these two cases separately:

(a) By (5.10), (5.14) and (5.15),

$$\begin{aligned} \ell_n &= \sum_{t=1}^{n} d_t \mathbf{1}\{C_t^*(\theta_0)\} + \sum_{t=1}^{n} d_t \mathbf{1}\{C_t^o(\theta_0)\} + \sum_{\theta > \theta_0} \sum_{t=1}^{n} d_t \mathbf{1}\{\overline{\theta}_t = \theta\} \\ &\le -\delta[N_n^o(\theta_0) + \sum_{\theta > \theta_0} N_n(\theta)] + 2M + M|\Theta|\eta n \,. \end{aligned}$$

However

$$N_n^o(\theta_0) + \sum_{\theta > \theta_0} N_n(\theta) = n - N_n^*(\theta_0) - \sum_{\theta < \theta_0} N_n(\theta) \ge \frac{1}{2} n - \eta n \,,$$

which is implied by $\overline{\Omega}_{4b}(n)$ and $\overline{\Omega}_1(n)$, so that in case (a):

$$\ell_n \le -\frac{1}{2} \delta n + \eta(\delta + M|\Theta|)n + 2M \,. \tag{5.16}$$

(b) Since $\overline{\Omega}_{4a}(n)$ is assumed, it follows from (5.11) that

$$\sum_{t=1}^{n} d_t \mathbf{1}\{C_t^*(\theta_0)\} < -\delta N_n^*(\theta_0) + M + M\eta n\,. \tag{5.17}$$

Proceeding as in case (a), with (5.17) used in place of (5.10), we get:

$$\ell_n \le -\delta[N_n(\theta_0) + \sum_{\theta>\theta_0} N_n(\theta)] + 2M + (M|\Theta| + M)\eta n\,. \tag{5.18}$$

Noting that $\overline{\Omega}_1(n)$ implies

$$N_n(\theta_0) + \sum_{\theta>\theta_0} N_n(\theta) \;=\; n - \sum_{\theta<\theta_0} N_n(\theta) \;\le\; n - \eta n\,,$$

it follows that in case (b):

$$\ell_n \le -\delta n + \eta(\delta + M|\Theta| + M)n + 2M\,. \tag{5.19}$$

It is readily seen that for some η small enough and n large enough, both (5.16) and (5.19) imply that $\ell_n \le 0$. □

Lemma 5.3 *Let $\Omega_4(n)$ be as defined in the previous lemma. Then*

$$E\sum_{n=1}^{\infty}(d_n)^+ \mathbf{1}\{\Omega_4(n), \overline{\theta}_n > \theta_0\} \;\le\; Q_3 \;<\; \infty\,. \tag{5.20}$$

Proof: By definition of $\Omega_4(n)$ and $G_o(\theta_0)$,

$$\begin{aligned}\mathbf{1}\{\Omega_4(n), \overline{\theta}_n > \theta_0\} &= \mathbf{1}\{\Omega_4(n), \overline{\theta}_n \in G_o(\theta_0)\} + \mathbf{1}\{\Omega_4(n), \overline{\theta}_n \in B(\theta_0)\} \\ &\le \mathbf{1}\{\Omega_{4a}(n), \overline{\theta}_n \in G_o(\theta_0)\} + \mathbf{1}\{\Omega_{4b}(n), \overline{\theta}_n \in B(\theta_0)\}\,.\end{aligned}$$

We first claim that, for some $Q' < \infty$,

$$E\sum_{n=1}^{\infty}(d_n)^+ \mathbf{1}\{\Omega_{4a}(n), \overline{\theta}_n \in G_o(\theta_0)\} \;\le\; \hat{D}\sum_{n=1}^{\infty} P\{\Omega_{4a}(n),\; \overline{\theta}_n \in G_o(\theta_0)\} \;\le\; Q'\,.$$

The proof of the last bound is the same as that of Lemma 6.4(ii) in [7]. Namely, by the union bound

$$P\{\Omega_{4a}(n),\; \overline{\theta}_n \in G_o(\theta_0)\} \;\le\; \sum_{\theta\in G_o(\theta_0)} P\{\sum_{t=1}^{n} I_{\theta_0,\theta}(t) > \eta n,\; \bar{\theta}_t = \theta\}\,,$$

and the rest is identical to the above-mentioned proof.

Thus, it remains to bound the term

$$J_3 = J_3(\tau) \stackrel{\Delta}{=} E\sum_{n=1}^{\infty}(d_n)^+ \mathbf{1}\{\Omega_{4b}(n),\; \overline{\theta}_n \in B(\theta_0)\}\,.$$

As established in [7, Lemma 5.2], there exists a constant $M < \infty$ such that $d_{\theta_0}[x^*_\theta, i) \le M\, I_{\theta_0,\theta}[x^*_\theta, j)$ for every j and $\theta > \theta_0$ (hence, in particular, for $\theta \in B(\theta_0)$). Replacing Ω_{4b} by its definition, it follows that

$$J_3 \le \sum_{\theta \in B(\theta_0)} E \sum_{n=1}^{\infty} M\, I_{\theta_0,\theta}(n)\, \mathbf{1}\{N^*_n(\theta_0) \ge \frac{n}{2},\, \overline{\theta}_n = \theta\}\,.$$

Now, $N^*_n(\theta_0) \ge \frac{1}{2}\, n$ implies that $C^*_m(\theta_0)$ is satisfied for some $\frac{1}{2}\, n \le m \le n$, which in turn implies that

$$U_n(\theta) \stackrel{\triangle}{=} \sum_{t=1}^{n-1} I_{\theta_0,\theta}(t)\, \mathbf{1}\{\overline{\theta}_t = \theta_o\} \;\ge\; \log K_{[n/2]} \;\ge\; \log K_n - \alpha\,,$$

where the last inequality follows from the definition of $\{K_n\}$ in (4.1) for some finite constant α (independent of n). Noting further that $\overline{\theta}_n = \theta > \theta_o$ implies $\Lambda_{n-1}(\theta_o, \theta) \le \log K_n$, we finally get

$$\begin{aligned} J_3 \;&\le\; M \sum_{\theta \in B(\theta_0)} E \sum_{n=1}^{\infty} I_{\theta_0,\theta}(t)\, \mathbf{1}\{U_n(\theta) \ge \log K_n - \alpha,\; \Lambda_{n-1}(\theta_0, \theta) \le \log K_n\} \\ &\le\; M \sum_{\theta \in B(\theta_0)} Q(\theta) \stackrel{\triangle}{=} Q'' < \infty\,, \end{aligned}$$

where the last bound follows by applying Lemma 3.3(v) of [7] (and the standard translation procedure as described there below equation (4.12)) to each $\theta \in B(\theta_0)$ separately. Thus, letting $Q_3 = Q' + Q''$, (5.20) is established. □

Lemma 5.4 *The following bound holds:*

$$J_4 = J_4(\tau) \stackrel{\triangle}{=} E \sum_{n=1}^{\infty} (d_n)^+\, \mathbf{1}\{\ell_n > 0,\, \overline{\theta}_n > \theta_0\} \;\le\; Q_4 \;<\; \infty\,.$$

Proof: By Lemma 5.2,

$$\begin{aligned} J_4 \;&\le\; \sum_{i=1}^{4} E \sum_{n=n_o}^{\infty} (d_n)^+ \mathbf{1}\{\Omega_i(n),\, \overline{\theta}_n > \theta_0\} + \hat{D} n_o \\ &\le\; \hat{D} \sum_{i=1}^{3} \sum_{n=1}^{\infty} P\left\{\Omega_i(n),\, \overline{\theta}_n > \theta_0\right\} + E \sum_{n=1}^{\infty} (d_n)^+ \mathbf{1}\{\Omega_4(n),\, \overline{\theta}_n > \theta_0\} + \hat{D}\, n_o\,. \end{aligned}$$

The three terms corresponding to $\Omega_1(n)$–$\Omega_3(n)$ have been bounded in [7, Lemma 6.4] (for any strategy σ of DM1). Therefore, the assertion follows by Lemma 5.3. □

The proof of Theorem 5.1 may now be concluded. By (5.3), Lemma 5.1 and Lemma 5.4, it follows that for every $\epsilon > 0$,

$$L_n^{\sigma^*,\tau}(\theta_0) \le \hat{D} Q_1 + (1+\epsilon) b(\theta_0) \log K_n + M_o(\epsilon) + Q_4\,,$$

where $b(\theta_0) = 0$ if $B(\theta_0) = \emptyset$. Thus:
If $B(\theta_0) = \emptyset$, then

$$L_n^{\sigma^*}(\theta_0) = \sup_{\tau} L_n^{\sigma^*,\tau}(\theta_0) \;\le\; \hat{D} Q_1 + M_o(1) + Q_4 \;<\; \infty\,, \quad \forall n \ge 1\,.$$

If $B(\theta_0) \neq \emptyset$, then

$$\limsup_{n\to\infty} \frac{L_n^{\sigma^*}(\theta_0)}{\log n} \leq (1+\epsilon) b(\theta_0)\,,$$

and the required result follows by letting $\epsilon \to 0$. □

Appendix

Proof of Proposition 4.3

Let $\langle \cdot, \cdot \rangle$ denote the standard Euclidean inner product, and let $d(\cdot,\cdot)$, $\|\cdot\|$ denote the corresponding distance and norm. Let

$$m_n = (\overline{A}_\theta[x_n,\, j_n),\; e_{j_n}) \in \mathbb{R} \times \mathbb{R}^{|J|}\,,$$
$$\overline{m}_n = \frac{1}{n}\sum_{t=1}^{n} m_n = (\overline{a}_n, \overline{y}_n)\,.$$

Let c_n be the closest point in Q to $\overline{m}_n$, and denote $\eta_n = c_n - \overline{m}_n$, $d_n = d(\overline{m}_n, Q)$.

We proceed to prove, by induction, that $d_n \leq B_1/\sqrt{n}$ for some constant $B_1 > 0$ (under the strategy $\overline{\sigma}(\theta)$ specified in the proposition). It is easily seen that this implies the assertion (i).

Assume first that $\overline{m}_n \in Q$, i.e. $d_n = 0$. Note that

$$\overline{m}_{n+1} = \overline{m}_n + \frac{m_{n+1} - \overline{m}_n}{n+1}\,. \tag{A.1}$$

Then

$$d_{n+1} = d(\overline{m}_{n+1}, Q) \leq \|\bar{m}_{n+1} - \bar{m}_n\| = \left\| \frac{m_{n+1} - \overline{m}_n}{n+1} \right\| \leq \frac{2B_2}{n+1}\,, \tag{A.2}$$

where B_2 is a constant which uniformly upper-bounds $\|m_n\|$.

Assume now that $\overline{m}_n \notin Q$, i.e. $d_n > 0$. Note that for every $\eta = (\alpha, \xi) \in \mathbb{R} \times \mathbb{R}^{|J|}$ with $\alpha \geq 0$,

$$\begin{aligned} \mathrm{val}(A^{(\eta)}) &= \min_{y\in\mathcal{P}(\mathcal{J})} \max_{x\in\mathcal{P}(\mathcal{I})} A^{(\eta)}[x,y] \\ &= \min_y (\alpha \max_x \overline{A}_\theta[x,y] + \langle \xi, y\rangle) \\ &= \min_y \langle \eta, (\max_x \overline{A}_\theta[x,y], y)\rangle \\ &\geq \min_{q\in Q} \langle \eta, q\rangle \end{aligned} \tag{A.3}$$

where we have used the definition of the payoff matrix $A^{(\eta)}$ and the fact that $(\max_x \overline{A}_\theta[x,y],\, y) \in Q$ for every $y \in \mathcal{P}(\mathcal{J})$.

Denote $(\alpha, \xi) \stackrel{\Delta}{=} \eta_n$. Recalling that x_{n+1} is optimal in $A^{(\eta_n)}$, noting that $\alpha > 0$ by the form of the set Q, and using equation (A.3), it follows that

$$\begin{aligned}
\langle c_n - \overline{m}_n, m_{n+1} \rangle &= \langle \eta_n, m_{n+1} \rangle = \alpha \overline{A}_\theta[x_{n+1}, j_{n+1}) + (\xi)_{j_{n+1}} \\
&= A^{(\eta_n)}[x_n, j_n) \geq \mathrm{val}(A^{(\eta_n)}) \\
&\geq \min_{q \in Q} \langle \eta_n, q \rangle = \min_{q \in Q} \langle c_n - \overline{m}_n, q \rangle \\
&= \langle c_n - \overline{m}_n, c_n \rangle
\end{aligned}$$

where the last equality holds since c_n is the closest point in Q to $\overline{m}_n$. It follows that

$$\langle c_n - \overline{m}_n,\ c_n - m_{n+1} \rangle \leq 0 .$$

Now, by (A.1) and the definition of c_n,

$$\begin{aligned}
d_{n+1}^2 &= \|\overline{m}_{n+1} - c_{n+1}\|^2 \leq \|\overline{m}_{n+1} - c_n\|^2 \\
&= \left\| \frac{1}{n+1}(n\overline{m}_n + m_{n+1}) - c_n \right\|^2 \\
&= \frac{1}{(n+1)^2} \|n(\overline{m}_n - c_n) + (m_{n+1} - c_n)\| \\
&\leq \frac{n^2}{(n+1)^2} \|\overline{m}_n - c_n\|^2 + \frac{1}{(n+1)^2} \|m_{n+1} - c_n\|^2 \\
&\leq \frac{n^2}{(n+1)^2} d_n^2 + \frac{(B_3)^2}{(n+1)^2} ,
\end{aligned} \tag{A.4}$$

where B_3 is a constant which uniformly upper-bounds $\|m_{n+1} - c_n\|$.

Letting $B_1 = \max(2B_2, B_3)$, it follows from (A.2) and (A.4) by simple calculation that $d_n \leq B_1/\sqrt{n}$ implies $d_{n+1} \leq B_1/\sqrt{n+1}$. □

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] D. Blackwell, "An analogue for the minimax theorem for vector payoffs," *Pacific J. Math.* **6**, pp. 1–8, 1956.

[3] D. Blackwell, "Controlled random walks," *Proc. Internat. Congress Math.* **3**, pp. 336–338, 1954.

[4] J. F. Hannan, "Approximation to Bayes risk in repeated play," in *Contributions to the Theory of games,* Vol. 3. Princeton Univ. Press, Princeton, NJ, pp. 97–139, 1957.

[5] T.L. Lai and H. Robbins, "Asymptotically efficient adaptive allocation rules," *Adv. in Appl. Math.* **6**, pp. 4–22, 1985.

[6] T. Parthasarathy and T.E.S. Ragahavan, *Some Topics in Two-Person Games,* American Elsevier Publishing Co., NY, 1971.

[7] N. Shimkin, "Asymptotically efficient adaptive strategies in repeated games. Part 1: Certainty Equivalence strategies," Technical Report, University of Minnesota, IMA preprint series no. 1092, December 1992.

[8] S. Sorin, "An introduction to two-person zero-sum repeated games with incomplete information," IMSS-Economics, Technical Report No. 312, Stanford University, 1980.

#	Author/s	Title
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, $D_n x O(2)$-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 governed 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

1094 **Robert F. Stengel, Laura R. Ray & Christopher I. Marrison**, Probabilistic evaluation of control system robustness

1095 **H.O. Fattorini & S.S. Sritharan**, Optimal chattering controls for viscous flow

1096 **Kathryn E. Lenz**, Properties of certain optimal weighted sensitivity and weighted mixed sensitivity designs

1097 **Gang Bao & David C. Dobson**, Second harmonic generation in nonlinear optical films

1098 **Avner Friedman & Chaocheng Huang**, Diffusion in network

1099 **Xinfu Chen, Avner Friedman & Tsuyoshi Kimura**, Nonstationary filtration in partially saturated porous media

1100 **Walter Littman & Baisheng Yan**, Rellich type decay theorems for equations $P(D)u = f$ with f having support in a cylinder

1101 **Satyanad Kichenassamy & Walter Littman**, Blow-up surfaces for nonlinear wave equations, II

1102 **Nahum Shimkin**, Extremal large deviations in controlled I.I.D. processes with applications to hypothesis testing

1103 **A. Narain**, Interfacial shear modeling and flow predictions for internal flows of pure vapor experiencing film condensation

1104 **Andrew Teel & Laurent Praly**, Global stabilizability and observability imply semi-global stabilizability by output feedback

1105 **Karen Rudie & Jan C. Willems**, The computational complexity of decentralized discrete-event control problems

1106 **John A. Burns & Ruben D. Spies**, A numerical study of parameter sensitivities in Landau-Ginzburg models of phase transitions in shape memory alloys

1107 **Gang Bao & William W. Symes**, Time like trace regularity of the wave equation with a nonsmooth principal part

1108 **Lawrence Markus**, A brief history of control

1109 **Richard A. Brualdi, Keith L. Chavey & Bryan L. Shader**, Bipartite graphs and inverse sign patterns of strong sign-nonsingular matrices

1110 **A. Kersch, W. Morokoff & A. Schuster**, Radiative heat transfer with quasi-monte carlo methods

1111 **Jianhua Zhang**, A free boundary problem arising from swelling-controlled release processes

1112 **Walter Littman & Stephen Taylor**, Local smoothing and energy decay for a semi-infinite beam pinned at several points and applications to boundary control

1113 **Srdjan Stojanovic & Thomas Svobodny**, A free boundary problem for the Stokes equation via nonsmooth analysis

1114 **Bronislaw Jakubczyk**, Filtered differential algebras are complete invariants of static feedback

1115 **Boris Mordukhovich**, Discrete approximations and refined Euler-Lagrange conditions for nonconvex differential inclusions

1116 **Bei Hu & Hong-Ming Yin**, The profile near blowup time for solution of the heat equation with a nonlinear boundary condition

1117 **Jin Ma & Jiongmin Yong**, Solvability of forward-backward SDEs and the nodal set of Hamilton-Jacobi-Bellman Equations

1118 **Chaocheng Huang & Jiongmin Yong**, Coupled parabolic and hyperbolic equations modeling age-dependent epidemic dynamics with nonlinear diffusion

1119 **Jiongmin Yong**, Necessary conditions for minimax control problems of second order elliptic partial differential equations

1120 **Eitan Altman & Nahum Shimkin**, Worst-case and Nash routing policies in parallel queues with uncertain service allocations

1121 **Nahum Shimkin & Adam Shwartz**, Asymptotically efficient adaptive strategies in repeated games, part II: Asymptotic optimality

1122 **M.E. Bradley**, Well-posedness and regularity results for a dynamic Von Kármán plate

1123 **Zhangxin Chen**, Finite element analysis of the 1D full drift diffusion semiconductor model

1124 **Gang Bao & David C. Dobson**, Diffractive optics in nonlinear media with periodic structure

1125 **Steven Cox & Enrique Zuazua**, The rate at which energy decays in a damped string

1126 **Anthony W. Leung**, Optimal control for nonlinear systems of partial differential equations related to ecology

1127 **H.J. Sussmann**, A continuation method for nonholonomic path-finding problems