UNIVERSITÄT HAMBURG

Controlled Markov fields with finite state space on graphs

R. K. Chornei , H. Daduna , P. S. Knopov

Preprint–No. 2000-08 November 2000

SCHWERPUNKT MATHEMATISCHE STATISTIK
UND STOCHASTISCHE PROZESSE

BUNDESSTR. 55, D – 20146 HAMBURG

Controlled Markov fields with finite state space on graphs*

Ruslan K. Chornei   Hans Daduna   Pavel S. Knopov

Summary: Discrete time Markov chains with multidimensional state space are considered where the coordinates are locally interacting and develop synchronously. The interaction structure of the process is given by some general graph. Decision makers control the system's behaviour on the coordinate level using only local information. In the class of local strategies there exist deterministic stationary strategies with achieve minimal asymptotic average expected costs. If the cost structure is separabel these strategies are even globally optimal.
Keywords: Markov random fields, Markov chains, synchronous transitions, local transitions, average expected costs, Markov policy, interacting processes, local strategies
MSC Classification: 60J10, 60K35, 93E20

1 Introduction

The aim of this note is to develop models for spatially distributed systems that are controlled by decision makers who act locally and have only information at hand about the system's state in the neighborhood around their position. To optimize such systems poses problems due to several restrictions on the information gain which arise from observing the system over time only locally. Other sources of difficulties may be cost functions which react on global behaviour of the system although only local control is possible by the decision makers. We shall consider only systems which are observed at equally spaced time instants, the time parameter of the underlying stochastic processes will be $\mathbb{N}$. Disrete time systems, especially queueing models have found increasing interest in recent years due to the occurence of models with a generic inherent discrete time scale. A typical example is the following:

Example 1.1. We consider a simple linear transmission line of successive transition channels numbered $1, 2, \dots, J$, with finite capacity. Messages arrive at channel 1 and after being transmitted there they proceed to channel 2, and so on, eventually leaving the line. The transmission protocol of the network is based on a common discrete time scale.

To protect the network from overload, where the performance is degraded due to strongly increasing protocol overhead, it is assumed that a window–flow control scheme is applied: If the line carries a certain amount of traffic no further messages are admitted, and for

*Most of the research reported in this paper was done when the third author visited the Department of Mathematics at Hamburg University. The visit was supported by the DEUTSCHE FORSCHUNGSGEMEINSCHAFT .

simplicity we assume that rejected messages are lost. So by applying such control scheme the overload is handled outside the critical path which the line is considered to be.

Considering the quality of service for such a transmission system, it is intuitively obvious, that, conditioned on being admitted for entrance, e.g. for such a message the transmission time can be reduced effectively by applying an overall small admission rate. But small admission rate is equivalent to a large loss rate, creating complementary degradation of the quality of service. Therefore, to find an optimal admission policy should balance carefully the costs due to lost arrivals against rewards from successful transmissions within prescribed due dates.

Clearly, an optimal admission policy will incorporate the complete state description of the line at arrival instants in making the decision. But from a practical point of view such a procedure is impossible especially in today's high speed networks. There admission decisions have to be made on the basis of the state information of the entrance node only, possibly with some information on the states of the nearest neighbor nodes. $\odot$

Steady state analysis of such a system with local admission control can be found in [16]. From this it becomes appearent, that local control as investigated here, often leads to models which at present are not analytically tractable, see the discussion in [7, section4]. (This is in contrast to application of some globally defined admission control policy for transmission lines, [1].) Before proceding with our general description we mention a field of research where spatio–temporal models are essentially needed as well, for more details see [2], [3], [14], and the references therein; some more information is presented here in example 3.1.

Example 1.2. Aspects of spatial localization that accompany dualism and pluralism in the use of production techniques and which arise from influences described by classical location theory as well as from innovation diffusion between firms that are geographically neighbored, but produce on different levels of production techniques, at present seem to be not well understood with respect to quantitative measures of such behaviour. Simple models, where effects of technologies which are established in some locally bounded destricts (enclaves) may be spill over to contiguous districts, are described in [3], using classical Ising models from Statistical Physics. These models allow to describe to a certain extend the interaction of firms located at the nodes of a subset of Z^2 and to prescribe simple evolution over time for a system which shows the Ising model's equilibrium behaviour. This behaviour is investigated in a simulation study in [3, section 6]. Although the simulation reproduce some of the aspects observed empirically, the authors point out that there is much need for more detailed models. Some problems mentioned are: Mimicking learning effects, which will lead to optimization of the firms' behaviour, which is just the need for closed loop control of the system – which appearently should be modeled in a first step on a firm's local level, including the neighborhood information for that firm (*localized learning effects*, [3, p. 23]; Feedback from the makro–state of the system, which e.g. can be incorporated into the model by applying globally determined cost functions; Allowing more than one site of the spatial lattice to change the state coordinate at the same time. $\odot$

As will be seen in the sequel most of these requirements are provided by the model developed in section 2 here under the heading of to control locally governed systems. To deal with these control problems we shall define neighborhood systems for (e.g.) transmission systems and controlled processes which evolve governed by controls being compatible with

these neighborhood structures. The general class of *local* policies obtained in this way gives rise to problems and classifications analoguously to those in the classical theory. The question then arises whether me may restrict our search for optimal policies from the very beginning to narrower classes of special *local* policies, say e.g. *stationary Markovian* policies as is well known in many situations where decisions are made on the basis of global information. This is answered positively in section 3.
On the other hand it is an interesting question under which conditions the locally defined policy classes contain strategies which are optimal even in the class of all admissible policies. We shall investigate this in Section 5 concentrating on the case where the cost functions are separable with respect to the neighborhood graph.

The standard stochastic model for spatially distributed systems are *random fields* over some multidimensional lattice Z^d, or subsets thereof, or over some general graph Γ. Control of stochastically driven systems over time is part of the theory of stochastic processes. Therefore our generic model in this note shall be a spatially distributed process evolving in time. The fundamental theory especially connected with the origins of those models in Statistical Physics can be found in [15]. There the class of stochastic processes used is that of Markovian processes in time and for the spatial structure: *Markov random fields.* Combining such sytems with a control structure we describe in section 2 Markovian processes, showing a locally spatial construction which mimics those prototype systems to a certain extend.

It is well known, that (in continuous time) Markov processes with local transition structure over some (non regular) graph must not have Markov random fields with respect to that graph as one–dimensional marginals nor as limiting random fields, see [13, setion 9]. Similar problems arise in our discrete time setting. We describe a class of processes in section 2 which to a certain extend resemble Kelly's *Spatial Processes* ([13, setion 9, p.189], see as well [21]). We do not impose the restriction that state changes are of strong local character: several coordinates of a state may change simultaneously, which is essential for discrete time systems, where *simultaneous events* usually occur.

Acknowledgement: We thank Gerhard Huebner for his remarks and helpful discussions on the subject of this paper.

2 Controlled Markov processes with locally interacting synchronous components and local control strategies

The interaction structure of our systems is defined by an interaction graph. We introduce the necessary prerequisits and notations.

Let $\Gamma = (V, B)$ be a finite undirected graph with set of vertices (nodes) V and set of edges B. Denote by $\{k, j\}$ the edge of the graph connecting the vertices k and j. The neighborhood of vertex k is the set $N(k) = \{j : \{k, j\} \in B\}$, i.e. the set of vertices connected by edges with the vertex k. The complete neighborhood of the vertex k is the set $\widetilde{N}(k) = N(k) \cup \{k\}$, i.e. the neighborhood of the vertex k including vertex k. For any $K \subset V$ we define the neighborhood $N(K) = \bigcup_{k \in K} N(k) - K$, and the complete neighborhood $\widetilde{N}(K) = N(K) \cup K$.

Let $X_i = \{x_i^1, \dots, x_i^{n_i}\}$ be the local state space of the system at node $i \in V$, i.e. there is a finite set of admissible values associated with vertex i. Then $X = \underset{i \in V}{\times} X_i$ is the global state space of the system. For every subset of nodes $K \subset V$ the vector $x_K := (x_k,\ k \in K) \in X_K = \underset{i\in K}{\times} X_i$ denotes the marginal state description at the nodes of K.

Throughout the paper we fix the underlying probability space $(\Omega, \mathcal{F}, P)$ for all random variables.

Definition 2.1. (a) A random variable ξ defined on $(\Omega, \mathcal{F}, P)$, taking its values in X is called a (discrete) random field on $\Gamma = (V, B)$ (or simply a random field on V). The marginal random variables with values in the space X_K are denoted by ξ_K. For $K = \{k\}$ we simply write ξ_k.

(b) A discrete random field ξ on $\Gamma = (V, B)$ is a (discrete) Markov field if the following holds:

$$P(\xi_k = x_k / \xi_{V-\{k\}} = x_{V-\{k\}}) = P(\xi_k = x_k / \xi_{N(k)} = x_{N(k)}), \qquad \forall\, x \in X. \tag{1}$$

$\odot$

Random fields on a graph or on a lattice (as usually used in Statistical Physics in connection with e.g. Gibbs fields) describe the spatial structure of the systems under consideration. In problems related to economical or engineering applications as we have in mind here the random fields explicitly undergo a time evolution, therefore describing the space–time structure of the system's development. It follows that we have to write $\xi = (\xi^t)$, for time instants t, to emphasize time-dependency of the random element ξ. Throughout the paper we assume that t takes discrete values $t = 0, 1, \ldots$.

In the following the subscript k in ξ_k^t refers to the vertex k, ξ_k^t therefore denotes the marginal distribution for time t and node k of some vector valued process $\xi = (\xi^t \colon t = 0, 1, \ldots)$.

If a Markov process ξ from its very construction is related to the neighborhood system $\{N(k) \colon k \in V\}$ of (V, B), it is natural to assume with respect to the evolution over time of the system that the value x_k of the k-th vertex at every moment depends on the previous state of the whole system only through the values of the vertices in $\tilde{N}(k)$ (including k) at the previous time instant.

To describe this we introduce the notion of Markov property in space and over time along the lines of Definition 2.1.

Definition 2.2. Let $\xi = \{\xi^t, t = 0, 1, \ldots\}$ be a discrete time Markov process as described above with finite state space X, which is the product space over Γ. The transition probabilities of X are said to be local [19, p. 100] if for all $k \in V, x^0, \ldots, x^{t+1} \in X$,

$$P\{\xi_k^{t+1} = x_k^{t+1} / \xi^t = x^t, \ldots, \xi^0 = x^0\} = P\{\xi_k^{t+1} = x_k^{t+1} / \xi^t_{\tilde{N}(k)} = x^t_{\tilde{N}(k)}\}, \tag{2}$$

holds, i.e. the state of vertex k depends only on state of its complete neighborhood at the previous time instant.

The transition probabilities of X are said to be synchronous [19, p. 100] if

$$P\{\xi_K^{t+1} = x_K^{t+1} / \xi^t = x^t\} = \prod_{k\in K} P\{\xi_k^{t+1} = x_k^{t+1} / \xi^t = x^t\}, \qquad K \subset V, \quad x^t, x^{t+1} \in X. \tag{3}$$

If ξ fulfills (2) and (3) it is called a Markov process with locally interacting synchronous components over (Γ, X), shortly, a (time dependent) Markov random field. $\odot$

An immediate consequence of the fundamental definition is a transition mechanism over time for the random fields under consideration of the following form:

Corollary 2.1. For any time $t \in \mathbb{N}$ we have:

$$\begin{aligned} & P\{\xi_K^{t+1} = x_K^{t+1} / \xi^t = x^t\} \\ = & \prod_{k \in K} P\{\xi_k^{t+1} = x_k^{t+1} / \xi^t_{\tilde{N}(k)} = x^t_{\tilde{N}(k)}\}, \quad K \subset V, \quad x^t, x^{t+1} \in X. \end{aligned} \tag{4}$$

$\odot$

For a general discussion of such processes see [8] and the references therein.

Example 2.1. Closed linear networks (cyclic networks) arise, e.g. if window flow control is applied for congestion control in circuit switching networks. Such flow control protocols are applied as well in high speed transmission networks. Under the ATM protocol these networks show an internal generic discrete time scale.

As a model for these networks we consider a closed cycle of single server nodes numbered $1, 2, \dots, J$, $J \geq 2$. Each node has an unlimited waiting room, the service discipline is First-Come-First-Served (FCFS). K undistinguishable customers are cycling in the system, $K \geq 1$. A customer leaving node j immediately proceeds to node $j+1$, $j = 1, \dots, J$; we set $J+1 := 1$, $1-1 := J$.

The evolution over time of the system is described by a discrete time stochastic process $\xi = (\xi(t) \colon t \in \{0, 1, \dots\})$ with state space $S(K, J) = \{(x_1, \dots, x_J) \in \mathbb{N}^J \colon x_1 + \dots + x_J = K\}$.

ξ records the development of the joint queue length vector, i.e. $\xi(t) = (\xi_1(t), \dots, \xi_J(t)) = (x_1, \dots, x_J)$ indicates that at time t there are x_j customers present at node j, including the customer just in service, if any $j = 1, \dots, J$.

The following assumptions imply that ξ is an ergodic Markov chain.

The nodes operate as independent Bernoulli servers under FCFS: if at time $t \in \mathbb{N}$ at node j a customer is in service and if there are $h-1 \geq 0$ other customers waiting, then in this time segment his service ends with probability $p_j(h) \in (0,1)$. He then leaves node j at time $(t+1)-$ and arrives at node $j+1$ just before time $(t+1)$. With probability $q_j(h) := 1 - p_j(h)$ this customer will stay at node j, requesting for at least one further quantum of service time there. The decision of a customer whether to stay or to leave is made independently of anything else. A customer leaving node j at time $(t+1)-$ will at time $t+1$ either join the queue of node $j+1$ or enter service immediately — if node $j+1$ was idle during $[t, t+1)$ or at time $(t+1)-$ the only customer present at node $j+1$ has left this node.

If at some node at the *same time point* a departure and an arrival occur we always assume that the departure event takes place first and both events happen just before the clock counts.(*Departure before Arrival — rule* (D/A), see [9], and *late arrival rule*, see [12]).

ξ is a time dependent Markov random field according to Definition 2.2 with respect to the interaction graph $\Gamma = (V, B)$ with $V = \{1, 2, \dots, J\}$, $B = \{\{j, j+1\} \colon j = 1, \dots, J\}$, which implies that for any node j we have $N(j) = \{j-1, j+1\}$. The transition probabilities of ξ obviously fulfill the locality condition (2) and are synchronous according to (3). $\odot$

We have the following theorem (see [1], Theorem 1).

Theorem 2.1 (Steady state). The joint queue length process ξ is ergodic on $S(K,J)$ and its unique limiting and stationary distribution is $\pi^{K,J} = (\pi^{K,J}(x)\colon x \in S(K,J))$ given by

$$\pi^{K,J}(x_1,\dots,x_J) = \prod_{j=1}^{J} \left(\frac{\prod_{h=1}^{x_j-1} q_j(h)}{\prod_{h=1}^{x_j} p_j(h)} \right) G(K,J)^{-1}, \qquad (x_1,\dots,x_J) \in S(K,J), \tag{5}$$

where $G(K,J)$ is the norming constant. $\odot$

Remark 2.1. The steady state distribution (5) is said to be of product form. Note that $\pi^{K,J}$ is not a Markovian random field with respect to Γ. Therefore in general, the one dimensional marginals in time of a time dependent Markov random field according to Definition 2.2 are not a Markov random field according to Definition 2.1. This holds in Example 2.1 although $\pi^{K,J}$ is conditional distribution of a vector with independent coordinates. $\odot$

In the following we shall develop the notion of a controlled Markov process with locally interacting synchronous components, shortly: a controlled time dependent Markov random field. For simplicity of the presentation we assume that a time dependent (uncontrolled) Markov random field $\xi = \{\xi^t\colon t \in \mathbb{N}\}$ according to Definition 2.2 is given as our starting point. ξ will be equipped in the sequel with a control structure which governs the transition kernel of ξ in a way to be described now. The standard definitions for this are as follows [6].

Definition 2.3. **(1)** The overall set of actions (control values) is $U = \underset{i\in V}{\times} U_i$ over Γ, where U_i is set of possible actions (decisions) for the decision maker at vertex i. U_i is finite and independent of time.
(2) The general unrestricted history dependent decisions at time t are
$\Delta^t = \Delta^t(x^0,\dots,x^t) = \{\Delta_i^t(x^0,\dots,x^t)\colon i \in V\} \in \underset{i\in V}{\times} U_i, \;\; t = 0,1,\dots$.
(3) Inadmissibility of actions dependent on time and state is introduced by fixing for any vertex $i \in V$ and any system configuration (state) $x \in X$ the set $U_i^t(x)$ of admissible actions (admissible decision values) for Δ_i^t at time t if $\xi^t = x$. $\odot$

We shall return to this general setting in section 5, but continue here with considering only controls, which are local in the following sense (for ease of introduction we first consider nonrestricted control sets):

Definition 2.4. Let an unrestricted control set U be given as in Definition 2.3 (1).
(1) If at any time $t = 0,1,\dots$ the decision Δ_i^t for node i is made on basis of the history of the complete neighborhood $\widetilde{N}(i)$ of i only, i.e. on the basis of $x^0_{\widetilde{N}(i)},\dots,x^t_{\widetilde{N}(i)}$, the sequence of decisions $\delta_i = \{\Delta_i^t, t \in \mathbb{N}\}$ is called local strategy for vertex i.

We therefore have for history dependent decision sequences $\Delta_i^t(x^0,\dots,x^t) = \Delta_i^t(x^0_{\widetilde{N}(i)},\dots,x^t_{\widetilde{N}(i)})$. These functions are defined on the space $X^{t+1}_{\widetilde{N}(i)} = \underbrace{X_{\widetilde{N}(i)} \times \cdots \times X_{\widetilde{N}(i)}}_{(t+1)-times}$ and take their values in U_i.

(2) A local strategy $\delta = (\delta_i\colon i \in V)$ is given by functions $\delta_i = \{\Delta_i^t,\; t \ge 0\}$, where

$$\Delta_i^0 = \Delta_i^0(x^0_{\widetilde{N}(i)}),\dots,\Delta_i^t = \Delta_i^t(x^0_{\widetilde{N}(i)},\dots,x^t_{\widetilde{N}(i)}),\dots,$$

is the local history dependent decision sequence at node i. $\odot$

From this definition it is natural to assume that the control of the time dependent random field ξ results in an evolution which is governed via modified transition probabilities

$$P\{\xi^{t+1} = x^{t+1} / \xi^t_{\widetilde{N}(i)} = y, \{(\Delta_i^t(\xi^0_{\widetilde{N}(i)}, \dots, \xi^t_{\widetilde{N}(i)}) \colon i \in V\} = u\}.$$

Note that the assumption of a local control structure is natural in our setting for defining locally interacting controlled systems. The possibility of making decisions which depend on the whole history of the system however is not dispensable, because, e.g. we want the decision maker being able to learn from the observed reactions of the system responding to control actions.

Incorporating a Markovian structure similar to Definition 2.2 into the construction of the process will then lead us to the assumption that the laws of motion for the system can be characterized by a time invariant set of transition probabilities. Namely, whenever the system is in state y and action u is taken, then, regardless of its history the next state is selected according to a transition law, which depends only on (y, u).

Using this, a strategy δ will define a probability measure on the space of sequences $(\xi^0, \xi^1, \dots)$ for every fixed initial state x^0. The pair (ξ, δ) will be called controlled version of ξ using strategy δ. The controlled process in general will not be Markovian, because the functions Δ_i^t, $i \in V$, depend not only on states $x^t_{\widetilde{N}(i)}$, $i \in V$, but on the previous (local) states $x^0_{\widetilde{N}(i)}, \dots, x^{t-1}_{\widetilde{N}(i)}$ as well.

Definition 2.5. A local strategy $\delta = (\delta_i \colon i \in V)$ is called local Markov strategy if for all $x^0, x^1, \dots, x^t \in X$ $\Delta_i^t(x^0_{\widetilde{N}(i)}, \dots, x^t_{\widetilde{N}(i)}) = \Delta_i^t(x^t_{\widetilde{N}(i)})$, $i \in V$, i.e. every local decision function Δ_i^t depends on the whole history only through the present local state at i. $\odot$

An immediate consequence is, that a Markov strategy δ applied as control strategy for the time dependent Markov random field defines a Markov process (ξ, δ). In general this version of ξ is not a time homogeneous Markov chain. This would be the case if the strategy is stationary in time.

Definition 2.6. A local Markov strategy $\delta = (\delta_i \colon i \in V)$ is called stationary local Markov strategy if $\Delta_i^{t'}(x_{\widetilde{N}(i)}) \equiv \Delta_i^{t''}(x_{\widetilde{N}(i)})$, $i \in V$, for all t' and t'', i.e. a stationary Markov strategy δ is completely determined by the functions Δ_i with

$$\begin{aligned} \Delta_i \colon X_{\widetilde{N}(i)} &\to U_i, \qquad\qquad i \in V; \\ x^t_{\widetilde{N}(i)} &\to \Delta_i(x^t_{\widetilde{N}(i)}). \end{aligned}$$

$\odot$

In considering local structures admissible actions for the decision makers to control their local parts of the system should reflect dependency on time and on the actual local state of the system.

Definition 2.7. (1) For any vertex $i \in V$ and any neighborhood configuration (state) $x_{\widetilde{N}(i)}$ at time t at node i
$U_i^t(x_{\widetilde{N}(i)})$ is the set of admissible actions (admissible decision values) for Δ_i^t at time t if $\xi^t_{\widetilde{N}(i)} = x_{\widetilde{N}(i)}$. We then have $U^t(x) = \underset{i \in V}{\times} U_i^t(x_{\widetilde{N}(i)})$.

(2) A local strategy δ is called admissible if for every $t \in \mathbb{N}$ $\Delta^t(x^0, \dots, x^{t-1}, x^t) = \{\Delta_i^t(x^0_{\widetilde{N}(i)}, \dots, x^{t-1}_{\widetilde{N}(i)}, x^t_{\widetilde{N}(i)})\colon i \in V\} \in U^t(x^t)$.

The class of all admissible deterministic local strategies is denoted by LD; the subclass of admissible local Markov strategies by LD_M.

(3) By LD_S we denote the class of admissible deterministic stationary local Markov strategies with time invariant restriction sets: $\Delta^t = \Delta^{t'}$ for all $t, t' \in \mathbb{N}$ and $U^t(x) = U(x) = \underset{i\in V}{\times} U_i(x_{\widetilde{N}(i)})$ for all t. $\odot$

Note, that from the very definition a deterministic stationary strategy is Markovian.

We summarize the principles and modeling features to control a system with locally interacting components which we have presented up to now in the following definition.

Definition 2.8. A pair (ξ, δ) is called a controlled process with locally interacting synchronous components with respect to the finite interaction graph $\Gamma = (V, B)$, if $\xi = (\xi^t\colon t \in \mathbb{N})$ is a stochastic process with state space $X = \underset{i\in V}{\times} X_i$, $\delta = (\delta_i\colon i \in V)$ is a local strategy, and the transitions of ξ are determined as follows:

$$\begin{aligned}
&P\{\xi_S^{t+1} = x_S / \xi^0 = x^0, \Delta^0(\xi^0) = u^0, \dots,\\
&\qquad\qquad \xi^{t-1} = x^{t-1}, \Delta^{t-1}(\xi^0, \dots, \xi^{t-1}) = u^{t-1}, \xi^t = y, \Delta^t(\xi^0, \dots, \xi^t) = u\}\\
&\overset{(1)}{=} P\{\xi_S^{t+1} = x_S / \xi^0 = x^0, \dots, \xi^{t-1} = x^{t-1}, \xi^t = y, \Delta^t(\xi^0, \dots, \xi^t) = u\}\\
&\overset{(2)}{=} P\{\xi_S^{t+1} = x_S / \xi^t = y, \Delta^t(\xi^0, \dots, \xi^t) = u\}\\
&\overset{(3)}{=} \prod_{j\in S} P\{\xi_j^{t+1} = x_j / \xi^t = y, \Delta^t(\xi^0, \dots, \xi^t) = u\}\\
&\overset{(4)}{=} \prod_{j\in S} P\{\xi_j^{t+1} = x_j / \xi^t_{\widetilde{N}(j)} = y_{\widetilde{N}(j)}, \Delta_j^t(\xi^0_{\widetilde{N}(j)}, \dots, \xi^t_{\widetilde{N}(j)}) = u_j\}\\
&\overset{(5)}{=} \prod_{j\in S} Q_j\{x_j / y_{\widetilde{N}(j)}, u_j\} \cdot 1_{\{u\}}(\Delta^t(\xi^0, \dots, \xi^t))\\
&\overset{(6)}{=} Q_S(x_S / y, u) \cdot 1_{\{u\}}(\Delta^t(\xi^0, \dots, \xi^t)), \qquad S \subseteq V, \quad y \in X, \quad u \in U(y),
\end{aligned}$$

where the $Q_j\{x_j / y_{\widetilde{N}(j)}, u_j\}$ are locally defined transition kernels which determine the time invariant laws of motion for the system. We shall write for $S = V$: $Q(x/y, u) = P\{\xi^{t+1} = x / \xi^t = y, \Delta^t = u\}$.

Some remarks on the modeling principle behind the definition may be in order:

It is assumed that the first expression is well defined, then $\overset{(1)}{=}$ is partly due to the Markov transition structure and on the other hand follows iteratively from the definition of the decision rules. $\overset{(2)}{=}$ is the part of the assumption of a Markovian transition law. $\overset{(3)}{=}$ expresses that the coordinates act synchronously, and $\overset{(4)}{=}$ is due to the locality of the transition law. Independence of time for the one–step transition probabilities then lead to the form $\overset{(5)}{=}$ of the Markov kernel, and $\overset{(6)}{=}$ again results from the coordinates acting synchronuosly.

ξ will be called shortly a controlled time dependent random field. $\odot$

Obviously, $\sum_{x\in X} Q(x/y,u) = 1, \qquad y \in X.$

Remark 2.2. If δ in the setting of Definition 2.8 is a stationary (Markov) strategy, with time invariant restriction sets $\Delta^t = \Delta^{t'}$ for all $t, t' \in \mathbb{N}$ and $U^t(x) = U(x) = \times_{i\in V} U_i(x_{\tilde{N}(i)})$ for all t, then ξ is a homogeneous Markov chain, and (because δ is local) according to Definition 2.2 a time dependent Markov random field. We call such pair (ξ, δ) (or the process ξ) a controlled Markov process with locally interacting synchronous components with respect to the finite interaction graph $\Gamma = (V, B)$, or shortly: a time dependent controlled Markov random field.

If at time $t \in \mathbb{N}$ the system is in state $\xi^t = x^t$ and a decision for action u^t is made a (one–step) cost $r(x^t, u^t) \geq 0$ is incurred to the system. The average expected cost up to time T when ξ is started with $\xi^0 = y$ and strategy δ is applied is then

$$Q_T^\delta(y) = E_y^\delta \frac{1}{T+1} \sum_{t=0}^{T} r(\xi^t, \Delta^t) := E_y^\delta \frac{1}{T+1} \sum_{t=0}^{T} r(\xi^t, \Delta^t(\xi^0, \ldots, \xi^t)), \tag{6}$$

where E_y^δ is expectation associated with the controlled process (ξ, δ) if $\xi^0 = y$. Note that we suppressed the dependency of Δ^t on $(\xi^0, \ldots, \xi^t))$, which we shall do similarly in the following as well. The problem is to find a strategy δ which minimizes the asymptotic average expected costs

$$R_y^\delta = \limsup_{T\to\infty} Q_T^\delta(y). \tag{7}$$

Denote by

$$\rho(y) = \inf_{\delta\in LD} R_y^\delta$$

the value which an optimal strategy should attain.

Definition 2.9. A strategy $\delta^\star \in LD$ (the class of admissible deterministic local strategies) is called optimal if $\rho(y) = R_y^{\delta^\star}$ for all $y \in X$. ⊙

3 Optimality of deterministic Markovian strategies in the class of local strategies

Similar to the case of Markov chains with finite state space and controls (see [20] or [6]) it is possible to prove that there exists an optimal strategy in the class of stationary local Markov strategies.

Theorem 3.1. Consider controlled processes (ξ, δ) with locally interacting synchronous components with respect to an interaction graph $\Gamma = (V, B)$ according to Definition 2.8 with finite state space X of ξ, and finite control space U. Let the sets of admissible actions $U^t(\cdot)$ be independent of t. Then in the class LD of admissible deterministic local strategies there exists a unique optimal strategy which is in the class LD_S of stationary Markovian policies. ⊙

Proof. Following the proof of the similar theorem for controlled Markov chains in [20], for every strategy $\delta \in LD$ we define for discount factor $\beta \in (0,1)$ the total expected discounted costs under δ when starting in y by

$$\mathfrak{R}_y^\delta(\beta) = E_y^\delta \sum_{t=0}^{\infty} \beta^t r(\xi^t, \Delta^t),$$

and by $\mathfrak{R}_y(\beta) = \inf_{\delta \in LD} \mathfrak{R}_y^\delta(\beta)$ the infimum over all admissible local policies.

Because X and U are finite, similar to the procedure in [20] (see [6], Chapter 3, as well) we can show that the functional $\mathfrak{R}_y^\delta$ attains its minimal value on LD, i.e. there exists $\delta^\star = \delta^\star(\beta) \in LD$:

$$\delta^\star(\beta) = \{\delta_i^\star(\beta),\ i \in V\}, \quad \delta_i^\star(\beta) = \{\overset{\star}{\Delta}{}_i^0(\beta)(x_{\widetilde{N}(i)}^0), \ldots, \overset{\star}{\Delta}{}_i^t(\beta)(x_{\widetilde{N}(i)}^0, \ldots, x_{\widetilde{N}(i)}^t), \ldots\},$$

such that for all $y \in X$

$$\mathfrak{R}_y^{\delta^\star}(\beta) = \mathfrak{R}_y(\beta).$$

(Note, that we wrote $\mathfrak{R}_y^{\delta^\star}(\beta)$ for $\mathfrak{R}_y^{\delta^\star(\beta)}(\beta)$. If from the context there is no ambiguity we shall treat similar expressions in the same way.)

Now we show that for fixed β and every strategy $\delta^\star(\beta)$ (which is optimal under the total expected discounted cost criterion $\mathfrak{R}_y^\delta(\beta)$) there exists some stationary Markov strategy $\widetilde{\delta}(\beta)$, which yields the same total discounted expected cost as $\delta^\star(\beta)$, when applied to control the system:

$$\mathfrak{R}_y^{\widetilde{\delta}}(\beta) = \mathfrak{R}_y^{\delta^\star}(\beta).$$

(We consider $\widetilde{\delta}(\beta)$ therefore to be equivalent to $\delta^\star(\beta)$ in the sense of generating the same total discounted expected cost.)

The strategy $\delta^\star(\beta)$ is characterized by the following property:

$$\mathfrak{R}_y^{\delta^\star}(\beta) \le \mathfrak{R}_y^\delta(\beta), \qquad \delta \in LD, \quad \forall\, y \in X. \tag{8}$$

Using conditional expectations we obtain for any $p \ge 0$

$$\mathfrak{R}_{x^0}^\delta(\beta) = E_{x^0}^\delta \Big\{ \sum_{t=0}^{p} \beta^t r(\xi^t, \Delta^t) + E^\delta \Big[\sum_{t=p+1}^{\infty} \beta^t r(\xi^t, \Delta^t) / \xi^0 = x^0, \ldots, \xi^p = x^p \Big] \Big\}. \tag{9}$$

We apply Bellman's principle and will show: if $x^0, \ldots, x^p$ are the first $p+1$ observations of the process $(\xi, \delta^\star(\beta))$ then for any strategy $\delta(\beta) \in LD$

$$E^{\delta^\star} \Big\{ \sum_{t=p+1}^{\infty} \beta^t r(\xi^t, \overset{\star}{\Delta}{}^t) / \xi^0 = x^0, \ldots, \xi^p = x^p \Big\} \le E^\delta \Big\{ \sum_{t=p+1}^{\infty} \beta^t r(\xi^t, \Delta^t) / \xi^0 = x^0, \ldots, \xi^p = x^p \Big\} \tag{10}$$

holds almost surely with respect to the process measure of $(\xi, \delta^\star(\beta))$.

Indeed, let $\mathfrak{X}_p$ denote the least sigma-field generated by $(\xi^0, \ldots, \xi^p)$. Suppose, there exists a strategy $\bar{\delta}$ and some set $\mathfrak{M}_p = \{\xi^0 = x^0, \ldots, \xi^p = x^p\} \in \mathfrak{X}_p$, with positive $P_{x^0}^{\delta^\star}$-measure of the process $(\xi, \delta^\star(\beta))$, such that

$$E^{\delta^\star} \Big\{ \sum_{t=p+1}^{\infty} \beta^t r(\xi^t, \overset{\star}{\Delta}{}^t) / \xi^0 = x^0, \ldots, \xi^p = x^p \Big\} > E^{\bar{\delta}} \Big\{ \sum_{t=p+1}^{\infty} \beta^t r(\xi^t, \bar{\Delta}^t) / \xi^0 = x^0, \ldots, \xi^p = x^p \Big\},$$

and the inequality (10) holds on the complementary set $\overline{\mathfrak{M}}_p$ of $\mathfrak{M}_p$. We then can construct a new strategy $\widetilde{\delta}(\beta) = \{\widetilde{\delta}_i(\beta),\ i \in V\}$, $\widetilde{\delta}_i(\beta) = \{\widetilde{\Delta}_i^t(\beta),\ t \geq 0\}$, by (for simplicity of presentation we neglect the dependence on β)

$$\widetilde{\Delta}_i^j(x^0_{\widetilde{N}(i)},\dots,x^j_{\widetilde{N}(i)}) = \overset{\star}{\Delta}{}_i^j(x^0_{\widetilde{N}(i)},\dots,x^j_{\widetilde{N}(i)}), \qquad j \leq p,$$

and for $j \geq 1$

$$\widetilde{\Delta}_i^{p+j}(x^0_{\widetilde{N}(i)},\dots,x^p_{\widetilde{N}(i)},\dots,x^{p+j}_{\widetilde{N}(i)}) = \begin{cases} \bar{\Delta}_i^{p+j}(x^0_{\widetilde{N}(i)},\dots,x^p_{\widetilde{N}(i)},\dots,x^{p+j}_{\widetilde{N}(i)}), & (x^0,\dots,x^p) \in \mathfrak{M}_p, \\ \overset{\star}{\Delta}{}_i^{p+j}(x^0_{\widetilde{N}(i)},\dots,x^p_{\widetilde{N}(i)},\dots,x^{p+j}_{\widetilde{N}(i)}), & (x^0,\dots,x^p) \in \overline{\mathfrak{M}}_p. \end{cases}$$

Thus, for the strategy $\widetilde{\delta}(\beta)$ we have

$$\begin{aligned}
\mathfrak{R}^{\widetilde{\delta}}_{x^0}(\beta) &= E^{\widetilde{\delta}}_{x^0}\Big\{\sum_{t=0}^{p}\beta^t r(\xi^t,\widetilde{\Delta}^t) + E^{\widetilde{\delta}}\Big[\sum_{t=p+1}^{\infty}\beta^t r(\xi^t,\widetilde{\Delta}^t)\,/\,\xi^0 = x^0,\dots,\xi^p = x^p\Big]\Big\} \\
&= E^{\widetilde{\delta}}_{x^0}\sum_{t=0}^{p}\beta^t r(\xi^t,\widetilde{\Delta}^t) + \\
&\quad + \int_{\mathfrak{M}_p} E^{\widetilde{\delta}}\Big[\sum_{t=p+1}^{\infty}\beta^t r(\xi^t,\widetilde{\Delta}^t)\,/\,\xi^0 = x^0,\dots,\xi^p = x^p\Big] dP^{\delta^\star}_{x^0}(\xi^0 = x^0,\dots,\xi^p = x^p) + \\
&\quad + \int_{\overline{\mathfrak{M}}_p} E^{\widetilde{\delta}}\Big[\sum_{t=p+1}^{\infty}\beta^t r(\xi^t,\widetilde{\Delta}^t)\,/\,\xi^0 = x^0,\dots,\xi^p = x^p\Big] dP^{\delta^\star}_{x^0}(\xi^0 = x^0,\dots,\xi^p = x^p) \\
&< E^{\delta^\star}_{x^0}\sum_{t=0}^{p}\beta^t r(\xi^t,\overset{\star}{\Delta}{}^t) + \\
&\quad + \int_{\mathfrak{M}_p} E^{\delta^\star}\Big[\sum_{t=p+1}^{\infty}\beta^t r(\xi^t,\overset{\star}{\Delta}{}^t)\,/\,\xi^0 = x^0,\dots,\xi^p = x^p\Big] dP^{\delta^\star}_{x^0}(\xi^0 = x^0,\dots,\xi^p = x^p) + \\
&\quad + \int_{\overline{\mathfrak{M}}_p} E^{\delta^\star}\Big[\sum_{t=p+1}^{\infty}\beta^t r(\xi^t,\overset{\star}{\Delta}{}^t)\,/\,\xi^0 = x^0,\dots,\xi^p = x^p\Big] dP^{\delta^\star}_{x^0}(\xi^0 = x^0,\dots,\xi^p = x^p) \\
&= \mathfrak{R}^{\delta^\star}_{x^0}(\beta),
\end{aligned}$$

i.e. $\mathfrak{R}^{\widetilde{\delta}}_{x^0}(\beta) < \mathfrak{R}^{\delta^\star}_{x^0}(\beta)$, which contradicts inequality (8). Therefore we henceforth assume that (10) is fulfilled for the optimal strategy $\delta^\star(\beta)$ for any realization $(x^0,\dots,x^p)$ with nonzero probability under $P^{\delta^\star}_{x^0}$ (because X is finite such realizations exist).

Setting

$$\Phi^{\delta^\star}(x^0,\dots,x^p) = E^{\delta^\star}\Big\{\sum_{t=p+1}^{\infty}\beta^t r(\xi^t,\overset{\star}{\Delta}{}^t)\,/\,\xi^0 = x^0,\dots,\xi^p = x^p\Big\}, \tag{11}$$

similar to [20] we shall prove that the function $\Phi^{\delta^\star}(x^0,\dots,x^p)$ depends on the whole history only through x^p, i.e., $\Phi^{\delta^\star}(x^0,\dots,x^{p-1},x^p) = \Phi^{\delta^\star}(y^0,\dots,y^{p-1},x^p)$ for all $x^m, y^m, m = 0,1,\dots,p-1$.

We fix time m and consider all finite time realizations $(x^0, \ldots, x^{m-1}, x^m)$ of ξ until time m which at time m attain the prescribed value x^m, and $\Phi^{\delta^\star}(x^0, \ldots, x^{m-1}, x^m)$ according to (11). Because X is finite we find only a finite number of such realizations and therefore $\Phi^{\delta^\star}(x^0, \ldots, x^{m-1}, x^m)$ attains its minimum on the set of all such realization in, say, $(\hat{x}^0, \ldots, \hat{x}^{m-1}, x^m) := (\hat{x}^0(x^m), \ldots, \hat{x}^{m-1}(x^m), x^m)$, and this sequence occurs with positive $P^{\delta^\star}_{\hat{x}^0}$–measure.

So for any other realization $(y^0, \ldots, y^{m-1}, x^m)$ which occurs with positive $P^{\delta^\star}_{y^0}$–measure, we have

$$\Phi^{\delta^\star}(\hat{x}^0, \ldots, \hat{x}^{m-1}, x^m) \leq \Phi^{\delta^\star}(y^0, \ldots, y^{m-1}, x^m). \tag{12}$$

We show, that in (12) equality holds. Assume we have strict inequality

$$\Phi^{\delta^\star}(\hat{x}^0, \ldots, \hat{x}^{m-1}, x^m) < \Phi^{\delta^\star}(y^0, \ldots, y^{m-1}, x^m) \tag{13}$$

for some realization $(y^0, \ldots, y^{m-1}, x^m)$ of the process. Then we construct a new admissible strategy $\delta^{\star\star}$ which acts as $\delta^\star$ until time m and which after time m acts on any sequence of states $(x^0, \ldots, x^{m-1}, x^m, x^{m+1}, \ldots, x^{m+k})$ in the same way as $\delta^\star$ would do when up to time m $(\hat{x}^0, \ldots, \hat{x}^{m-1}, x^m)$ would have happened:

$$\overset{\star\star}{\Delta}{}^{m+k}(x^0, \ldots, x^{m-1}, x^m, x^{m+1}, \ldots, x^{m+k}) := \overset{\star}{\Delta}{}^{m+k}(\hat{x}^0, \ldots, \hat{x}^{m-1}, x^m, x^{m+1}, \ldots, x^{m+k}). \tag{14}$$

For the strategy $\delta^{\star\star}$ we therefore obtain

$$E^{\delta^{\star\star}}\Big\{\sum_{t=m+1}^{\infty} \beta^t r(\xi^t, \overset{\star\star}{\Delta}{}^t)/\xi^0 = x^0, \ldots, \xi^p = x^p\Big\} < E^{\delta^\star}\Big\{\sum_{t=p+1}^{\infty} \beta^t r(\xi^t, \overset{\star}{\Delta}{}^t)/\xi^0 = x^0, \ldots, \xi^p = x^p\Big\}, \tag{15}$$

which contradicts (10), and thus we have proved that $\Phi^{\delta^\star}(x^0, \ldots, x^m)$ depends on the whole history only through x^m.

Therefore we apply for any realization $(x^0, \ldots, x^{m-1}, x^m, x^{m+1}, \ldots, x^{m+k}, \ldots)$ of ξ (with our fixed x^m at time m) from time m on the control decisions

$$\overset{\star}{\Delta}{}^{m+k}(\hat{x}^0, \ldots, \hat{x}^{m-1}, x^m, x^{m+1}, \ldots, x^{m+k})\text{instead of}$$

$$\overset{\star}{\Delta}{}^{m+k}(x^0, \ldots, x^{m-1}, x^m, x^{m+1}, \ldots, x^{m+k}), \qquad k \geq 1.$$

As we have shown, this does not change the value $\Phi^{\delta^\star}(x^0, \ldots, x^m)$. Therefore the strategy $\delta^{(m)}$, defined to be equal to $\delta^\star$ until time m, and then defined by $\overset{(m)}{\Delta}{}^{m+k}(x^0, \ldots, x^{m-1}, x^m, x^{m+1}, \ldots, x^{m+k}) := \overset{(\star)}{\Delta}{}^{m+k}(\hat{x}^0, \ldots, \hat{x}^{m-1}, x^m, x^{m+1}, \ldots, x^{m+k})$, $\quad k \geq 0$, yields the same conditional expectation $\Phi^{\delta^{(m)}}(x^0, \ldots, x^m) = \Phi^{\delta^\star}(x^0, \ldots, x^m)$ as the optimal control $\delta^\star$.

Note, that by definition $\delta^{(m)}$ is a local strategy because of the locality of $\delta^\star$ and that

$$\overset{(m)}{\Delta}{}^{m+k}(x^0, \ldots, x^{m-1}, x^m, x^{m+1}, \ldots, x^{m+k}) \in U^t(x^{m+k}) = \underset{i \in V}{\times} U_i^t(x^{m+k}_{\tilde{N}(i)})$$

holds. Therefore $\delta^{(m)}$ is admissible according to Definition 2.7 (2).

Having observed for any fixed m the independence property of $\Phi^{\delta^\star}(x^0,\dots,x^{m-1},x^m)$ from $(x^0,\dots,x^{m-1})$, we now construct a sequence of policies $(\delta^\star_{(m)} = (\overset{(m)}{\Delta}{}^i : i \in \mathbb{N}), m \in \mathbb{N})$ such that for fixed m a similar independence holds. For $m=1$ we start with

$$\delta^\star(x^0,x^1,\dots,x^m,\dots) = (\overset{\star}{\Delta}{}^0(x^0), \overset{\star}{\Delta}{}^1(x^0,x^1), \overset{\star}{\Delta}{}^m(x^0,x^1,\dots,x^m),\dots).$$

According to the construction performed above we can find a control $\overset{\star}{\delta}_1$ with

$$\overset{\star}{\delta}_{(1)}(x^0,x^1,\dots,x^m,\dots) = (\overset{(1)}{\Delta}{}^0(x^0), \overset{(1)}{\Delta}{}^1(x^1), \overset{(1)}{\Delta}{}^2(x^1,x^2),\dots,\overset{(1)}{\Delta}{}^m(x^1,x^2,\dots,x^m),\dots),$$

which yields the same conditional expectation

$$\Phi^{\overset{\star}{\delta}_{(1)}}(x^0) = \Phi^{\delta^\star}(x^0)$$

as the optimal control $\delta^\star$. We further have $\overset{(1)}{\Delta}{}^0(x^0) = \overset{\star}{\Delta}{}^0(x^0)$.

Continueing in the same way for $m=2$, we apply the same procedure to strategy $\overset{\star}{\delta}_{(1)}$ and obtain a new strategy

$$\overset{\star}{\delta}_{(2)}(x^0,x^1,\dots,x^m,\dots) = (\overset{(2)}{\Delta}{}^0(x^0), \overset{(2)}{\Delta}{}^1(x^1), \overset{(2)}{\Delta}{}^2(x^2), \overset{(2)}{\Delta}{}^2(x^2,x^3),\dots,\overset{(2)}{\Delta}{}^m(x^2,x^3,\dots,x^m),\dots),$$

and observe $\overset{(2)}{\Delta}{}^0(x^0) = \overset{(1)}{\Delta}{}^0(x^0)$, $\overset{(2)}{\Delta}{}^1(x^1) = \overset{(1)}{\Delta}{}^1(x^1)$.

We continue similarly and construct a sequence of strategies $\overset{\star}{\delta}_{(n)}$, $n \in \mathbb{N}$, which is the starting point for the last step of our construction. Following [20], and using Weierstrass' diagonalization method we can construct a Markov strategy $\overset{\infty}{\delta}(\beta) = \{\overset{\infty}{\delta}_i(\beta),\ i \in V\}$ such that

$$\overset{\star}{\delta}_{(\infty)}(x^0,x^1,\dots,x^m,\dots) = (\overset{(\infty)}{\Delta}{}^0(x^0), \overset{(\infty)}{\Delta}{}^1(x^1), \overset{(\infty)}{\Delta}{}^2(x^2), \overset{(\infty)}{\Delta}{}^2(x^3),\dots,\overset{(\infty)}{\Delta}{}^m(x^m),\dots),$$

holds and additionally $\overset{(\infty)}{\Delta}{}^i(x^i) = \overset{(i)}{\Delta}{}^i(x^i)$, $\forall\, i \in \mathbb{N}$.

The construction further showes, that the total expected discounted costs for all strategies $\overset{\star}{\delta}_{(n)}$, $n \in \mathbb{N}$, are the same and equal that of the optimal strategy:

$$\mathfrak{R}_{x^0}^{\overset{\star}{\delta}_{(n)}}(\beta) = \mathfrak{R}_{x^0}^{\delta^\star}(\beta), \qquad x_0 \in X. \tag{16}$$

From

$$\lim_{n\to\infty} \mathfrak{R}_{x^0}^{\overset{\star}{\delta}_{(n)}}(\beta) = \mathfrak{R}_{x^0}^{\delta^\infty}(\beta), \qquad x_0 \in X, \tag{17}$$

(see [20]) and because $\overset{\infty}{\delta} \in LD_M$ we have

$$\mathfrak{R}_{x^0}^{\overset{\infty}{\delta}}(\beta) = \mathfrak{R}_{x^0}^{\delta^\star}(\beta). \tag{18}$$

We shall show that it is possible to construct this Markov strategy (satisfying (18)) in the class LD_S of stationary Markov strategies. Bellman's equation for $\mathfrak{R}_x(\beta)$ is

$$\mathfrak{R}_{x^0}(\beta) = \min_{\substack{\Delta^0:\, \Delta^0(x^0)=u^0\in U(x^0)\\ \Delta^0 \ local}} \Big\{ r(x^0,u^0) + \beta \sum_{y\in X} \mathfrak{R}_y(\beta) Q(y/x^0,u^0)) \Big\}. \tag{19}$$

Because for the strategy $\overset{\infty}{\delta}(\beta)$ equality (18) holds, determining decision $\overset{\infty}{\Delta}{}_i^0(x^0)$ is reduced to solving equation (19) in which $U(x^0) \subseteq U$. Suppose now that the decisions $\overset{\infty}{\Delta}{}_i^0(x^0), \dots, \overset{\infty}{\Delta}{}_i^p(x^p)$, $i \in V$, are found and they depend on the respective observations $x^0, \dots, x^p$ only. From inequality (10) the continuation $\overset{\infty}{\delta}{}^{p+1}(\beta) = \big\{ \overset{\infty}{\delta}{}_i^{p+1}(\beta) = \{\overset{\infty}{\Delta}{}_i^{p+1}(x^{p+1}), \overset{\infty}{\Delta}{}_i^{p+2}(x^{p+2}), \dots\},\ i \in V \big\}$ of strategy $\overset{\infty}{\delta}(\beta)$ is such that

$$E^{\overset{\infty}{\delta}{}^p}\Big\{ \sum_{t=p+1}^{\infty} \beta^t r(\xi^t, \overset{\infty}{\Delta}{}^t) / \xi^p = x^p \Big\} \le E^{\delta^p}\Big\{ \sum_{t=p+1}^{\infty} \beta^t r(\xi^t, \Delta^t) / \xi^p = x^p \Big\}$$

for each continuation $\delta^{p+1}(\beta) = \big\{ \delta_i^{p+1}(\beta) = \{\Delta_i^{p+1}(x^{p+1}), \Delta_i^{p+2}(x^{p+2}), \dots\},\ i \in V \big\}$ of the Markov strategy $\delta(\beta)$. It is easy to see that

$$E^{\overset{\infty}{\delta}{}^{p+1}_{x^p}}\Big\{ \sum_{t=p}^{\infty} \beta^t r(\xi^t, \overset{\infty}{\Delta}{}^t) / \xi^p = x^p \Big\} = \beta^{p+1} \mathfrak{R}_{x^p}(\beta)$$

and that the optimal decisions $\overset{\infty}{\Delta}{}_i^p$, $i \in V$, for the strategy $\delta(\beta)$ for every fixed x^p can be found as a solution of the equation

$$\mathfrak{R}_{x^p}(\beta) = \min_{\substack{\Delta^p:\Delta^p(x^p)=u^p\in U(x^p)\\ \Delta^p \ local}} \Big\{ r(x^p,u^p) + \beta \sum_{y\in X} \mathfrak{R}_y(\beta) Q(y/x^p,u^p) \Big\}$$

which coincides with equation (19) because neither the decision space $U(x^p)$, nor the probabilities $Q(y/x^p,u^p)$ and costs $r(x^p,u^p)$ depend on time. Thus, we have proved that in the class of Markov strategies there exists some stationary strategy $\widetilde{\delta}(\beta)$ for which

$$\mathfrak{R}_{x^0}^{\widetilde{\delta}}(\beta) = \min_{\delta\in LD} \mathfrak{R}_{x^0}^{\delta}(\beta).$$

The strategy $\widetilde{\delta}(\beta) = \{\widetilde{\delta}_i(\beta),\ i \in V\}$ is by construction determined through the functions $\widetilde{\delta}_i(\beta) = (\widetilde{\Delta}_i^t(\beta) \colon t \in \mathbb{N})$, $i \in V$, where $(\widetilde{\Delta}_i^t(\beta) \colon t \in \mathbb{N})$, $(x_{\widetilde{N}(i)})$ are defined on finite sets $X_{\widetilde{N}(i)}$, $i \in V$, attaining therefore only a finite number of values. Therefore it is possible to find an infinite sequence of discount factors $\beta_\alpha \to 1$ as $\alpha \to \infty$ such that for all of them the optimal strategies $\widetilde{\delta}(\beta_\alpha)$ coincide: $\widetilde{\delta}(\beta_\alpha) = \delta^\star$, $\delta^\star = \{\delta_i^\star = \overset{\star}{\Delta}_i(x), i \in V\}$.

Let us consider this strategy $\delta^\star$. The process ξ associated with $\delta^\star$ is a homogeneous Markov chain (in fact: a time dependent controlled Markov random field according to Remark 2.2) with transition probabilities

$$P\{x/y, \overset{\star}{\Delta}_i(y_{\widetilde{N}(i)}),\ i \in V\} =: Q(x/y).$$

For this fixed strategy $\delta^\star$ we consider

$$\limsup_{T\to\infty} \frac{1}{T+1} \sum_{t=0}^{T} E_y^{\delta^\star} r(\xi^t, \overset{\star}{\Delta}) = R_y^{\delta^\star}.$$

Denoting $r(x) := r(x, \overset{\star}{\Delta}(x))$ we obtain

$$\frac{1}{T+1} \sum_{t=0}^{T} E_y^{\delta^\star} r(\xi^t, \overset{\star}{\Delta}) = \sum_{x\in X} r(x)\varphi_{T,y}(x),$$

where

$$\varphi_{T,y}(x) := \frac{1}{T+1} \sum_{t=0}^{T} P_y^{\delta^\star}\{\xi^t = x\}.$$

Similarly to the case of controlled Markov chains with finite state space ([6] and [20]) we have for the fixed strategy $\delta^\star$ convergence

$$\lim_{T\to\infty} \varphi_{T,y}(x) =: \phi_y^\star(x),$$

and therefore the limit

$$\lim_{T\to\infty} \frac{1}{T+1} \sum_{t=0}^{T} E_y^{\delta^\star} r(\xi^t, \overset{\star}{\Delta}) = R_y^{\delta^\star}$$

exists. If, independent of the initial value y of ξ, the Markov chains controlled by $\delta^\star$ have only one and the same ergodic class, we have

$$\lim_{T\to\infty} \varphi_{T,y}(x) =: \phi^\star(x),$$

independent of the initial values. It follows

$$R^{\delta^\star} = R_y^{\delta^\star}, \qquad y \in X.$$

Following the arguments in [20] the optimality of the stationary Markov strategy $\delta^\star$ in the class LD of all local admissible strategies is proved. □

Remark 3.1. We have restricted our considerations to the study of nonrandomized strategies. Using Bellman's optimality equation and standard techniques it can be shown that due to the finiteness of the decision space U for proving optimality considerations it is not necessary to consider randomized strategies because some extremum will always be attained in the class of nonrandomized strategies. ⊙

From Theorem 3.1 we conclude that determining an optimal strategy can be done using linear programming techniques, see [5] or [11] for details.

Example 3.1. We sketch an example from the area economical decisions, for details see [2]. Let $\Gamma = (V, B)$ be a finite undirected graph. The set of vertices denotes a group of organizations (or companies) and the edges indicate interactions between the organizations. Let $X_i = \{x_i^1, ..., x_i^{n_i}\}$ be the set of possible states for i, state x_i in X_i indicating the usage of a specific standard of technology by the i-th company. The global states of the group are $X = \underset{i\in V}{\times} X_i$.

$\widetilde{N}(k)$ is the subgroup of organizations interacting with the k- th company and k itself. The Markov property according to Definition 2.2 implies that decisions of organization k at time t depend only on the states and decisions of organizations interacting with the organization k and of that of k itself at time t. This is a natural assumption put on the behavior of many real systems.

In case that a company can chose between two technologies only at any time instant, we may describe this situation by a variant of the stochastic Ising model from statistical physics [18, 17]. Transition probabilities for this model are of the form [14]

$$P\{\xi_j^{t+1} = x_j / \xi_{\widetilde{N}(j)} = y_{\widetilde{N}(j)}\} = \frac{\exp\left[-\beta \sum_{j\in\widetilde{N}(k)} y_k x_j\right]}{\sum_{k\in V} \exp\left[-\beta \sum_{j\in\widetilde{N}(k)} y_k x_j\right]},$$

where y_k and x_j attain values ± 1 and $\beta > 0$.

For subgroup $S \subset V$ of companies we want to determine an optimal strategy based on the previous behavior of companies in their neighborhood only. Let $U_i = \{-1, 1\}$ be the decision set for company $i \in S$. -1 corresponds to choosing the standard -1 and 1 corresponds to the choosing the standard 1. The set of admissible strategies when being in state x describes restrictions on the possible decisions by $U(x) = \underset{i\in S}{\times} U_i(x_{\widetilde{N}(i)}$. If $r(x, u)$ be the cost per time unit when we are in a state x and decide for u, then the expected average costs per time unit is generally accepted as a criterion for evaluating production standards. From the results derived above there exists an optimal stationary Markov nonrandomized strategy in the set of local strategies.

To reduce determining an optimal strategy to solving a linear program we usually need explicitly given the stationary probabilities of the system, which for the Ising model may be found in [15].

4 Computational example: Cyclic queues

We reconsider the cyclic network of queues described in Example 2.1 and assume that the local decision makers at the nodes may adapt the service probabilities at their nodes according to the load of their own node and of the neighbored nodes. So their strategies are of local nature. According to Remark 3.1 optimal behaviour of the local decision makers can be achieved already by using deterministic policies only.

The local decision spaces $U_i = \{u^1, u^2, \ldots, u^{n_i}\}$, $i = 0, 1, \ldots, J$, are finite and do not depend on t. Then globally observed joint actions of the decision makers are from $U = \underset{i=0,1,\ldots,J}{\times} U_i$.

From Theorem 3.1 we know that an optimal local strategy $\delta^\star$, (with respect to asymptotic average expected costs (7),) can be found in the class LD_S of stationary Markovian policies, i.e. $\delta^\star = (\delta_i^\star(x_{\widetilde{N}(i)}) \colon i = 1, 2, \ldots, J)$. We therefore consider in the following at time instant t for vertex j decision $\Delta_j^t(x_{\widetilde{N}(j)})$ according to Definition 2.6, which depends on states of its neighborhood only. It follows that at the end of a time slot at node j a customer's

service ends with probability $p_j(h,u) \in (0,1)$ and with probability $q_j(h,u) := p_j(h,u)$ that customer will stay there for at least on further time slot, $h \geq 1$, $u \in U_j$. Consequently, we obtain a local Markov strategy $\delta = (\delta_i \colon i = 1,2,\ldots,J)$ according to Definition 2.4 **(2)** and a controlled process (ξ, δ) which is according to Definition 2.8 a controlled time dependent random field.
For any strategy $\delta \in LD_S$ which results in service probabilities $p_j(h,u) \in (0,1)$ at the nodes the controlled process (ξ, δ) is ergodic on state space $S(K,J)$ with some stationary distribution $\pi^{K,J,\delta} =: \pi^\delta = (\pi^\delta(x) \colon x \in S(K,J))$. But note, that in general (5) of Theorem 2.1 does not apply. The one-step costs occuring at time $t \in \mathbb{N}$ when being in state $\xi^t = x^t$ and applying action u^t are $r(x^t, u^t) \geq 0$. To determine an optimal strategy we shall use the *strategy improvement procedure* [6, section 6]:

Select some strategy δ, and consider for some (unknown) function $v = (v(x) : x \in S(K,J))$ the equations [6, p. 56, 66]:

$$R_y^\delta + v(y) = r(y, \delta(y)) + \sum_{x \in S(K,J)} Q(x/y, \delta(y)) v(x), \qquad y \in S(K,J); \tag{20}$$

and

$$\sum_{x \in S(K,J)} \pi^\delta(x) v(x) = 0. \tag{21}$$

Solving for $\{(v(x), R_y^\delta) : x, y \in S(K,J)\}$ yields the costs R_y^δ, when using δ, which turns out to be independent of initial state y. The policy improvement algorithm is [6, p. 70]:
For each $y \in S(K,J)$, define U^y to be the set of actions u in state y for which

$$\sum_{x \in S(K,J)} Q(x/y, u) R_x^\delta < R_y^\delta,$$

or, if no actions satisfy the inequality, the set that satisfies

$$\sum_{x \in S(K,J)} Q(x/y, u) R_x^\delta = R_y^\delta \tag{22}$$

and

$$\begin{aligned} r(y,u) + \sum_{x \in S(K,J)} Q(x/y, u) v^\delta(x) &< r(y, \delta(y)) + \sum_{x \in S(K,J)} Q(x/y, \delta(y)) v^\delta(x) \\ &= R_y^\delta + v^\delta(y). \end{aligned} \tag{23}$$

Starting with the prescribed strategy δ we define some local strategy $\delta' \in LD_S$ which takes some action $u \in U^y$ in at least one state y for which U^y is nonempty; otherwise, the action taken is the one dictated by δ.

Theorem 4.1. (see [6, p. 70, 71]) (a) If $\delta' \neq \delta$, then $R_y^{\delta'} \leq R_y^\delta, \qquad y \in S(K,J)$.

(b) The strategy improvement procedure leads to an optimal strategy within a finite number of iterations. The optimal strategy is obtained if U^y is empty for all y and then the actual policy fulfills: $\delta' =: \delta^\star$ is optimal. $\odot$

To demonstrate the application of the procedure we specialize now the data.

Example 4.1. Consider the cyclic network of queues described in Example 2.1 with If $\delta' \neq \delta$, then $R_y^{\delta'} \leq R_y^{\delta}, \qquad y \in S(K,J)$. node set $\{1,2,3\}$, and $K = 2$ customers. Selecting service probabilities is according to binary alternatives with $U_j \equiv \{0,1\} \quad \forall\, j \in \{1,2,3\},$, where u_j is indicates decision for a specific service capacity. We assume $p_j(h,u)$ and $r(x,u)$ as

$$\begin{pmatrix} (p_1(1,0);p_1(1,1)) & (p_1(2,0);p_1(2,1)) \\ (p_2(1,0);p_2(1,1)) & (p_2(2,0);p_2(2,1)) \\ (p_3(1,0);p_3(1,1)) & (p_3(2,0);p_3(2,1)) \end{pmatrix} = \begin{pmatrix} (1/3;1/2) & (2/3;3/4) \\ (1/4;1/2) & (1/3;2/3) \\ (1/4;2/3) & (1/2;3/4) \end{pmatrix} \tag{24}$$

$$\text{and } r(x,u) = \sum_{j \in S} \big[(c_j - u_j)x_j + u_j b_j\big], \tag{25}$$

where $c = (1,2,3)$, $b = (3,1,2)$, i.e. b_j is a cost associated to using additional service capacity for node j. We abbreviate the states of $S(3,2)$ by

$$\begin{array}{lll} x^1 = (0,0,2), & x^2 = (0,1,1), & x^3 = (0,2,0), \\ x^4 = (1,0,1), & x^5 = (1,1,0), & x^6 = (2,0,0), \end{array}$$

and the decision vectors

$$\begin{array}{llll} u^1 = (0,0,0), & u^2 = (0,0,1), & u^3 = (0,1,0), & u^4 = (0,1,1), \\ u^5 = (1,0,0), & u^6 = (1,0,1), & u^7 = (1,1,0), & u^8 = (1,1,1). \end{array}$$

Then the transition kernels $Q(x/y,u)$ are given by

Q	$(0,0,2)$	$(0,1,1)$	$(0,2,0)$	$(1,0,1)$	$(1,1,0)$	$(2,0,0)$
x^1	$q_3(2,u_3)$	$p_2(1,u_2)q_3(1,u_3)$	0	0	0	0
x^2	0	$q_2(1,u_2)q_3(1,u_3)$	$p_2(2,u_2)$	$p_1(1,u_1)q_3(1,u_3)$	$p_1(1,u_1)p_2(1,u_2)$	0
x^3	0	0	$q_2(2,u_2)$	0	$p_1(1,u_1)q_2(1,u_2)$	0
x^4	$p_3(2,u_3)$	$p_2(1,u_2)p_3(1,u_3)$	0	$q_1(1,u_1)q_3(1,u_3)$	$q_1(1,u_1)p_2(1,u_2)$	0
x^5	0	$q_2(1,u_2)p_3(1,u_3)$	0	$p_1(1,u_1)p_3(1,u_3)$	$q_1(1,u_1)q_2(1,u_2)$	$p_1(2,u_1)$
x^6	0	0	0	$q_1(1,u_1)p_3(1,u_3)$	0	$q_1(2,u_1)$

We start with strategy $\delta(x) \equiv u^1 = (0,0,0)$. From (24)–(25) we have for this δ: $p_j(x_j) = p_j(x_j, \Delta(x_j))$ and $r(x) = r(x,\delta(x))$ as follows:

$$\begin{pmatrix} p_1(1) & p_1(2) \\ p_2(1) & p_2(2) \\ p_3(1) & p_3(2) \end{pmatrix} = \begin{pmatrix} 1/3 & 2/3 \\ 1/4 & 1/3 \\ 1/4 & 1/2 \end{pmatrix}$$

$r(x^1)$	$r(x^2)$	$r(x^3)$	$r(x^4)$	$r(x^5)$	$r(x^6)$
6	5	4	4	3	2

For the prescribed δ Theorem 2.1 applies and we obtain the ergodic distribution $\pi^\delta = (\pi^\delta(x) \colon x \in S(K,J))$ of ξ and costs R^δ :

$$\pi^\delta = \left(\frac{3}{29}; \frac{8}{29}; \frac{9}{58}; \frac{6}{29}; \frac{6}{29}; \frac{3}{58}\right), \quad \textit{and} \quad R^\delta = \sum_{x \in S(K,J)} \pi^\delta(x) r(x) = \frac{121}{29}.$$

Applying (20)-(21) we have $v^\delta(x)$:

$v^\delta(x^1)$	$v^\delta(x^2)$	$v^\delta(x^3)$	$v^\delta(x^4)$	$v^\delta(x^5)$	$v^\delta(x^6)$
$\frac{7469}{3364}$	$\frac{31023}{16820}$	$\frac{22323}{16820}$	$-\frac{4827}{3364}$	$-\frac{31153}{16820}$	$-\frac{85963}{16820}$

We determine the sets U^y, $y \in S(K,J)$ at this stage of the algorithm. Equation (22) is satisfied for each $y \in S(K,J)$. From (23) we have that $U^{x^1} \ni u^2$, $U^{x^2} \ni u^2$, $U^{x^3} \ni u^3$, $U^{x^4} \ni u^2$, $U^{x^5} = \emptyset$, $U^{x^6} = \emptyset$.

Thus, we may define a new strategy δ' as

$\delta'(x^1)$	$\delta'(x^2)$	$\delta'(x^3)$	$\delta'(x^4)$	$\delta'(x^5)$	$\delta'(x^6)$
u^2	u^2	u^3	u^2	u^1	u^1

or shortly $\delta'(x) = (0, x_2 \operatorname{div} 2, \operatorname{sign} x_3)$, where $a \operatorname{div} b$ is the integral part obtained from dividing a by b, $\operatorname{sign} a$ is the sign of number a ($\operatorname{sign} a \in \{-1, 0, 1\}$).

For the new strategy $\delta'(x)$ (24)–(25) yield $p'_j(x_j) = p_j(x_j, \Delta'(x_j))$ and $r'(x) = r(x, \delta(x))$:

$$\begin{pmatrix} p'_1(1) & p'_1(2) \\ p'_2(1) & p'_2(2) \\ p'_3(1) & p'_3(2) \end{pmatrix} = \begin{pmatrix} 1/3 & 2/3 \\ 1/4 & 2/3 \\ 2/3 & 3/4 \end{pmatrix}$$

$r(x^1)$	$r(x^2)$	$r(x^3)$	$r(x^4)$	$r(x^5)$	$r(x^6)$
6	6	3	5	3	2

Theorem 2.1 yields the ergodic distribution $\pi^{\delta'} = (\pi^{\delta'}(x) \colon x \in S(K,J))$ and costs $R^{\delta'}$:

$$\pi^{\delta'} = \left(\frac{1}{46}; \frac{9}{46}; \frac{27}{184}; \frac{27}{184}; \frac{9}{23}; \frac{9}{92}\right), \quad \textit{and} \quad R^{\delta'} = \sum_{x \in S(K,J)} \pi^{\delta'}(x) r'(x) = \frac{177}{46}.$$

Solving (20)-(21) for $v^{\delta'}(x)$ we have:

$v^{\delta'}(x^1)$	$v^{\delta'}(x^2)$	$v^{\delta'}(x^3)$	$v^{\delta'}(x^4)$	$v^{\delta'}(x^5)$	$v^{\delta'}(x^6)$
$\frac{103227}{42320}$	$\frac{212523}{84640}$	$\frac{104883}{84640}$	$-\frac{36429}{84640}$	$-\frac{67801}{84640}$	$-\frac{302401}{84640}$

Equation (22) is now satisfied for all $y \in S(K,J)$. From (23) we conclude that $U^y = \emptyset$ for all $y \in S(K,J)$. So the strategy $\delta^\star = \big(\delta'(x) = (0, x_2 \operatorname{div} 2, \operatorname{sign} x_3), x \in S(K,J)\big)$ is optimal and $R^{\delta^\star} = \frac{177}{46}$. $\odot$

5 Separable cost functions and global optimality of local strategies

Up to now we were mainly concerned with finding optimal strategies in classes of local strategies. Clearly this is an essential problem if (as it is often the case) local strategies are the only possible policies which can be applied because of practical limitations.
But there is a related problem which deals with the question whether local strategies can be globally optimal. We investigate in this section conditions which imply that locally defined policy classes contain strategies which are optimal even in the class of all admissible policies.

For now we consider cost functions $r(x,u)$ which are separabel (additive) functions with respect to the neighborhood graph as follows.

$$r(x,u) = \sum_{i \in V} r_i(x_{\widetilde{N}(i)}, u_i), \tag{26}$$

and we shall assume that according to Definition 2.3 the controls sets are globally dependent on state and time.

Definition 5.1. (1) For any vertex $i \in V$ and any system configuration (state) x^t let $U_i^t(x^t)$ be the set of admissible actions (admissible decision values) for Δ_i^t at time t if $\xi^t = x^t$. We denote $U^t(x^t) = \underset{i \in V}{\times} U_i^t(x^t)$.

(2) A strategy δ according to Deinition 2.3 (2) is called admissible if for every $t \in \mathbb{N}$ $\Delta^t(x^0, \ldots, x^{t-1}, x^t) = \{\Delta_i^t(x^0, \ldots, x^{t-1}, x^t) \colon i \in V\} \in U^t(x^t)$ given $\xi^t = x^t$.

A strategy $\delta = (\delta_i \colon i \in V)$ is called Markov strategy if for all $x^0, x^1, \ldots, x^t \in X$ $\Delta_i^t(x^0, \ldots, x^t) = \Delta_i^t(x^t)$, $i \in V$, i.e. every decision function Δ_i^t depends on the whole history only through the present state (x^t).
The class of all admissible deterministic strategies is denoted by D; the subclass of admissible Markov strategies by D_M.

(3) By D_S we denote similarly to Definition 2.7 (3) the class of admissible deterministic stationary Markov strategies with time invariant restriction sets: $\Delta^t = \Delta^{t'}$ for all $t, t' \in \mathbb{N}$ and $U^t(x) = U(x) = \underset{i \in V}{\times} U_i(x)$ for all t. $\odot$

Theorem 5.1. Suppose there exist a constant g and Borel function $v(x)$ on X such that for the transition function Q (Definition 2.8) of the controlled process with locally interacting synchronuous components holds

$$g + v(x) = \min_{u \in U(x)} \Big\{ r(x,u) + \sum_{y \in X} v(y) Q(y/x,u) \Big\}, \qquad x \in X. \tag{27}$$

Then

$$\inf_{\delta \in D} R_x^\delta \geq g, \qquad x \in X. \odot$$

Moreover, if for a strategy $\delta^\star \in LD_S$ holds

$$g + v(x) = r(x, \delta^\star(x)) + \sum_{y \in X} v(y) Q(y/x, \delta^\star(x)), \qquad x \in X, \tag{28}$$

then $\delta^\star$ is optimal and

$$R_x^{\delta^s tar} \equiv g. \tag{29}$$

Proof. It is obvious that for any strategy $\delta \in D$ the equality takes the place

$$E^{\delta}_{x^0}\Big\{\sum_{t=1}^{T}\big[v(\xi^t) - E^{\delta}\{v(\xi^t)/\xi^0 = x^0, \Delta^0 = u^0, \dots, \xi^{t-1} = x^{t-1}, \Delta^{t-1} = u^{t-1}\}\big]\Big\} = 0.$$

From the definition and from the Markov property follows

$$\begin{aligned} E^{\delta}\{v(\xi^t)/\xi^0 &= x^0, \Delta^0 = u^0, \dots, \xi^{t-1} = x^{t-1}, \Delta^{t-1} = u^{t-1}\} \\ &= \sum_{y\in X} v(y)Q(y/x^{t-1}, u^{t-1}) \\ &= r(x^{t-1}, u^{t-1}) + \sum_{y\in X} v(y)Q(y/x^{t-1}, u^{t-1}) - r(x^{t-1}, u^{t-1}) \\ &\geq g + v(x^{t-1}) - r(x^{t-1}, u^{t-1}). \end{aligned}$$

Therefore

$$E^{\delta}_{x^0}\sum_{t=1}^{T}\big[r(\xi^{t-1}, \Delta^{t-1}) + v(\xi^t) - v(\xi^{t-1}) - g\big] \geq 0,$$

and consequently

$$E^{\delta}_{x^0}\Big\{\sum_{t=1}^{T} r(\xi^{t-1}, \Delta^{t-1}) + v(\xi^T) - v(x^0)\Big\} \geq Tg. \tag{30}$$

Because X is finite and therefore $v(x)$ is bounded, from (30) we obtain

$$R^{\delta}_{x} \geq g, \qquad x \in X.$$

The second part of the Theorem follows from the fact that for strategy $\delta^{\star}$ in (28) and (29) equalitiy holds. □

In general local strategies cannot be expected to be globally optimal. But under mild technical conditions we can prove such a statement. We assume henceforth in this section

Assumption I. There exist non-negative measures μ_i on X_i with $\mu_i(X_i) > 0, \quad \forall\, i \in V$ such that:
$\mu_i(x_i) \leq Q(x_i/y_{\tilde{N}(i)}, u_i)\,, \quad x_i \in X_i\,, \quad y_{\tilde{N}(i)} \in X_{\tilde{N}(i)}\,, \quad u_i \in U_i\,, \quad i \in V\,.$ ⊙

Let $S(X)$ be Banach space of (bounded) functions $v(x) = \sum_{i\in V} v_i(x_i)$, whith norm $||v|| = \sum_{i\in V}\max_{x_i\in X_i}|v_i(x_i)|$. Denote the operator $\mathfrak{U}$ on $S(X)$ by:

$$\mathfrak{U}v(x) = \min_{u\in U}\Big\{\sum_{i\in V} r_i(x_{\tilde{N}(i)}, u_i) + \sum_{y\in X} v(y)Q'(y/x, u)\Big\},$$

where

$$Q'(y/x, u) = Q(y/x, u) - \prod_{i\in V}\mu_i(y_i).$$

Let

$$U'_v(x) = \Big\{u \in U: \quad \mathfrak{U}u(x) = \sum_{i\in V} r_i(x_{\tilde{N}(i)}, u_i) + \sum_{y\in X} v(y)Q'(y/x, u)\Big\}$$

and define the local operators

$$U_i^{v_i}(x_{\widetilde{N}(i)}, u_i) = r_i(x_{\widetilde{N}(i)}, u_i) + \sum_{y_i \in X_i} v_i(y_i)\big[Q_j(y_i/x_{\widetilde{N}(i)}, u_i) - \mu_i(y_i) \prod_{j:\, j \in V,\ j \neq i} \mu_j(X_j)\big].$$

Theorem 5.2. Under Assumption I in class LD_S a globally optimal strategy exists. $\odot$

Proof. Using standard arguments from [10], it follows that the operator $\mathfrak{U}$ is a contraction on $S(X)$. By Banach's fixed point theorem in $S(X)$ exists a fixed point $v^\star$ such that

$$\mathfrak{U}v^\star(x) = v^\star,$$

which is

$$v^\star(x) = \min_{u \in U}\Big\{\sum_{i \in V} r_i(x_{\widetilde{N}(i)}, u_i) + \sum_{y \in X} v^\star(y) Q'(y/x, u)\Big\}, \qquad x \in X.$$

Using the structure of the operator $\mathfrak{U}$ we have

$$\mathfrak{U}v^\star(x) = \sum_{i \in V} \min_{u_i \in U_i} U_i^{v_i^\star}(x_{\widetilde{N}(i)}, u_i).$$

From finiteness of U_i there exists the measurable functions $\Delta_i^\star = \Delta_i^\star(x_{\widetilde{N}(i)})$ such that

$$\min_{u_i \in U_i} U_i^{v_i^\star}(x_{\widetilde{N}(i)}, u_i) = U_i^{v_i^\star}(x_{\widetilde{N}(i)}, \Delta_i^\star(x_{\widetilde{N}(i)})).$$

Therefore there exists a function $\delta^\star = \{= \Delta_i^\star(x_{\widetilde{N}(i)}) : i \in V\} \in LD_S$ such that

$$\mathfrak{U}v^\star(x) = \sum_{i \in V} r_i(x_{\widetilde{N}(i)}, \Delta_i^\star(x_{\widetilde{N}(i)})) + \sum_{y \in X} v^\star(y) Q'(y/x, \delta^\star(x))$$

and

$$v^\star(x) = \sum_{i \in V} r_i(x_{\widetilde{N}(i)}, \Delta_i^\star(x_{\widetilde{N}(i)})) + \sum_{y \in X} v^\star(y) Q'(y/x, \delta^\star(x))$$

or

$$g + v^\star(x) = \sum_{i \in V} r_i(x_{\widetilde{N}(i)}, \Delta_i^\star(x_{\widetilde{N}(i)})) + \sum_{y \in X} v^\star(y) Q(y/x, \delta^\star(x)),$$

where

$$g = \sum_{i \in V} \sum_{y \in X} v_i^\star(y_i) \mu_i(y_i) \prod_{j:\, j \in V,\ j \neq i} \mu_j(X_j).$$

So the conditions of the Theorem 5.1 are satisfy and strategy $\delta^\star \in LD_S$ is optimal. $\square$

References

[1] Daduna H (1997) Some results for steady–state and sojourn time distributions in open and closed linear networks of Bernoulli servers with state–dependent service and arrival rates. Performance Evaluation 30: 3–18

[2] David PA, Foray D (1992) Percolation structures, Markov random fields. The economics of EDI standards diffusion. Center for Economical Policy Research, Stanford University

[3] David PA, Foray D, Dalle J-M (1996) Marshallian externalities and the emergence and spatial stability of technological enclaves. Preprint Center for Economical Policy Research, Stanford University

[4] Derman C (1963) Stable sequential control rules and Markov chains. J. Math. Anal.and Appl. 6: 257–265

[5] Derman C (1964) Markovian sequential decision processes. Proc. Symposia Appl. Math. 16: 281–289

[6] Derman C (1970) Finite State Markovian Decision Processes. Academic Press, New York

[7] Desert B, Daduna H (1999) Discrete time tandem networks of state dependent queues: The effect of different regulation schemes for simultaneous events on customer oriented performance measures. Preprint 99-07, Inst. Math. Stochastics, University of Hamburg

[8] Dobrushin RL, Kryukov VI, Toom AL (ed.) (1978) Locally Interacting Systems and Their Application in Biology. Lecture Notes in Mathematics 653, Springer Berlin

[9] Gravey A, Hebuterne G (1992) Simultaneity in discrete–time single server queues with Bernoulli inputs. Performance Evaluation 14: 123–131

[10] Gubenko LG, Shtatland ES (1972) On discrete time Markov decision processes. Teor. Verojatnost. i Mat. Stat.7: 51–64, English translation in: Theory of Probability and Mathematical Statistics 7: 47–61 (1975)

[11] Howard R A (1960) Dynamic Programming and Markov Processes. Wiley, New York

[12] Hunter J (1983) Mathematical Techniques in Applied Probability Vol.I, II. Academic Press, New York

[13] Kelly F P (1979) Reversibility and Stochastic Networks. Wiley, New York

[14] Knopov PS (1996) Markov fields and their applications in economics. Obchysljuvalna ta prykladna matematyka 80: 33–46 [in Russian]

[15] Liggett T (1989) Interacting particle systems. Springer-Verlag, New York

[16] Pujolle G, Claude JP, Seret D (1986) A discrete–time queueing system with a product form solution. in: Hasegava T, Takagi H, Takahashi V (eds.) Computer Networking and Performance Evaluation 139 - 147 Elsevier, Amsterdam

[17] Ruelle D (1969) Statistical mechanics. WA Benjamin Inc., New York Amsterdam

[18] Sinay YaG (1980) The Theory of phase transitions. Nauka, Moscow [in Russian]

[19] Vasilyev, NB (1978) Bernoulli and Markov stationary measures in discrete local interactions. in: Dobrushin, RL, Kryukov, VI, Toom,AL (ed.) Locally Interacting Systems and Their Application in Biology. Lecture Notes in Mathematics 653: 99–112, Springer, Berlin

[20] Viskov OV, Shiryaev AN (1964) Controls leading to optimal stationary modes. In: Math. Inst. of Acad. of Sci. USSR Collect. for Theory of Probab. and Math. Statist., Nauka, Moscow, pp. 35–45

[21] Whittle P (1986) Systems in Stochastic Equilibrium Wiley , New York