UNIVERSITÄT HAMBURG

Local control of interacting Markov processes on graphs with compact state space

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

Preprint–No. 2000-09 November 2000

SCHWERPUNKT MATHEMATISCHE STATISTIK
UND STOCHASTISCHE PROZESSE

BUNDESSTR. 55, D – 20146 HAMBURG

Local control of interacting Markov processes on graphs with compact state space *

Ruslan K. Chornei Hans Daduna Pavel S. Knopov

Summary: Discrete time Markov processes with multidimensional compact state space are considered where the coordinate processes are locally interacting and change their states synchronously. The interaction structure of the process is determined by some general graph. Decision makers control synchronuosly the system's behaviour on the coordinate level using only local information. Conditions are given which guarantee that in the class of local randomized strategies there exist deterministic stationary Markov strategies which achieve maximal asymptotic average expected rewards. If the reward structure is separabel these strategies are even globally optimal.

Keywords: Markov random fields with compact state space, Markov processes in discrete time, synchronous transitions, local transitions, synchronuos control, asymptotic average expected rewards, Markov policy, local strategies

MSC Classification: 60J05, 60K35, 93E20

1 Introduction

In this note we investigate interacting Markov processes with compact state space over some (in general nonregular) undirected graph. The graph determines the neighborhood structure for the system: If vertices are connected by an edge, they are assumed to influence one another vice versa with respect to their evolution. These processes are equipped with a control structure of local character in such a way, that decision makers which are located at the nodes of the graph, make their decisions according to randomized policies (strategies) which take into consideration the information provided by the history of the neighbored vertices.

We assume the system to be observed at discrete time points only, and that these time points are control instants as well. To optimize such systems poses problems due to restrictions on the information gain which arise from observing the system over time only locally. (Clearly, an optimal policy should incorporate knowledge about the complete information provided by the system's space–time development up to the decision instant.) Other sources of difficulties are e.g. cost or reward functions which react on global behaviour of the system although only local control is possible by the decision makers.

*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 .

We describe in section 2 the classes of controlled space–time processes of interest. With prescribed neighborhood systems we fix the interaction structure for the coordinate processes and determine the transition mechanism for the controlled processes which then evolve being governed by controls which are compatible with the neighborhood structures. The general class of *randomized local* policies obtained in this way gives rise to problems and classifications analoguously to those in the classical theory. The question then arises whether we 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 question is answered positively in section 3 by providing conditions under which in the class of local randomized policies the (local) optimum is attained by stationary deterministic policies.
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 4 concentrating on the case where the reward functions are separable with respect to the neighborhood graph.

Spatially distributed systems with stochastic influences are usually modeled using *random fields* over some multidimensional lattice Z^d, or subsets thereof, or over some general graph. Control of stochastically driven systems over time is part of the theory of stochastic processes. Therefore our generic model in this note is 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 [10]. The class of stochastic processes used there is that of *Markovian processes* in time and for the spatial structure is that of *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.

The processes defined here generalize the controlled space–time processes from [2] with discrete time and finite state space. The processes described in section 2 resemble to a certain extend Kelly's *Spatial Processes* ([7, setion 9, p.189], see as well [13]). We do not impose the restriction posed by Kelly that state changes are of strong local character, i.e., only one coordinate may change its local state at a time instant: We allow here that several coordinates of a state may change simultaneously, which is essential e.g. for queueing network, where by a transition of a customer simultaneous state changes of at least two coordinate processes occur.

Compactness and continuity conditions as imposed here on state space and action space are standard assumptions for many investigations in the literature on stochastic dynamic optimization or dynamic programming, see [11], [1], and the references there. Our new point of view is the local structure imposed on these spaces in addition.

These locally defined properties can be found in many old as well as recent applications in economics [3], [4], Biology [5], Physics [10], Operations Research and Engineering [7].

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

2 Description of the system

For the systems under consideration with locally interacting coordinates the interaction structure is defined via an undirected finite neighborhood graph $\Gamma = (V, B)$ with set of vertices (nodes) V and set of edges $B \subset V^2 - diag(V)$. Denote by $\{k, j\}$ the edge of the graph connecting vertices k and j. The neighborhood of vertex k is the node set $N(k) = \{j\colon \{k, j\}) \in B\}$. The complete neighborhood of vertex k is $\widetilde{N}(k) = N(k) \cup \{k\}$, i.e. the neighborhood of the vertex k, and including k. For any $K \subset V$ we define the neighborhood $N(K) = \bigcup_{k\in K} N(k) - K$, and the complete neighborhood $\widetilde{N}(K) = N(K) \cup K$. For every subset of nodes $K \subset V$ we denote the marginal vector of state x by $x_K = (x_k,\ k \in K) \in X_K = \underset{i\in K}{\times} X_i$.

For every node $i \in V$ let X_i be some compact metric space with countable basis, the local state space at that node. Then $X := \underset{i\in V}{\times} X_i$ is the global state space of the system. Let $\mathfrak{X}_i$ be the sigma-field of Borel subsets of X_i and $\mathfrak{X}_K = \sigma\{\underset{i\in K}{\times} \mathfrak{X}_i\}$ the sigma-field generated by $\underset{i\in K}{\times} \mathfrak{X}_i$, $K \subseteq V$, $\mathfrak{X}_V = \mathfrak{X}$.

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

Definition 2.1. (1) A random variable ξ defined on $(\Omega, \mathcal{F}, P)$ with values in X is called a random field over (V, B) (or simply a random field over V). The marginal random variables with values in the space X_K are denoted by ξ_K. For $K = \{k\}$ we write ξ_k.

(2) A random field ξ on (V, B) is a Markov field if for all $k \in V$ the following holds:

$$P(\xi_k \in C_k/\xi_{V-\{k\}}) = x_{V-\{k\}}) = P(\xi_k \in C_k/\xi_{N(k)} = x_{N(k)}), \quad \forall\ x \in X, C_k \in \mathfrak{X}_k$$

where $P(\xi_k \in C_k/\xi_{N(k)} = x_{N(k)})$, is the regular conditional probability on $(X_k, \mathfrak{X}_k)$ given $\xi_{N(k)} = x_{N(k)}$. $\odot$

In many problems related to economical or engineering applications random fields describe the state of a system for some fixed time instant. The evolution over time of that system is then described by a stochastic process with random fields as one-dimensional marginals in time. In this case we write $\xi = (\xi^t)$ to emphasize the time-dependence. We assume that t takes discrete values $t = 0, 1, \dots$. 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 $\xi = (\xi^t\colon t = 0, 1, \dots)$.

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 that with respect to the evolution over time the value $\xi_k^t = x_k$ of the k-th vertex depends the on the previous states of the whole system only through the values of the vertices in $\widetilde{N}(k)$ (including k) at $t - 1$. To describe this we introduce the notion of Markov property in space and in time along the lines of Definition 2.1.

Definition 2.2. Let $\xi = \{\xi^t, t = 0, 1, \dots\}$ be a discrete time Markov process with state space X. The transition probabilities of X are said to be *local* [12, p. 100] if for all $k \in V, x^0, \dots, x^{t+1} \in X, C_k \in \mathfrak{X}_k$

$$P\{\xi_k^{t+1} \in C_k/\xi^t = x^t, \dots, \xi^0 = x^0\} = P\{\xi_k^{t+1} \in C_k/\xi^t_{\widetilde{N}(k)} = x^t_{\widetilde{N}(k)}\} \tag{1}$$

holds, i.e. the state of vertex k depends on the state of its complete neighborhood at the previous time instant only. The transition probabilities of X are said to be *synchronous* [12, p. 100] if

$$P\{\xi_K^{t+1} \in C_K/\xi^t = x^t\} = \prod_{k\in K} P\{\xi_k^{t+1} \in C_k/\xi^t = x^t\}, \qquad \forall K \subset V, \quad x^t \in X, \tag{2}$$

where $C_K = \underset{k\in K}{\times} C_k \in \mathfrak{X}_K$.

If ξ fulfills (1) and (2) it is called a *Markov process with locally interacting synchronous components over* (Γ, X), shortly, a *(time dependent) Markov random field.* $\odot$

In the following we shall describe the notion of a controlled Markov process with locally interacting synchronous components, shortly: a controlled time dependent Markov random field, as it was developed in [2] for the case of finite state spaces. For simplicity of the presentation we assume that some 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. Before presenting the details we recall some Definitions and Theorems from [9].

Definition 2.3. For nonempty sets X and A denote by $(2)^A_{set}$ the set of all subsets of A, and let $F\colon X \to (2)^A_{set}$ be a set valued function which associates with $x \in X$ a nonempty closed set $F(x) =: A(x) \subseteq A$ such that the set $\Delta = \{(x,a)\colon x \in X,\ a \in A(x)\} \subset X \times A$ is Borel measurable. (The function F will later determine the admissible decision (control) set $A(x)$ for all states x.)
The map $F\colon X \to (2)^A_{set}$ is called open, closed or Borel measurable if $\{x\colon F(x)\cap E \neq \emptyset\} \in \mathfrak{X}$, for all open, closed or Borel subsets E of A, respectively.
The function $f\colon X \to A$ is called a selector for the map F if $f(x) \in F(x)$ for all $x \in X$.

Theorem of Choice. [9, p. 74]. Any measurable map according to Definition 2.3 has a Borel measurable selector. $\odot$

Theorem of Choice for Semicontinuous Maps. [9, p.71]. If A is a compact metric space then any semicontinuous map $F\colon X \to (2)^A_{set}$ has a selector belonging to Baire class 1. $\odot$

We introduce classes of admissible policies to control the interacting coordinates. We shall mainly consider controls, which are local in the sence to be specified now.

Definition 2.4. (1) The set of actions (control values) usable at control instants is $A = \underset{i\in V}{\times} A_i$ over Γ, where A_i is a set of possible actions (decisions) for vertex i. We assume that A_i is a compact metric space with countable basis.

(2) If for the decision maker at node i at time t with state $\xi^t = x$ the set of control actions is restricted to $A_i^t(x) \subset A_i$, we call $A_i^t(x)$ the set of admissible actions (decisions) at time t in state x.

(3) We assume that the so defined set valued maps

$$F_i^t\colon X \to A_i, \quad x \to F_i^t(x) = A_i^t(x), \quad i \in V,$$

depend on $x \in X$ only through $x_{\widetilde{N}(i)} \in X_{\widetilde{N}(i)}$ and are measurable in the sense of Definition 2.3. The F_i^t determine therefore admissibility as a local property. We also assume the sets

$$\Delta_i^t = \{(x_{\widetilde{N}(i)}, a_i) \colon x_{\widetilde{N}(i)} \in X_{\widetilde{N}(i)}, a_i \in A_i^t(x_{\widetilde{N}(i)})\}$$

to be Borel measurable sets of the product space $X_{\widetilde{N}(i)} \times A_i$ and $\Delta^t = \underset{i \in V}{\times} \Delta_i^t$ to be Borel measurable in the product space $\widetilde{X} \times A$, $\widetilde{X} = \underset{i \in V}{\times} X_{\widetilde{N}(i)}$ and $\widetilde{\mathfrak{X}} = \sigma\{\underset{i \in V}{\times} X_{\widetilde{N}(i)}\}$. $\odot$

Definition 2.5. (1) Let α_i^t denote the action chosen by the decision maker at node i at time t, $\alpha^t := (\alpha_i^t : i \in V)$ the joint decision vector of all decision makers at time $t \in N$.

(2) A strategy (policy) δ to control the system with interacting components is defined as vector of coordinate policies $\delta = (\delta_i, i \in V)$, where for node i $\delta_i = \{\pi_i^0, \dots, \pi_i^t, \dots\}$ is a sequence of transition probabilities $\pi_i^t = \pi_i^t(\cdot / x^0, a^0, \dots, x^{t-1}, a^{t-1}, x^t)$. So π_i^t is a probability measure on $(A_i, \mathfrak{A}_i)$ for any $(x^0, a^0, \dots, x^{t-1}, a^{t-1}, x^t)$ and measurably dependent on the history $h^t = (x^0, a^0, \dots, x^{t-1}, a^{t-1}, x^t)$ of the system up to time t. We therefore have

$$\begin{aligned} & P(\alpha_i^t \in B_i / \xi^0 = x^0, \alpha^0 = a^0, \dots, \xi^{t-1} = x^{t-1}, \alpha^{t-1} = a^{t-1}, \xi^t = x^t) \\ = \; & \pi_i^t(B_i / x^0, a^0, \dots, x^{t-1}, a^{t-1}, x^t) \end{aligned}$$

for all $B_i \in \mathfrak{A}_i$. $\odot$

Definition 2.6. (1) If at any time $t = 0, 1, \dots$ the decision α_i^t for node i is made according to the probability π_i^t on basis of the history of the complete neighborhood $\widetilde{N}(i)$ of i only, i.e. on the basis of $h_i^t = (x^0_{\widetilde{N}(i)}, a_i^0, \dots, x^{t-1}_{\widetilde{N}(i)}, a_i^{t-1}, x^t_{\widetilde{N}(i)})$, and if $\pi_i^t(F_i^t(x)/h_i^t) = \pi_i^t(A_i^t(x^t_{\widetilde{N}(i)}/h_i^t) = 1$, $x^s_{\widetilde{N}(i)} \in X_{\widetilde{N}(i)}, a_i^s \in A_i^s(x^s_{\widetilde{N}(i)})$, i.e. π_i^t is locally admissible, then the sequence of transition probabilities (decisions) $\delta_i = \{\pi_i^t, t \in \mathbb{N}\}$ is called admissible local strategy for vertex i.

(2) An admissible local strategy $\delta = (\delta_i, i \in V)$ is called admissible local Markov strategy if

$$\pi_i^t(\cdot / x^0_{\widetilde{N}(i)}, a_i^0, \dots, x^{t-1}_{\widetilde{N}(i)}, a_i^{t-1}, x^t_{\widetilde{N}(i)}) = \pi_i^t(\cdot / x^t_{\widetilde{N}(i)}), \qquad i \in V.$$

(3) An admissible local Markov strategy $\delta = (\delta_i, i \in V)$ is called admissible stationary local Markov strategy if $\pi_i^{t'}(\cdot / x_{\widetilde{N}(i)}) = \pi_i^{t''}(\cdot / x_{\widetilde{N}(i)})$, $i \in V$, for all t' and t'' and all x.

(4) An admissible stationary local Markov strategy $\delta = (\delta_i, i \in V)$ is called admissible stationary local Markov deterministic (nonrandomized) strategy if $\pi_i(\cdot / x_{\widetilde{N}(i)})$, $i \in V$, are singular measures on $A_i^t(x_{\widetilde{N}(i)})$, $i \in V$, accordingly for each $x \in X$.

The class of all admissible local strategies is denoted by LS; the subclass of admissible local Markov strategies by LS_M. By LS_S we denote the class of admissible stationary local Markov deterministic strategies with time invariant restriction sets: $\pi^t = \pi^{t'}$ for all $t, t' \in \mathbb{N}$ and $A^t(x) = A(x) = \underset{i \in V}{\times} A_i(x_{\widetilde{N}(i)})$ for all t. $\odot$

Similar to (1) we shall construct our process in a way that a Markov property holds: The law of motion of the system is characterized by a set of time invariant transition probabilities. Whenever the state of the system is $\xi^t = x^t$ and decision $\alpha^t = a^t$ is taken the transition probability is $P\{\xi^{t+1} \in C / \xi^t = x^t, \alpha^t = a^t\} =: Q(C / x^t, a^t)$, and therefore independent of the past given the present (generalized) state. We further assume, that this transition

probability is independent of t, the motion is homogeneous in time. Applying a control policy to a time dependent Markov random field as defined in definition 2.2 we shall call the pair (ξ, δ) a controlled version of ξ using strategy δ. It is evident that a controlled process $\xi = (\xi^t)$ in general is not Markovian because the sequence (α^t) of decisions according to π_i^t depends not only on states $x^t_{\widetilde{N}(i)}$, $i \in V$, but on the previous (local) states $x^0_{\widetilde{N}(i)}, \ldots, x^{t-1}_{\widetilde{N}(i)}$ as well. An immediate consequence of the definition will be, that if a Markov strategy δ is applied as control strategy for the time dependent Markov random field then we obtain a Markov process ξ from (ξ, δ).

Definition 2.7. 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 $\xi = (\xi^t : t \in \mathbb{N})$ is a stochastic process with state space $X = \underset{i \in V}{\times} X_i$, $\delta = (\delta_i : i \in V)$ is an admissible local strategy, and the transitions of ξ are determined as follows:

$$
\begin{aligned}
&P\{\xi_K^{t+1} \in C_K / \xi^0 = x^0, \alpha^0 = a^0, \ldots, \xi^{t-1} = x^{t-1}, \alpha^{t-1} = a^{t-1}, \xi^t = y, \alpha^t = a\} \\
&\quad \overset{(1)}{=} P\{\xi_K^{t+1} \in C_K / \xi^0 = x^0, \ldots, \xi^{t-1} = x^{t-1}, \xi^t = y, \alpha^t = a\} \\
&\quad \overset{(2)}{=} P\{\xi_K^{t+1} \in C_K / \xi^t = y, \alpha^t = a\} \\
&\quad \overset{(3)}{=} \prod_{j \in K} P\{\xi_j^{t+1} \in C_j / \xi^t = y, \alpha^t = a\} \\
&\quad \overset{(4)}{=} \prod_{j \in K} P\{\xi_j^{t+1} \in C_j / \xi^t_{\widetilde{N}(j)} = y_{\widetilde{N}(j)}, \alpha_j^t = a_j\} \\
&\quad \overset{(5)}{=} \prod_{j \in K} Q_j\{C_j / y_{\widetilde{N}(j)}, a_j\} \\
&\quad \overset{(6)}{=} Q_K(C_K / y, a), \qquad K \subseteq V, \quad y \in X, \quad a_j \in A_j(y_{\widetilde{N}(j)}).
\end{aligned} \tag{3}
$$

If $K = V$ we shall write $Q_V(C_V / y, a) = Q(C / y, a)$.
The Markov kernel $Q = \prod_{j \in V} Q_j$ is said to be local and synchronuos.
The decisions to control the process are governed by a synchronuos control kernel

$$
\begin{aligned}
& P(\alpha^t \in \underset{i \in V}{\times} B_i / \xi^0 = x^0, \alpha^0 = a^0, \ldots, \xi^{t-1} = x^{t-1}, \alpha^{t-1} = a^{t-1}, \xi^t = x^t) \\
= \; & \prod_{j \in V} P(\alpha_i^t \in B_i / \xi^0 = x^0, \alpha^0 = a^0, \ldots, \xi^{t-1} = x^{t-1}, \alpha^{t-1} = a^{t-1}, \xi^t = x^t) \\
= \; & \prod_{j \in V} \pi_i^t(B_i / x^0, a^0, \ldots, x^{t-1}, a^{t-1}, x^t), \qquad B_i \in \mathfrak{A}_i, a^s \in A, x^s \in X.
\end{aligned} \tag{4}
$$

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

Some remarks on the modeling principle behind this definition may be in order:
It is assumed that the first expression is well defined, then $\overset{(1)}{=}$ follows iteratively from the definition of the decision rules. $\overset{(2)}{=}$ is 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.

If at time $t \in \mathbb{N}$ the system is in state $\xi^t = x^t$ and a decision for action $\alpha^t = a^t$ is made a (one–step) reward $r(x^t, a^t)$ is incurred to the system. Suppose $r(x, a)$ is a bounded Borel measurable function on $X \times A$.

The average expected reward up to time T when ξ is started with $\xi^0 = x$ and strategy δ is applied is then

$$\phi_T(x, \delta) = E_x^\delta \frac{1}{T+1} \sum_{t=0}^{T} r(\xi^t, \alpha^t),$$

where E_x^δ is expectation associated with the controlled process (ξ, δ) if $\xi^0 = x$.

The problem is to find a strategy δ which maximizes the asymptotic average expected reward

$$\phi(x, \delta) = \liminf_{T \to \infty} \phi_T(x, \delta).$$

Definition 2.8. A strategy $\delta^\star \in LS$ is called optimal in the class LS of local strategies if

$$\phi(x, \delta^\star) = \sup_{\delta \in LS} \phi(x, \delta), \qquad \forall\ x \in X.$$

Such a $\delta^\star$ will shortly be said to be locally optimal. $\odot$

3 Existence of best strategies in LS

We first prove in this section statements on the existence of best policies in subsets of the class of all policies for a given problem. The statements of the theorems are specialized to the case of local policies but we point out that the proofs work for general classes as well. Clearly such best policies in a subclass of all policies may be globally suboptimal, although the local optimum is attained within the prescribed class as will be shown.

Theorem 3.1. If there exist bounded Borel functions $g(x)$ and $v(x)$ on X such that for all $t \in \mathbb{N}$ and admissible control sets $A^t(x) := \underset{i \in V}{\times} A_i^t(x_{\widetilde{N}(i)})$

$$g(x) \geq \sup_{a \in A^t(x)} \int g(y) Q(dy/x, a), \qquad x \in X, \tag{5}$$

$$g(x) + v(x) \geq \sup_{a \in A^t(x)} \Big\{ r(x, a) + \int v(y) Q(dy/x, a) \Big\}, \qquad x \in X, \tag{6}$$

then

$$\sup_{\delta \in LS} \phi(x, \delta) \leq g(x), \qquad x \in X. \tag{7}$$

If additionally for a strategy $\delta^\star \in LS$

$$g(x) = \int g(y) Q(dy/x, \delta^\star(x)), \qquad x \in X,$$

$$g(x) + v(x) = r(x, \delta^\star(x)) + \int v(y) Q(dy/x, \delta^\star(x)), \qquad x \in X,$$

holds, then $\phi(x, \delta^\star) \equiv g(x)$, and $\delta^\star$ is optimal in LS (locally optimal). $\odot$

Proof. For any strategy $\delta \in LS$

$$E^{\delta}_{x^0}\Big\{\sum_{t=1}^{T}\big[v(\xi^t) - E^{\delta}\{v(\xi^t)/\xi^0 = x^0, \alpha^0 = a^0, \ldots, \xi^{t-1} = x^{t-1}, \alpha^{t-1} = a^{t-1}\}\big]\Big\} = 0. \tag{8}$$

From (3) and (6) we have

$$\begin{aligned} &E^{\delta}\{v(\xi^t)/\xi^0 = x^0, \alpha^0 = a^0, \ldots, \xi^{t-1} = x^{t-1}, \alpha^{t-1} = a^{t-1}\} \\ &\qquad = \int v(y)Q(dy/x^{t-1}, a^{t-1}) \\ &\qquad = r(x^{t-1}, a^{t-1}) + \int v(y)Q(dy/x^{t-1}, a^{t-1}) - r(x^{t-1}, a^{t-1}) \\ &\qquad \leq g(x^{t-1}) + v(x^{t-1}) - r(x^{t-1}, a^{t-1}). \end{aligned} \tag{9}$$

Inserting (9) into (8) we obtain

$$E^{\delta}_{x^0}\Big\{\sum_{t=1}^{T} r(\xi^{t-1}, a^{t-1}) + v(\xi^T) - v(x^0)\Big\} \leq E^{\delta}_{x^0}\sum_{t=1}^{T} g(\xi^{t-1}). \tag{10}$$

From (5) we have

$$\begin{aligned} E^{\delta}_{x^0} g(\xi^t) &= E^{\delta}_{x^0}\big\{E^{\delta}[g(\xi^t)/\xi^0 = x^0, \alpha^0 = a^0, \ldots, \xi^{t-1} = x^{t-1}, \alpha^{t-1} = a^{t-1}]\big\} \\ &= E^{\delta}_{x^0}\Big\{\int g(y)Q(dy/\xi^{t-1}, \alpha^{t-1})\Big\} \\ &\leq E^{\delta}_{x^0} g(\xi^{t-1}) \leq \cdots \leq E^{\delta}_{x^0} g(\xi^0) \\ &= g(x^0), \end{aligned} \tag{11}$$

and, consequently,

$$E^{\delta}_{x^0}\sum_{t=1}^{T} g(\xi^{t-1}) \leq T g(x^0). \tag{12}$$

Inserting (12) in (10) we obtain from the boundedness of $g(x)$ and $v(x)$ inequality (7).

To prove of the second part of the Theorem we note that for the strategy $\delta^\star \in LS$ as described the inequalities (9)–(12) are equalities. □

Corollary 3.1. If (5) and (6) hold and a strategy $\delta^\star \in LS_S \subseteq LS$ fulfills

$$\begin{aligned} g(x) &= \int g(y)Q(dy/x, \delta^\star(x)), \qquad x \in X, \\ g(x) + v(x) &= r(x, \delta^\star(x)) + \int v(y)Q(dy/x, \delta^\star(x)), \qquad x \in X, \end{aligned}$$

then there exist admissible deterministic local policies which are optimal in LS (locally optimal). ⊙

From Theorem 3.1 we obtain directly the following criterion.

Theorem 3.2. Assume there exist a constant g and a bounded Borel function $v(x)$ on X such that for all $t \in \mathbb{N}$ and admissible control sets $A^t(x) := \underset{i\in V}{\times} A_i^t(x_{\widetilde{N}(i)})$

$$g + v(x) = \sup_{a\in A^t(x)} \Big\{ r(x,a) + \int v(y) Q(dy/x,a) \Big\}, \qquad x \in X. \tag{13}$$

Then

$$\sup_{\delta\in LS} \phi(x,\delta) \le g, \qquad x \in X.$$

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

$$g + v(x) = r(x,\delta^\star(x)) + \int v(y) Q(dy/x,\delta^\star(x)), \qquad x \in X,$$

then $\phi(x,\delta^\star) \equiv g$ and $\delta^\star$ is optimal in LS. $\odot$

Remark 3.1. Corollary 3.1 applies here again. $\odot$

The following regularity assumptions on the transition kernels provide conditions for proving optimality properties of local strategies.

Assumption I. $Q = \prod_{j\in V} Q_j$ is according to Definition 2.7 a local and synchronuos Markov kernel for (ξ,δ) and the admissible control sets are independent of the time points of decisions: $A(x) := A^t(x) = \underset{i\in V}{\times} A_i^t(x_{\widetilde{N}(i)}) =: \underset{i\in V}{\times} A_i(x_{\widetilde{N}(i)})$.
There exists a non-negative measure μ on $(X,\mathfrak{X})$ with $\mu(X) > 0$ such that:
$\mu(C) \le Q(C/x,a), \qquad C \in \mathfrak{X}, \quad (x,a) \in \Delta.$ $\odot$

Assumption II. $Q = \prod_{j\in V} Q_j$ is according to Definition 2.7 a local and synchronuos Markov kernel for (ξ,δ) and the admissible control sets are independent of the time points of decisions: $A(x) := A^t(x) = \underset{i\in V}{\times} A_i^t(x_{\widetilde{N}(i)}) =: \underset{i\in V}{\times} A_i(x_{\widetilde{N}(i)})$.
There exists a non-negative measure η on $(X,\mathfrak{X})$ with $\eta(X) < 2$ such that:
$\eta(C) \ge Q(C/x,a), \qquad C \in \mathfrak{X}, \quad (x,a) \in \Delta.$ $\odot$

Let $M(X)$ be the Banach space of bounded Borel measurable functions on X with norm $||u|| = \sup_{x\in X} |u(x)|$. We define operators $\mathfrak{U}'$ and $\mathfrak{U}''$ on $u \in M(X)$ by:

$$\mathfrak{U}'u(x) = \sup_{a\in A(x)} \Big\{ r(x,a) + \int u(y) Q'(dy/x,a) \Big\},$$

$$\mathfrak{U}''u(x) = -\sup_{a\in A(x)} \Big\{ r(x,a) + \int u(y) Q''(dy/x,a) \Big\},$$

where

$$Q'(C/x,a) = Q(C/x,a) - \mu(C),$$
$$Q''(C/x,a) = \eta(C) - Q(C/x,a).$$

For every function $u \in M(X)$ we denote

$$A'_u(x) = \Big\{a\colon a \in A(x), \quad \mathfrak{U}'u(x) = r(x,a) + \int u(y)Q'(dy/x,a)\Big\},$$
$$A''_u(x) = \Big\{a\colon a \in A(x), \quad -\mathfrak{U}''u(x) = r(x,a) + \int u(y)Q''(dy/x,a)\Big\}.$$

Using these definitions we define maps

$$F'_u : X \to (2)^A_{set}, \qquad x \to F'_u(x) = A'_u(x)$$

and

$$F''_u : X \to (2)^A_{set}, \qquad x \to F''_u(x) = A''_u(x).$$

Theorem 3.3. Let Assumption I (Assumption II) hold and assume further:
1) the operator $\mathfrak{U}'$ (operator $\mathfrak{U}''$) transforms some metric subspace $S(X) \subseteq M(X)$ (with metric ρ induced by the norm of $M(X)$) into itself;
2) the map F'_u (map F'_u) is measurable (Definition 2.3) for every function $u \in S(X)$.
Then there exists a strategy in LS_S which is optimal in LS (locally optimal). $\odot$

The proof is similar to [6] (Theorems 2 and 2′)and therefore we only sketch the main ideas: The operator $\mathfrak{U}'$ (operator $\mathfrak{U}''$) is a contraction on $S(X) \subseteq M(X)$, so the equation $\mathfrak{U}'u = u$ ($\mathfrak{U}''u = u$) has a unique solution $u^\star(x)$ in $S(X)$. Explicitly this is

$$u^\star(x) = \sup_{a\in A(x)} \Big\{r(x,a) + \int u^\star(y)Q'(dy/x,a)\Big\}, \tag{14}$$

$$u^\star(x) = -\sup_{a\in A(x)} \Big\{r(x,a) + \int u^\star(y)Q''(dy/x,a)\Big\}, \quad \textit{respectively.} \tag{15}$$

(14) ((15), respectively) are the optimality equation (13) with $v = u^\star$ ($v = -u^\star$) and $g = \int u^\star(x)\,\mu(dx)$ ($g = \int u^\star(x)\,\eta(dx)$). Therefore there exists $a^\star \in A'_{u^\star}(x)$ ($a^{\star\star} \in A''_{u^\star}(x)$) such that

$$\sup_{a\in A(x)} \Big\{r(x,a) + \int u^\star(y)Q'(dy/x,a)\Big\} = r(x,a^\star) + \int u^\star(y)Q'(dy/x,a^\star)$$
$$\Big(\sup_{a\in A(x)} \Big\{r(x,a) + \int u^\star(y)Q''(dy/x,a)\Big\} = r(x,a^{\star\star}) + \int u^\star(y)Q''(dy/x,a^{\star\star})\Big).$$

From measurability by the Theorem of Choice the map F'_u (map F''_u) has a Borel measurable selector $\delta^\star\colon X \to A$ ($\delta^{\star\star}\colon X \to A$) such that

$$r(x,a^\star) + \int u^\star(y)Q'(dy/x,a^\star) = r(x,\delta^\star(x)) + \int u^\star(y)Q'(dy/x,\delta^\star(x))$$
$$\Big(r(x,a^{\star\star}) + \int u^\star(y)Q''(dy/x,a^{\star\star}) = r(x,\delta^{\star\star}(x)) + \int u^\star(y)Q''(dy/x,\delta^{\star\star}(x))\Big).$$

We assumed $\delta^\star(x) = (\delta^\star_i(x_{\widetilde{N}(i)}), i \in V) \in LS$, where $\delta^\star_i(x_{\widetilde{N}(i)}) \in A_i(x_{\widetilde{N}(i)})$. Therefore, because the values of the selector are deterministic, we have proved $\delta^\star(x) \in LS_S$. Similarly, $\delta^{\star\star}(x) \in LS_S$.

The next lemmata can be found in [6].

Lemma 3.1. If A is a compact metric space, $F\colon X \to (2)^A_{set}$ is lower semicontinuous, and the function $u(x,a)$ is bounded and continuous on Δ, then the function $u^\star(x) := \sup_{a\in A(x)} u(x,a)$ is lower semicontinuous on X. $\odot$

Lemma 3.2. If A is a compact metric space, $F\colon X \to (2)^A_{set}$ is upper semicontinuous, and the function $u(x,a)$ is bounded and upper semicontinuous on Δ, then the function $u^\star(x) := \sup_{a\in A(x)} u(x,a)$ is upper semicontinuous. $\odot$

Theorem 3.4. Let A be a compact metric space, assume that $F\colon X \to (2)^A_{set}$ is continuous, and Assumption I or II holds. Assume further that

1) the function $r(x,a), (x,a) \in \Delta$ is continuous;

2) the transition probability $Q(\cdot/x,a), (x,a) \in \Delta$ is weakly continuous.

Then an optimal strategy $\delta^\star$ exists in LS_S. The function $\delta^\star\colon X \to A$ can be selected as being in Baire class 1. $\odot$

Proof. Our proof of the theorem borrows arguments from Theorem 3 in [6]: Conditions 1, 2 and Lemmas 3.1, 3.2 imply that the operator $\mathfrak{U}'$ (operator $\mathfrak{U}''$) transforms the Banach space $C(X) \subset M(X)$ of bounded continuous functions on X into itself. We have to show that the map F'_u is upper semicontinuous for every function $u \in C(X)$ (semicontinuity of F''_u will be proved similarly). For this purpose it suffices to check according to [9, p. 61] that for every function $u \in C(X)$ the following holds: if $(x^n, a^n) \to (x,a)$ as $n \to \infty$ and $a^n \in A'_u(x^n)$ ($n = 1, 2, \dots$) then $a \in A'_u(x)$. (That the sets $A'_u(x)$ ($A''_u(x)$) are nonempty and closed is obvious). From upper semicontinuity of the map F it follows that $a \in A(x)$. From $a^n \in A'_u(x^n)$ we have

$$r(x^n,a^n) + \int u(y)Q'(dy/x^n,a^n) = \max_{a'\in A(x^n)} \Big\{ r(x^n,a') + \int u(y)Q'(dy/x^n,a') \Big\}.$$

Both, the left and the right hand side of this equality are continuous functions. Passing to the limit for $n \to \infty$ we obtain

$$r(x,a) + \int u(y)Q'(dy/x,a) = \max_{a'\in A(x)} \Big\{ r(x,a') + \int u(y)Q'(dy/x,a') \Big\},$$

i.e. $a \in A'_u(x)$.

From Theorem 3.3 an optimal strategy exists and according to the Theorem of Choice for semicontinuous maps a function $\delta^\star\colon X \to A$ can be selected from Baire class 1. We assumed $\delta^\star(x) = (\delta^\star_i(x_{\widetilde{N}(i)}), i \in V) \in LS$, where $\delta^\star_i(x_{\widetilde{N}(i)}) \in A_i(x_{\widetilde{N}(i)})$, and the measurable selector is a deterministic function. Therefore $\delta^\star(x) \in LS_S$. $\square$

4 Separable reward functions

In this section we consider the control problem for Markov processes with locally interacting synchronous components over $\Gamma = (V,B)$) when the reward function $r(x,a)$ is separable in the sense of the following definition:

Definition 4.1. A reward function $r : \Delta = \{(x,a) \colon x \in X,\ a \in A(x)\} \subset X \times A \to I\!R$ is separable (with respect to $\Gamma = (V,B)$,) if

$$r(x,a) = \sum_{i \in V} r_i(x_{\widetilde{N}(i)}, a_i), \tag{16}$$

i.e., if it is a sum of locally defined measurable bounded reward functions $r_i(x_{\widetilde{N}(i)}, a_i)$. $\odot$

In this section we assume that the control sets (see Definition 2.4,(2)) are not necessarily local:
$A^t(x) = A(x) = \underset{i \in V}{\times} A_i(x)$ for all t, i.e. $A_i(x)$ may depend on the complete state x and is independent of time. We recall the accordingly defined nonlocal subclasses of general policies.

Definition 4.2. (1) For any vertex $i \in V$ and any system configuration (state) $\xi^t = x$ let $A_i(x)$ be the time invariant set of admissible actions (admissible decision values) for π_i^t at time t.

(2) A strategy δ is called admissible if $\pi_i^t(\cdot / x^0, a^0, \ldots, x^{t-1}, a^{t-1}, x^t)$ is a probability measure on $\big(A_i(x^t), \mathfrak{A}_i(x^t)\big)$ and measurably dependent on the history $h_i^t = (x^0, a^0, \ldots, x^{t-1}, a^{t-1}, x^t)$ for every $t \in I\!N$.

The class of all admissible strategies is denoted by S; the subclass of admissible Markov strategies by S_M.

(3) By S_S we denote the class of admissible stationary Markov deterministic strategies with time invariant restriction sets: $\pi^t = \pi^{t'}$ for all $t, t' \in I\!N$ on $A(x) = \underset{i \in V}{\times} A_i(x)$. $\odot$

We first state an analogue to Theorem 3.1, the proof of which is similar to the one presented there.

Theorem 4.1. Let there exist a constant g and a bounded Borel function $v(x)$ on X such that

$$g + v(x) = \sup_{a \in A(x)} \Big\{ r(x,a) + \int_X v(y) Q(dy/x,a) \Big\}, \qquad x \in X. \tag{17}$$

Then

$$\sup_{\delta \in S} \phi(x,\delta) \le g, \qquad x \in X.$$

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

$$g + v(x) = r(x, \delta^\star(x)) + \int_X v(y) Q(dy/x, \delta^\star(x)), \qquad x \in X,$$

then $\phi(x, \delta^\star) \equiv g$, and $\delta^\star$ is optimal. $\odot$

Corollary 4.1. If (17) holds and a strategy $\delta^\star \in LS_S \subseteq LS$ fulfills

$$g + v(x) = r(x, \delta^\star(x)) + \int v(y) Q(dy/x, \delta^\star(x)), \qquad x \in X,$$

then there exist admissible deterministic local policies which are optimal in S. $\odot$

In the present context Assumption I takes the following form:

Assumption I′. $Q = \prod_{j\in V} Q_j$ is according to Definition 2.7 a local and synchronuos Markov kernel for (ξ, δ) and the admissible control sets are independent of the time points of decisions: $A(x) := A^t(x) = \underset{i\in V}{\times} A_i(x)$.
There exist non-negative measures μ_i on $(X_i, \mathfrak{X}_i)$ with $\mu_i(X_i) > 0, \quad \forall\, i \in V$, such that:
$\mu_i(B_i) \le Q(B_i/x_{\widetilde{N}(i)}, a_i)\,, \quad B_i \in \mathfrak{X}_i\,, \quad (x_{\widetilde{N}(i)}, a_i) \in \Delta_i\,, \quad i \in V.$ $\odot$

Let $S(X) \subseteq M(X)$ denote the metric space of real valued separable functions $v(x) = \sum_{i\in V} v_i(x_i)$ according to Definition 4.1 with norm $||v|| = \sum_{i\in V} \sup_{x_i\in X_i} |v_i(x_i)|$. Define the operator $\widetilde{\mathfrak{U}}'$ on $S(X)$:

$$\widetilde{\mathfrak{U}}'v(x) = \sup_{a\in A}\Big\{\sum_{i\in V} r_i(x_{\widetilde{N}(i)}, a_i) + \int_X v(y)Q'(dy/x, a)\Big\},$$

where

$$Q'(B/x, a) = Q(B/x, a) - \prod_{i\in V} \mu_i(B_i).$$

Let

$$\widetilde{A}'_v(x) = \Big\{a\colon a \in A, \quad \widetilde{\mathfrak{U}}'v(x) = \sum_{i\in V} r_i(x_{\widetilde{N}(i)}, a_i) + \int_X v(y)Q'(dy/x, a)\Big\}$$

and for measurable $v_i : X_i \to I\!R$

$$U_i^{v_i}(x_{\widetilde{N}(i)}, a_i) = r_i(x_{\widetilde{N}(i)}, a_i) + \int_{X_i} v_i(y_i)\Big[Q_i(dy_i/x_{\widetilde{N}(i)}, a_i) - \mu_i(dy_i) \prod_{j\colon j\in V,\ j\ne i} \mu_j(X_j)\Big].$$

Theorem 4.2. Let the decision set A be a compact metric space with countable basis and assume that $F\colon X \to (2)^A_{set}$ is continuos. Assume that Assumption I′ holds and additionally
1) the functions $r_i(x_{\widetilde{N}(i)}, a_i)$ are continuous on Δ_i.
2) the transition probabilities $Q(B_i/x_{\widetilde{N}(i)}, a_i)$ are weakly continuous on Δ_i.
Then an optimal strategy in class LS_S exists. $\odot$

Proof. The operator $\widetilde{\mathfrak{U}}'$ is monoton increasing:
$v^1(x) \le v^2(x)\,,\ x \in X \Longrightarrow \widetilde{\mathfrak{U}}'v^1(x) \le \widetilde{\mathfrak{U}}'v^2(x), \qquad x \in X;$
and from

$$\begin{aligned}\widetilde{\mathfrak{U}}'(v(x) + c) &= \sup_{a\in A(x)} \Big\{\sum_{i\in V} r_i(x_{\widetilde{N}(i)}, a_i) + \int_X v(y)Q'(dy/x, a) + c\int_X Q'(dy/x, a)\Big\}\\ &= \widetilde{\mathfrak{U}}'v(x) + \alpha c.\end{aligned}$$

$\widetilde{\mathfrak{U}}'$ fulfills: for any nonnegative constant c is

$$\widetilde{\mathfrak{U}}'(v(x) + c) = \widetilde{\mathfrak{U}}'v(x) + \alpha c, \qquad x \in X, \quad \alpha = 1 - \prod_{i\in V} \mu_i(X_i) < 1.$$

As $\alpha < 1$, these properties guarantee that $\widetilde{\mathfrak{U}}'$ is a contraction. Indeed, applying the operator $\widetilde{\mathfrak{U}}'$ on both sides of the inequality

$$v^1(x) \leq v^2(x) + \rho(v^1, v^2)$$

we have

$$\widetilde{\mathfrak{U}}'v^1(x) \leq \widetilde{\mathfrak{U}}'(v^2(x) + \rho(v^1, v^2)) = \widetilde{\mathfrak{U}}'v^2(x) + \alpha\rho(v^1, v^2).$$

Therefore

$$\widetilde{\mathfrak{U}}'v^1(x) - \widetilde{\mathfrak{U}}'v^2(x) \leq \alpha\rho(v^1, v^2).$$

Interchanging v^1 and v^2, we have

$$|\widetilde{\mathfrak{U}}'v^1(x) - \widetilde{\mathfrak{U}}'v^2(x)| \leq \alpha\rho(v^1, v^2), \qquad x \in X,$$

i.e.

$$\rho(\widetilde{\mathfrak{U}}'v^1, \widetilde{\mathfrak{U}}'v^2) \leq \alpha\rho(v^1, v^2),$$

and $\widetilde{\mathfrak{U}}'$ is a contraction. Banach's fixed point theorem implies that in $S(X)$ exists a function $v^\star(x)$ such that

$$v^\star(x) = \sup_{a \in A(x)} \Big\{ \sum_{i \in V} r_i(x_{\widetilde{N}(i)}, a_i) + \int_X v^\star(y) Q'(dy/x, a) \Big\}, \qquad x \in X.$$

The definition of $\widetilde{\mathfrak{U}}'$ and the separability of $r(.,.)$ yield

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

From condition 1) and 2) of the Theorem it follows that there exists measurable functions $a_i^\star = a_i^\star(x_{\widetilde{N}(i)})$ such that

$$\sup_{a_i \in A_i(x)} U_i^{v_i^\star}(x_{\widetilde{N}(i)}, a_i) = U_i^{v_i^\star}(x_{\widetilde{N}(i)}, a_i^\star(x_{\widetilde{N}(i)})).$$

Therefore there exists a function $\delta^\star = \{\delta_i^\star(x_{\widetilde{N}(i)})\} \in LS_S$ such that

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

and

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

or

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

where

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

Therefore the conditions of the Theorem 4.1 are satisfied and strategy $\delta^\star \in LS_S$ is optimal. □

References

[1] Balder EJ (1989) On compactness of the space of policies in stochastic dynamic programming. Stoch. Processes and Their Appl. 32, 141-150

[2] Chornei RK, Daduna H, Knopov PS (2000) Controlled Markov fields with finite state space on graphs. Preprint 00-08, Inst. Math. Stochastics, University of Hamburg

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

[4] 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

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

[6] 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)

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

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

[9] Kuratowski K (1969) Topology 2. Academic Press, New York

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

[11] Schael M (1975) On dynamic programming: Compactness of the space of policies. Stoch. Processes and Their Appl. 3, 345-364

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

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