UNIVERSITÄT HAMBURG

**Controlled semi-Markov fields with graph-structured compact state space**

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

**Preprint–No. 2002-08 Dezember 2002**

FACHBEREICH MATHEMATIK

SCHWERPUNKT MATHEMATISCHE STATISTIK
UND STOCHASTISCHE PROZESSE

BUNDESSTR. 55, D – 20146 HAMBURG

# Controlled semi-Markov fields with graph-structured compact state space

Ruslan K. Chornei Hans Daduna Pavel S. Knopov *

December 16, 2002

**Summary:** We introduce the concept of locally acting distributed decision makers into semi-Markov decision process theory for systems with general compact state and action spaces. Such decision makers are characterized by making the decisions on the basis of information gathered in their local neighborhood only. The state transition jump function for the system is of local structure as well. The holding times of the systems are general and this results in semi-Markov properties in time, and the neighborhood structure of the system resembles in space the Markov property of spatial processes. We prove that under some technical regularity assumptions we can reduce optimality problems within the set of local strategies to only considering deterministic Markov strategies.

**MSC (2000) subject classification:** 60K15, 60K35, 90C40

**Key Words:** Semi-Markov processes, Markov renewal processes, optimal control, average asymtotic reward, renewal reward processes,, random fields, local strategies.

## 1 Introduction

Optimal sequential control of stochastic systems is in most cases described by Markov process models. But there is a rich literature on the control of processes which do not show the memoryless property of Markovian systems. Introducing nonexponential, resp. nongeometrical sojourn times of the processes leads to considering semi-Markov processes and Markov renewal processes. Early papers on stochastic dynamic optimization for semi-Markov processes (sometimes called renewal reward processes) are [8], [9], [7]. For further references to early works see [6]. More recent work on semi-Markov decision processes can be found e.g. in [10], [17], [18], [13] and the references given therein. The general construction of semi-regenerative control processes and semi-Markov control processes with application to queueing systems can be found in the recent book of Kitaev and Rykov [11].

Introductory definitions of semi-Markov processes and Markov rewal processes as we shall consider in this paper and fundamental investigations on the subject may be found in [1], [2], and with a special emphasis on optimization in [11].

In this note we consider stochastic processes in continuous time with general compact state spaces which are structured by an underlying graph. At any node of the graph there

*Part of this research was done when the third author held a grant of the DFG at Hamburg University.

is a local state space, such that the global state space of the process is the product (indexed by the graph) of the local state spaces. The graph then defines a neighborhood structure for the states of the described systems. These neighborhoods determine the local interactions of the spatial process coordinates. Then for a fixed time instant the random state of the system, respectively of the describing stochastic process, is a random field with respect to the neighborhood graph.

We assume that the process has the semi-Markov property in time and that the transition kernels of the process have the spatial Markov property with respect to the underlying graph. Additionally we construct control structures to optimize the system's development over time. Optimality here is meant with respect to the asymptotic time averaged reward criterion.

Our aim is to find optimal strategies to control the system when the decision makers located at the nodes of the network can use in decision making only information gathered in their respective neighborhoods. We find conditions under which in the class of such local strategies there exist optimal stationary deterministic policies.
With respect to this our results extend those of [5] in the direction of interacting semi-Markov processes.

The necessary definitions on stochastic processes which develop in space and time are recalled in Section 2.1, in Section 2.2 we introduce local structures for the policies and the transition mechanism of the processes. We assume time to be continuous but allow decisions only at jump instants of the process. The random policies show as usual a conditional independence structure. We assume that a similar structure can be found in the transition kernels of the jump chains (synchronized kernels). For the Markov case we then are in the setting of [3]. In Section 3 we first provide abstract conditions for the existence of stationary deterministic strategies in the class of local structures and then show that under some weak smoothness assumptions these conditions are fulfilled.

# 2 Description of the system

## 2.1 Preliminaries

For systems 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 $B$. 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 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. For every subset of nodes $K \subset V$ we denote the marginal vector of state $x$ by $x_K = (x_k,\ k \in K) \in X_K = \underset{i \in K}{\times} X_i$.

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 $\underset{i\in K}{\times}\mathfrak{X}_i$, and $K\subseteq V$, $\mathfrak{X}_V=\mathfrak{X}$.

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

**Definition 2.1 (Random fields). (1)** A random variable $\xi$ 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 $\xi_K$. For $K=\{k\}$ we write $\xi_k$.

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

$$P(\xi_k\in C_k|\xi_{V-\{k\}}=x_{V-\{k\}})=P(\xi_k\in C_k|\xi_{N(k)}=x_{N(k)}),\qquad \forall\, x\in X,\quad C_k\in\mathfrak{X}_k,\qquad (1)$$

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

**Remark 2.1.** Equation (1) is to be read as follows:

$$P(\xi_k\in C_k|\xi_{V-\{k\}}=x_{V-\{k\}})=P(\xi_k\in C_k|\xi_{V-\{k\}}=y_{V-\{k\}})\quad \text{if } x_{N(k)}=y_{N(k)} \text{ holds.}$$

Similar readings will apply throughout the paper. ⊙

In many problems related to biological, 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 $\eta$ with random fields as one-dimensional marginals in time. In this case we write $\eta=(\eta^t)$ to emphasize the time-dependence. We assume that time is continouos, $t\in\mathbb{R}_+=[0,\infty)$. The subscript $k$ in $\eta_k^t$ refers to the vertex $k$, $\eta_k^t$ therefore denotes the marginal distribution for time $t$ and node $k$ of some vector valued process $\eta=(\eta^t\colon t\geq 0)$. Such processes which vary in space and time, with the space variable being stuctured by some graph,were investigated in [3], using Markov chain techniques. It is the aim here to remove the assumption that in time the holding times of the processes are memoryless, i.e., we investigate semi-Markov processes and Markov renewal processes.

**Definition 2.2 (Markov renewal process and Semi-Markov process).** Consider a stochastic jump process $\eta=(\eta^t:t\geq 0)$ with state space $X$ and paths which are right continuous and have lefthand limits (cadlag paths) defined by a sequence $(\xi,\tau)=\{(\xi^n,\tau^n),n=0,1,\ldots,\}$. The $\mathbb{R}_+$-valued random variables $\tau^n$ are the interjump times and the sequence $\{\xi^n,n=0,1,\ldots\}$ is the sequence of states of the process entered just after the jump instants. To be more precise:
Let $\sigma=\{\sigma^n:n=0,1,\ldots\}$ be given as $\sigma^0=0,$ and $\sigma^n=\sum_{i=0}^{n-1}\tau^i,n>0$, the increasing sequence of jump times. Then if $t\in[\sigma^n,\sigma^{n+1})$ we have $\eta^t=\xi^n$, $n\in\mathbb{N}$.
If the conditional distributions of $(\xi,\tau)=\{(\xi^n,\tau^n),n=0,1,\ldots\}$ fullfill

$$\begin{aligned}&P(\xi^n\in C,\tau^{n-1}\leq s|\xi^k=x^k,k=0,1,\ldots,n-1,\sigma^0=0,\tau^k\leq s^k,k=0,1,\ldots,n-2)\\ =\;&P(\xi^n\in C,\tau^{n-1}\leq s|\xi^{n-1}=x^{n-1}),\qquad n>0,\end{aligned}$$

and are independent of $n$, then $(\xi,\tau)$ is called a homogeneous Markov renewal process. The process $\eta=(\eta^t:0\leq t<\sum_{i=0}^{\infty}\tau^i)$ is called homogeneous semi-Markov process. The transition function

$$\widetilde{Q}(C,s|x)=P(\xi^{n+1}\in C,\tau^n\leq s|\xi^n=x)$$

is called the semi-Markov kernel of both, the Markov renewal process and the semi-Markov process. We shall henceforth assume that

$$\widetilde{Q}(C,0|x) = 0 \ \ \forall x, C, \quad \text{and} \quad \sum_{i=0}^{\infty} \tau^i = \infty.$$

holds and that the sequence $\sigma = \{\sigma^n : n = 0, 1, \dots\}$ has no finite accumulation points. The transition probability of the embedded Markov chain $\xi = (\xi^n = \eta(\sigma^n) : n \in \mathbb{N})$ is

$$Q(C|x) = \widetilde{Q}(C,\infty|x) = P(\xi^{n+1} \in C, \tau^n \leq \infty | \xi^n = x) \ \ \forall x \in X, C \in \mathfrak{X},$$

and by

$$\mathrm{T}(\cdot|x,y) = P(\tau^n \leq \cdot | \xi^n = x, \xi^{n+1} = y), \quad x, y \in X$$

the conditional sojourn time distribution of the sojourn time $\tau^n$ given the process is in state $x$ and will be next in $y$ when $\tau^n$ has expired. $\odot$

Before we shall present the details of the control structures for the semi-Markov processes we recall some Definitions and Theorems from [12].

**Definition 2.3.** For topological spaces $X$ and $A$ with Borel $\sigma$–algebras $\mathfrak{X}$ and $\mathfrak{A}$ 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 $\{x\colon F(x) \cap E \neq \varnothing\} \in \mathfrak{X}$, 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 $x \in X$. $\odot$

**Definition 2.4.** In the framework of Definition 2.3 a closed (open) measurable mapping $F$ is called upper (lower) semicontinuous if for any closed (open) set $E \subseteq A$ the set $\{x \in X : F(x) \cap E \neq \emptyset\}$ is closed (open).
$F$ is continuous if it is both upper and lower semicontinuous.

**Theorem of Choice.** [12, p. 74]. In the setting of Definition 2.3 any Borel measurable map $F\colon X \to (2)^A_{set}$ has a Borel measurable selector $f\colon (X,\mathfrak{X}) \to (A,\mathfrak{A})$. $\odot$

**Theorem of Choice for Semicontinuous Maps.** [12, 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. $\odot$

## 2.2 Semi-Markov processes with local and synchronous transition mechanisms

The aim of this section is to introduce a controlled semi-Markov process with locally interacting synchronous components, shortly: a controlled semi-Markov random field, similar to that constructed in the pure Markovian framework in [3] as controlled Markov random field.

If a semi-Markov process $\eta$ and its associated Markov renewal process $(\xi,\tau)$ from their very construction are 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 $\xi_k^n=\eta_k(\sigma^n)=x_k$ of the $k$-th vertex depends on the previous states of the whole system only through the values of the vertices in $\widetilde{N}(k)$ (including $k$) after transition moment $\sigma^{n-1}$. To describe this we introduce the Markov property in space for the kernel of the embedded jump chain for a semi-Markov process along the lines of Definition 2.1.
Because we concentrate on the local behavior of the embedded jump chain we arrive at the definition of synchronous and local kernels as described in [3][Section 2]. Note, that the sojourn time distributions of the semi-Markov Process remain globally determined.

**Definition 2.5 (Synchronized and local transition kernels).** Let $\eta=(\eta^t:t\geq 0)$ with state space $X$ be a semi-Markov process with associated Markov renewal process $(\xi,\tau)=\{(\xi^n,\tau^n),n=0,1,\ldots\}$. The jump transition probabilities of $\eta$, resp. $\xi$ are said to be *local* [16, p. 100] if for all $k\in V$, $x^0,\ldots,x^{n+1}\in X$, $C_k\in\mathfrak{X}_k$

$$P\{\xi_k^{n+1}\in C_k|\xi^n=x^n,\ldots,\xi^0=x^0\}=P\{\xi_k^{n+1}\in C_k|\xi^n_{\widetilde{N}(k)}=x^n_{\widetilde{N}(k)}\}\tag{2}$$

holds, i.e. the state of vertex $k$ just after the jump depends on the state of its complete neighborhood at the previous transition instant only. The transition probabilities of $X$ are said to be *synchronous* [16, p. 100] if

$$P\{\xi_K^{n+1}\in C_K|\xi^n=x^n\}=\prod_{k\in K}P\{\xi_k^{n+1}\in C_k|\xi^n=x^n\},\qquad\forall K\subset V,\quad x^n\in X,\tag{3}$$

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

If $\eta$, resp. $\xi$ or $(\xi,\tau)$, fulfill (2) and (3) then $\eta$ is called a *semi-Markov process with locally interacting synchronous components over* $(\Gamma,X)$, shortly, a *semi-Markov random field.* Similarly we call $(\xi,\tau)$ a *Markov renewal process with locally interacting synchronous components over* $(\Gamma,X)$, shortly, a *Markov renewal field* $\odot$

We introduce classes of admissible policies to control the interacting coordinates of the stochastic processes under consideration. For simplicity of the presentation we assume that some (uncontrolled) semi-Markov random field $\eta=(\eta^t:t\geq 0)$, resp. $(\xi,\tau)=\{(\xi^n,\tau^n)\colon n\in\mathbb{N}\}$ according to Definition 2.2 is given as our starting point. $\eta$ will be equipped in the sequel with a control structure which governs the jump transition kernel and the sojourn time distributions of $(\xi,\tau)$. We mainly consider controls, which are local in the sence to be specified now.

**Definition 2.6 (Action spaces and local restrictions).** The sequence of decision instants (control instants) is $\tau=\{\tau^n,n=0,1,\ldots\}$.

**(1)** The set of actions (control values) usable at control instants is $A=\mathop{\times}\limits_{i\in V}A_i$ over $\Gamma$, 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 and with Borel $\sigma$ - algebra $\mathcal{A}_i$. $\mathcal{A}$ is the Borel $\sigma$ -algebra over $A$.

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

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

$$F_i^n \colon X \to (2)_{set}^{A_i}, \quad x \to F_i^n(x) = A_i^n(x), \qquad i \in V,$$

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

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

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 2.7 (Strategies). (1)** Let $\alpha_i^n$ denote the action chosen by the decision maker at node $i$ at the $n$-th transition instant $\sigma^n$, $\alpha^n := (\alpha_i^n : i \in V)$ the joint decision vector at time $\sigma^n, n \in \mathbb{N}$.

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

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

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

**(3)** In parallel to the synchronuos transitions and the locality of the transition kernels we always assume that the decision makers located at the nodes act conditionally independent given the history of the system. This leads to control of the process governed by a synchronuos control kernel

$$\begin{aligned} & P(\alpha^n \in \underset{i \in V}{\times} B_I | \xi^0 = x^0, \alpha^0 = a^0, \tau^0 = t^0, \ldots \\ & \qquad \ldots, \xi^{n-1} = x^{n-1}, \alpha^{n-1} = a^{n-1}, \tau^{n-1} = t^{n-1}, \xi^n = x^n) \\ = & \prod_{j \in V} P(\alpha_i^n \in B_i | \xi^0 = x^0, \alpha^0 = a^0, \tau^0 = t^0, \ldots \\ & \qquad \ldots, \xi^{n-1} = x^{n-1}, \alpha^{n-1} = a^{n-1}, \tau^{n-1} = t^{n-1}, \xi^n = x^n) \\ = & \prod_{j \in V} \pi_i^n(B_i | x^0, a^0, t^0, \ldots, x^{n-1}, a^{n-1}, t^{n-1}, x^n), \qquad B_i \in \mathfrak{A}_i, a^s \in A, x^s \in X, t^s \in \mathbb{R}_+. \end{aligned}$$

$\odot$

**Definition 2.8 (Local strategies). (1)** If at transition times $\sigma^n, n = 0, 1, \ldots,$ the decision $\alpha_i^n$ at node $i$ is made according to the probability $\pi_i^n$ on basis of the local history $h_i^n = (x_{\widetilde{N}(i)}^0, a_i^0, t^0 \ldots, x_{\widetilde{N}(i)}^{n-1}, a_i^{n-1}, t^{n-1}, x_{\widetilde{N}(i)}^n)$, of the neighborhood $\widetilde{N}(i)$ of $i$ only, and if $\pi_i^n(F_i^n(x) | h_i^n) = \pi_i^n(A_i^n(x_{\widetilde{N}(i)}^n) | h_i^n) = 1$, $x_{\widetilde{N}(i)}^s \in X_{\widetilde{N}(i)}$, $a_i^s \in A_i^s(x_{\widetilde{N}(i)}^s)$, $t^s \in \mathbb{R}^+$, then

$\pi_i^n$ is said to be locally admissible, and the sequence of transition probabilities (decisions) $\delta_i = \{\pi_i^n, n \in \mathbb{N}\}$ 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

$$\pi_i^n(\cdot|x^0_{\widetilde{N}(i)}, a_i^0, t^0 \ldots, x^{n-1}_{\widetilde{N}(i)}, a_i^{n-1}, t^{n-1}, x^n_{\widetilde{N}(i)}) = \pi_i^n(\cdot|x^n_{\widetilde{N}(i)}), \qquad i \in V.$$

**(3)** An admissible local Markov strategy $\delta = (\delta_i, i \in V)$ is called admissible local stationary (Markov) strategy if $\pi_i^{n'}(\cdot|x_{\widetilde{N}(i)}) = \pi_i^{n''}(\cdot|x_{\widetilde{N}(i)})$, $i \in V$, for all $n'$, $n''$ and all $x$.

**(4)** An admissible local stationary strategy $\delta = (\delta_i, i \in V)$ is called admissible local stationary deterministic (nonrandomized) strategy if $\pi_i(\cdot|x_{\widetilde{N}(i)})$, $i \in V$, are one-point measures on $A_i^n(x_{\widetilde{N}(i)})$, $i \in V$, accordingly for all $x \in X$.

**(5)** We always assume that for local strategies the restriction sets are time invariant: $A^n(x) = A(x) = \times_{i \in V} A_i(x_{\widetilde{N}(i)})$ for all $n$. The class of all admissible local strategies (with time invariant restriction sets) is denoted by $LS$; the subclass of admissible local Markov strategies by $LS_M$.
By $LS_S$ we denote the class of admissible local stationary strategies with time invariant restriction sets: $\pi^n = \pi^{n'}$ for all $n, n' \in \mathbb{N}$.
By $LS_D$ we denote the class of admissible local stationary deterministic strategies with time invariant restriction sets. $\odot$

We now incorporate into our semi-Markov process framework synchronization and locality of the transition kernel (similar to (2) and (3)) and the decision rules, and then the decision dependent transition mechanism and the decision dependent sojourn time behavior of the system.

The law of motion of the system is characterized by a set of time invariant transition probabilities. Whenever the state of the system is $\xi^n = x^n$ and decision $\alpha^n = a^n$ is made, the transition probability is $P\{\xi^{n+1} \in C|\xi^n = x^n, \alpha^n = a^n\} =: Q(C|x^n, a^n)$, which is assumed to be independent of the past given the present generalized state, which includes the present actions made. Then, given $\{\xi^n = x^n, \alpha^n = a^n\}$ the next state $\xi^{n+1} = x^{n+1}$ of the system is sampled according to $Q(\cdot|x^n, a^n)$, and thereafter the sojourn time in $x^n$, given $x^n, a^n, x^{n+1}$, is sampled according to some distribution function $\mathrm{T}(\cdot|x^n, a^n, x^{n+1})$, which is Borel measurable on $\Delta \times X$.
We further assume, that this transition probability and the sojourn time distributions are independent of $t$, the motion is homogeneous in time.
Applying a control policy $\delta$ to a semi-Markov process with interacting components $\eta$ (semi-Markov random field), respectively to the associated Markov renewal process $(\xi, \tau)$, as defined in Definition 2.5 we shall call the pair $(\eta, \delta)$, respectively the triple $(\xi, \tau, \delta)$ a controlled version of $\eta$, resp. $(\xi, \tau)$, using strategy $\delta$.

It should be noted that for such controlled process in general even the embedded jump chain is not Markovian because the sequence $(\alpha^n)$ of decisions according to $\pi_i^n$ depends not only on states $x^n_{\widetilde{N}(i)}$, $i \in V$, but on the previous (local) states $x^0_{\widetilde{N}(i)}, \ldots, x^{n-1}_{\widetilde{N}(i)}$ as well. An immediate consequence of the definition will be, that if a Markov strategy $\delta$ is applied as control strategy for the semi-Markov random field then we obtain an embedded Markov jump chain $\xi$ from $(\eta, \delta)$, respectively $(\xi, \tau, \delta)$.

**Definition 2.9.** A pair $(\eta, \delta)$, respectively a triple $(\xi, \tau, \delta)$, is called a controlled stochastic jump process with locally interacting synchronous components with respect to the finite interaction graph $\Gamma = (V, B)$, if the following holds:
$\xi = (\xi^n \colon n \in \mathbb{N})$ is a stochastic process with state space $X = \underset{i \in V}{\times} X_i$, $\tau = (\tau^n : n \in \mathbb{N})$ is stochastic process with state space $\mathbb{R}_+$, $\eta = (\eta^t : t \geq 0)$ is a stochastic process with state space $X = \underset{i \in V}{\times} X_i$, and these processes are connected pathwise as follows:
The $\mathbb{R}_+$-valued random variables $\tau^n$ are the inter jump times and the sequence $\{\xi^n, n = 0, 1, \dots\}$, is the sequence of states of the process entered just after the jump instants.
$\sigma = \{\sigma^n : n = 0, 1, \dots\}$ with $\sigma^0 = 0$, and $\sigma^n = \sum_{i=0}^{n-1} \tau^i, n > 0$ is the increasing sequence of jump times. Then, if $t \in [\sigma^n, \sigma^{n+1})$, we have $\eta^t = \xi^n$, $n \in \mathbb{N}$.
We assume that the conditional distribution function of $\tau^n$ given $\{\xi^{n+1} = x^{n+1}, \xi^n = x^n, \alpha^n = a^n\}$ is $\mathrm{T}(\cdot | x^n, a^n, x^{n+1})$ defined as Borel measurable function on $\Delta \times X$. $\delta = (\delta_i \colon i \in V)$ is an admissible local strategy, and the transitions of $\xi$ are determined as follows:

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

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 synchronous.
This construction results in a semi-Markov kernel

$$\widetilde{Q}(C, s | x, a) = P\{\xi^{n+1} \in C, \tau^n \leq s | \xi^n = x^n, \alpha^n = a^n\} =: Q(C | x^n, a^n).$$

$\eta$, resp. $(\xi, \tau)$ will then be called shortly a controlled time dependent semi-Markov random field, resp. a controlled time dependent Markov renewal 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 $\overset{(1)}{=}$ introduces memoryless (Markov) properties with respect to the decision process given the state process history. $\overset{(2)}{=}$ is the similar assumption of a Markovian transition law. $\overset{(3)}{=}$ expresses that the coordinates act synchronously, and $\overset{(4)}{=}$ is due to the locality of the state transition law. Independence of time for the one–step transition probabilities then leads to the form $\overset{(5)}{=}$ of the Markov kernel, and $\overset{(6)}{=}$ again results from the coordinates acting synchronously.

## 2.3 Criteria for optimality

Consider a time dependent semi-Markov random field $\eta$. If the $k$-th state of the system is $x^k$ and the joint decision $a^k$ is made and if the duration of the sojourn time in $x^k$ is $t^k$ then a random reward $r(t^k, x^k, a^k)$ is earned. The function $r(s, x, a)$ is assumed to be Borel measurable on $[0, +\infty) \times \Delta$.

We shall evaluate the quality, resp. optimality of the applied strategies with respect to the following long time average reward measure:

$$\phi(x, \delta) = \liminf_{n \to \infty} \frac{E_x^\delta \sum_{k=0}^{n} r(\tau^k, \xi^k, \alpha^k)}{E_x^\delta \sum_{k=0}^{n} \tau^k}, \tag{5}$$

where $E_x^\delta$ is expectation associated with the controlled process $(\eta, \delta)$, resp. $(\xi, \tau, \delta)$, if $\xi^0 = x$. The aim of our investigation is to find conditions for the existence of optimal Markovian policies.

**Definition 2.10.** 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. ⊙

The quality criterion (5) was invented by Ross [15] and used in [4] as well. In [15] were only stationary policies allowed which make decisions on the basis of the present state only. In this case it can be argued that the value of (5) depends only on on the transition kernels $Q(\cdot|x, a)$, $\pi(\cdot|x)$ and the conditional expectations

$$\tau(x, a) = \int_X \int_0^\infty t \, d\mathrm{T}(t|x, a, y) Q(dy|x, a),$$

$$r(x, a) = \int_X \int_0^\infty r(t|x, a) \, d\mathrm{T}(t|x, a, y) Q(dy|x, a).$$

Then one is in a position to restrict proofs to special forms of sojourn time distributions, say one point distributions:

$$\mathrm{T}(t|x, a, y) = \begin{cases} 1, & t \geq \tau(x, a), \\ 0, & t < \tau(x, a); \end{cases}$$

$$r(t, x, a) = \begin{cases} 0, & t < \tau(x, a), \\ r(x, a), & t \geq \tau(x, a), \end{cases}$$

Because we allow general policies such a property does not hold for the asymptotic average reward in general.

# 3 Existence of optimal strategies in $LS$

We shall need in the following some technical assumptions, which we collect here:

**Assumption I.** $\tau(x,a) \geq m > 0, \qquad (x,a) \in \Delta\,,$

**Assumption II.** $\tau(x,a) \leq M < \infty, \qquad (x,a) \in \Delta\,;$

**Assumption III.** there exists a non-negative measure $\mu$ on $(X, \mathfrak{X})$ such that:

a) $\mu(C) \leq Q(C|x,a), \qquad (x,a) \in \Delta, \qquad C \in \mathfrak{X}\,;$

b) $\mu(X) > 0\,.$

Let us denote by $\kappa^n$ a random history of the system up to time $\sigma^n$. So $\kappa^n$ takes values of the form $h^n = (x^0, a^0, t^0, \ldots, x^{n-1}, a^{n-1}, t^{n-1}, x^n)$. Recall that the restriction sets for the local strategies are time independent.
We define the conditional expectations

$$\begin{aligned} \tau(x^k, a^k) &= E_x^\delta\{\tau^k | \kappa^k = h^k, \alpha^k = a^k\} \\ &= \int_X \int_0^\infty t \, d\mathrm{T}(t|x^k, a^k, y) Q(dy|x^k, a^k), \end{aligned}$$

and

$$\begin{aligned} r(x^k, a^k) &= E_x^\delta\{r(\tau^k, \xi^k, \alpha^k) | \kappa^k = h^k, \alpha^k = a^k\} \\ &= \int_X \int_0^\infty r(t, x^k, a^k) \, d\mathrm{T}(t|x^k, a^k, y) Q(dy|x^k, a^k). \end{aligned}$$

Using this notation it follows

$$E_x^\delta \sum_{k=0}^n r(\tau^k, \xi^k, \alpha^k) = E_x^\delta \sum_{k=0}^n r(\xi^k, \alpha^k), \tag{6}$$

and

$$E_x^\delta \sum_{k=0}^n \tau^k = E_x^\delta \sum_{k=0}^n \tau^k(\xi^k, \alpha^k),$$

**Theorem 3.1.** If Assumption I holds and if there exist a constant $q$ and a bounded function $v(x)$ on $X$ such that

$$v(x) = \sup_{a \in A(x)} \Big\{ r(x,a) + \int v(y) Q(dy|x,a) - q\tau(x,a) \Big\}, \qquad x \in X. \tag{7}$$

Then

$$\sup_{\delta \in LS} \phi(x, \delta) \leq q, \qquad x \in X. \tag{8}$$

If thus

$$v(x) = \max_{a \in A(x)} \Big\{ r(x,a) + \int v(y) Q(dy|x,a) - q\tau(x,a) \Big\}, \qquad x \in X.$$

and for some strategy $\delta^\star \in LS_D$

$$v(x) = r(x, \delta^\star(x)) + \int v(y) Q(dy|x, \delta^\star(x)) - q\tau(x, \delta^\star(x)), \qquad x \in X, \tag{9}$$

holds, then $\delta^\star$ is locally optimal and

$$\phi(x, \delta^\star) \equiv q.$$

$\odot$

*Proof.* For any strategy $\delta$ we have

$$E_x^\delta \Big\{ \sum_{k=1}^{n} \big[ v(\xi^k) - E^\delta \{ v(\xi^k) | \kappa^{k-1}, \alpha^{k-1} \} \big] \Big\} = 0. \tag{10}$$

From (4) and (7) we have

$$\begin{aligned} E^\delta\{v(\xi^k)|\kappa^{k-1} = h^{k-1}, \alpha^{k-1} = a^{k-1}\} &= \int v(y) Q(dy|x^{k-1}, a^{k-1}) \qquad (11) \\ &= r(x^{k-1}, a^{k-1}) + \int v(y) Q(dy|x^{k-1}, a^{k-1}) - \\ &\quad - q\tau(x^{k-1}, a^{k-1}) - r(x^{k-1}, a^{k-1}) + q\tau(x^{k-1}, a^{k-1}) \\ &\leq v(x^{k-1}) - r(x^{k-1}, a^{k-1}) + q\tau(x^{k-1}, a^{k-1}). \end{aligned}$$

Substituting (11) in (10) we obtain

$$E_x^\delta \Big\{ \sum_{k=1}^{n} r(\xi^{k-1}, \alpha^{k-1}) + v(\xi^n) - v(x^0) \Big\} \leq q E_x^\delta \sum_{k=1}^{n} \tau(\xi^{k-1}, \alpha^{k-1}). \tag{12}$$

From boundedness of $v(x)$, Assumption I, and (6), we obtain

$$\phi(x, \delta) \leq q, \qquad x \in X, \tag{13}$$

so inequality (8) is proved.

To prove the second part of the theorem we notice that for the strategy $\delta^\star$ in relations (11)–(13) equality holds. $\square$

**Remark 3.1.** The proof of Theorem 3.1 is similar to the proof of Theorem 2 in [15], which deals with the class of stationary strategies only. Further, our Assumption I is weaker than that used there. $\odot$

Let $M(X)$ be Banach space of bounded Borel measurable functions on $X$ with sup-norm $||u|| = \sup_{x \in X} |u(x)|$. Denote by $\rho$ the metric on $M(X)$ induced by that norm. We define an operator $\mathfrak{U}$ on $M(X)$ by

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

where

$$Q'(C|x, a) = Q(C|x, a) - \frac{1}{M} \mu(C) \tau(x, a). \tag{15}$$

**Theorem 3.2.** Let Assumptions I–III hold and assume further that

1) the operator $\mathfrak{U}$ maps some metric subspace $S(X) \subseteq M(X)$ (with metric $\rho$ induced by norm of $M(X)$) into itself;

2) for every function $u \in S(X)$ the map $F_u : X \to (2)^A_{set}$, which determines for every $x \in X$ the set

$$F_u(x) := \Big\{a\colon a \in A(x), \quad \mathfrak{U}u(x) = r(x,a) + \int u(y)Q'(dy|x,a)\Big\},$$

is Borel measurable.

Then there exists a strategy in $LS_D$ which is optimal in $LS$ (locally optimal). $\odot$

*Proof.* We will show that the operator $\mathfrak{U}$ has the following properties:

a) $\mathfrak{U}$ is isotone, i.e., if $u_1(x) \geq u_2(x)$, $x \in X$, then $\mathfrak{U}u_1(x) \geq \mathfrak{U}u_2(x), \qquad x \in X$;

b) for any nonnegative constant $c$

$$\mathfrak{U}(u(x)+c) \leq \mathfrak{U}u(x) + \alpha c, \qquad x \in X, \qquad \text{with} \qquad \alpha = 1 - \frac{m}{M}\mu(X).$$

Indeed, property a) follows directly from Assumption III a), and property b) is seen as follows:

$$\begin{aligned}
\mathfrak{U}(u(x)+c) &= \sup_{a\in A_x}\Big\{r(x,a) + \int u(y)Q'(dy|x,a) + c\int Q'(dy|x,a)\Big\} \\
&= \sup_{a\in A_x}\Big\{r(x,a) + \int u(y)Q'(dy|x,a) + c(1 - \frac{1}{M}\tau(x,a)\mu(X))\Big\} \\
&\leq \sup_{a\in A_x}\Big\{r(x,a) + \int u(y)Q'(dy|x,a) + c(1 - \frac{m}{M}\mu(X))\Big\} \\
&= \mathfrak{U}u(x) + \alpha c.
\end{aligned}$$

Properties a) and b) guarantee that $\mathfrak{U}$ is a contraction operator with contraction coefficient $\alpha = 1 - \frac{m}{M}\mu(X) \in [0,1)$, because $\mu$ is substochastic and $m \leq M$. Indeed, apply $\mathfrak{U}$ to both sides of the inequality $u_1(x) \leq u_2(x) + \rho(u_1,u_2)$. From a), b) we have

$$\mathfrak{U}u_1(x) \leq \mathfrak{U}(u_2(x) + \rho(u_1,u_2)) \leq \mathfrak{U}u_2(x) + \alpha\rho(u_1,u_2).$$

Consequently,

$$\mathfrak{U}u_1(x) - \mathfrak{U}u_2(x) \leq \alpha\rho(u_1,u_2).$$

Exchanging $u_1$ and $u_2$ we have

$$|\mathfrak{U}u_1(x) - \mathfrak{U}u_2(x)| \leq \alpha\rho(u_1,u_2), \qquad x \in X,$$

i.e.

$$\rho(\mathfrak{U}u_1, \mathfrak{U}u_2) \leq \alpha\rho(u_1,u_2).$$

Banach's fixed point theorem then implies that in $S(X)$ there exists a function $u^\star(x)$ such that

$$u^\star(x) = \sup_{a\in A(x)}\Big\{r(x,a) + \int u^\star(y)Q'(dy|x,a)\Big\}, \qquad x \in X. \tag{16}$$

From (16) and (15) we obtain the optimality equation (7) with $v(x) = u^\star(x)$ and $q = \frac{1}{M}\int u^\star(y)\mu(dy)$.
Assumption 2) guarantees that the map $F_{u^\star}$ is Borel measurable. This by Definition 2.3 especially yields $F_{u^\star}(x) \neq \emptyset$ for all $x \in X$, and from the definition $F_{u^\star}(x) \subseteq A(x)$ holds. So, the set $F_{u^\star}(x)$ of maximizers of $\mathfrak{U}u^\star(x)$ contains only local policies.
Now, by the Theorem of Choice, there exists a Borel measurable $A(x)$-valued selector $\delta^\star$, i.e. we have

$$\delta^\star(x) \in \{a \in A(x) : \mathfrak{U}u^\star(x) = r(x,a) + \int u^\star(y)Q'(dy|x,a)\}$$

which means

$$\mathfrak{U}u^\star(x) = r(x,\delta^\star(x)) + \int u^\star(y)Q'(dy|x,\delta^\star(x)),$$

i.e. for the strategy $\delta^\star \in LS_D$ (9) is fulfilled and therefore, by Theorem 3.1, $\delta^\star$ is a locally optimal strategy. □

**Remark 3.2.** Under the conditions and assumptions of Theorem 3.2 we now conclude that in the set of optimal strategies there is at least one deterministic strategy, which makes decisions on the basis of the actual state only. The optimal value $\phi(x,\delta^\star)$ under this strategy depends according to the observation of Ross [15] only on the transition kernels $Q(\cdot|x,a)$ and $\pi^n(\cdot,x)$ and on mean conditional sojourn times $\tau(x,a)$ and mean conditional rewards $r(x,a)$. Thus $\phi(x,\delta^\star)$ is insensitive against variations of the shape of the conditional sojourn times distribution and the conditional rewards distribution as long as their conditional means remain invariant. Therefore we can restrict the computation of $\phi(x,\delta^\star)$ without loss of generality to processes with deterministic sojourn times and special reward functions as given at the end of Section 2.3.

Our next theorem provides easy to check smoothness conditions for the data of the semi-Markov model such that the condition 1) and 2) of Theorem 3.2 are fulfilled.

**Theorem 3.3.** Consider a semi-Markov random field according to Definition 2.9. Recall that the local state spaces $X_i$ and local decision spaces $A_i$, and therefore the global state space $X$ and the global decision space $A$ as well, are compact metric spaces with countable basis. Let the map $F$, which associates with any state $x$ the time invariant restriction set $A(x)$, i.e.

$$F : X \longrightarrow (2)^A_{set}, \qquad x \to F(x) = \prod_{i\in V} F_i(x) = A(x),$$

be continuous and let Assumptions I, II, and III hold. Assume further that
1) the mean value functions $r(x,a)$ and $\tau(x,a)$ are continuous in $(x,a) \in \Delta$, and
2) the transition probability $Q(\cdot|x,a)$ is weakly continuous in $(x,a) \in \Delta$,
then there exists an optimal strategy $\delta^\star$ in $LS_D$ and the function $\delta^\star\colon X \to A$ can be selected from Baire class 1. ⊙

*Proof.* The proof uses ideas from Theorem 3 in [5] to apply our Theorem 3.2. To do this we need the following property which combines [5][Lemma 1, 2]:
Let $A$ be compact and the mapping $F : X \to (2)^A_{set}$ continuous. Let $u : \Delta \subseteq X \times A \to \mathbb{R}$

be bounded and continuous. Then the function $u^*(x) = \sup_{a \in F(x) = A(x)} u(x,a)$ is continuous.

Applying the conditions 1) and 2) and this property to the maximization operator $\mathfrak{U}$ in (14) yields that $\mathfrak{U}$ maps the space $B(X) \subseteq M(X)$ of real valued bounded continuous functions on $X$ into itself. So condition 1) of Theorem 3.2 is fulfilled. For condition 2) of Theorem 3.2 it suffices to show that for any $u \in C(X)$ the function $F_u : X \to (2)^A_{set}, \quad x \to F_u(x) = A(x)$ is upper semicontinuous. According to [12][p. 61] it is sufficient to prove that for fixed $u \in C(X)$ the following holds:
If for $n \to \infty$ we have $(x^n, a^n) \to (x,a)$ with $x^n \in X$ and $a^n \in F_u(x^n)$, then it follows $a \in F_u(x)$.
By definition $a^n \in F_u(x^n)$ means

$$r(x^n, a^n) + \int u(y) Q'(dy|x^n, a^n) = \sup_{\tilde{a} \in A(x^n)} \Big\{ r(x^n, \tilde{a}) + \int u(y) Q'(dy|x^n, \tilde{a}) \Big\},$$

Now the left side of the equation is a continuous function in $(x,a)$, while the right side is a continuous functions in $x$. Letting $n \to \infty$ we have

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

i.e., $a \in F_u(x)$. From Theorem 3.2 we conclude the existence of an optimal local stationary determististic strategy $\delta^\star \in LS_D$. $\square$

# References


[1] Cinlar E (1975) Introduction to Stochastic Processes. Prentice–Hall, Inc. Englewood Cliffs, New Jersey

[2] Disney RR, Kiessler PC (1987) Traffic processes in queueing networks. A Markov renewal approach. The Johns Hopkins University Press London

[3] Daduna H, Knopov PS, Chornei RK (2001) Local Control of Interacting Markov Processes on Graphs with Compact State Space. Kibernetika i Sistemny Analiz 3:62–77

[4] Gubenko LG, Shtatland ES (1972) On controlled semi-Markov processes. Kibernetika 2:26–29, English translation in: Controlled Semi-Markov Processes. Cybernetics 8:200–205

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

[6] Hinderer K (1970) Foundations of Non-stationary Dynamic Programming with Discrete Time Parameter. Lecture Notes in Operations Research and Mathematical Systems 33, Springer Berlin

[7] Howard RA (1964) Research in semi-Markov decision structures. Journal of the Operational Research Society of Japan 6:163-199

[8] Jewell WS (1963) Markov renewall programming. I. Formulation, finite return models. Operations Research 11: 938-948

[9] Jewell WS (1963) Markov renewall programming. I. Infinite return models, example. Operations Research 11: 948-971

[10] Kitaev M (1987) Elimination of Randomization in Semi-Markov Decision Models with Average Cost Criterion. Optimization 18:439–446.

[11] Kitaev MY, Rykov VV (1995) Controlled queueing systems. CRC Press Boca Raton

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

[13] Luque Vasquez F, Robles Alcaraz MT (1994) Controlled Semi-Markov Models with Discounted Unbounded Costs. Bol. Soc. Math. Mex. 39:51–68

[14] Rockafellar RT (1969) Measurable Dependence of Convex Sets and Functions on Parameters. J. Math. Anal. and Appl. 80:45–57

[15] Ross SM (1970) Average cost semi-Markov processes. Journal of Applied Probability 7: 649–656

[16] 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

[17] Vega-Amaya O (1993) Average Optimality in Semi-Markov Control Models on Borel Spaces: Unbounded Cost and Controls. Bol. Soc. Math. Mex. 38:47–60.

[18] Wakuta K (1987) Arbitrary State Semi-Markov Decision Processes with Unbounded Rewards. Optimization 18:447–454.