Markov Processes Relat. Fields **3**, 129–150 (1997)

Markov Processes and **Related Fields**

Fluid Approximations of Markov Decision Chains

A.S. Gajrat[1,2] **, A. Hordijk**[1,*]**, V.A. Malyshev**[2,3] **and F.M. Spieksma**[1,†]

[1] Department of Mathematics and Computer Science, University of Leiden, Niels Bohrweg 1, 2333 CA Leiden, The Netherlands
[2] Laboratory of Large Random Systems, Moscow State University, Moscow, 119899, Russia
[3] INRIA — Domaine de Voluceau, Rocquencourt, BP105–78153, Le Chesnay, France

Received December 16, 1996

Abstract. Euler limits or fluid approximations have proved useful in obtaining ergodicity conditions and more detailed information on random walks and communication networks. Until now the similar question for controlled Markov chains remained open. For a basic example we establish convergence of Markov decision chains to their fluid approximation. This allows to explicitly obtain the asymptotically optimal solution to our control problem.

KEYWORDS: asymptotic optimality, fluid model, controlled Markov chains, random walk

AMS SUBJECT CLASSIFICATION: Primary 93E20

1. Introduction

Most papers on Markov decision chains essentially consider the case of a finite state space in the following sense. They use typical restrictions like uniform Lyapunov functions for all strategies, or conditions ensuring the existence of such uniform Lyapunov function for a suitable subclass of strategies containing the optimal one. From a practical point of view this means that the process mainly moves in a finite or bounded part of the state space and so it is stationary

*The research of this author was completed while he was on sabbatical leave at INRIA, Sophia-Antipolis; it has been partially supported by the Ministère Français de l'Éducation Nationale et de l'Enseignement Supérieur et de la Recherche

†The research of this author has been supported by a fellowship from the Royal Netherlands Academy for Arts and Sciences

or close to stationary. The transient period then is sufficiently small with respect to the scale of time.

It is quite remarkable that some practical situations demand a completely different time scale. We shall call this scale "state of emergency scale" or fluid model scale. Roughly speaking, this means the following: assume that we are far from equilibrium, which means that there is an emergency situation, a catastrophe or a temporary lack of supply. Then the problem is to find the optimal path that could bring us somewhere sufficiently close to equilibrium. The strategy to achieve this, can be quite different from the optimal strategy during calm periods.

We should explain what we mean by saying that the initial state is far from equilibrium. The only way to do this, is to scale the initial state in the "space" metrics. Then the time scale for reaching equilibrium that we should consider, should be related to the time to reach the vicinity of the most probable states, starting from the thus chosen initial state.

This formulation already suggests the use of fluid approximations for the Markov chains under consideration. This approach was recommended in [9]. But even earlier, there were papers for example by G. Weiss [14] on this topic. In these papers, control problems for deterministic fluid approximations for queueing networks were considered, as purely deterministic problems. However, we do not know of any papers, where convergence to the controlled fluid approximation was proven. Here it is our objective to solve such kind of problem.

Note that there are two approaches to "fluid approximations". One was recommended in papers on networks, see for example [4], and, independently, in some papers on deterministic queues. The second one had been developed earlier in the framework of random walks (see references in [3, 8]). For Markov decision chains the situation is far from obvious: there is no direct analogy with the two above approaches to fluid approximations. We will now explain some of our intuitions on which this paper is based.

Assume that we have two queues and so the state space is $\mathbf{Z}_+^2$. Let $c(n,a), n \in \mathbf{Z}_+^2$, be the immediate cost that is paid in state n when using action a. Assume that at time 0 we are at the point $[rN] = ([xN],[yN]) \in \mathbf{Z}_+^2$. We are looking for an optimal policy $a(n)$ minimising the total cost $\mathsf{E}\sum_{t=0}^{\tau} c(a(\xi_t))$, where τ is the first hitting time of zero and $c(a)$ a positive cost. Thus we assume stationarity from the beginning.

Which assumptions on the transition probabilities $p_a(n,m)$ could we use? One of the simplest assumptions (which is by the way a sufficiently difficult case because of possible discontinuities of the mean jump vector field) is maximal possible homogeneity, that is homogeneity inside the quarter plane and on each of the two axes. This problem will be considered in a subsequent paper. As a first step, we will illustrate our main ideas by considering the simpler problem of Markov decision chains on $\mathbf{Z}^2$ with homogeneous transition probabilities $p_a(n,m) = p_a(n-m)$. This approach is the essence of the ideology concerning

random walks with boundaries: to first solve the necessary problems for the induced chains.

We guess that the asymptotics for the optimal cost should be cN and we will try to find the coefficient c. Let $N(a)$ be the mean number of times that we use action a till we reach zero from $[rN]$. Then we would like to minimise $\sum_a N(a)c(a)$ under the restriction that $\sum_a N(a)v_a = -[rN]$ where v_a is the mean drift from a point under action a. It is quite plausible that $N(a) \equiv b(a)N$, and this is an assumption that should be proved, and which is reminiscent of fluid approximations. This should give us an asymptotically optimal solution up to terms of order $o(N)$.

This idea amounts to solving a Linear Programming problem (LP), which we will call the fluid-LP (FLP). Our main theorem shows that the optimal value of the FLP as a function of the state x, $C_F(x)$ say, is asymptotically equal to the value of the optimal control problem with starting state x, say $C(x)$. More precisely, $\lim_{N\to\infty} C(Nx)(C_F(Nx)) = 1$, for all $x \in \mathbf{Z}^2$. Moreover, the control constructed from the optimal solution of FLP, π_{LP}, is asymptotically optimal (see Theorem 2.1).

We will end this section by discussing other applications of LP techniques in Markov decision problems and their connection with our approach. In MDPs with a finite state space LPs have been used to obtain the optimal control for all cost criteria. For the expected total cost criterion relevant results have been obtained in [5]. In Section 5 we give the precise formulation of the LP in [5], which we will call the control-LP (CLP), in order to distinguish it from the FLP studied in this paper. Theorem 17 of [5] deals with the following three cases: (i) the CLP is infeasible, then there is no policy with a finite expected hitting time of state 0; (ii) the CLP has an unbounded solution, and then there is no policy with finite expected cost until state 0 is reached; (iii) if there is a finite extreme optimal solution of CLP, then an optimal policy for the control problem is easily derived from this (see Section 5). These results can be generalised to our MDP with state space $\mathbf{Z}^2$ corresponding to the control problem of this paper. However, the resulting CLP is infinite in this case and generally an optimal solution can not be computed. Also, in case an optimal control could be determined, it is of little value in practice, since it is too complicated to be implemented. Indeed, the optimal actions depend on the state, mostly in a complex way, and any implementation needs an infinite list of state-action-pairs. Whereas the coefficient matrix in the CLP consists of the transition probabilities, the coefficient matrix in the FLP consists of the drifts. This enormously simplifies the LP. Indeed, the dimension reduces from infinity to dimension 2, the dimension of our state space. As a consequence, the FLP is extremely easy to solve and the resulting asymptotically optimal policy has a simple and easily implementable structure.

1.1. Notations

We use here the standard terminology in countable Markov decision chains (see e.g. [11,12]). We will consider Markov chains $\mathcal{L}_\pi$ on a common state space $\mathbf{Z}^2$. Each of these chains is defined by a (deterministic, stationary) policy $\pi(x) \in \mathcal{A}$, $x \in \mathbf{Z}^2$, which is a function defined on the state space and taking values in a finite set of actions $\mathcal{A}$. Actions will be usually denoted by a, b, c.

The Markov chain $\mathcal{L}_\pi$ has transition probabilities $p(x,y) = p_{\pi(x)}(y-x)$ for $x \neq 0$, and $p(0,y) = \mathbf{1}_0(y)$. So 0 is an absorbing state under any policy π. By $c(a)$ we denote the immediate cost of using action $a \in \mathcal{A}$, and $p_a(x,y) = p_a(y-x)$ are the transition probabilities, when using action $a \in \mathcal{A}$. Further, we denote by $v_a = \sum_y (y-x)p_a(y-x)$ the mean drift from the point x under action $a \in \mathcal{A}$.

For this model we will study the expected total cost $C_\pi(x)$ of policy π, when the system starts in state x, defined by

$$C_\pi(x) = \mathsf{E}_x \sum_{t=0}^{\infty} c(\pi(\xi_t)) 1(\xi_t \neq 0).$$

Finally, we will use the notation $\|x\| = \sqrt{x_1^2 + x_2^2}$, $x = (x_1, x_2)^T \in \mathbf{R}^2$,

$$\operatorname{dist}(x, U) = \inf_{y \in U} \|y - x\|,$$

where $x \in \mathbf{R}^2, U \subset \mathbf{R}^2$, and $cU = \{cx : x \in U\}$, for $U \subset \mathbf{R}^2$, $c \in \mathbf{R}$.

1.2. Assumptions

1. All possible Markov chains have uniformly bounded jumps i.e., there exists $d > 0$ such that for all $a \in \mathcal{A}$ $p_a(y-x) = 0$ if $\|y - x\| > d$.
2. For all $a \in \mathcal{A}$ and x, y $\|y - x\| \leq d$ $p_a(y-x) > 0$.
3. The cone, $\{\sum_a -t_a v_a, t_a \geq 0\} = \mathbf{R}^2$.
4. The immediate cost of any action is positive, i.e. $c(a) > 0$ for all a.

Boundedness of jumps has been required in Assumptions 1 and 2 mainly for reasons of convenience and clarity of the exposition. It clearly can be relaxed. Assumption 3 has been introduced to ensure that for each point there is a policy, for which the origin can be reached from this point. Assumption 4 is used to avoid trivialities. For example, if $c(a) = 0$ for some a, then the policy that plays this action a in each point of the plane has 0 associated cost, but the Markov chain under this policy may be transient.

2. Results

First note that by the results in [13] on negative dynamic programming, there exists an optimal stationary, deterministic policy for our control problem, such that the corresponding value function solves the dynamic programming equations.

One of the possible approaches to this problem (realised in [7] for a particular case) is therefore to directly write down these discrete dynamic programming equations. This approach, besides being quite general, is nevertheless too formal and the task of solving this system is still formidable. Here we use a completely different and more intuitive approach.

Our main concern here will be to find a (stationary, deterministic) policy π minimising $C_\pi(x)$ for given x. Intuitively, this means that we would like to reach zero as quick and as cheap as possible. Theorem 2.1 below reduces this stochastic control problem to a deterministic one (up to lower order terms in the large N asymptotics) that can be explicitly solved.

2.1. The FLP

For any $x \in \mathbf{R}^2$ we consider the following LP.

FLP. Determine

$$C_{\text{opt}}(x) = \inf \sum_{a \in \mathcal{A}} c(a) t_a,$$

subject to

$$\begin{aligned} \sum_{a \in \mathcal{A}} v_a t_a &= -x, \\ t_a &\geq 0, \quad a \in \mathcal{A}. \end{aligned} \tag{2.1}$$

Let π_{opt} be the optimal policy for the control problem. The next theorem shows that the asymptotic cost of the stochastic problem for large initial states equals the optimal value of FLP.

Theorem 2.1. *Assume that Assumptions* 1–4 *hold. Then*

$$\lim_{N \to \infty} \frac{C_{\pi_{\text{opt}}}(x_N)}{N} = C_{\text{opt}}(x),$$

where $x_N = [xN]$, *and* $[x_1, x_2]$ *is* $([x_1], [x_2])$, *with* $[x_i]$ *the largest integer smaller than or equal to* $x_i, i = 1, 2$.

The proof of the theorem consists of the following two steps.

Step 1. Construct a policy π_{LP} such that

$$\lim_{N \to \infty} \frac{C_{\pi_{LP}}(x_N)}{N} = C_{\text{opt}}(x). \tag{2.2}$$

Step 2. Show that

$$\liminf_{N\to\infty} \frac{C_{\pi_{\rm opt}}(x_N)}{N} \geq C_{\rm opt}(x). \tag{2.3}$$

These two facts together obviously immediately imply the theorem.

3. Proofs

3.1. Step 1

3.1.1. Useful facts for the FLP

Here we show how to solve the FLP.

From elementary Linear Programming it follows that under Assumptions 3 and 4 for any $x \in \mathbf{R}^2$ there exists a (in general not unique) solution $t_a(x) \geq 0, a \in \mathcal{A}$ of FLP, such that

$$C_{\rm opt}(x) = \sum_{a\in\mathcal{A}} c(a)t_a(x). \tag{3.1}$$

We will denote by $\{t_a(x)\}_{a\in\mathcal{A}}$ an optimal solution of FLP.

Let $\{a_0, \ldots, a_{n-1}\} \in \mathcal{A}$ be the subset of actions such that for each a_k there exists $x \in \mathbf{R}^2$ with $t_{a_k}(x) \neq 0$. By elementary Linear Programming also the following simple results hold.

1) If $x = -tv_{a_k}, t \geq 0$ $(0 \leq k < n)$, then

$$\begin{aligned} t_a(x) &= t \text{ if } a = a_k, \\ t_a(x) &= 0 \text{ if } a \neq a_k \end{aligned}$$

is an optimal solution. Hence, the direction $-v_{a_k}$ is optimal in FLP.

2) Divide the plane $\mathbf{R}^2$ into n cones by the rays $\{-tv_{a_k}, t \geq 0\}_{k=0}^{n-1}$, as has been shown in Figure 1. Define the ray $R_k = \{-tv_{a_k}, t \geq 0\}$. For two rays R_k, R_l we define the cone $[R_k, R_l) = \{-t_1 v_{a_k} - t_2 v_{a_l}, t_1 > 0, t_2 \geq 0\}$. For $0 \leq k, l < n$ define $k \oplus l = (k + l) \bmod n$ and $k \ominus l = (n + k - l) \bmod n$.

Let us draw the vectors v_a with starting point the origin. By renumbering we can enumerate them in clockwise order: $v_1, \ldots, v_k, \ldots$ Rays R_k are numbered correspondingly (see Figure 1). Let $U_k = [R_k, R_{k\oplus 1})$. Then

$$\bigcup_{k=0}^{n-1} U_k = \mathbf{R}^2 \backslash \{0\}$$

and, for any $x \in U_k$, we can choose an optimal solution $t_a(x)$ of the following form: $t_a(x) = 0$, if $a \notin \{a_k, a_{k\oplus 1})\}$, where $t_{a_k}(x)$, $t_{a_{k\oplus 1}}(x)$ is defined by the equation

$$-x = t_{a_k}(x)v_{a_k} + t_{a_{k\oplus 1}}(x)v_{a_{k\oplus 1}}. \tag{3.2}$$

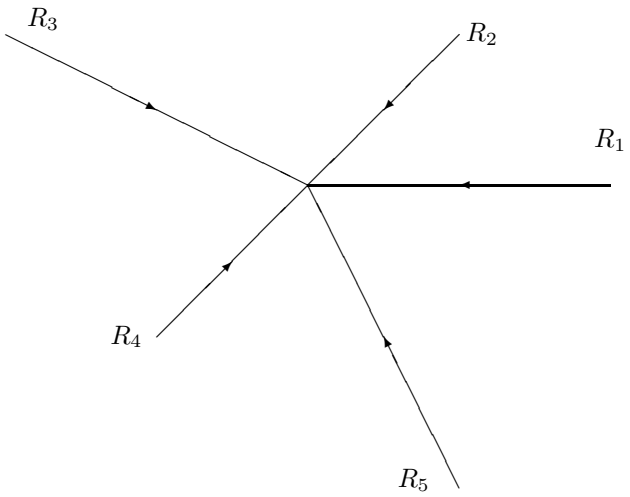

Figure 1.

Thus, for any fixed x two actions are sufficient to construct an asymptotically optimal path and this path consists of two linear parts.

3.1.2. Construction of the π_{LP} policy and its ergodicity

Define the policy π_{LP} for the optimal control problem by

$$\pi_{LP}(x) = a_k, \text{ if } x \in [R_k, R_{k\oplus 1}). \tag{3.3}$$

Then π_{LP} is defined for all points $x \in \mathbf{Z}^2$.

Let us prove that (2.2) holds for this policy. First we will prove that the Markov chain $\mathcal{L} = \mathcal{L}_{\pi_{LP}}$ is "ergodic" (more exactly, we reach zero with probability one in finite mean time) and that the following lemma holds.

Lemma 3.1. *Let τ be the hitting-time of state 0 in the Markov chain $\mathcal{L}$. Then there exist constants $C_1, C_2, T, \delta > 0$, such that for any $x \in \mathbf{Z}^2$*

$$\mathsf{E}_x \tau < C_1 \|x\|,$$

$$\mathsf{P}_x\{\tau > T\|x\|\} < C_2 e^{-\delta\|x\|}.$$

Proof. To prove the theorem, we construct a positive function $f(x)$, $x \in \mathbf{R}^2$, such that

$$B_1\|x\| < f(x) < B_2\|x\|, \tag{3.4}$$

$$|f(x+y) - f(x)| \le B_3\|y\|,$$

for some constants $B_1, B_2, B_3 > 0$, and there exists $\epsilon > 0$ and $r_0 > 0$, such that for any $x \in \mathbf{Z}^2$, $\|x\| > r_0$

$$\mathsf{E}(f(\xi_{t+1}) - f(\xi_t) \mid \xi_t = x) \leq -\epsilon. \tag{3.5}$$

Then one can easily get the first assertion of the theorem (see Theorem 2.1.1 in [3]). We explain the second in much more detail. To this end we need the following theorem (see [3], Theorem 2.1.7). We have changed a little bit the formulation of this theorem to make it simpler.

Theorem 3.1. (Exponential inequality for supermartingales). *Let $(\Omega, \mathcal{F}, \mathsf{P})$ be a given probability space and $\{\mathcal{F}_t, t \geq 0\}$ an increasing family of σ-algebras $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots \subset \mathcal{F}_t \subset \cdots \subset \mathcal{F}$. Let $\{S_t, t \geq 0\}$ be a sequence of real random variables, such that S_t is $\mathcal{F}_t$-measurable, for all $t \geq 0$. Moreover, S_0 will be taken constant. If there exist positive numbers ϵ, d such that for all $k \geq 0$*

$$\begin{aligned} |S_{k+1} - S_k| &\leq d, \\ \mathsf{E}(S_{k+1} - S_k \mid \mathcal{F}_k) &\leq -\epsilon \quad \text{a.s.}, \end{aligned}$$

then, for any $\delta_1 < \epsilon$, there also exist constants $D = D(S_0, \epsilon, d)$ and $\delta_2 > 0$, such that for any $t \geq 0$

$$\mathsf{P}\{S_t > -\delta_1 t\} < De^{-\delta_2 t}.$$

Proof. See [3]. □

In order to use Theorem 3.1, we will introduce the random moment $\tilde{\tau} = \min\{t : f(\xi_t) \leq r_0\}$ and the sequence of random variables $\{S_t, t \geq 0\}$ with

$$\begin{aligned} S_t &= f(\xi_t) - f(\xi_0), \quad \text{if } t < \tilde{\tau}, \\ S_t &= S_{t-1} - \epsilon, \quad \text{if } t \geq \tilde{\tau}. \end{aligned}$$

The conditions of Theorem 3.1 hold for S_t. Hence for $\tilde{\epsilon} < \epsilon$ there exist $C > 0$, $\delta > 0$ such that

$$\mathsf{P}_x\{S_t > -\tilde{\epsilon}t\} < Ce^{-\delta t}.$$

Remark that

$$\begin{aligned} \mathsf{P}_x\{\tilde{\tau} > t\} &= \mathsf{P}_x\{\tilde{\tau} > t, f(\xi_t) > r_0\} \\ &\leq \mathsf{P}_x\{S_t > r_0 - f(x)\} \\ &\leq \mathsf{P}_x\{S_t > -B_2\|x\|\} \\ &< Ce^{-\delta t}, \quad \text{for } t \geq B_2\|x\|/\tilde{\epsilon}. \end{aligned}$$

In particular, for $t = B_2\|x\|/\tilde{\epsilon}$ we obtain

$$\mathsf{P}_x\{\tilde{\tau} > B_2\|x\|/\tilde{\epsilon}\} < Ce^{-\delta B_2\|x\|/\tilde{\epsilon}}.$$

From this inequality it is easy to get the second assertion of the lemma. To this end one can get exponential inequality for $\mathsf{P}_x\{\tau - \tilde{\tau}\}$, for x, $f(x) \leq r_0$. We leave the proof for reader.

To construct the function f we use the following idea. Let $L_r = \{x \in \mathbf{R}^2 : f(x) = r\}$ be the set of points at the same level for function f. If f is not "too bad", then L_r would be a smooth curve. If in any point on L_r the vector of mean drift (of random walk ξ_t) "looks inside" of curve L_r then we can hope to get inequality (3.5). Following this idea, it is natural to define $f(x)$ for

$$x = (x_1, x_2)^T = (r(x)\cos\alpha(x), r(x)\sin\alpha(x))^T, r(x) = \|x\|,$$

in polar coordinates (r, α) as,

$$f(x) = \frac{r(x)}{l(\alpha(x))},$$

where $l(\alpha)$ is a positive smooth function on the segment $[0, 2\pi]$, i.e. $l(\alpha) \in C^1([0, 2\pi])$.

This uses the so-called ϵ-linearity construction (see [3]).

Define $s(\alpha) = (l(\alpha)\cos\alpha, l(\alpha)\sin\alpha)$, $\dot{s} = ds/d\alpha$. So $L_r = \{rs(\alpha), \alpha \in [0, 2\pi]\}$ and $\dot{s}$ is tangent to curve L_1.

Suppose that for any $x \in L_1$

$$v_x \times \dot{s}(\alpha(x)) < 0, \tag{3.6}$$

where $\times$ denotes the vector product, and $v_x = v_{a_k}$, if $x \in U_k$. In other words, v_x "looks inside" L_1 (see Figure 2). Remark that $v_x \times \dot{s}(\alpha(x))$ can be extended to a continuous function in each region $\bar{U}_k$. Therefore there is $\epsilon > 0$ such that for any $x \in L_1$

$$v_x \times \dot{s}(\alpha(x)) < -\epsilon.$$

Now we can compute (3.5). Let $x \in U_k$, $\|x\| = 1$, $p(y) = p_{a_k}(y)$. Then for $r > 0$ such that $rx \in \mathbf{Z}^2$

$$\mathsf{E}(f(\xi_{t+1}) - f(\xi_t) \mid \xi_t = rx) = \sum_{\|y\| \leq d} p(y)(f(rx + y) - f(rx)).$$

Taking into account the boundedness of jumps and using a Taylor expansion, we get

$$f(rx + y) = rf(x + \frac{y}{r}) = r\left(f(x) + \frac{d}{dx}f(x) \cdot \frac{y}{r} + O(1/r^2)\right),$$

where $\frac{d}{dx}f(x) = (\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2})$. Hence,

$$\mathsf{E}(f(\xi_{t+1}) - f(\xi_t) \mid \xi_t = rx) = \frac{d}{dx}f(x)v_x + O(1/r),$$

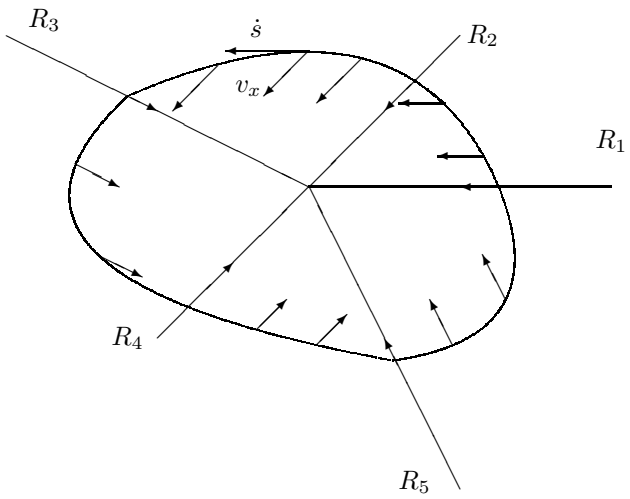

Figure 2.

$$\frac{d}{dx}f(x)v_x = \Big(\frac{\partial f}{\partial r}, \frac{\partial f}{\partial \alpha}\Big) \begin{pmatrix} \partial r/\partial x_1 & \partial r/\partial x_2 \\ \partial \alpha/\partial x_1 & \partial \alpha/\partial x_2 \end{pmatrix} v_x,$$

$$\begin{pmatrix} \partial r/\partial x_1 & \partial r/\partial x_2 \\ \partial \alpha/\partial x_1 & \partial \alpha/\partial x_2 \end{pmatrix} = \frac{1}{r} \begin{pmatrix} r\cos\alpha & r\sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix}.$$

One can easily check that,

$$\frac{d}{dx}f(x)v_x = \frac{1}{l^2(\alpha(x))} v_x \times \dot{s}(\alpha(x)).$$

Hence, there is r_0 such that for $x : \|x\| > r_0$ (3.5) holds. So we have reduced our problem to the construction of function $l(\alpha)$ such that (3.6) holds. This can be done in the following way. To simplify the notation we write $v_k = v_{a_k}$, $0 \le k < n$. By Assumption 3 and step 1 we have that

$$v_k \times v_{k\oplus 1} > 0,$$

i.e. the angle between these vectors is smaller than π, otherwise the plane $\mathbf{R}^2$ could not be obtained by negative linear combinations of the v_k.

Let V be the oriented polygon with vertices $\{V_k\}_{k=0}^{n-1}$, $V_k = -v_k$ and sides $\{(V_k, V_{k\oplus 1})\}_{k=0}^{n-1}$.

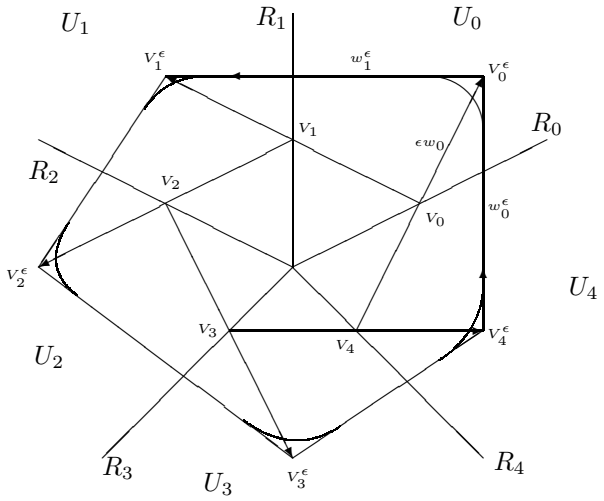

Figure 3.

Let w_k be a vector corresponding to side $(V_{k\ominus 1}, V_k)$ (see Figure 3), i.e.

$$w_k = v_{k\ominus 1} - v_k.$$

Then

$$\begin{aligned} v_k \times w_k &= v_k \times v_{k\ominus 1} < 0, \\ v_k \times w_{k\oplus 1} &= -v_k \times v_{k\oplus 1} < 0. \end{aligned}$$

For $\epsilon > 0$ we define the oriented polygon V^ϵ with vertexes $\{V_k^\epsilon\}_{k=0}^{n-1}$, $V_k^\epsilon = -v_k + \epsilon w_k$ and sides $\{(V_k^\epsilon, V_{k\oplus 1}^\epsilon)\}_{k=0}^{n-1}$. As above let w_k^ϵ be the vector corresponding to side $(V_{k\ominus 1}^\epsilon, V_k^\epsilon)$

$$w_k^\epsilon = v_{k\ominus 1} - v_k + \epsilon w_k - \epsilon w_{k\ominus 1}.$$

We can choose ϵ sufficiently small such that for all $0 \le k < n$

$$V_k^\epsilon \in U_k$$

and

$$\begin{aligned} v_k \times w_k^\epsilon &< 0, \\ v_k \times w_{k\oplus 1}^\epsilon &< 0. \end{aligned} \tag{3.7}$$

The polygon V^ϵ is a good basis for constructing the function $l(\alpha)$. Indeed, from (3.7) it follows that v_x "looks into" V^ϵ (see (3.6)). We should only make

from this polygon a smooth curve. This is easy. Let us take a circle in each angle of the polygon V^ϵ (see Figure 4). Denote the curve that we get by L_1. Let $r = l(\alpha)$ be equation of this curve in polar coordinates. Let us check (3.6). Put $x \in U_1$. If x is in the linear part of L_1, then $\dot{s}(\alpha(x))$ is equal to cw_0^ϵ or to cw_1^ϵ ($c > 0$) and (3.6) follows from (3.7). If x belongs to one of the circle segments, then one can easily show that

$$\dot{s}(\alpha(x)) = c_0 w_0^\epsilon + c_1 w_1^\epsilon, c_0, c_1 > 0.$$

Hence by again using (3.7), we get (3.6). □

3.1.3. Asymptotics of the total cost for π_{LP}

The proof of step 1 of Theorem 2.1 will be completed by the following lemma.

Lemma 3.2. *Let π_{LP} be defined by* (3.3). *Then* (2.2) *holds.*

Proof. Define the following random variables

$$\tau_a = \#\{t : \pi_{LP}(\xi_t) = a\}, \quad a \in \mathcal{A}.$$

From Lemma 3.1 we know that for some $C > 0$ and any $a \in \mathcal{A}$

$$\mathsf{E}_{x_N} \tau_a \le \mathsf{E}_{x_N} \tau < NC.$$

For any $t \ge 0$ we have

$$\mathsf{E}(\xi_{t+1} - \xi_t \mid \xi_t) = v_{\pi_{LP}(\xi_t)}.$$

Hence from the ergodicity of $\mathcal{L}_{\pi_{LP}}$ we get

$$-x_N = \sum_{t=0}^{\infty} \mathsf{E}(\xi_{t+1} - \xi_t) = \sum_{a \in \mathcal{A}} v_a \mathsf{E}_{x_N} \tau_a. \tag{3.8}$$

Suppose that $x \in U_1$. The proof for x an element of one of the other regions U_k is similar. Assume for the moment, that for any $a \neq a_1, a_2$

$$\lim_{N \to \infty} \mathsf{E}_{x_N} \tau_a / N = 0. \tag{3.9}$$

From this fact and (3.8) we get that

$$-x = \lim_{N \to \infty} \frac{1}{N} (v_{a_1} \mathsf{E}_{x_N} \tau_{a_1} + v_{a_2} \mathsf{E}_{x_N} \tau_{a_2}).$$

Hence, (see (3.2))

$$\lim_{N \to \infty} \frac{1}{N} \mathsf{E}_{x_N} \tau_{a_i} = t_{a_i}(x), \quad i = 1, 2.$$

Then (2.2) immediately follows from

$$C_{\pi_{LP}}(x_N) = \sum_{a \in \mathcal{A}} c_a \mathsf{E}_{x_N} \tau_a.$$

So, it remains to prove that (3.9) holds.

Intuitively, it is clear that (3.9) is true. Indeed, the scaled trajectory ξ_{tN}/N starting at the point x_N tends to a deterministic path (see Figure 4). In other words, the random walk ξ_t leaves the region $U_1 \cup U_2$ only, when it is close to 0. But in this case ξ_t hits 0 after some finite time (see Lemma 3.1). Hence the time period during which it could "use" action a, will be bounded. To prove this, one can use results from [1], but we prefer to give a more simple proof. Denote by $(y_1(x), y_2(x))$ the coordinates of the point $x \in \mathbf{R}^2$ with respect to the basis $(-v_{a_1}, -v_{a_2})$, i.e.

$$x = -y_1(x)v_{a_1} - y_2(x)v_{a_2}.$$

Let $\epsilon > 0$ be fixed. Consider the rectangle T (in the (y_1, y_2) coordinate system) with vertices $\{(0, y_2(x)-\epsilon), (y_1(x)+\epsilon, y_2(x)-\epsilon), (y_1(x)+\epsilon, y_2(x)+\epsilon), (0, y_2(x)+\epsilon)\}$, and the pentagon $\tilde{T}$ with vertices $\{(0,0), R_3 \cap \{y_1 = -\epsilon\}, (-\epsilon, y_2(x) + 2\epsilon), (\epsilon, y_2(x) + 2\epsilon), (\epsilon, 0)\}$. We also define the following sets

$$\begin{aligned} I_1 &= T \cap U_1, \\ I_0 &= T \cap R_2, \\ J_1 &= \tilde{T} \cap (U_1 \cup U_2), \\ J_0 &= \tilde{T} \cap (R_1 \cup R_3), \end{aligned}$$

see Figure 4.

For any subset $U \in \mathbf{R}^2$, we define the random time τ_U as the first crossing time of the set U, i.e.

$$\begin{aligned} \tau_U &= \min\{t : [\xi_t, \xi_{t-1}] \cap U \neq \emptyset\}, \\ \tau_U &= \infty, \text{ if for all } t \geq 1 \quad [\xi_t, \xi_{t-1}] \cap U = \emptyset. \end{aligned}$$

Next, define the event A_N by

$$A_N = \{\tau_{NI_0} < \infty, \tau_{NJ_0} < \infty, \tau_{NI_0} \leq \tau_{NI_1}, \tau_{NJ_0} \leq \tau_{NJ_1}\}.$$

This event corresponds to the deviation of the trajectory of ξ_t from the "drift-path" is not "too large". We will bound

$$\mathsf{E}_{x_N} \tau_a = \mathsf{E}_{x_N} \tau_a 1(A_N) + \mathsf{E}_{x_N} \tau_a 1(\bar{A}_N).$$

The first term can be bounded easily by using Lemma 3.1.

$$\mathsf{E}_{x_N} \tau_a 1(A_N) \leq \sum_{\substack{z \in \mathrm{Z}^2, \\ \mathrm{dist}(z, NJ_0) \leq d}} \mathsf{P}_{x_N}\{\xi_{\tau_{NJ_0}} = z\} \mathsf{E}_z \tau_a$$

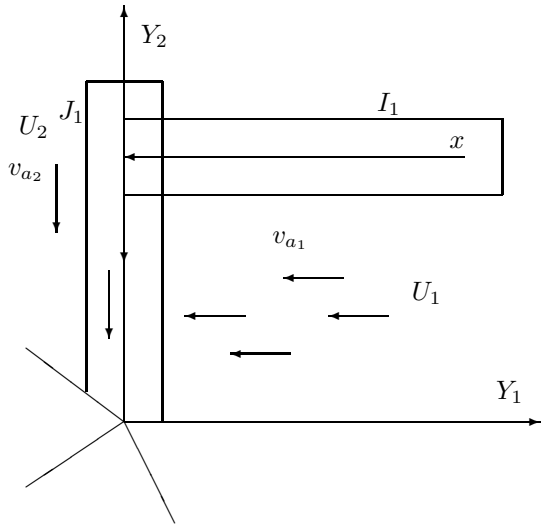

Figure 4.

$$\begin{aligned} &\leq \mathsf{P}(A_N) \max_{\substack{z\in\mathbb{Z}^2,\\ \mathrm{dist}(z,NJ_0)\leq d}} \mathsf{E}_z\tau \leq C\mathsf{P}(A_N) \max_{\substack{z\in\mathbb{Z}^2,\\ \mathrm{dist}(z,NJ_0)\leq d}} \|z\| \\ &\leq \tilde{C}\epsilon N, \end{aligned} \tag{3.10}$$

where the constant $\tilde{C}$ does not depend on ϵ. In the same way, we have for some $C > 0$

$$\mathsf{E}_{x_N}\tau_a 1(\bar{A}_N) \leq CN\mathsf{P}_{x_N}(\bar{A}_N).$$

By Lemma 3.1, there exist $C_1, C_2, \delta > 0$ such that

$$\mathsf{P}_{x_N}\{\tau > C_1 N\} < C_2 e^{-\delta N}.$$

Denote the event $\{\tau \leq C_1 N\}$ by B_N. Then

$$\begin{aligned} \mathsf{P}_{x_N}(\bar{A}_N) &= \mathsf{P}_{x_N}(\bar{A}_N \cap \bar{B}_N) + \mathsf{P}_{x_N}(\bar{A}_N \cap B_N) \\ &\leq C_2 e^{-\delta N} + \mathsf{P}_{x_N}(\bar{A}_N \cap B_N), \end{aligned}$$

and

$$\begin{aligned} \mathsf{P}_{x_N}(\bar{A}_N \cap B_N) \leq\ & \mathsf{P}_{x_N}\{B_N, \tau_{NI_0} > \tau_{NI_1}\} \\ & + \mathsf{P}_{x_N}\{B_N, \tau_{NI_0} < \infty, \tau_{NI_0} \leq \tau_{NI_1}, \tau_{NJ_0} > \tau_{NJ_1}\}. \end{aligned} \tag{3.11}$$

Obviously $\tau \geq \min(\tau_{NI_0}, \tau_{NI_1})$ and so

$$\mathsf{P}_{x_N}\{B_N, \tau_{NI_0} > \tau_{NI_1}\} \leq \mathsf{P}_{x_N}\{\tau_{NI_0} > \tau_{NI_1}, \tau_{NI_1} \leq C_1 N\}.$$

Let $\tilde{\xi}_t$ be the homogeneous random walk in $\mathbf{Z}^2$ with transition probabilities $p_{a_1}(x,y)$. Define $\tilde{\tau}_U$ for the process $\tilde{\xi}_t$ in the same way, as we defined τ_U for the process ξ_t. It is clear that

$$\begin{aligned}
&\mathsf{P}_{x_N}\{B_N, \tau_{NI_0} > \tau_{NI_1}\} \\
&\quad\leq \mathsf{P}_{x_N}\{\tau_{NI_0} > \tau_{NI_1}, \tau_{NI_1} \leq C_1 N\} \\
&\quad\leq \mathsf{P}_{x_N}\{\tilde{\tau}_{NI_1} \leq C_1 N\} \\
&\quad\leq 2\mathsf{P}_{x_N}\{\text{there exists } t \leq C_1 N : y_2(\tilde{\xi}_t) - y_2(\tilde{\xi}_0) > \epsilon N\ \} \qquad (3.12)\\
&\qquad + \mathsf{P}_{x_N}\{\text{there exists } t \leq C_1 N : y_1(\tilde{\xi}_t) - y_1(\tilde{\xi}_0) > \epsilon N\}. \qquad (3.13)
\end{aligned}$$

But $S_t = y_2(\tilde{\xi}_t) - y_2(\tilde{\xi}_0)$ is a martingale with bounded jumps, $S_0 = 0$. Therefore, the probabilities in 3.12 are exponentially decreasing in N. Indeed, by virtue of Theorem 3.1 there exist $C, \delta > 0$, such that

$$\mathsf{P}_0\Big\{S_t > \frac{\epsilon}{C_1} t\Big\} < Ce^{-\delta t}.$$

Hence

$$\begin{aligned}
&\mathsf{P}_{x_N}\{\text{there exists } t \leq C_1 N : y_2(\tilde{\xi}_t) - y_2(\tilde{\xi}_0) > \epsilon N\} \\
&\quad= \mathsf{P}_0\{\text{there exists } t \leq C_1 N : S_t > \epsilon N\} \\
&\quad\leq \sum_{t \geq \epsilon N/\tilde{d}}^{C_1 N} \mathsf{P}_0\Big\{S_t > \frac{\epsilon}{C_1} t\Big\} \leq C_1 N C e^{-N\delta/\tilde{d}},
\end{aligned}$$

where $\tilde{d}$ is the maximum jumpsize of S_t. In the same way we can estimate (3.13).

Let us consider the second term in (3.11). If $\tau_{NI_0} < \infty$, then

$$\tau \geq \min(\tau_{NJ_0}, \tau_{NJ_1}).$$

We obtain that

$$\begin{aligned}
&\mathsf{P}_{x_N}\{B_N, \tau_{NI_0} < \infty, \tau_{NI_0} \leq \tau_{NI_1}, \tau_{NJ_0} > \tau_{NJ_1}\} \\
&\quad\leq \mathsf{P}_{x_N}\{\tau_{NI_0} < \infty, \tau_{NI_0} \leq \tau_{NI_1}, \tau_{NJ_0} > \tau_{NJ_1}, \tau_{NJ_1} \leq C_1 N\} \\
&\quad\leq \max_{\substack{z \in \mathbf{Z}^2, \\ \operatorname{dist}(z, NI_0) \leq d}} \mathsf{P}_z\{\tau_{NJ_0} > \tau_{NJ_1}, \tau_{NJ_1} \leq C_1 N\}.
\end{aligned}$$

A trajectory of ξ_t can cross NJ_1 on the "left side", on the "right side" and on the "top", i.e.

$$\begin{aligned}
&\mathsf{P}_z\{\tau_{NJ_0} > \tau_{NJ_1}, \tau_{NJ_1} \leq C_1 N\} \\
&\quad= \mathsf{P}_z\{\tau_{NJ_0} > \tau_{NJ_1}, \tau_{NJ_1} \leq C_1 N, y_1(\xi_{\tau_{NJ_1}}) < -\epsilon N\} \\
&\qquad + \mathsf{P}_z\{\tau_{NJ_0} > \tau_{NJ_1}, \tau_{NJ_1} \leq C_1 N, y_1(\xi_{\tau_{NJ_1}}) > \epsilon N\} \\
&\qquad + \mathsf{P}_z\{\tau_{NJ_0} > \tau_{NJ_1}, \tau_{NJ_1} \leq C_1 N, y_2(\xi_{\tau_{NJ_1}}) > y_2(x_N) + 2\epsilon N\}.
\end{aligned}$$

The second and the third terms can be estimated in the same way as in (3.12). Let us consider the first term. This term should be also exponentially small in N, because in order to cross the "left side", the trajectory should pass through a region where the drift in the direction of the "left side" is zero. We will treat this more accurately. To this end, note that the trajectory consists of a number of parts contained in U_2 or in $U_1 \cup Y_2$. To obtain the exponential estimate, we split up the trajectory into two pieces, where the second piece is the last part of the trajectory that is completely contained in U_2 (and starts close to the axis $Y_2 = \{z : y_1(z) = 0\}$). This yields

$$\begin{aligned}
&\mathsf{P}_z\{\tau_{NJ_0} > \tau_{NJ_1}, \tau_{NJ_1} \le C_1 N, y_1(\xi_{\tau_{NJ_1}}) < -\epsilon N\} \\
&\le \sum_{\substack{w \in \mathrm{Z}^2 \cap U_2, \\ \mathrm{dist}(w, Y_2) \le d}} \sum_{k=0}^{[C_1 N]} \mathsf{P}_z\{\xi_{t-k} = w\} \mathsf{P}_w\{\xi_s \in U_2, \text{ for } s \le \tau_{NJ_1}, \tau_{NJ_1} \le k, \tau_{NJ_0} > \tau_{NJ_1}\} \\
&\le C_3 N^2 \max_{\substack{w \in \mathrm{Z}^2 \cap U_2, \\ \mathrm{dist}(w, Y_2) \le d}} \mathsf{P}_w\{\xi_s \in U_2, \text{ for } s \le \tau_{NJ_1}, \tau_{NJ_1} \le k, \tau_{NJ_0} > \tau_{NJ_1}\}.
\end{aligned}$$

The latter probability is determined by the behaviour of a homogeneous random walk. Hence, by Theorem 3.1 it can be estimated in the usual way (see the proof of (3.12)).

Finally, we get the existence of $\tilde{C} > 0$ (see (3.10)) such that for any $\epsilon > 0$

$$\limsup_{N \to \infty} \frac{1}{N} \mathsf{E}_{x_N} \tau_a \le \tilde{C} \epsilon.$$

This proves (3.9) and so the proof of step 1 is complete. □

3.2. Step 2

This part is rather easy to show, due to the fact that the immediate costs are non-negative. Let $\mathcal{L}_{\pi_{\mathrm{opt}}}$ be the Markov chain induced by the optimal policy π_{opt}. As above, let τ_a be the number of times that ξ_t used action $a \in \mathcal{A}$, i.e.

$$\tau_a = \#\{t : \pi_{\mathrm{opt}}(\xi_t) = a\}, a \in \mathcal{A}.$$

Obviously,

$$C_{\pi_{\mathrm{opt}}}(x_N) = \sum_{a \in \mathcal{A}} c(a) \mathsf{E}_{x_N} \tau_a$$

and (3.8) holds as well.

By the definition of the optimal policy we have

$$C_{\pi_{\mathrm{opt}}}(x_N) \le C_{\pi_{LP}}(x_N).$$

From step 1 we know that $C_{\pi_{LP}}(x_N) < CN$ for some $C > 0$ and all N. Hence, for all $a \in \mathcal{A}$

$$\mathsf{E}_{x_N} \tau_a < \frac{C}{\min_{a \in \mathcal{A}} c(a)} N.$$

So we can choose a subsequence N_k, such that for any $a \in A$ the following limits exist

$$\lim_{k \to \infty} \frac{\mathsf{E}_{x_{N_k}} \tau_a}{N_k} = \tilde{t}_a(x)$$

and

$$\sum_{a \in \mathcal{A}} c(a) \tilde{t}_a(x) = \lim_{k \to \infty} \frac{C_{\pi_{\text{opt}}}(x_{N_k})}{N_k} = \liminf_{N \to \infty} \frac{C_{\pi_{\text{opt}}}(x_N)}{N}.$$

From (3.8) we get that $\{t_a(x)\}_{a \in \mathcal{A}}$ satisfy (3.1). So (2.3) follows immediately. This completes the proof of Theorem 2.1. □

3.3. Other policies

Denote the optimal directions, i.e. the directions of the vectors v_k, by O_k (see Figure 5).

Let R_k be some ray between $O_{k \ominus 1}$ and O_k. In the same way as before, we define the regions U_k through the rays R_k. Consider the partially homogeneous random walk on $\mathbf{Z}^2$ induced by the policy $\pi(R_1, \ldots, R_n)$ defined by

$$\pi(R_1, \ldots, R_N)(x) = a_k, \quad x \in [R_k, R_{k \oplus 1}).$$

The "induced chain" perpendicular to the axis R_k (it can be defined in a standard way, see [3]) is ergodic and it is also clear that along R_k one is forced to go linearly to 0.

Theorem 3.2. *If the U_k are not degenerate, i.e. $U_k \neq O_k$, then*

$$\lim_{N \to \infty} \frac{C_{\pi(R_1, \ldots, R_n)}(xN)}{N} = C_{\text{opt}}(x).$$

In other words, the terms of order N in $C_{\pi(R_1, \ldots, R_n)}(x_N)$ do not depend on $\{R_k\}_{k=1}^n$. The case we considered before is a special version of this construction with $R_k = O_k$. The general case can be proved in the same way. The reader can repeat the with some small modifications.

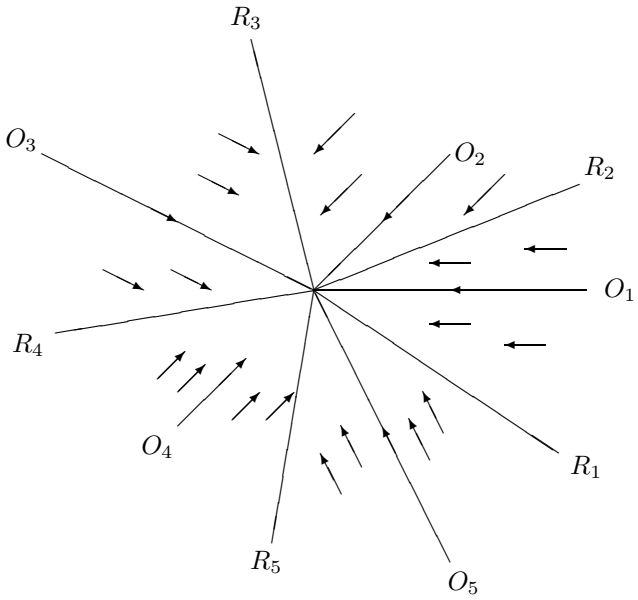

Figure 5.

4. More Problems

Smooth Cost Functions. Assume that in our case we would like to delete the homogeneity assumption on the cost function. The simplest case then is a "smooth" one: for some smooth, bounded functions $f_a(x) : \mathbf{R}^2 \to \mathbf{R}$ we have cost functions

$$c^N(n,a) = f_a\Big(\frac{n}{N}\Big).$$

The same smoothness we assume for transition probabilities but in fact we need it only for mean jumps

$$v^N(n,a) = v_a\Big(\frac{n}{N}\Big)$$

for some smooth bounded functions $v_a(x) : \mathbf{R}^2 \to \mathbf{R}$. Then the LLN takes place for fixed strategy $\pi_a(x) \equiv a$, i.e. we move along the dynamical system

$$\frac{ds(t)}{dt} = v_a(s(t))$$

where

$$s(t) = \lim_{N\to\infty} \frac{\xi^N(t)}{N}.$$

We conjecture that the following special strategy that we will call *piecewise fluid strategy*, will give an optimal solution: in the large N limit, the time interval (starting from the initial point till 0 is hit) will be subdivided in a finite number of subintervals and in each subinterval a fixed strategy π_a should be used. These asymptotically optimal solutions, which do not need to be unique, should obey some kind of deterministic Hamilton – Jacobi – Bellman equation.

Discontinuous Problems. Assume now that we are in $\mathbf{Z}_+^2$ instead of $\mathbf{Z}^2$. The transition probabilities are assumed to be maximally homogeneous for each action a (i.e. homogeneous in the interior, and on each of the axes). That is, we have a family of random walks in the quarter plane, see [3] for the theory of such random walks. Also the cost functions are assumed to be maximally homogeneous. As far as we know, such strong discontinuities were never considered in the literature. It seems plausible that the methods of [3] can be also used in such type of Markov decision chains.

We present here some ideas how to do this. First, we conjecture that the asymptotically optimal "fluid" path should again be piecewise linear.

- We should calculate asymptotically the optimal cost of reaching the boundary for the first time in *some fixed point*, starting from some point $[rN]$, with both coordinates of r being positive. For example, in the point xN on the x-axis. This is can be done in the same way as in this paper. Denote this optimal cost by $D(x)$.

- Next we should determine the movement along the boundary to 0. Here we consider strategies to reach 0 that are stationary in the x-direction. This means that $a = a(y)$ depends only on the y-coordinate of the current point. We get an induced chain and everything is similar to the one-dimensional case. Denote this optimal cost by $D_1(x)$.

- Then minimise $D(x) + D_1(x)$. This will give the still unknown point x. It seems that the problem of finding the best x is essentially a two-stage shortest path problem.

- There are of course many technical points; for example, to prove that we cannot move from one axis to the other, etc.

Diffusion terms. For the case we discussed in this paper, now assume that we would like to get more a exact approximation. What is the next term of the asymptotic expansion? Is it of order $\sqrt{N}$? Is it related to the central limit theorem along one of the asymptotically optimal paths?

Large deviations. If in our case Assumption 3 does not hold, then it can appear that we can only reach 0 along some large deviation path with expo-

nentially small probability. None did this for Markov decision chains, but there exists some techniques which seems to be helpful in this problem: [1, 2, 6, 10].

5. Applications

Suppose that in a flexible manufacturing system two different types of products can be produced. With probability $p_a(i)$ there is a successful production of i_1 items of product 1 and i_2 items of product 2 per time unit, by adjustment or action a of the machine park. Here $i = (i_1, i_2)$ and we assume that at most one item of each product per unit time can be produced (i.e. $i_k = 0, 1$). The production cost under adjustment a is $c(a)$ per time unit. The probability of a demand of i_1 items of product 1 and i_2 items of product 2 per time unit is $q(i)$, and we again assume that the demand for each product per unit time can be at most 1 (i.e. $i_k = 0, 1$). Shortage of demands is back-logged; assuming that the holding and shortage costs are small with respect to the production costs, they can be neglected. The states of the flexible manufacturing system are then given by $i = (i_1, i_2)$, where i_k is the number of items of product k in stock. A negative stock corresponds to backlog of the corresponding product.

It is easily seen, that if we always use action a, the induced process is a random walk with transition probabilities: $p_a(x, y) = \sum_{x+i-l=y} p_a(i)q(l)$. We are interested in stabilising the production system, especially for those cases in which there is a large backlog or inventory of one or both of the products. Hence, we are interested in the minimal total expected cost until the system hits a certain bounded set, for example the set consisting of the empty state only.

The optimal control problem for starting state $i = (i_1, i_2)$, corresponding to this flexible manufacturing model can be found via an extreme optimal solution x^* of the following CLP:

$$\min \left\{ c(a)x_{ia} \;\middle|\; \begin{array}{rcll} \sum_{i,a} \big(\delta_{ij} - p_a(i,j)\big)x_{ia} & = & \beta_j, & \text{for all } j \\ x_{ia} & \geq & 0, & \text{for all } a, i \end{array} \right\},$$

with β_j equals 1 or 0 if $j = i$ or $j \neq i$ respectively. The variables x_{ia} of the CLP denote the expected number of visits to the state-action pair (i, a), for all pairs (i, a) in the state-action space. Hence the dimension of the LP is the number of states times the number of actions, which in our model is infinite.

We get an optimal policy by taking in state i an action a_i such that $x^*_{ia_i} > 0$ (see [5] for more results and details on the CLP for a finite state space). This approach determines the optimal policy in all states. In this sense it is the best that we can get. However, since the LP has infinite dimension, it is only solvable for very specific models. Moreover, the optimal policy is too complicated to implement in practice.

Solving the simple FLP, as has been formulated in Section 2, we find a policy, which we called π_{LP}, having a very simple cone-shaped structure (see Figure 1).

Theorem 2.1 also implies that the optimal cost $C([x_N])$ and the cost $C_{\pi_{LP}}(x_N)$ under the simple policy π_{LP} both equal c^*N plus a small order term of N in the large N limit, where c^* is the optimal value of the LP-problem.

This implies, that the value function $C(x)$ is asymptotically linear in x, and moreover, that the difference between the optimal value and that of the simple policy π_{LP} is of small order of x, for x tending to ∞. This asymptotically optimal policy is more useful in practice, because of its simple structure.

We computed the optimal control of the flexible manufacturing model for several parameters and, remarkably, it turned out that the optimal policies have this cone-shaped structure in the whole state space (note that since the state space is a discrete lattice, we cannot have an "exact" cone-shaped structure). This seems mainly due to our boundedness assumptions on the demand and production, causing that under any action there is only "nearest neighbour interaction" in the corresponding random walks.

References

[1] V.M. Blinovskii and R.L. Dobrushin (1994) Process level large deviations for a class of piecewise homogeneous random walks. In: *The Dynkin Festschrift* in celebration of E.B. Dynkin's 70th birthday, *Progress in Probability*, **34**, Birkhäuser, 1–59.

[2] P. Dupuis, H. Ishii and H. Mete Soner (1990) A viscosity solution approach to the asymptotic analysis of queueing systems. *Ann. Prob.* **18**, 226–255.

[3] G. Fayolle, V.A. Malyshev and M.V. Menshikov (1995) *Topics in Constructive Theory of Countable Markov Chains.* Cambridge University Press.

[4] S.G. Foss and A.N. Rybko (1996) Stability of multiclass Jackson-type networks. *Markov Processes Relat. Fields* **2**, 461–486.

[5] A. Hordijk and L.C.M. Kallenberg (1984) Transient policies in discrete dynamic programming: linear programming including suboptimality tests and additional constraints. *Math. Programm.* **30**, 46–70.

[6] I.A. Ignatyuk, V.A. Malyshev and V.V. Scherbakov (1994) Boundary effects in large deviation problems. *Usp. Mat. Nauk* **49** (2), 43–102. (English translation in Russian Math. Surveys).

[7] H. Kesten and F. Spitzer (1975) Controlled Markov Chains. *Ann. Prob.* **3**, 32–40.

[8] V.A. Malyshev (1993) Networks and dynamical systems. *Adv. Appl. Probab.* **25**, 140–175.

[9] A.V. Filin, V.A. Malyshev and A.D. Manita (1997) Probabilistic models for computer architectures. *Fund. Appl. Math.* **3** (1), 263–301.

[10] A.A. Pukhalskii (1993) On the theory of large deviations. *Theory Probab. and Appl.* **38**, 490–497.

[11] M.L. Puterman (1994) *Markov Decisions Chains.* J. Wiley, New York.

[12] Sh.M. Ross (1970) *Applied Probability Models and Optimization Applications*. Holden-Day, San Francisco.

[13] R.E. Strauch (1966) Negative dynamic programming. *Ann. Math. Stat.* **37**, 871–890.

[14] G. Weiss (1995) Optimal draining of a re-entrant fluid lines. In: *Stochhastic Networks,* IMA Volumes in Mathematics and its Applications, F. Kelly and R.Williams (eds.), Springer-Verlag, New York, 93–105.