10.2478/v10170-011-0039-6

Archives of Control Sciences
Volume 23(LIX), 2013
No. 1, *pages 5–18*

Discrete homing problems

MARIO LEFEBVRE and MOUSSA KOUNTA

We consider the so-called *homing problem* for discrete-time Markov chains. The aim is to optimally control the Markov chain until it hits a given boundary. Depending on a parameter in the cost function, the optimizer either wants to maximize or minimize the time spent by the controlled process in the continuation region. Particular problems are considered and solved explicitly. Both the optimal control and the value function are obtained.

Key words: discrete-time Markov chains, optimal control, principle of optimality, absorption problems

1. Introduction

Let $\{X(t), t \geqslant 0\}$ be a one-dimensional controlled diffusion process defined by the stochastic differential equation

$$dX(t) = m[X(t)]\,dt + b[X(t)]\,u[X(t)]\,dt + \{v[X(t)]\}^{1/2}\,dB(t),$$

where $u(\cdot)$ is the control variable, $m(\cdot)$, $b(\cdot) \neq 0$ and $v(\cdot) > 0$ are real functions, and $\{B(t), t \geqslant 0\}$ is a standard Brownian motion.

The problem of finding the control u^* that minimizes the expected value of the cost function

$$C(x) = \int_0^{T_d(x)} \left\{ \frac{1}{2} q[X(t)]\,u^2[X(t)] + \lambda \right\} dt, \tag{1}$$

where $q(\cdot)$ is strictly positive, $\lambda \neq 0$ is a constant, and $T_d(x)$ is the first-passage time defined by

$$T_d(x) = \inf\{t > 0 : X(t) = d \text{ or } -d \mid X(0) = x\},$$

where $x \in (-d, d)$, is a particular case of what has been termed *LQG homing* by Whittle (1982, p. 289).

M. Lefebvre, the corresponding author, is with Département de mathématiques et de génie industriel, École Polytechnique, C.P. 6079, Succursale Centre-ville, Montréal, Québec, Canada H3C 3A7; e-mail: mlefebvre@polymtl.ca. M. Kounta is with Département de mathématiques et de statistique, Université de Montréal, C.P. 6128, Succursale Centre-ville, Montréal, Québec, Canada H3C 3J7; e-mail: kounta@DMS.UMontreal.CA.

Received 5.02.2013.

In the general formulation, $\{\mathbf{X}(t), t \geqslant 0\}$ is an n-dimensional process that obeys the first-order stochastic differential equation

$$\dot{\mathbf{X}}(t) = \mathbf{A}[\mathbf{X}(t), t] + \mathbf{B}[\mathbf{X}(t), t]\,\mathbf{u}[X(t)] + \varepsilon(t),$$

where $\varepsilon(t)$ is a vector white noise having zero mean and power matrix $\mathbf{N}[\mathbf{X}(t), t]$. Moreover, the cost function is

$$\mathbb{C}(x,t) = \int_t^T \left\{ \frac{1}{2}\mathbf{u}'\mathbf{Q}[\mathbf{X}(\tau), \tau]\,\mathbf{u} + g[\mathbf{X}(\tau), \tau] \right\} d\tau + \mathbb{K}(\mathbf{X}(T), T),$$

where T is the first time $(\mathbf{X}(t), t)$, starting from $\mathbf{X}(0) = x$, enters a prescribed stopping set D, and $\mathbb{K}$ is a general termination cost function.

In Whittle (1990, p. 222) (see also Kuhn (1985)), the homing problem is given a risk-sensitive formulation. That is, the optimizer wants to minimize the expected value of

$$\mathbb{C}_\theta(x,t) := -\frac{2}{\theta} \log\left(E\left[e^{-\theta \mathbb{C}(x,t)/2}\right]\right).$$

The case when $\theta = 0$ corresponds to the risk-neutral criterion $\mathbb{C}(x,t)$ defined above.

When the parameter λ in (1) is positive (respectively, negative), the optimizer wants to minimize (respectively, maximize) the survival time of the controlled process in the interval $(-d, d)$, taking the quadratic control costs into account.

We can take λ as large as we want. However, in general, λ cannot take any negative value. If the absolute value of λ (< 0) becomes too large, then the expected reward becomes infinite.

Whittle has shown that it is sometimes possible to obtain the optimal control u^* by considering the uncontrolled process $\{\xi(t), t \geqslant 0\}$ that corresponds to $\{X(t), t \geqslant 0\}$. Indeed, if the relation

$$\alpha v[X(t)] = \frac{b^2[X(t)]}{q[X(t)]}$$

holds for some positive constant α, and if $P[\tau_d(x) < \infty] = 1$, where τ_d is the same as $T_d(x)$, but for the uncontrolled process $\{\xi(t), t \geqslant 0\}$, then the value function

$$G(x) := \inf_{u[X(t)],\, 0 \leqslant t \leqslant T_d(x)} E[C(x)]$$

can be expressed in terms of a mathematical expectation for $\{\xi(t), t \geqslant 0\}$. Moreover, the optimal control is given by

$$u^*(x) = -\frac{b(x)}{q(x)} G'(x).$$

Whittle's result was used, and generalized, in a number of papers by the first author (see, for instance, Lefebvre (2011)), and by Makasu (2009).

As Whittle pointed out, it is not possible to extend his result to the case of discrete-time and discrete-space controlled processes. However, we can nevertheless consider homing problems for Markov chains. Here, we will set up and solve explicitly such problems for various random walks.

Let $\{X_n, n = 0, 1, \ldots\}$ be a controlled Markov chain, starting at $X_0 = x$, defined by

$$X_{n+1} = X_n + u_n + \varepsilon_n, \tag{2}$$

where u_n can take a finite number of values, and ε_n is the noise term that can take either of two values with probability 1/2.

Our aim will be to minimize the expected value of the cost function

$$J(x) = \sum_{n=0}^{T(x)-1} (u_n^2 + \lambda), \tag{3}$$

where $T(x)$ is a first-passage time defined with respect to the controlled Markov chain.

Three particular problems will be solved explicitly in Sections 2-4. Finally, we will make a few concluding remarks, including possible extensions, in Section 5.

2. Maximizing the survival time

The first problem that we consider is the one for which the state space of the Markov chain $\{X_n, n = 0, 1, \ldots\}$ is the set $\{-k, \ldots, 0, \ldots, k\}$, where $k \in \mathbb{N}$. Moreover, the control u_n must belong to $\{-1, 0, 1\}$, and $\varepsilon_n = +1$ or -1 with probability 1/2. Finally, the parameter λ in the cost function (3) is assumed to be negative, and

$$T(x) := \min\{n \geqslant 0 : |X_n| \geqslant k \mid X_0 = x\}. \tag{4}$$

Hence, the optimizer wants to maximize the survival time of the controlled Markov chain in $C := \{-k+1, \ldots, 0, \ldots, k-1\}$.

Let $F(x)$ be the value function defined by

$$F(x) = \min_{u_n,\, n=0,\ldots,T(x)-1} E[J(x)]. \tag{5}$$

By making use of the principle of optimality, we obtain the following lemma.

Lemma 1 *The value function $F(x)$ defined in Eq.* (5) *satisfies the dynamic programming equation*

$$F(x) = \min_{u_0} \left\{ u_0^2 + \lambda + \frac{1}{2}\left[F(x + u_0 - 1) + F(x + u_0 + 1)\right] \right\}, \tag{6}$$

subject to the boundary condition

$$F(x) = 0 \quad \text{if } |x| \geqslant k. \tag{7}$$

Now, by symmetry, we can state that $u_0^*(-x) = -u_0^*(x)$. Therefore, we can assume that $x \in \{0, \ldots, k\}$. Furthermore, because we want the controlled process to remain in the continuation region as long as possible, taking the quadratic control costs into account, it is clear that $u_0^*(x)$ should be either equal to -1 or 0 if x is positive.

Next, notice that the value function $F(x)$ is negative for any $x \in C$. Indeed, if we choose $u_n \equiv 0$, then we have

$$E[J(x)] = E\left[\sum_{n=0}^{T(x)-1} \lambda\right] = \lambda E[T(x)] < 0.$$

Actually, using the well-known results on the *gambler's ruin problem*, we can write that if $u_n \equiv 0$, then $P[T(x) < \infty] = 1$ and (see, for instance, Feller (1968, p. 348))

$$E[T(x)] = k^2 - x^2 \quad \text{for } x = -k, \ldots, k.$$

Thus,

$$F(x) \leqslant \lambda(k^2 - x^2).$$

Not only is the value function negative, it is equal to $-\infty$ (that is, the expected reward becomes infinite) if the absolute value of λ is too large, as we will prove in the following lemma.

Lemma 2 *If the parameter λ in the cost function* (3) *is smaller than -1, then we have $F(x) = -\infty$.*

Proof. When $\lambda < -1$, we receive a reward for surviving in the continuation region C, even if we take $u_n = \pm 1$. But, because $\varepsilon_n = \pm 1$, by choosing $u_0 = -1$ (respectively, $+1$) when $X_0 = x \in C$ is positive (respectively, negative), we can remain in C as long as we want. It follows that $F(x) = -\infty$. $\square$

We will henceforth assume that $\lambda \in [-1, 0)$. Then the value function is necessarily finite, because there is a positive cost when we control the Markov chain, and if we choose $u_n \equiv 0$, then the process will hit either boundary in finite time.

Now, whatever the value of λ (< 0), we can state that $u_0^*(0) = 0$. Indeed, the ideal position to maximize the profit is when the random walk is at the origin, so that the optimizer should try to remain near there. Since $E[X_1] = 0$ if we take $u_0(0) = 0$, this is surely the best option to choose.

Therefore, we deduce from the dynamic programming equation (6) that

$$F(0) = \lambda + \frac{1}{2}\left[F(-1) + F(1)\right].$$

Moreover, by symmetry, $F(-x) = F(x)$ for any $x \in C$. It follows that

$$F(1) = F(0) - \lambda.$$

Next, we can write that

$$F(1) = \min_{u_0} \left\{ u_0^2 + \lambda + \frac{1}{2} [F(u_0) + F(u_0 + 2)] \right\}.$$

Thus, we have either (with $u_0 = 0$)

$$F(1) = \lambda + \frac{1}{2} [F(0) + F(2)] \tag{8}$$

or (with $u_0 = -1$)

$$F(1) = 1 + \lambda + \frac{1}{2} [F(-1) + F(1)].$$

However, this last equation leads to $0 = 1 + \lambda$, which contradicts the hypothesis that we made above (namely, $\lambda \in [-1, 0)$), unless $\lambda = -1$. It follows that, when $\lambda \in (-1, 0)$, we must choose $u_0^*(1) = 0$. Hence, substituting the value of $F(1)$ in terms of $F(0)$ into Eq. (8), we obtain that

$$F(2) = F(0) - 4\lambda.$$

Proposition 1 *For the problem set up in this section, the optimal control is* $u_0^*(x) \equiv 0$ *and the value function is given by*

$$F(x) = \lambda(k^2 - x^2).$$

Moreover, if $\lambda = -1/(2x - 1)$ *(with* $x > 0$*), then we can choose* $u_0^*(x) = 0$ *or* -1 *indifferently.*

Proof. We will prove, by induction, that

$$F(x) = F(0) - x^2\lambda \quad \text{for } x \in \{-k, \ldots, k\}.$$

Indeed, we deduce from the dynamic programming equation (6) that, if $x \in C$, then we have

$$F(x) = \lambda + \frac{1}{2} [F(x-1) + F(x+1)] \quad \text{(with } u_0 = 0) \tag{9}$$

or

$$F(x) = 1 + \lambda + \frac{1}{2} [F(x-2) + F(x)] \quad \text{(with } u_0 = -1). \tag{10}$$

This last equation can be rewritten as

$$F(0) - x^2\lambda - [F(0) - (x-2)^2\lambda] = 2(1 + \lambda),$$

which is only possible if x is such that

$$\lambda = -\frac{1}{2x - 1}.$$

For any $\lambda \in [-1,0)$ different from this value, we must take $u_0^* = 0$, and we then deduce from Eq. (9) that

$$\begin{aligned} F(x+1) &= 2[F(x) - \lambda] - F(x-1) \\ &= 2[F(0) - x^2\lambda - \lambda] - [F(0) - (x-1)^2\lambda] \\ &= F(0) - (x+1)^2\lambda. \end{aligned}$$

Since λ is a constant, we can conclude that $u_0^*(x) = 0$ for any x, except that if there exists a value of x for which $\lambda = -1/(2x-1)$, then $u_0^*(x)$ is also equal to -1.

Finally, as we mentioned above, in the uncontrolled case we know that

$$E[J(x)] = \lambda(k^2 - x^2) \quad \text{for } x = -k, \ldots, k.$$

This is actually the maximal expected reward when the process starts from $X_0 = x$. That is,

$$F(x) = \lambda(k^2 - x^2) \quad \text{for } x = -k, \ldots, k,$$

which completes the proof. □

Remark. Notice that, with $F(x) = \lambda(k^2 - x^2)$, Eq. (9) is satisfied for any $\lambda \in [-1,0)$. Moreover, both Eq. (9) and Eq. (10) are satisfied at the same time if and only if $\lambda = -1/(2x-1)$.

3. Minimizing the time spent in the continuation region

In this section, the state space of the Markov chain $\{X_n, n = 0, 1, \ldots\}$ is the set $\{0, \ldots, k\}$, the optimizer must choose $u_n = 0$ or $u_n = 1$, the random variable ε_n is equal to 0 or 1 with probability 1/2, the parameter λ in the cost function (3) is strictly positive, and the first-passage time $T(x)$ is defined by

$$T(x) := \min\{n \geqslant 0 : X_n \geqslant k \mid X_0 = x\}. \tag{11}$$

Thus, the objective of the optimizer is to minimize the time spent in the continuation region by the controlled Markov chain, taking the quadratic control costs into account.

The value function $F(x)$ defined in Eq. (5) now satisfies the dynamic programming equation

$$F(x) = \min_{u_0} \left\{ u_0^2 + \lambda + \frac{1}{2}\left[F(x+u_0) + F(x+u_0+1)\right] \right\}, \tag{12}$$

subject to the boundary condition $F(x) = 0$ if $x \geqslant k$. Since $u_0 = 0$ or $u_0 = 1$, this equation can be rewritten as follows:

$$F(x) = \min\left\{ \lambda + \frac{1}{2}\left[F(x) + F(x+1)\right], 1 + \lambda + \frac{1}{2}\left[F(x+1) + F(x+2)\right] \right\}. \tag{13}$$

To solve our problem, we will first consider the case when $x = k-1$. Then, Eq. (13) becomes

$$\begin{aligned} F(k-1) &= \min\left\{\lambda + \frac{1}{2}[F(k-1)+F(k)], 1+\lambda+\frac{1}{2}[F(k)+F(k+1)]\right\} \\ &= \min\left\{\lambda+\frac{1}{2}F(k-1), 1+\lambda\right\}. \end{aligned}$$

Therefore, if the optimizer chooses $u_0 = 0$, then $F(k-1) = 2\lambda$, whereas $F(k-1) = 1+\lambda$ if $u_0 = 1$ is chosen. It follows that

$$u_0^*(k-1) = \begin{cases} 0 & \text{if } \lambda \leqslant 1, \\ 1 & \text{if } \lambda > 1. \end{cases} \tag{14}$$

Next, when $x = k-2$, we have

$$F(k-2) = \min\left\{\lambda+\frac{1}{2}[F(k-2)+F(k-1)], 1+\lambda+\frac{1}{2}F(k-1)\right\}.$$

If $0 < \lambda \leqslant 1$, then $F(k-1) = 2\lambda$, whereas $F(k-1) = 1+\lambda$ if $\lambda > 1$. Therefore, we must consider two cases: firstly, if $0 < \lambda \leqslant 1$, we can write that

$$F(k-2) = \min\left\{2\lambda+\frac{1}{2}F(k-2), 1+2\lambda\right\}.$$

Hence, if the optimizer chooses $u_0 = 0$ (respectively, $u_0 = 1$), the value function $F(k-2)$ is equal to 4λ (respectively, $1+2\lambda$). So, we conclude that $u_0^*(k-2) = 0$ if $0 < \lambda \leqslant 1/2$, and $u_0^*(k-2) = 1$ if $1/2 < \lambda \leqslant 1$.

Secondly, in the case when $\lambda > 1$, we have

$$F(k-2) = \min\left\{\lambda+\frac{1}{2}[F(k-2)+(1+\lambda)], 1+\lambda+\frac{1}{2}(1+\lambda)\right\}.$$

Thus, we must compare $F(k-2) = 1+3\lambda$ (with $u_0 = 0$) to $F(k-2) = \frac{3}{2}(1+\lambda)$ (with $u_0 = 1$). Since $\lambda > 1$, we find that $u_0^*(k-2) = 1$. Hence, we can write that

$$u_0^*(k-2) = \begin{cases} 0 & \text{if } \lambda \leqslant 1/2, \\ 1 & \text{if } \lambda > 1/2. \end{cases} \tag{15}$$

It turns out that this optimal value of the control variable is actually valid for any $x \in \{0,\ldots,k-2\}$, as we will prove by induction.

Proposition 2 *For the problem set up in this section, the optimal control is given by Eq.* (14) *if $x = k-1$, and by*

$$u_0^*(x) = \begin{cases} 0 & \text{if } 0 < \lambda \leqslant 1/2, \\ 1 & \text{if } \lambda > 1/2 \end{cases}$$

for $x \in \{0,\ldots,k-2\}$.

Proof. In general, to determine whether $u_0^*(x) = 0$ or 1, we must compare

$$F(x) = \lambda + \frac{1}{2}[F(x) + F(x+1)] \quad (\text{if } u_0 = 0) \tag{16}$$

to

$$F(x) = 1 + \lambda + \frac{1}{2}[F(x+1) + F(x+2)] \quad (\text{if } u_0 = 1). \tag{17}$$

First, we will show that $u_0^*(x) \equiv 0$ if $0 < \lambda \leqslant 1/2$. We already found (see Eq. (14)) that $u_0^*(k-1) = 0$ if $0 < \lambda \leqslant 1/2$. If we assume that $u_0^*(y) = 0$ for $y = x, x+1, \ldots, k-2$, then we deduce from Eq. (16) that the value function $F(x)$ satisfies the first-order difference equation

$$F(x) = 2\lambda + F(x+1).$$

We easily find that

$$F(x) = c_0 - 2\lambda x.$$

Making use of the boundary condition $F(k) = 0$, we obtain that the constant c_0 must be equal to $2\lambda k$, so that

$$F(x) = 2\lambda(k - x). \tag{18}$$

Next, with $X_0 = x - 1$, Eq. (16) and Eq. (17) would respectively become

$$F(x-1) = \lambda + \frac{1}{2}[F(x-1) + 2\lambda(k-x)] \quad \Longrightarrow \quad F(x-1) = 2\lambda(k-x+1)$$

and

$$F(x-1) = 1 + \lambda + \frac{1}{2}[2\lambda(k-x) + 2\lambda(k-x-1)] = 1 + 2\lambda(k-x).$$

We find at once that

$$2\lambda(k-x+1) \leqslant 1 + 2\lambda(k-x) \quad \Longleftrightarrow \quad \lambda \leqslant 1/2,$$

so that we can conclude that $u_0^*(x-1) = 0$, which proves, by induction, that $u_0^*(x) \equiv 0$ if $0 < \lambda \leqslant 1/2$.

Now, if we assume that $u_0^*(x) \equiv 1$, then the function $F(x)$ satisfies Eq. (17) for all values of x. The general solution of this second-order linear difference equation can be written as follows:

$$F(x) = c_0 + c_1(-2)^x - \frac{2}{3}(1+\lambda)(x+1). \tag{19}$$

The boundary condition $F(k) = 0$ yields that

$$F(x) = c_1\left[(-2)^x - (-2)^k\right] - \frac{2}{3}(1+\lambda)(x-k).$$

To determine the value of the constant c_1, we may use the fact that $F(k+1)$ is also equal to zero, which implies that

$$F(x) = -\frac{2}{9}(1+\lambda)\left[(-2)^{x-k} - 1 + 3(x-k)\right]. \tag{20}$$

Remark. When we choose $u_0(x) \equiv 1$, the controlled process can cross the boundary at $x = k$. Therefore, we must set $F(k+1)$ equal to 0. However, with $u_0(x) \equiv 0$, there is no overshoot. Actually, there is no overshoot as long as $u_0(k-1) = 0$.

If $\lambda > 1$, we know (see Eq. (14)) that $u_0^*(k-1) = 1$. If we make the hypothesis that $u_0^*(y) = 1$ for $y = x, x+1, \ldots, k-2$, then the expression in Eq. (20) for the value function $F(x)$ is valid for $y = x, x+1, \ldots, k-1$. Substituting this expression into Eq. (16) and Eq. (17), with x replaced by $x-1$, we find (after simplifying) that the value of $F(x-1)$ when $u_0(x-1) = 0$ is larger than or equal to the value of $F(x-1)$ when $u_0(x-1) = 1$ if and only if

$$\lambda - 1 \geqslant -\frac{1}{3}(1+\lambda)\left[1-(-2)^{x-k}\right].$$

Since

$$1 \geqslant (-2)^{x-k} \quad \text{for } x = k, k-1, k-2, \ldots, 0,$$

we can state that $u_0^*(x-1) = 1$. Hence, by induction, $u_0^*(x) \equiv 1$ if $\lambda > 1$.

Finally, when $1/2 < \lambda \leqslant 1$, we found (see Eq. (14)) that $u_0^*(k-1) = 0$, but (see Eq. (15)) $u_0^*(k-2) = 1$. We find that $u_0^*(k-3) = u_0^*(k-4) = 1$ as well. If we assume that $u_0^*(y) = 1$ for $y = x, x+1, \ldots, k-3$, then Eq. (20) is valid for $y = x, x+1, \ldots, k-2$. Proceeding as above, we can show that $u_0^*(x-1) = 1$. Thus, by mathematical induction, $u_0^*(x) \equiv 1$ for $x = 0, \ldots, k-2$ if $1/2 < \lambda \leqslant 1$, which completes the proof. □

Remark. We deduce from Eq. (16) and Eq. (17) that $u_0^*(x) = 1$ if and only if

$$F(x) \geqslant 2 + F(x+2).$$

Therefore, to prove that $u_0^*(x) \equiv 1$ if $\lambda > 1$, we can also use the fact that the cheapest cost to move from x to $x+2$ is either 2λ (if $u_0 = u_1 = 0$ and $\varepsilon_0 = \varepsilon_1 = 1$) or $1+\lambda$ (if $u_0 = \varepsilon_0 = 1$). Hence, $F(x)$ will indeed be greater than or equal to $2 + F(x+2)$ if $\lambda > 1$.

Corollary 1 *For the problem considered in this section, the value function is given by Eq.* (18) *if* $0 < \lambda \leqslant 1/2$, *by Eq.* (20) *if* $\lambda > 1$, *and by*

$$F(x) = \frac{1}{3}\left[2\lambda - \frac{2}{3}(1+\lambda)\right](-2)^{x+1-k} + \frac{4}{3}\lambda - \frac{2}{3}(1+\lambda)\left(x + \frac{2}{3} - k\right) \tag{21}$$

if $1/2 < \lambda \leqslant 1$.

Proof. The expressions for $F(x)$ when $0 < \lambda \leqslant 1/2$ and when $\lambda > 1$ have been derived in the proof of Proposition 2. To prove Eq. (21), we can solve Eq. (17), but subject to the boundary conditions $F(k) = 0$ and $F(k-1) = 2\lambda$ (obtained above). Making use of the general solution of Eq. (17) given in Eq. (19), we obtain the formula in Eq. (21). □

4. A problem with non-constant u_0^*

In Section 2, the optimal control $u_0^*(x)$ was identical to zero for all admissible values of the parameter λ. In Section 3, we had $u_0^*(x) \equiv 0$ if $0 < \lambda \leqslant 1/2$, and $u_0^*(x) \equiv 1$ if $\lambda > 1$. The only case for which $u_0^*(x)$ was not a constant is when $1/2 < \lambda \leqslant 1$. Then, $u_0^*(x) \equiv 1$ for $x = 0, \ldots, k-2$, but $u_0^*(k-1) = 0$. In this section, the optimal control will depend more strongly on x and λ.

Assume that the state space of the controlled Markov chain $\{X_n, n = 0, 1, \ldots\}$ is, as in Section 2, the set $\{-k, \ldots, k\}$, the control variable u_n belongs to $\{-2, -1, 1, 2\}$, the noise term ε_n is equal to ± 1 with probability 1/2, the parameter λ in Eq. (3) is strictly positive, and the first-passage time $T(x)$ is defined as in Eq. (4). Therefore, as in Section 3, the optimizer wants to minimize the time spent by the controlled Markov chain in $C = \{-k+1, \ldots, k-1\}$, with the quadratic control costs being taken into account.

The dynamic programming equation satisfied by the value function $F(x)$ defined in Eq. (5) is

$$F(x) = \min_{u_0} \left\{ u_0^2 + \lambda + \frac{1}{2} \left[F(x + u_0 - 1) + F(x + u_0 + 1) \right] \right\}. \tag{22}$$

It is subject to the boundary condition $F(x) = 0$ if $|x| \geqslant k$.

Because λ is positive, we can state, by symmetry, that $u_0^*(-x) = -u_0^*(x)$. Therefore, we can limit ourselves to the case when $x \in \{0, 1, \ldots, k\}$. Moreover, the fact that λ is positive also implies that $u_0^*(x) > 0$ when $x > 0$. When $x = 0$, the optimizer can choose $u_0^*(x)$ positive or negative indifferently. That is, the sign of $u_0^*(x)$ is irrelevant.

Since $E[T(x)]$ is finite in the uncontrolled case (see Section 2), it is intuitively clear that the optimizer should choose $u_n(x) \equiv 1$ if λ is close enough to zero and $x > 0$. Similarly, if λ is very large, compared to u_n^2, then the optimizer should choose $u_n(x) \equiv 2$ (if x is positive) to leave the continuation region as soon as possible.

The optimal control being equal to $+1$ or $+2$ when x is non-negative, the dynamic programming equation (22) can be written as follows:

$$F(x) = \min \left\{ 1 + \lambda + \frac{1}{2} \left[F(x) + F(x+2) \right], 4 + \lambda + \frac{1}{2} \left[F(x+1) + F(x+3) \right] \right\}. \tag{23}$$

We will try to determine the value of $u_0^*(x)$, starting from $x = k-1$. We have then

$$F(k-1) = 1 + \lambda + \frac{1}{2} \left[F(k-1) + F(k+1) \right] \quad (\text{if } u_0 = 1),$$

and

$$F(k-1) = 4 + \lambda + \frac{1}{2} \left[F(k) + F(k+2) \right] \quad (\text{if } u_0 = 2).$$

Hence, if $u_0 = 1$, we obtain that

$$F(k-1) = 2(1 + \lambda),$$

whereas

$$F(k-1) = 4 + \lambda$$

when $u_0 = 2$. It follows that

$$u_0^*(k-1) = \begin{cases} 1 & \text{if } \lambda \leqslant 2, \\ 2 & \text{if } \lambda > 2. \end{cases}$$

Remark. When $X_0 = k-1$, the optimizer is certain to leave the continuation region by choosing $u_0 = 2$, but not with $u_0 = 1$.

Next, when $x = k-2$, we must compare the value of $F(k-2)$ obtained from

$$F(k-2) = 1 + \lambda + \frac{1}{2}\left[F(k-2) + F(k)\right] \quad (\text{if } u_0 = 1) \tag{24}$$

to that deduced from

$$F(k-2) = 4 + \lambda + \frac{1}{2}\left[F(k-1) + F(k+1)\right] \quad (\text{if } u_0 = 2). \tag{25}$$

Equation (24) yields that

$$F(k-2) = 2(1+\lambda) \quad (\text{if } u_0 = 1).$$

In the case of Eq. (25), we can write that

$$F(k-2) = \begin{cases} 4 + \lambda + (1+\lambda) & \text{if } 0 < \lambda \leqslant 2, \\ \frac{3}{2}(4+\lambda) & \text{if } \lambda > 2 \end{cases} \qquad (\text{if } u_0 = 2).$$

Therefore, this time we conclude that

$$u_0^*(k-2) = \begin{cases} 1 & \text{if } \lambda \leqslant 8, \\ 2 & \text{if } \lambda > 8. \end{cases}$$

Remark. Notice that $F(k-2) = F(k-1)$ if $0 < \lambda \leqslant 2$.

When $x = k-3$, we have either

$$F(k-3) = 1 + \lambda + \frac{1}{2}\left[F(k-3) + F(k-1)\right] \quad (\text{if } u_0 = 1)$$

or

$$F(k-3) = 4 + \lambda + \frac{1}{2}\left[F(k-2) + F(k)\right] \quad (\text{if } u_0 = 2).$$

We must then consider three cases: $0 < \lambda \leqslant 2$, $2 < \lambda \leqslant 8$ and $\lambda > 8$. With $u_0 = 1$, we find that

$$F(k-3) = \begin{cases} 4(1+\lambda) & \text{if } 0 < \lambda \leqslant 2, \\ 2(1+\lambda) + (4+\lambda) = 6 + 3\lambda & \text{if } \lambda > 2. \end{cases}$$

When $u_0 = 2$, we obtain that

$$F(k-3) = \begin{cases} 4+\lambda+\frac{1}{2}[2(1+\lambda)] = 5+2\lambda & \text{if } 0<\lambda \leqslant 8, \\ 4+\lambda+\frac{1}{2}\left[\frac{3}{2}(4+\lambda)\right] = 7+\frac{7}{4}\lambda & \text{if } \lambda > 8. \end{cases}$$

Comparing the various expressions above, we find that

$$u_0^*(k-3) = \begin{cases} 1 & \text{if } \lambda \leqslant 1/2, \\ 2 & \text{if } \lambda > 1/2. \end{cases}$$

To obtain the value of $u_0^*(x)$ when $x = k-4$, we must now consider four cases: $0 < \lambda \leqslant 1/2$, $1/2 < \lambda \leqslant 2$, $2 < \lambda \leqslant 8$ and $\lambda > 8$. Proceeding as above, we can show that

$$u_0^*(k-4) = \begin{cases} 1 & \text{if } \lambda \leqslant 3, \\ 2 & \text{if } \lambda > 3. \end{cases}$$

For k small, we can compute the optimal control $u_0^*(x)$ for all values of $x \in \{-k+1,\dots,k-1\}$. However, from what precedes, we must conclude that it is difficult to obtain a general formula for $u_0^*(x)$.

In the case of the value function, we can at least give a difference equation that it satisfies.

Proposition 3 *The value function $F(x)$ in the problem set up in this section satisfies the non-linear third order difference equation*

$$\begin{aligned} 0 \quad = \quad & 2F^2(x) + F(x)\left[-F(x+1)+2F(x+2)-F(x+3)-12+6\lambda\right] \\ & +2(4+\lambda)F(x+2)+2(1+\lambda)\left[F(x+1)+F(x+3)\right] \\ & +F(x+1)F(x+2)+F(x+2)F(x+3) \end{aligned} \tag{26}$$

for $x = 0,1,\dots,k-1$. The boundary conditions are $F(x) = 0$ if $x = k, k+1, k+2$.

Proof. Since u_0 is equal to either 1 or 2, we can make use of the formula

$$\min\{x,y\} = \frac{1}{2}\{x+y-|x-y|\}$$

in the dynamic programming equation (22). We have

$$\begin{aligned} F(x) \quad = \quad & \frac{1}{2}\left\{\left[1+\lambda+\frac{1}{2}[F(x)+F(x+2)]\right] + \left[4+\lambda+\frac{1}{2}[F(x+1)+F(x+3)]\right]\right. \\ & \left. -\left|-3+\frac{1}{2}[F(x)+F(x+2)-F(x+1)-F(x+3)]\right|\right\}, \end{aligned}$$

which implies that

$$\begin{aligned}&\frac{3}{2}F(x)-(5+2\lambda)-\frac{1}{2}[F(x+1)+F(x+2)+F(x+3)]\\&=-\left|-3+\frac{1}{2}[F(x)+F(x+2)-F(x+1)-F(x+3)]\right|.\end{aligned}$$

Squaring each side of the previous equation and simplifying, we obtain Eq. (26). □

When $x=k-1$, Eq. (26) reduces to

$$F^2(k-1)-3(2+\lambda)F(k-1)+2(\lambda^2+5\lambda+4)=0.$$

It follows that

$$F(k-1)=4+\lambda \quad \text{or} \quad F(k-1)=2(1+\lambda),$$

so that the value function is given by $F(k-1)=2(1+\lambda)$ if $0<\lambda\leqslant 2$ and by $F(k-1)=4+\lambda$ if $\lambda>2$, as we had obtained above.

5. Concluding remarks

We have extended the homing problem proposed by Whittle to the case of controlled discrete-time and discrete-space Markov chains. In Sections 2 and 3, we were able to obtain general expressions, valid for any integer k, for both the optimal control $u_0^*(x)$ and the value function $F(x)$. Although the problems considered in these sections might seem simple, even in the continuous case homing problems that were solved explicitly so far are only in one or two dimensions, or possess some suitable symmetry property that enables one to reduce them to much simpler problems.

It is quite straightforward to see why we cannot expect to be able to derive explicit general formulas in most cases. For instance, in Section 4, we could not find a general formula for either $u_0^*(x)$ or $F(x)$. As we have shown in Proposition 4.1, to obtain the value function $F(x)$ in that section, we would have to solve a non-linear third order difference equation. Now, suppose that we complexity this problem by assuming that ε_n is equal to ± 1 or ± 2 with probability 1/4. Then, $F(x)$ is such that

$$\begin{aligned}F(x) \quad = \quad &\frac{1}{2}\left\{\left[1+\lambda+\frac{1}{4}[F(x-1)+F(x)+F(x+2)+F(x+3)]\right]\right.\\&+\left[4+\lambda+\frac{1}{4}[F(x)+F(x+1)+F(x+3)+F(x+4)]\right]\\&\left.-\left|-3+\frac{1}{4}[F(x-1)+F(x+2)-F(x+1)-F(x+4)]\right|\right\}.\end{aligned}$$

Proceeding as above, we obtain a non-linear fifth order difference equation. In general, if $\varepsilon_n\in\{\pm 1,\pm 2,\ldots,\pm m\}$ with probability $1/(2m)$, where $m\in\{3,4,\ldots\}$, then the difference equation that the value function satisfies is non-linear and of order $2m+1$. Solving such an equation is possible when the number of possible states is small, but almost surely not for a general integer k. Therefore, one must limit oneself to seemingly simple cases. In fact, what is important is the usefulness of the results, not the difficulty of the calculations involved. The problems that we were able to solve explicitly in Sections 2 and 3 could serve as models in various applications.

We could obviously generalize both Eq. (2) and Eq. (3). For instance, we could define the controlled stochastic process $\{X_n, n = 0, 1 \ldots\}$ by

$$X_{n+1} = aX_n + bu_n + \varepsilon_n,$$

where a and b are non-zero constants, and take

$$J(x) = \sum_{n=0}^{T(x)-1} (q_0 u_n^2 + \lambda),$$

where $q_0 > 0$.

For all the problems considered in this paper, the optimizer had to choose between two possible values of u_0. The optimal control problem would of course be even more complicated if there were three or more candidates for the optimal control. For example, in Section 4, if we assume that $u_n \in \{-2, -1, 0, 1, 2\}$, then $u_0^*(x)$ could be equal to 0, 1 or 2 when x is positive. In particular, $u_0^*(x)$ should be equal to 0 (and not to 1, as above) when λ is very close to zero.

Finally, we could assume that ε_n has a Gaussian distribution with mean 0 and variance σ^2, rather than being a discrete random variable. Then, the control variable u_n should also be a continuous variable.

References

[1] W. FELLER: An Introduction to Probability Theory and its Applications. Vol. I. Wiley, New York, 1968.

[2] J. KUHN: The risk-sensitive homing problem. *J. Appl. Prob.*, **22** (1985), 796-803.

[3] M. LEFEBVRE: Maximizing the mean exit time of a Brownian motion from an interval. *Int. J. Stoch. Anal.*, vol. 2011, Article ID 296259, 5 pages, 2011. doi: 10.1155/2011/296259

[4] C. MAKASU: Risk-sensitive control for a class of homing problems. *Automatica J. IFAC*, **45** (2009), 2454-2455.

[5] P. WHITTLE: Optimization over Time. Vol. I. Wiley, Chichester, 1982.

[6] P. WHITTLE: Risk-sensitive Optimal Control. Wiley, Chichester, 1990.