Optimal decision rules for marked point process models

M.N.M. van Lieshout

CWI, P.O. Box 94079, NL-1090 GB Amsterdam, The Netherlands

Department of Applied Mathematics, University of Twente, P.O. Box 217
NL-7500 AE Enschede, The Netherlands

Abstract

We study a Markov decision problem in which the state space is the set of finite marked point configurations in the plane, the actions represent thinnings, the reward is proportional to the mark sum which is discounted over time, and the transitions are governed by a birth-death-growth process. We show that thinning points with large marks is optimal when births follow a Poisson process and marks grow logistically. Explicit values for the thinning threshold and the discounted total expected reward over finite and infinite horizons are also provided. When the points are required to respect a hard core distance, upper and lower bounds on the discounted total expected reward are derived.

2020 Mathematics Subject Classification: 60G55, 90C40.

1 Introduction

The classic Markov decision process [3, 8, 22] on a finite state space $\mathcal{X}$ and action set A is defined as follows. Write $A(x)$ for the subset of A which contains all actions that may be taken in state $x \in \mathcal{X}$. Then, a policy ϕ is a procedure for the selection of an action at each decision epoch $i \in \mathbb{N}_0$. Such a policy could be random or deterministic, and in principle take into account the entire history of the process. Often though, one may restrict attention to the class of deterministic Markov policies. Such a policy $\phi = (\phi_i)_{i=0}^{\infty}$ is a sequence of mappings $\phi_i : \mathcal{X} \to A$ that, at time i, assign an action $a = \phi_i(x) \in A(x)$ to the current state x. In doing so, a direct reward $r(x, a)$ is earned and a probability mass function $p(\cdot|x, a)$ on $\mathcal{X}$ governs the next state of the process. Being Markovian, only the current state and action are important; the past history is irrelevant. A policy $(\phi_i)_i$ is said to be stationary if its members ϕ_i do not depend on the time i.

Let (X_i, Y_i) denote the stochastic process of states X_i and actions Y_i. Write $\mathbb{E}_x^{\phi}$ for its expectation when the initial state $X_0 = x$ and the transitions are driven by policy ϕ. Then an optimal policy maximises the discounted total expected reward

$$v_{\alpha}^{\phi}(x) = \mathbb{E}_x^{\phi}\left[\sum_{i=0}^{\infty} \alpha^i r(X_i, Y_i)\right], \tag{1}$$

$0 \leq \alpha < 1$. The reward function is usually assumed to be bounded, in which case (1) is well-defined. When the state and action spaces are both finite, it is well known [22, Theorem 5.5.3b] that it suffices to consider only Markov policies and, by [22, Theorem 6.2.10], one may restrict oneself even further to the class of Markov policies that are deterministic and stationary. The maximal discounted total expected reward can be found by policy iteration [22, Theorem 6.4.2] or value iteration, also known as successive approximation or dynamic programming.

When the cardinality of the state or action space is infinite, policy iteration is not guaranteed to converge in a finite number of steps. The dynamic programming approach on the other hand is amenable to generalisation to more general state and action spaces. Results in this direction include [4, 7, 10, 27]. The tutorial by Feinberg [6] provides an exhaustive overview with particular emphasis on inventory control problems.

In this paper, we concentrate on the case where the state space $\mathcal{X}$ consists of finite simple marked point patterns in two-dimensional Euclidean space. Markov decision theory using spatial point process models has found many applications in mobile network optimisation. However, the role of the point process is auxiliary in that it is used to model the spatial distribution of users, base stations and so on, from which coverage probabilities and other performance characteristics of the network can be calculated [1, 13, 16, 11]. Spatial point process models are also convenient in multi-target tracking [14] and their void probabilities or divergence measures can form the basis for observer trajectory optimisation [2].

Our focus of interest here is to assume that the actions operate directly on the point process. More precisely, we assume that, at decision epoch i, an action $\phi_i(\mathbf{x})$ maps $\mathbf{x}$ into a subset of $\mathbf{x}$. In other words, the action set $A(\mathbf{x})$ is the finite power set of $\mathbf{x}$. When the decision to retain a point or not is based on the mark or the inter-point distances, it can be interpreted as a (mark-)dependent thinning [17, 18]. The set-up described above is appropriate for harvesting problems in forestry [19]. Here, the classical strategy is to use discretised stand based growth tables and dynamic programming [26]. Point pattern based policies have been rarer due to 'a lack of models and to difficulties in selecting trees to be removed' [20] and tend to be simulation based [9, 21, 23, 24]. One example is German thinning, which enhances natural selection by picking trees whose diameter at breast height is at most d and fells a fraction of them. More formally, if each X_i consists of tree locations marked by diameter at breast height, German thinning fells a fraction of the set $\{(x, m) \in X_i : m \leq d\}$ and the Markov transition kernel governs the growth of the remaining trees (for example using the logistic growth curve or extensions such as the Richards curve [25]) as well as natural births and deaths (e.g. a hardcore model [12], the asymmetric soft core models of [15] or the dynamic models of [23]). French thinning is similar, except that a fraction of trees with large rather than small sizes is removed to stimulate forest rejuvenation. In either case, picking a policy amounts to choosing the level d. Simulations suggest that French thinning might be the better strategy [9].

The paper's plan is as follows. In Section 2, we study a decision process in which the actions consist of deleting a subset of the current points and the reward is proportional to the marks. The stochastic process that governs the dynamics is a birth-and-death process with independent deaths and a Poisson process of births; the marks grow logistically. We calculate the discounted total expected reward function over finite and infinite horizons and show that French thinning is an optimal policy. An explicit expression for the mark threshold d is derived too. In Section 3, we move on to allow interaction between the points and replace the Poisson birth process by one in which no point is allowed to come too close to another point. In this setting, we provide upper and lower bounds on the discounted total expected reward function over finite and infinite horizons. The tightness of the bounds is investigated by means of some simulated examples. We conclude by mentioning some topics for further research.

2 Marked Poisson process model with logistic growth

2.1 Definition of the model

To define a Markov decision process [22, Section 2.3.2], let the state space $\mathcal{X}$ consists of finite simple marked point patterns on a compact set $W \subset \mathbb{R}^2$ with marks in $L = [0, K]$ for some $K > 0$. When $\mathcal{X}$ is equipped with the Borel σ-algebra of the weak topology, by the discussion below [5, Prop 9.1.IV], $\mathcal{X}$ is Polish. When at time $i \in \mathbb{N}_0$ the process is in state $\mathbf{x}$, a thinning action is carried out, resulting in a new state $\mathbf{a}$ that consists of all retained points $\mathbf{a} \subset \mathbf{x}$. Thus, the action space $A(\mathbf{x})$ is finite and contains all subsets of $\mathbf{x}$. Define a stationary reward function $r_i(\mathbf{x}, \mathbf{a}) = r(\mathbf{x}, \mathbf{a})$ by

$$r(\mathbf{x}, \mathbf{a}) = R \sum_{(x,m) \in \mathbf{x} \setminus \mathbf{a}} m, \quad \mathbf{x} \in \mathcal{X}, \mathbf{a} \subset \mathbf{x}. \tag{2}$$

Thus, the reward is proportional to the sum of the marks of all removed points. When $R > 0$, $r(\cdot, \cdot) \geq 0$. Since the mark content in an $\mathbb{R}^+$-marked point process is a random variable [5, Proposition 6.4.V], r is well-defined.

Upon taking action $\mathbf{a}$ in state $\mathbf{x}$, the dynamics that lead to the next state are modelled as a birth-death-growth process. Specifically, the marks of the retained points $(x, m) \in \mathbf{a}$ grow according to the well-known logistic model that was proposed around 1840 by Verhulst and Quetelet [25]. In this model, when the mark at time 0 is $m_0 > 0$, the mark at time $n \in \mathbb{N} \cup \{0\}$ is

$$g^{(n)}(m_0) = \frac{K}{1 + \left(\frac{K}{m_0} - 1\right) e^{-\lambda n}}. \tag{3}$$

By convention, $g^{(n)}(0) = 0$. The parameter $\lambda > 0$ governs the rate of growth and $K \geq m_0 \geq 0$ is an upper bound on the size. In combination with independent births and deaths, the next state is defined by the following dynamics:

- delete $\mathbf{x} \setminus \mathbf{a}$;
- independently of other points, let each $(x_i, m_i) \in \mathbf{a}$ die with probability $p_d \in (0, 1)$ (natural deaths) and otherwise grow to $(x_i, g^{(1)}(m_i))$ as in (3);
- add a Poisson process on W with intensity $\beta > 0$ and mark its points independently according to a probability measure ν on $[0, K]$.

Write $(X_i, Y_i)_{i=0}^{\infty}$ for the sequence of successive states X_i and actions Y_i. A randomised policy $\phi = (\phi_i)_{i=0}^{\infty}$ is a sequence of conditional probability kernels $\phi_i(\cdot | X_0, Y_0, \ldots, X_{i-1}, Y_{i-1}, X_i)$ on A to generate Y_i based on the history of the process such that $\phi_i(A(\mathbf{x}_i) | \mathbf{x}_0, \mathbf{a}_0, \ldots, \mathbf{x}_i) =$

1. If the policy is Markov and deterministic, Y_i is simply a function of X_i, and one may write $Y_i = \phi_i(X_i)$. Then, for $0 \le \alpha < 1$, the infinite horizon α-discounted total expected reward function (1) under policy $\phi = (\phi_i)_{i=0}^\infty$ with initial state $X_0 = \mathbf{x}$ is

$$v_\alpha^\phi(\mathbf{x}) = \mathbb{E}^\phi \left[\sum_{i=0}^\infty \alpha^i \left(R \sum_{(x,m) \in X_i \setminus Y_i} m \right) \mid X_0 = \mathbf{x} \right]. \tag{4}$$

The following Lemma shows that the model is well-defined for the birth-death-grow dynamics defined above.

Lemma 1 *The infinite horizon α-discounted total expected reward function $v_\alpha^\phi(\mathbf{x})$, $\mathbf{x} \in \mathcal{X}$, defined in (4) is finite for all $0 \le \alpha < 1$, all $R > 0$ and all policies ϕ.*

Proof: Pick $\mathbf{x} \in \mathcal{X}$ and write $n(\mathbf{x}) < \infty$ for its cardinality. Since the growth function (3) is bounded by K,

$$\mathbb{E} \left[\sum_{(x,m) \in X_0 \setminus Y_0} m \mid X_0 = \mathbf{x} \right] \le K n(\mathbf{x}).$$

For $i > 0$, X_i is the union of survivors from $\mathbf{x}$, from subsequent generations starting with $X_1 \setminus X_0$ up to $X_{i-1} \setminus X_{i-2}$ and points born in the last decision epoch. Therefore, recalling the birth and death dynamics,

$$\mathbb{E} \left[\sum_{(x,m) \in X_i \setminus Y_i} m | X_0 = \mathbf{x} \right] \le K n(\mathbf{x})(1 - p_d)^i + K\beta|W| \sum_{k=0}^{i-1} (1 - p_d)^k$$

where $|W|$ denotes the area of W. Hence

$$v_\alpha^\phi(\mathbf{x}) \le RKn(\mathbf{x}) \sum_{i=0}^\infty \alpha^i (1 - p_d)^i + RK\beta|W| \sum_{i=1}^\infty \alpha^i \sum_{k=0}^{i-1} (1 - p_d)^k.$$

For all $p_d \in (0, 1)$, the first series in the right hand side converges to $1/(1 - \alpha(1 - p_d))$. Since

$$\sum_{i=1}^\infty \alpha^i \sum_{k=0}^{i-1} (1 - p_d)^k = \sum_{i=1}^\infty \alpha^i \frac{1 - (1 - p_d)^i}{p_d} \le \frac{1}{p_d} \sum_{i=1}^\infty \alpha^i < \infty$$

for all $p_d \in (0, 1)$, $v_\alpha^\phi(\mathbf{x})$ is finite. □

The reward function r itself is not bounded, so the (N) regime of [4, Chapter 9] applies.

2.2 Optimal policy and reward

The optimal α-discounted total expected reward $v_\alpha^*(\mathbf{x})$ is defined as the supremum of the $v_\alpha^\phi(\mathbf{x})$ over all policies, including randomised ones. In this section, we will show that French thinning is optimal and give an explicit expression for the corresponding reward.

By [4, Proposition 9.1], the supremum in the definition of $v^*_\alpha(\mathbf{x})$ may be taken over the class of Markov policies, and, by [4, Proposition 9.8], satisfies the equation

$$v^*_\alpha(\mathbf{x}) = \max_{\mathbf{a}\subset\mathbf{x}} \left\{ R \sum_{(x,m)\in\mathbf{x}\setminus\mathbf{a}} + \alpha\mathbb{E}\left[v^*_\alpha(X) \mid \mathbf{x}, \mathbf{a}\right] \right\} \tag{5}$$

where X is distributed according to the one step birth-death-growth dynamics from state $\mathbf{x}$ under action $\mathbf{a}$. Observe that the optimality equations (5), $\mathbf{x} \in \mathcal{X}$, are not sufficient conditions for v^*_α. Nevertheless, $v^*_\alpha(\mathbf{x})$ can be calculated as the limit of an iterative procedure [4, Proposition 9.14] known as the dynamic programming algorithm. Set $v_0(\mathbf{x}) = 0$ for all $\mathbf{x} \in \mathcal{X}$ and set $n = 1$. Define, for every $\mathbf{x} \in \mathcal{X}$,

$$v_n(\mathbf{x}) = \max_{\mathbf{a}\subset\mathbf{x}} \left\{ R \sum_{(x,m)\in\mathbf{x}\setminus\mathbf{a}} m + \alpha\mathbb{E}\left[v_{n-1}(X) \mid \mathbf{x}, \mathbf{a}\right] \right\}.$$

Then set $n = n + 1$ and repeat. This algorithm converges to $v^*_\alpha(\mathbf{x})$ as $n \to \infty$ by [4, Proposition 9.14] but – in general – is of little help in constructing an optimal policy, let alone a stationary one. Given a stationary policy ϕ, a necessary and sufficient condition for it to be optimal is [4, Prop. 9.13]

$$v^\phi_\alpha(\mathbf{x}) = \max_{\mathbf{a}\subset\mathbf{x}} \left\{ R \sum_{(x,m)\in\mathbf{x}\setminus\mathbf{a}} + \alpha\mathbb{E}\left[v^\phi_\alpha(X) | \mathbf{x}, \mathbf{a}\right] \right\}. \tag{6}$$

For our model, the dynamic programming algorithm does suggest an optimal deterministic and stationary Markov policy.

Theorem 1 *Consider the Markov decision process with state space $\mathcal{X}$, action spaces $A(\mathbf{x}) = \{\mathbf{y} \in \mathcal{X} : \mathbf{y} \subset \mathbf{x}\}$, $\mathbf{x} \in \mathcal{X}$, reward function (2) with $R > 0$, and birth-death-growth dynamics based on independent deaths with probability $p_d \in (0, 1)$, a Poisson birth process with intensity $\beta > 0$ marked independently according to probability measure ν on $[0, K]$ for $K > 0$ and logistic growth function (3). Then, for $0 \le \alpha < 1$,*

$$v^*_\alpha(\mathbf{x}) = R\beta|W| \sum_{k=1}^{\infty} \alpha^k \int_0^K s(m)\, d\nu(m) + R \sum_{(x,m)\in\mathbf{x}} s(m),$$

where $|W|$ is the area of W and

$$s(m) = \sup_{n\in\mathbb{N}_0} \left\{ \frac{K\alpha^n(1-p_d)^n}{1 + \left(\frac{K}{m} - 1\right) e^{-\lambda n}} \right\}, \quad m \in [0, K].$$

Furthermore, the optimal α-discounted total expected reward corresponds to a French thinning that removes all points with a mark that is at least

$$d^*_\alpha = \sup_{n\in\mathbb{N}_0} \left\{ \frac{K}{1 - e^{-n\lambda}} \left(\alpha^n (1-p_d)^n - e^{-n\lambda} \right) \right\}.$$

For $\alpha = 1$, the total expected reward $v_1^*(\mathbf{x})$ is infinite.

Proof: After initialising $v_0(\mathbf{x}) = 0$ for all $\mathbf{x} \in \mathcal{X}$, clearly the optimal expected reward at time 0 is $v_1(\mathbf{x}) = R \sum_{(x,m)\in\mathbf{x}} m$, which is attained for action $\mathbf{a} = \emptyset$, or, in other words, by removing all points with mark greater than or equal to $d_1 = 0$. The proof proceeds by induction. Set, for $n \in \mathbb{N}$,

$$d_n = \max\left\{0, K\frac{\alpha(1-p_d) - e^{-\lambda}}{1-e^{-\lambda}}, \ldots, K\frac{\alpha^{n-1}(1-p_d)^{n-1} - e^{-(n-1)\lambda}}{1-e^{-(n-1)\lambda}}\right\} \tag{7}$$

and suppose that the optimal α-discounted expected reward over n steps is attained by French thinning at level d_n and given by

$$v_n(\mathbf{x}) = R\beta|W| \sum_{k=1}^{n-1} \alpha^k \int_0^K s_{n-k}(m)\, d\nu(m) + R \sum_{(x,m)\in\mathbf{x}} s_n(m) \tag{8}$$

where, for $1 \leq k \leq n$,

$$s_k(m) = \max\left\{m, \alpha\,(1-p_d)\, g^{(1)}(m), \ldots, \alpha^{k-1}\,(1-p_d)^{k-1}\, g^{(k-1)}(m)\right\}.$$

Now, for $n+1$, the optimal finite horizon α-discounted expected reward is

$$v_{n+1}(\mathbf{x}) = \max_{\mathbf{a}\subset\mathbf{x}} \left\{ R \sum_{(x,m)\in\mathbf{x}\setminus\mathbf{a}} m + \alpha\mathbb{E}\left[v_n(X) \mid \mathbf{x}, \mathbf{a}\right] \right\}.$$

By the induction assumption, the discounted expectation $\alpha\mathbb{E}\left[v_n(X) \mid \mathbf{x}, \mathbf{a}\right]$ is the sum of

$$\alpha R\beta|W| \sum_{k=1}^{n-1} \alpha^k \int_0^K s_{n-k}(m)\, d\nu(m) = R\beta|W| \sum_{k=2}^{n} \alpha^k \int_0^K s_{n+1-k}(m)\, d\nu(m)$$

and contributions from the points in $\mathbf{a}$ that survive a decision epoch as well as from points born in the interval between time n and $n+1$. These contributions are, respectively,

$$\alpha R \sum_{(x,m)\in\mathbf{a}} (1-p_d)\, s_n(g(m))$$

and, using the Campbell–Mecke formula [5, Section 6.1],

$$\alpha R\beta|W| \int_0^K s_n(m)\, d\nu(m).$$

The optimal policy assigns a point $(x, m) \in \mathbf{x}$ to $\mathbf{x} \setminus \mathbf{a}$ if and only if $m \geq \alpha(1-p_d)s_n(g^{(1)}(m))$. By the induction assumption and (3), this is the case if and only if

$$m \geq \alpha^k(1-p_d)^k g^{(k)}(m) \Leftrightarrow m \geq K\frac{\alpha^k(1-p_d)^k - e^{-k\lambda}}{1-e^{-k\lambda}} \tag{9}$$

for all integers $1 \leq k \leq n$. Consequently, d_{n+1} has the required form. For this allocation rule, the reward is $\max\left\{m, \alpha\,(1-p_d)\, s_n(g^{(1)}(m))\right\} = s_{n+1}(m)$ and the induction step is complete.

Next, let n go to infinity and fix $m \in [0, K]$. Then $s(m)$ is finite for all $p_d \in (0,1)$ and $0 \leq \alpha < 1$. Additionally, $\lim_{n\to\infty} s_n(m) = s(m)$. Thus, for any $\mathbf{x} \in \mathcal{X}$,

$$R \sum_{(x,m)\in\mathbf{x}} s_n(m) \to R \sum_{(x,m)\in\mathbf{x}} s(m)$$

as $n \to \infty$. Furthermore,

$$\sum_{k=1}^{n-1} \alpha^k \int_0^K s_{n-k}(m)\, d\nu(m) \to \sum_{k=1}^{\infty} \alpha^k \int_0^K s(m)\, d\nu(m), \quad n \to \infty,$$

because of dominated convergence applied to the doubly indexed sequence $a_{k,n}$ defined by $\mathbf{1}\{k \leq n-1\}\, \alpha^k \int s_{n-k}\, d\nu$. In conclusion, for each $\mathbf{x} \in \mathcal{X}$, $\lim_{n\to\infty} v_n(\mathbf{x}) = v_\alpha^*(\mathbf{x})$, the optimal α-discounted total expected reward [4, Proposition 9.14], and $v_\alpha^*(\mathbf{x})$ has the claimed form.

To complete the proof, we need to show that $v^*(\mathbf{x})$ is attained by the stationary deterministic policy that retains all points with mark smaller than d_α^*. Denote its infinite horizon α-discounted total expected reward by

$$v_\alpha^{d^*}(\mathbf{x}) = \mathbb{E}\left[R \sum_{i=0}^{\infty} \alpha^i \sum_{(x,m)\in X_i} m\, \mathbf{1}\{m \geq d_\alpha^*\} \mid X_0 = \mathbf{x}\right]$$

and focus on the contributions of each generation of points. A point $(x, m) \in \mathbf{x}$, the initial generation, yields a reward $R\,\alpha^n (1-p_d)^n g^{(n)}(m)$ precisely when $g^{(n-1)}(m)$ is less than d_α^* but $g^{(n)}(m) \geq d_\alpha^*$. Since, as in (9), $g^{(n)}(m) \geq d_\alpha^*$ if and only if

$$g^{(n)}(m) \geq \alpha^k (1-p_d)^k g^{(n+k)}(m)$$

for all $k \in \mathbb{N}_0$, we conclude that every point of $\mathbf{x}$ contributes $R\, s(m)$. The points that are born in the first decision epoch (generation 1) yield the same total expected reward, but this is discounted by α due to the later birth date. Similarly, the total expected reward of points belonging to the second generation is discounted by α^2, and so on. Tallying up, the α-discounted total expected reward of generations $k = 1, 2, \ldots$ is

$$R\beta|W| \sum_{k=1}^{\infty} \alpha^k \int_0^K s(m)\, d\nu(m)$$

on application of the Campbell–Mecke formula. Finally add the contribution from the initial generation to conclude that the threshold d_α^* defines an optimal policy. Condition (6) is readily verified. □

As a by-product, the proof of Theorem 1 derives the optimal α-discounted total expected reward (8) for finite time horizons too, and French thinning with threshold (7) is an optimal policy. The suprema in $s(m)$ and d_α^* are attained, which can be seen by considering the limit for $n \to \infty$.

3 Hard core models with logistic growth

3.1 Bounds for the optimal discounted total expected reward

In this section, we refine the Poisson model of the previous section to the case where births are governed by a hard core process. Thus, the state space $\mathcal{X}_K$ consists of all finite simple marked point patterns on a compact set W in the plane that contain no pair $\{x_1, x_2\}$ such that $||x_1 - x_2|| \leq K$ with marks in $L = [0, K]$. For the motivating example from forestry in which the marks correspond to the diameter at breast height, the condition ensures that all trees can grow to their maximal size.

As in Section 2.1, when at time $i \in \mathbb{N}_0$ the process is in state $\mathbf{x}$, a thinning action is carried out, resulting in a new state $\mathbf{a}$ that consists of all retained points. The reward is defined in (2).

The dynamics are modified in such a way that the hard core is respected. Specifically, suppose that action $\mathbf{a}$ is taken in state $\mathbf{x} \in \mathcal{X}_K$. The next state is then governed by the following birth-death-growth process:

- delete $\mathbf{x} \setminus \mathbf{a}$;
- independently of other points, let each $(x_i, m_i) \in \mathbf{a}$ die with probability $p_d \in (0, 1)$ and otherwise grow to $(x_i, g(m_i))$ for some bounded, continuous function $g : [0, K] \to [0, K]$ satisfying $m \leq g(m)$ for $m \in [0, K]$;
- add a hard core process on W with hard core distance K and intensity $\beta > 0$; mark its points independently according to a probability measure ν on $[0, K]$ and remove all points that fall within distance K to a point in $\mathbf{a}$.

In this framework, the reward function is bounded since the hard core condition implies an upper bound on the number of points that can be alive at any time. We are therefore in the (D) regime of [4, Chapter 9].

For $\mathbf{x} \in \mathcal{X}_K$, define $v^*_\alpha(\mathbf{x})$ as the supremum of (4) over all policies ϕ. By [4, Proposition 9.1] it suffices to consider Markov policies only, and $v^*_\alpha(\mathbf{x})$ is the limit of the dynamic programming algorithm [4, Proposition 9.14]. The optimality condition (6) applies. Moreover, since the action sets are finite, Corollary 9.17.1 in [4] guarantees the existence of an optimal deterministic stationary policy. An explicit expression seems hard to obtain. However, the following bounds are available.

Theorem 2 *Consider the Markov decision process with state space $\mathcal{X}_K$, action spaces $A(\mathbf{x}) = \{\mathbf{y} \in \mathcal{X}_K : \mathbf{y} \subset \mathbf{x}\}$, $\mathbf{x} \in \mathcal{X}_K$, reward function (2) with $R > 0$, and birth-death-growth dynamics based on independent deaths with probability $p_d \in (0, 1)$, a hard core birth process with intensity $\beta > 0$ marked independently according to probability measure ν on $[0, K]$ for $K > 0$ and growth function g. Write $g^{(n)}(m)$ for the n-fold composition of g.*

For $\alpha \in [0, 1)$, initialise $v_0(\mathbf{x}) = 0$ for all $\mathbf{x} \in \mathcal{X}_K$. Define, for $n \in \mathbb{N}$ and $\mathbf{x} \in \mathcal{X}_K$,

$$v_n(\mathbf{x}) = \max_{\mathbf{a} \subset \mathbf{x}} \left\{ R \sum_{(x,m) \in \mathbf{x} \setminus \mathbf{a}} m + \alpha \mathbb{E} \left[v_{n-1}(X) \mid \mathbf{x}, \mathbf{a} \right] \right\}.$$

Then $\tilde{v}_n(\mathbf{x}) \leq v_n(\mathbf{x}) \leq \hat{v}_n(\mathbf{x})$ *where*

$$\begin{aligned}
\tilde{v}_n(\mathbf{x}) &= R \sum_{(x,m)\in\mathbf{x}} \tilde{s}_n(x,m) + R\beta \sum_{k=1}^{n-1} \alpha^k \int_W \int_0^K \tilde{s}_{n-k}(w,l)\, d\nu(l) dw \\
\hat{v}_n(\mathbf{x}) &= R \sum_{(x,m)\in\mathbf{x}} \hat{s}_n(m) + R\beta |W| \sum_{k=1}^{n-1} \alpha^k \int_0^K \hat{s}_{n-k}(l)\, d\nu(l)
\end{aligned}$$

with $\tilde{s}_0 = \hat{s}_0 = 0$ *and, for* $n \in \mathbb{N}$,

$$\hat{s}_n(m) = \max\left\{m, \alpha(1-p_d)g^{(1)}(m),\, \ldots,\, \alpha^{n-1}(1-p_d)^{n-1} g^{(n-1)}(m)\right\}$$

and, writing $b(x,K)$ *for the closed ball centred at* x *with radius* K,

$$\begin{aligned}
&\tilde{s}_n(x,m) = \max\{m, \alpha(1-p_d)g^{(1)}(m) - \alpha K\beta |b(x,K)\cap W|,\, \ldots, \\
&\alpha^{n-1}(1-p_d)^{n-1} g^{(n-1)}(m) - \alpha K \beta |b(x,K)\cap W| \sum_{i=0}^{n-2} \alpha^i (1-p_d)^i\}.
\end{aligned}$$

When the growth function is logistic,

$$\begin{aligned}
\tilde{s}_n(x,m) &= \max_{i=0,\ldots,n-1} \left\{ \frac{K\alpha^i(1-p_d)^i}{1+\left(\frac{K}{m}-1\right)e^{-\lambda i}} - \alpha K \beta |W\cap b(x,K)| \frac{1-\alpha^i(1-p_d)^i}{1-\alpha(1-p_d)} \right\}; \\
\hat{s}_n(m) &= \max_{i=0,\ldots,n-1} \left\{ \frac{K\alpha^i(1-p_d)^i}{1+\left(\frac{K}{m}-1\right)e^{-\lambda i}} \right\}.
\end{aligned}$$

Proof: The proof proceeds by induction. For $n = 0$, evidently $v_0 \leq \tilde{v}_0$. Assume that $\tilde{v}_k(\mathbf{x}) \leq v_k(\mathbf{x}) \leq \hat{v}_k(\mathbf{x})$ for all $k \leq n$ and all $\mathbf{x} \in \mathcal{X}_K$ and that $\tilde{v}_k$, $\hat{v}_k$ have the required form. Since

$$v_{n+1}(\mathbf{x}) = \max_{\mathbf{a}\subset\mathbf{x}} \left\{ R \sum_{(x,m)\in\mathbf{x}\setminus\mathbf{a}} m + \alpha \mathbb{E}\left[v_n(X) \mid \mathbf{x},\mathbf{a}\right] \right\} \tag{10}$$

and $v_n(X) \geq \tilde{v}_n(X)$, let us consider the expectation of $\tilde{v}_n(X)$ under the hard core birth-death-growth dynamics when action $\mathbf{a}$ is taken in state $\mathbf{x}$. By the definition of $\tilde{v}_n$ and distinguishing between surviving and new-born points,

$$\begin{aligned}
\mathbb{E}\left[\tilde{v}_n(X) \mid \mathbf{x},\mathbf{a}\right] &= R\,\mathbb{E}\left[\sum_{(x,m)\in X} \tilde{s}_n(x,m) \mid \mathbf{x},\mathbf{a} \right] + R\beta \sum_{k=1}^{n-1} \alpha^k \int_W \int_0^K \tilde{s}_{n-k}(w,l)\, d\nu(l) dw \\
&= R \sum_{(x,m)\in\mathbf{a}} (1-p_d)\, \tilde{s}_n(x, g^{(1)}(m)) + R\beta \sum_{k=1}^{n-1} \alpha^k \int_W \int_0^K \tilde{s}_{n-k}(w,l)\, d\nu(l) dw \\
&\quad + R\beta \int_W \int_0^K \tilde{s}_n(w,l)\, \mathbf{1}\{w \notin U_K(\mathbf{a})\}\, d\nu(l) dw
\end{aligned}$$

where the symbol $U_K(\mathbf{a})$ signifies the union of closed balls with radius K around the points in $\mathbf{a}$. The calculation of the last term above relies on the Campbell–Mecke formula [5, Section 6.1]. Now, the integral in the last line above can be written as

$$R\beta \int_W \int_0^K \tilde{s}_n(w,l)\, d\nu(l) dw - R\beta \int_W \int_0^K \tilde{s}_n(w,l)\, \mathbf{1}\{w \in U_K(\mathbf{a}\}\, d\nu(l) dw$$

and is bounded from below by

$$R\beta \int_W \int_0^K \tilde{s}_n(w,l)\, d\nu(l) dw - RK\beta \sum_{(x,m)\in\mathbf{a}} \int_W \int_0^K \mathbf{1}\{w \in b(x,K)\}\, d\nu(l) \tag{11}$$

where the induction assumption is invoked for the inequality $\tilde{s}_n \leq K$. Next, return to (10). The bound on $\mathbb{E}[\tilde{v}_n(X) \mid \mathbf{x}, \mathbf{a}]$ implies

$$v_{n+1}(\mathbf{x}) \geq \max_{\mathbf{a}\subset\mathbf{x}} \left\{ R \sum_{(x,m)\in\mathbf{x}\setminus\mathbf{a}} m + \alpha \mathbb{E}[\tilde{v}_n(X) \mid \mathbf{x}, \mathbf{a}] \right\} \geq \max_{\mathbf{a}\subseteq\mathbf{x}} \{ R \sum_{(x,m)\in\mathbf{x}\setminus\mathbf{a}} m +$$

$$\alpha R \sum_{(x,m)\in\mathbf{a}} \Big[(1-p_d)\tilde{s}_n(x, g^{(1)}(m)) - K\beta |b(x,K)\cap W| \Big] + R\beta \sum_{k=1}^{n} \alpha^k \int_W \int_0^K \tilde{s}_{n+1-k}(w,l) d\nu(l) dw.$$

The policy that assigns (x,m) to $\mathbf{x}\setminus\mathbf{a}$ if and only if

$$m \geq \alpha \Big[(1-p_d)\, \tilde{s}_n(x, g^{(1)}(m)) - K\beta |b(x,K)\cap W| \Big]$$

optimises the right hand side and, with

$$\tilde{s}_{n+1}(x,m) = \max\Big\{ m, \alpha\,(1-p_d)\, \tilde{s}_n(x, g^{(1)}(m)) - \alpha K\beta |b(x,K)\cap W| \Big\},$$

one sees that

$$v_{n+1}(\mathbf{x}) \geq \tilde{v}_{n+1}(\mathbf{x}) = R \sum_{(x,m)\in\mathbf{x}} \tilde{s}_{n+1}(x,m) + R\beta \sum_{k=1}^{n} \alpha^k \int_W \int_0^K \tilde{s}_{n+1-k}(w,l)\, d\nu(l) dw,$$

an observation that completes the induction argument and therefore the proof of the lower bound.

For the upper bound $v_n \leq \hat{v}_n$, as in the proof of Theorem 1, an induction proof applies based on $\hat{s}_n$ but with (11) replaced by the upper bound

$$R\beta \int_W \int_0^K \hat{s}_n(w,l)\, d\nu(l) dw.$$

□

Over an infinite time horizon, the optimal α-discounted total expected reward is bounded by the same functional forms, which coincide if $\alpha = 0$.

Corollary 1 *The functions $\hat{s}_n$ and $\tilde{s}_n$ defined in Theorem 2 take values in $[0, K]$ and increase monotonically to*

$$\hat{s}(m) = \sup_{n \in \mathbb{N}_0} \left\{ \alpha^n (1 - p_d)^n g^{(n)}(m) \right\}, \quad m \in [0, K],$$

and, for $x \in W$ and $m \in [0, K]$,

$$\tilde{s}(x, m) = \sup_{n \in \mathbb{N}_0} \left\{ \alpha^n (1 - p_d)^n g^{(n)}(m) - \alpha K |b(x, K) \cap W| \sum_{i=0}^{n-1} \alpha^i (1 - p_d)^i \right\}.$$

3.2 Simulation study

To assess the tightness of the bounds in Theorem 2, we calculated $\hat{v}_n(\mathbf{x})$ and $\tilde{v}_n(\mathbf{x})$ in two regimes, a dense one and a sparse one. For the inital pattern $\mathbf{x}$, a sample from a Strauss process [12] on $W = [0, 5]^2$ with interaction parameter set to zero was chosen. The activity parameter was set to give the required intensity: $\beta = 1.0$ in the sparse regime and $\beta = 4.3$ in the dense regime. For the mark dynamics, we used a logistic growth function with $\lambda = 2$ and maximal size $K = 0.1$; the initial marks were sampled from a Beta distribution with shape parameters $\lambda_1 = 2$ and $\lambda_2 = 20$. The death rate was set to $p_d = 0.05$. Finally, we used discount factor $\alpha = 0.9$ and reward parameter $R = 1$.

The results are plotted in Figure 1. The left panels show the pattern $\mathbf{x}$. In the right panels, the solid lines are the graphs of $\hat{v}_n(\mathbf{x})$ as a function of n, the dotted lines show $\tilde{v}_n(\mathbf{x})$ plotted against n. Integrals were estimated by the Monte Carlo method with $1,000$ samples. In the sparse regime, the approximation is quite good, for the denser regime, the gap between the two graphs is quite wide except for very small n. In both cases, the dynamic programming algorithm converges rapidly.

4 Conclusion

In this paper we considered optimal policies for Markov decision problems inspired by forest harvesting. We proved that French thinning is optimal when births follow a Poisson process and marks grow logistically. When the points are required to respect a hard core distance, we derived upper and lower bounds on the discounted total expected reward for general birth-death-growth dynamics. Although we focused on a homogeneous birth process, the results carry over to the case where the birth process is governed by some spatially varying intensity function.

In future it would be of interest to study configuration dependent asymmetric birth and growth models [14, 15, 24]. Indeed, in a forestry setting, the growth of well-established, large trees may hardly be hampered by the emergence of saplings close by, while it would be harder for young and small trees to flourish near large ones. Moreover, the natural environment, such as the availability of nutrients, might play a role. Finally, refinements of the action space that allow for different thresholds in different mark strata could be investigated.

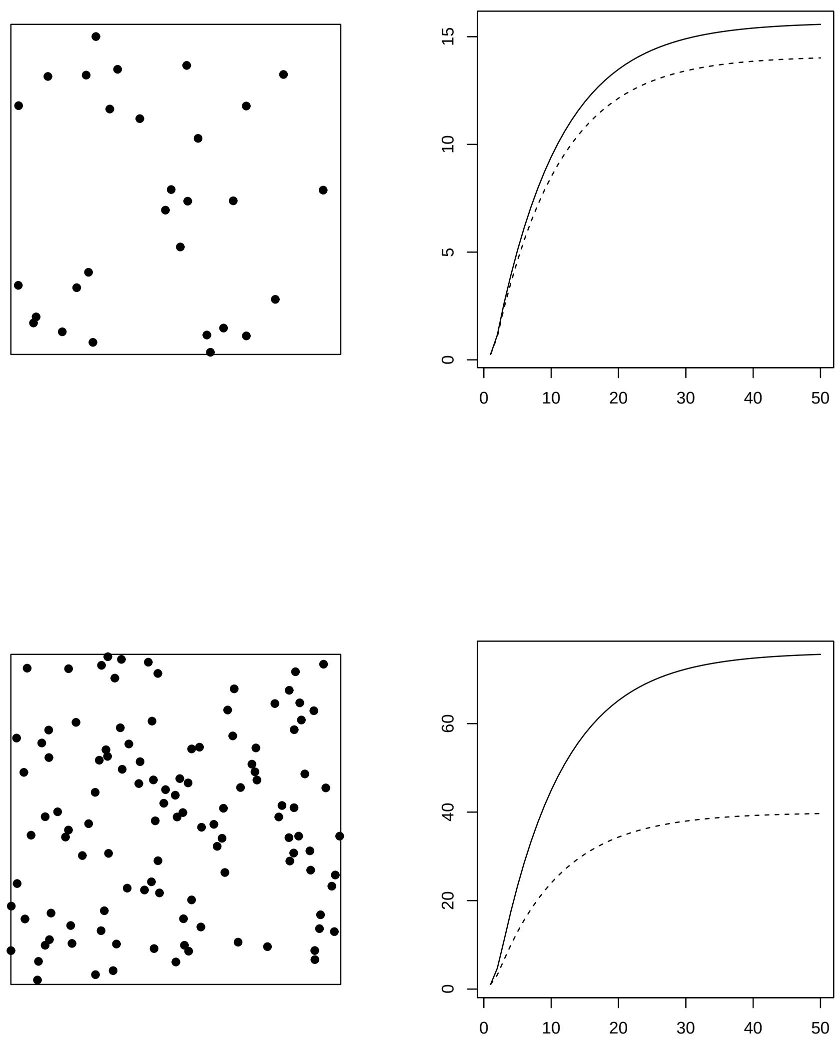

Figure 1: Left panels: samples $\mathbf{x}$ from a Strauss hard core process with intensity $\beta = 1.0$ (top) and $\beta = 4.3$ (bottom) on $[0, 5]^2$. Right panels: graphs of $\hat{v}_n(\mathbf{x})$ (solid lines) and $\tilde{v}_n(\mathbf{x})$ (dotted lines) against n for the birth-death-growth dynamics of Section 3.2.

References

[1] Baccelli, F. and Blaszczyszyn, B. (2009). *Stochastic geometry and wireless networks*, in two volumes. NOW.

[2] Beard, M., Vo, B.T., Vo, B.N. and Arulampalam, S. (2017). Void probabilities and Cauchy–Schwarz divergence for generalized labeled multi-Bernoulli models. *IEEE Trans. Signal Process.*, **65**, 5047–5061.

[3] Bertsekas, D.P. (1995). *Dynamic programming and optimal control.* Prentice and Hall.

[4] Bertsekas, D.P. and Shreve, S.E. (1978). *Stochastic optimal control: The discrete time case.* Academic Press.

[5] Daley, D.J. and Vere–Jones, D. (2003, 2008). *An introduction to the theory of point processes*, second edition in two volumes. Springer.

[6] Feinberg, E.A. (2016). *Optimality conditions for inventory control*. Tutorials in Operations Research, INFORMS 2016, pp. 14–44.

[7] Feinberg, E.A. and Lewis, M.E. (2007). Optimality inequalities for average cost Markov decision processes and the stochastic cash balance problem. *Math. Oper. Res.*, **32**, 769–783.

[8] Feinberg, E.A. and Schwartz, A. (2002). *Handbook of Markov decision processes*. Springer.

[9] Fransson, P., Franklin, O., Lindroos, O., Nilsson, U. and Brännström, Å. (2020). A simulation-based approach to a near optimal thinning strategy: Allowing for individual harvesting times for individual trees. *Can. J. For. Res.*, **50**, 320–331.

[10] Hernández–Lerma, O. and Lasserre, B.J. (1996). *Discrete-time Markov control processes: Basic optimality criteria*. Springer.

[11] Khloussy, E., Gelabert, X. and Jiang, Y. (2015). Investigation on MDP-based radio access technology selection in heterogeneous wireless networks. *Comput. Netw.*, **91**, 57–67.

[12] Kelly, F.P. and Ripley, B.D. (1976). On Strauss's model for clustering. *Biometrika*, **63**, 357–360.

[13] Lee, W., Jung, B.C. and Lee, H. (2020). DeCoNet: Density clustering-based base station control for energy-efficient cellular IoT networks. *IEEE Access*, **8**, 120881.

[14] Lieshout, M.N.M. van (2008). Depth map calculation for a variable number of moving objects using Markov sequential object processes. *IEEE Trans. Pattern Anal. Mach. Intell.*, **30**, 1308–1312.

[15] Lieshout, M.N.M. van (2009). Sequential spatial processes for image analysis. In *Stereology and Image Analysis. ECS10–Proceedings of the 10th European Congress of ISS*, V. Capasso *et al.* (Eds.), 6 pages. Bologna.

[16] Lu, X., Salehi, M., Haenggi, M., and Hossain, E. (2021). Stochastic geometry analysis of spatial-temporal performance in wireless networks: A tutorial. *IEEE Commun. Surveys & Tutorials*, **23**, 2753–2801.

[17] Matérn, B. (1986). *Spatial variation*. Springer.

[18] Myllimäki, M. (2009). *Statistical models and inference for spatial point patterns with intensity-dependent marks*. PhD thesis, University of Jyväskylä.

[19] Pretzch, H. (2009). *Forest dynamics, growth and yield*. Springer.

[20] Pukkala, T. and Miina, J. (1998). Tree-selection algorithms for optimizing thinning using a distance-dependent growth model. *Can. J. For. Res.*, **28**, 693–702.

[21] Pukkala, T., Lähde, E. and Laiho, O. (2015). Which trees should be removed in thinning treatments? *For. Ecosyst.*, **2**, 1–12.

[22] Puterman, M.L. (1994). *Markov decision processes*. Wiley.

[23] Renshaw, E. and Särkkä, A. (2001). Gibbs point processes for studying the development of spatial-temporal stochastic processes. *Comput. Stat. Data Anal.*, **36**, 85–105.

[24] Renshaw, E., Comas, C. and Mateu, J. (2009). Analysis of forest thinning strategies through the development of space-time growth-interaction simulation models. *Stoch. Environ. Res. Risk Assess.*, **23**, 275–288.

[25] Richards, F.J. (1959). A flexible growth function for empirical use. *J. Exp. Bot.*, **10**, 290–300.

[26] Rönnqvist, M. (2003). Optimization in forestry. *Math. Program. Ser. B*, **97**, 267–284.

[27] Schäl, M. (1993). Average optimality in dynamic programming with general state space. *Math. Oper. Res.*, **18**, 163–172.