Risk-sensitive Policies for Sustainable Renewable Resource Allocation

Stefano Ermon
Computer Science Department
Cornell University
ermonste@cs.cornell.edu

Jon Conrad
Applied Economics Department
Cornell University
jmc16@cornell.edu

Carla Gomes, Bart Selman
Computer Science Department
Cornell University
{gomes,selman}@cs.cornell.edu

Abstract

Markov Decision Processes arise as a natural model for many renewable resources allocation problems. In many such problems, high stakes decisions with potentially catastrophic outcomes (such as the collapse of an entire ecosystem) need to be taken by carefully balancing social, economic, and ecologic goals. We introduce a broad class of such MDP models with a risk averse attitude of the decision maker, in order to obtain policies that are more balanced with respect to the welfare of future generations. We prove that they admit a closed form solution that can be efficiently computed.

We show an application of the proposed framework to the Pacific Halibut marine fishery, obtaining new and more cautious policies. Our results strengthen findings of related policies from the literature by providing new evidence that a policy based on periodic closures of the fishery should be employed, in place of the one traditionally used that harvests a constant proportion of the stock every year.

1 Introduction

In this paper, we consider a broad class of MDPs that can be used as a model for the management of a renewable resource. While the most common examples are probably living resources, such as animal populations or forests, other types of resources such as time, energy, and financial resources can be also considered [Ermon *et al.*, 2010]. In fact, the problem of managing the stock of a resource dynamically changing over time is quite general and it arises in many different fields under different names. In particular, in a *renewable* resource allocation problem, the stock of a (valuable) resource is dynamically *growing* over time. Such growth processes might occur on a continuous or discrete time scale, they are subject to nonlinear dynamics and they are affected by randomness. When we also factor in social and economic aspects, these multiple levels of complexity can easily lead to very challenging sequential decision making problems.

Most of the work [Clark, 1990; Conrad, 1999] in the literature on the management of such resources is based on *risk-neutral* formulations, where the goal is that of maximizing the sum of (discounted) economic rewards over time. However, in many domains arising in the emerging field of Computational Sustainability [Gomes, 2009], this approach can be inadequate. In fact in many such problems, high stakes decisions with potentially catastrophic outcomes (such as the collapse of an entire ecosystem) need to be taken by carefully balancing social, economic and ecologic goals. More in general, a fundamental challenge of *sustainability science* [Clark, 2007] is how to take decisions that balance the welfare of current and future generations and to devise policies that provide guarantees on the welfare of future generations. In this context, *risk-neutral* approaches are usually not capable of providing such guarantees. For instance, in the case of the management of an animal population, policies that are too aggressive may cause the collapse of an entire population.

As a partial answer to these questions, we propose the introduction of decision theoretic concepts into the optimization frameworks, by leveraging the significant advancements made by the AI community. In general, we can model and solve MDPs with general utility functions [Liu and Koenig, 2006], so that we can capture the behavior of almost every rational decision maker [Von Neumann *et al.*, 2007] and cast resource allocation problems into a much richer framework. In particular, in the context of natural resources, it is critical to introduce *risk-averse* attitudes of the decision makers by choosing suitable utility functions, as opposed to standard *risk-neutral* approaches. By enforcing risk-aversion, we obtain policies that are more cautious and less prone to those catastrophic outcomes that might endanger the welfare of future generations.

With similar goals, Ermon et al. [Ermon *et al.*, 2010] propose the use of a *worst-case scenario* analysis (which is analogous to a game against nature) in order to minimize the risk of collapse in resource allocation problems. However, while a risk-neutral approach might be too optimistic, the worst-case scenario is often too pessimistic. To overcome these problems, we propose the use of an exponential utility function of the form $U(x) = \exp(\gamma x)$ that maps revenues to utilities. Our approach is significantly more general, because by choosing the right value of γ we are able to capture a broad range of risk-attitudes between the two extreme cases of *risk-neutral* and *worst-case* approaches. In fact, when the population becomes dangerously small, the revenue will drop and the exponential scaling will significantly penalize the utility

of such scenarios. While on one hand we lose the theoretical lower-bound guaranteed by the game against nature approach, on the other hand in any practical application there is always another level of uncertainty affecting the estimated support of the probability distributions involved (i.e. the set of available moves in the game against nature framework). Therefore, in reality, such theoretical lower-bounds are less reliable than one might expect because of the uncertainty in the model parameters. Moreover, exponential utility functions are particularly suitable for sequential decision making problems because they are the only class of utility functions (together with linear ones) that satisfy the so-called *delta property*, that allows for optimal policies that are not history-dependent and hence are easy to describe and to implement in real-world problems.

Previously, risk-sensitive policies for harvesting biological populations were considered in [Horwood, 1996]. The author extended some known optimality results to the risk-sensitive case for some continuous-time models which were originally introduced in [Clark, 1990]. However, except in the linear dynamics and quadratic cost case, the obtained results were heuristic or characterizations of the *local* optimality of the policies (with necessary but not sufficient conditions based on Pontryagin's Maximum Principle). In contrast, our main result (Theorem 2) proves that a general class of discrete-time MDPs with exponential utility functions admits a closed form optimal solution (i.e. with a given structure that allows for a compact analytical representation) that can be easily implemented by policy makers. These results are especially interesting because there are only a few MDPs with known closed form solutions. The proof is based on a generalized notion of concavity known as K-concavity originally introduced to study inventory control problems [Scarf, 1960]. In order to prove Theorem 2, we first show some useful results on the composition of K-concave functions that generalize some fundamental properties on the composition of standard concave functions.

We demonstrate the practical applicability of the proposed framework by studying the Pacific Halibut fishery management problem. This domain of applications is particularly important because marine fishery resources are one of the most poorly managed and most endangered ones due to over-exploitation [FAO, 2005]. Our results provide new evidence that a new policy based on periodic closures of the fishery should be employed, instead of the traditional policy that harvests a constant proportion of the stock every year.

2 Preliminaries

2.1 Risk-sensitive MDP formulation

We consider a class of Markov Decision Processes that arises as a general model for many renewable resources allocation problems. Examples of such real-world problems are the management of natural resources such as animal species or forests, as described in detail in [Clark, 1990; Conrad, 1999; Ermon *et al.*, 2010]. However, our model is more general and our results apply to a broad class of so called generalized inventory control problems [Scarf, 1960] that are extended with the introduction of a stock-dependent internal growth function. This framework is quite general and finds applications in many different fields. In fact, the stock variable can represent resources such as time, energy, natural and financial resources, depending on the application domains. These include natural resources management, supply chain management, vaccine distribution and pollution control [Ermon *et al.*, 2010].

We consider a discrete-time continuous-space problem where the state variable $x_n \in \mathbb{R}$ represents the available stock of the resource at time step n. The stock x_n is modeled as a Markov Chain with non-linear dynamics evolving according to the following difference equation

$$x_{n+1} = f(x_n - h_n, w_n), \tag{1}$$

where h_n is the control variable, representing the amount of resource that is extracted or harvested at time step n. The internal growth processes are modeled using a *stock recruitment function* $f(x)$ that is, in general, density dependent, in order to capture phenomena such as competition for limited food and habitat in animal populations. We suppose that there is a finite maximum stock size denoted by m. Uncertainty is introduced into the model using the random variables w_n, assumed to be *discrete* and independent. As an example, in a fishery model w_n are used to model factors such as weather conditions, climate change or the temperature of the water, all uncontrollable factors that affect the growth rate of the stock.

To complete the description of the MDP, we introduce a standard economic model used by [Clark, 1990; Conrad, 1999; Ermon *et al.*, 2010] and others. We suppose that a harvest of size h generates a revenue equal to hp, where p is a fixed constant selling price. We assume that there is a marginal harvesting cost $g(x)$ per unit harvested when the stock size is x and that each time a harvest is undertaken there is a fixed set-up cost K, independently of the size of the harvest. The net revenue (selling profits minus costs) associated with a harvest h from an initial population x is then given by

$$ph - \int_{x-h}^{x} g(y)dy - K \triangleq R(x) - R(x-h) - K, \tag{2}$$

where

$$R(x) = px - \int_0^x g(y)dy.$$

The marginal harvesting cost $g(x)$ is assumed to be a decreasing function of x, so that harvesting becomes cheaper when the stock is abundant. We define x_0 to be the stock size such that $g(x_0) = p$, that is the *zero profit level*. As a consequence for all $x > x_0$ we have that the function $R(x)$ defined in Equation (2) is non decreasing and convex.

2.2 Risk-sensitive optimization

According to Utility Theory [Von Neumann *et al.*, 2007], every rational decision maker who accepts a small number of axioms has a utility function U that maps their real-valued wealth levels w into finite real-valued utilities $U(w)$ so that they always choose the course of action that maximizes their expected utility.

In decision theory, the attitude towards risk of a decision maker is represented by the utility function U. Given a random variable v (often called a *lottery*), in general not all outcomes of v are equally significant: the decision maker ranks them according to the utility function U. In particular, a decision maker is *risk-averse* if $\mathbb{E}[U(v)] \leq U(\mathbb{E}[v])$ meaning that he prefers the expected value of the lottery over the lottery itself.

While MDPs with general utility functions can be solved using the method proposed in [Liu and Koenig, 2006], the resulting optimal policy is defined in an extended state-space and is in general history-dependent. Of particular interest for the case of sequential decision making is the case of exponential (and linear) utility functions, that as shown in [Howard and Matheson, 1972] have a constant aversion to risk that provides the kind of separability required to implement dynamic programming and lead to optimal policies that are not history dependent. Instead of considering the traditional case of linear utility function (risk-neutral approach), in this paper we consider the case of exponential utility functions, so that the resulting policies can be easily described and implemented in practice.

Since we are interested in fully observable closed loop optimization approaches, where decisions are made in stages and the manager is allowed to gather information about the system between stages, we introduce the concept of *policy*. A *policy* is a sequence of rules used to select at each period a harvest level for each stock size that can possibly occur. In particular, an *admissible policy* $\pi = \{\mu_1, \ldots, \mu_N\}$ is a sequence of functions, each one mapping stocks sizes x to harvests h, so that for all x and for all i

$$0 \leq \mu_i(x) \leq x. \tag{3}$$

We consider the problem of finding an *admissible policy* $\pi = \{\mu_i\}_{i\in[1,N]}$ that minimizes

$$\mathbb{E}^{\pi}[\exp(-\gamma J_N^{\pi}(x))]\,, \gamma > 0 \tag{4}$$

where the *lottery* is the total discounted net revenue

$$J_N^{\pi}(x) = \sum_{n=1}^{N} \alpha^n (R(x_n) - R(x_n - h_n) - K\delta_0(h_n))$$

where x_n is subject to (1) and $h_n = \mu_n(x_n)$, with initial condition $x_1 = x$ and

$$\delta_0(x) = \begin{cases} 1 & \text{if } x > 0, \\ 0 & \text{otherwise.} \end{cases}$$

and $0 < \alpha < 1$ is the discount factor. The parameter γ is called *risk sensitiveness*, and the larger γ the more risk averse the decision maker is. Intuitively, this is because small realizations of $J_N^{\pi}(x)$ are heavily penalized by the exponentiation. In the limit $\gamma \to \infty$, the smallest realization of $J_N^{\pi}(x)$ dominates the expectation in (4), as happens in the worst-case analysis situation [Ermon *et al.*, 2010].

3 Properties of K-concave Functions

To study the optimal policies for this risk-sensitive MDP model, we will make use of a property known as K-concavity. The standard definition of K-concavity is the following:

Definition 1. *A function $\beta(\cdot)$ is K-concave on an interval I if for all x, y, $x < y$, and for all $b > 0$ such that $y + b \in I$*

$$\beta(x) - \beta(y) - (x - y)\frac{\beta(y + b) - \beta(y)}{b} \leq K. \tag{5}$$

Equivalently, $\beta(\cdot)$ is K-concave if for all $x < z \in I$ and $\lambda \in [0, 1]$

$$\lambda\beta(x) + (1 - \lambda)\beta(z) \leq K\lambda + \beta(\lambda x + (1 - \lambda)z)$$

The fundamental result on K-concavity is the following Lemma:

Lemma 1. *The following properties hold:*

a) A concave function is 0-concave and hence K-concave for all $K \geq 0$.

b) If $\beta_1(q)$ and $\beta_2(q)$ are K_1-concave and K_2-concave, respectively, for constants $K_1 \geq 0$ and $K_2 \geq 0$, then $a\beta_1(q) + b\beta_2(q)$ is $(aK_1 + bK_2)$-concave for any scalars $a > 0$ and $b > 0$.

c) If $\beta(\cdot)$ is nondecreasing and concave on I and ψ is nondecreasing and K-concave on $[\inf_{x\in I} \beta(x), \sup_{x\in I} \beta(x)]$ then the composition $\psi \circ \beta$ is K-concave on I

d) If $\beta(\cdot)$ is a continuous, K-concave function on the interval $[0, m]$, then there exists scalars $0 \leq S \leq s \leq m$ such that

- *$\beta(S) \geq \beta(q)$ for all $q \in [0, m]$.*
- *Either $s = m$ and $\beta(S) - K \leq \beta(m)$ or $s < m$ and $\beta(S) - K = \beta(s) \geq \beta(q)$ for all $q \in [s, m)$.*
- *$\beta(\cdot)$ is a decreasing function on $[s, m]$.*
- *For all $x \leq y \leq s$, $\beta(x) - K \leq \beta(y)$.*

This is a dual version of the properties of K-convex functions proved for example in [Bertsekas, 1995, Section 4.2].

In order to prove our main results, we use the following novel theoretical results on the vector composition of K-concave functions, that significantly generalize the standard results on the composition of regular concave functions.

Theorem 1. *Let $\phi : \mathbb{R}^n \to \mathbb{R}$ be a concave function non-decreasing in each of its arguments such that $\phi(K + x_1, \ldots, K + x_n) \leq K + \phi(x_1, \ldots, x_n)$ for all $x_1, \ldots, x_n, K$ and let $\beta_1(x), \ldots, \beta_n(x)$ be a family of functions $\mathbb{R} \to \mathbb{R}$ such that $\beta_i(x)$ is K-concave and non-decreasing for $i = 1, \ldots, n$. Then $\gamma(x) = \phi(\beta_1(x), \ldots, \beta_n(x))$ is K-concave.*

Proof. We have

$$\begin{aligned}
\gamma(x) + (1 - \lambda)\gamma(z) = \\
\phi(\beta_1(x), \ldots, \beta_n(x)) + (1 - \lambda)\phi(\beta_1(z), \ldots, \beta_n(z)) \leq \\
\phi(\lambda\beta_1(x) + (1 - \lambda)\beta_1(z), \ldots, \lambda\beta_n(x) + (1 - \lambda)\beta_n(z)) \leq \\
\phi(K\lambda + \beta_1(\lambda x + (1 - \lambda)z), \ldots, K\lambda + \beta_n(\lambda x + (1 - \lambda)z)) \leq \\
K\lambda + \phi(\beta_1(\lambda x + (1 - \lambda)z), \ldots, \beta_n(\lambda x + (1 - \lambda)z)) = \\
K\lambda + \gamma(\lambda x + (1 - \lambda)z)
\end{aligned}$$

□

Lemma 2. *The log-sum-exp function $\phi : \mathbb{R}^n \to \mathbb{R}$ defined as $\phi(z_1, \ldots, z_n) = -\log \sum e^{-z_i}$ is a concave function non-decreasing in each of its arguments such that $\phi(K + x_1, \ldots, K + x_n) \leq K + \phi(x_1, \ldots, x_n)$ for all $x_1, \ldots, x_n, K \in \mathbb{R}$.*

Proof. From [Boyd and Vandenberghe, 2004] we have that

$$f(x) = \log(e^{x_1} + \ldots + e^{x_n})$$

is convex and non decreasing in each of its arguments, so

$$\phi(z_1, \ldots, z_n) = -\log \sum e^{-z_i}$$

is a concave function non-decreasing in each of its arguments. Moreover

$$\phi(K + x_1, \ldots, K + x_n) = K - \log\left(\sum e^{-z_i}\right)$$

□

4 Main Results

4.1 Optimal policy

A policy π is called an *optimal* N-period policy if $\mathbb{E}^\pi[\exp(-\gamma J_N^\pi(x))]$ attains its infimum over all admissible policies at π for all x. We call

$$S_N^\gamma(x) = \inf_{\pi \in \Pi} \mathbb{E}^\pi[\exp(-\gamma J_N^\pi(x))]$$

the *optimal expected utility function*, where Π represents the set of all admissible policies. As shown in [Marcus *et al.*, 1997], by making use of the *Delta property* the problem can be decomposed and solved by Dynamic Programming using Bellman-like equations:

$$S_t^\gamma(x) = \min_{0 \le h \le x} \mathbb{E}[e^{-\gamma\alpha(R(x)-R(x-h)-K\delta_0(h))} S_{t-1}^{\gamma\alpha}(f(x-h, w_t))] \quad (6)$$

for all $t > 0$, with boundary condition $S_0^\gamma(x) = 1$. Equation (6) can also be rewritten in terms of the escapement $z = x - h$ as

$$S_t^\gamma(x) = \min_{0 \le z \le x} e^{-\gamma\alpha(R(x)-R(z)-K\delta_0(x-z))} \mathbb{E}[S_{t-1}^{\gamma\alpha}(f(z, w_t))] \quad (7)$$

Given the presence of fixed costs K, a harvest is undertaken if and only if there exists $0 \le z \le x$ such that

$$e^{-\gamma\alpha(R(x)-R(z)-K)} \mathbb{E}[S_{t-1}^{\gamma\alpha}(f(z, w_t))] < \mathbb{E}[S_{t-1}^{\gamma\alpha}(f(x, w_t))]$$

or equivalently if and only if

$$-\gamma\alpha(R(x) - R(z) - K) + \log \mathbb{E}[S_{t-1}^{\gamma\alpha}(f(z, w_t))] < \log \mathbb{E}[S_{t-1}^{\gamma\alpha}(f(x, w_t))] \quad (8)$$

We introduce

$$P_t^\gamma(x) = -R(x) - \frac{1}{\gamma\alpha} \log \mathbb{E}[S_{t-1}^{\gamma\alpha}(f(x, w_t))] \quad (9)$$

so from (8) we get that a harvest is undertaken at time t if and only if there exists $0 \le z \le x$ such that

$$P_t^\gamma(x) < P_t^\gamma(z) - K. \quad (10)$$

If we prove that $P_t^\gamma(x)$ is continuous and strictly K-concave, then using Lemma 1-d we can characterize the optimal policy at time step t. In fact by Lemma 1-d there exists two thresholds $S_t \le s_t$ such that condition (10) is met if and only if $x > s_t$. If $x > s_t$, the optimal escapement z in Equation (7) is exactly the other threshold S_t by the properties proved in Lemma 1-d. This policy is known in the Operations Research community [Scarf, 1960] as a nonstationary $S - s$ policy, because the levels S_t and s_t depend on the time index t.

Our main result is that we can prove that a nonstationary $S - s$ policy is optimal for the risk-sensitive MDP model we considered. In particular, if the marginal cost function g satisfies

$$\tau = G(m) - mg(m) < K\left(\frac{1-\alpha}{\alpha}\right) \quad (11)$$

(meaning that g does not decrease too rapidly) where G is the integral of g and m is the maximum stock size, then the following theorem holds:

Theorem 2. *For any setup cost $K > 0$, risk sensitiveness $\gamma > 0$ and length of the management horizon N, if w_n are independently distributed, $f(\cdot, w)$ is nondecreasing and concave for any w, and g is non increasing and satisfies condition (11), then the functions $P_n^\gamma(x)$ defined as in (9) are continuous and K-concave for all $n = 1, \ldots, N$. Therefore, there exists a non-stationary $S - s$ policy that is optimal for the risk-sensitive optimization problem. The corresponding optimal expected log-utility functions $-1/\gamma \log S_n^\gamma(x)$ are continuous, nondecreasing and K-concave for all $n = 1, \ldots, N$.*

Proof. From Equation (11) there exists $k \in \mathbb{R}$ such that

$$(K + \tau)\alpha < k < K \quad (12)$$

The proof is by induction on the length of the control horizon N. The base case $N = 0$ is trivial because $S_0^\gamma(x) = S_N^\gamma(x) = 1$ for all x, and therefore $-\log(S_N^\gamma(x)) = 0$ is continuous, nondecreasing and $k\gamma$-concave. Now we assume that $-\log S_{n-1}^\gamma(x)$ is continuous, nondecreasing and $k\gamma$-concave, and we show that $P_n^\gamma(x)$ is continuous and K-concave, and that $-\log S_n^\gamma(x)$ is continuous, nondecreasing and $k\gamma$-concave.

Using the composition result of Lemma 1, since $f(\cdot, w)$ is nondecreasing and concave for all w and by the inductive hypothesis, we have that $-\log S_{n-1}^\gamma(f(x, w))$ is $k\gamma$-concave. From the definition of expectation we obtain

$$\log \mathbb{E}[S_{n-1}^\gamma(f(x, w_n))] = \log\left(\sum_i e^{\log p(w_n^i) + \log S_{n-1}^\gamma(f(x, w_n^i))}\right)$$

so that using Theorem 1 and Lemma 2, we get that $-\log \mathbb{E}[S_{n-1}^\gamma(f(x, w_n))]$ is $k\gamma$-concave as well.

By Equation (9) using Lemma 1-b, since $-R(x)$ is concave and by induction $-\log \mathbb{E}[S_{t-1}^{\gamma\alpha}(f(x, w_t))]$ is $k\gamma\alpha$-concave, $P_n^\gamma(x)$ is k-concave and therefore also K-concave. By Equation (9) we get that $P_n^\gamma(x)$ is continuous, because by the inductive hypothesis $S_{n-1}^{\gamma\alpha}(x)$ and $R(x)$ are continuous.

Since $P_n^\gamma(x)$ is K-concave, by Lemma 1-d there exists two thresholds $S_n \le s_n$ associated with $P_n^\gamma(x)$ with the properties of Lemma 1-d. From the Dynamic Programming Equation (7) and what we have shown on condition (10), we have that

$$S_n^\gamma(x) = \begin{cases} e^{-\gamma\alpha(P_n^\gamma(x)+R(x))} & \text{if } x \le s_n \\ e^{-\gamma\alpha(P_n^\gamma(S_n)+R(x)-K)} & \text{if } x > s_n \end{cases} \quad (13)$$

$$-\log S_n^\gamma(x) = \begin{cases} \gamma\alpha(P_n^\gamma(x) + R(x)) & \text{if } x \le s_n \\ \gamma\alpha(P_n^\gamma(S_n) + R(x) - K) & \text{if } x > s_n \end{cases} \quad (14)$$

From Equation (13) we get $S_n^\gamma(x)$ is continuous because $P_n(x)$ and $R(x)$ are continuous and by definition $P_n^\gamma(s_n) + R(s_n) = P_n^\gamma(S_n) + R(s_n) - K$. To show $-\log S_n^\gamma(x)$ is nondecreasing, we consider several cases. First notice that since harvesting below the zero profit level x_0 is not profitable and reduces the marginal growth of the stock, it must be the case that $s_n \ge S_n \ge x_0$. Given $x_2 > x_1 > s_n \ge x_0$ we have:

$$-\log S_n^\gamma(x_2) + \log S_n^\gamma(x_1) = \gamma\alpha(R(x_2) - R(x_1)) \ge 0,$$

because R is nondecreasing for $x \ge x_0$. In the case $0 \le x_1 < x_2 \le s_n$ we obtain:

$$-\log S_n^\gamma(x_2) + \log S_n^\gamma(x_1) = \log \frac{\mathbb{E}[S_{n-1}^{\gamma\alpha}(f(x_1, w_n))]}{\mathbb{E}[S_{n-1}^{\gamma\alpha}(f(x_2, w_n))]}.$$

Then is must be the case that $-\log S_n^\gamma(x_2) + \log S_n^\gamma(x_1) \ge 0$ because if $-\log S_{n-1}^{\gamma\alpha}(x)$ is nondecreasing , then $S_{n-1}^{\gamma\alpha}(x)$ must be non-increasing, so for each ω, $S_{n-1}^{\gamma\alpha}(f(x_1, \omega)) \ge S_{n-1}^{\gamma\alpha}(f(x_2, \omega))$.

In order to show that $-\log S_n^\gamma(x)$ is $k\gamma$-concave, by Equation (12) it is sufficient to show that $-\log S_n^\gamma(x)$ is $(K + \tau)\alpha\gamma$-concave using definition (5). When $s_n < x < y$, Equation (5) holds because $R(\cdot)$ is τ-concave. When $x < y \le s_n$, $-\log S_n^\gamma(x) = \gamma\alpha(P_n^\gamma(x) + R(x))$ and therefore using by Lemma 1-b Equation (5) holds because $P_n^\gamma(x)$ is K-concave and $R(\cdot)$ is τ-concave. Finally, when $x \le s_n < y$ Equation (5) can be written as

$$\log S_n^\gamma(y)/S_n^\gamma(x) - (x-y)\frac{-\log S_n^\gamma(y+b) + \log S_n^\gamma(y)}{b} =$$
$$\gamma\alpha(P_n^\gamma(x) - P_n^\gamma(S_n) + K + R(x) - R(y) - (x-y)\frac{R(y+b) - R(y)}{b}) \le$$
$$\gamma\alpha\left(K + R(x) - R(y) - (x-y)\frac{R(y+b) - R(y)}{b}\right)$$
$$\le \gamma\alpha(K + \tau) \le \gamma k.$$

because $P_n^\gamma(x) \le P_n^\gamma(S_n)$ and $R(\cdot)$ is τ-concave. □

Theorem 2 completely characterizes the structure of the optimal policy, but it does not provide an analytical solution for the values of S_n, s_n. These values can be computed numerically by discretizing the continuous domains and solving the resulting MDP by Dynamic Programming. Our knowledge on the structure of the optimal policy guarantees the *consistency* of the method, that is a (uniform) convergence of the discretized solution to the true one as the discretization step goes to zero. Moreover, the a priori knowledge on the structure of the optimal policy reduces the computational complexity of the algorithm used to compute the numerical values of the $S - s$ thresholds characterizing the policy. In fact the policy for a given time step t is completely characterized by the corresponding threshold s_t (that can be computed for example by bisection) and by the optimal escapement S_t associated with any state $x > s_t$.

5 The Pacific Halibut Marine Fishery

The North American Pacific halibut fishery is one of the most important fisheries of the western coast of North America. It is jointly managed by the governments of U.S. and Canada through the International Pacific Halibut Commission (IPHC) to provide rational management and avoid overfishing. In order to do that, every year the IPHC decides the *total allowable catch* (TAC), represented in our model by the decision variable h_n.

5.1 Management problem formulation

We consider the biological model presented in [Ermon *et al.*, 2010] to study Area 3A of the Pacific Halibut fishery, one of the 10 major regulatory areas in which waters are divided. Their model is based on 33 years of data from 1975 to 2007, extracted from IPHC reports, and provides a very good fit with the historical stock sizes. We briefly summarize their model.

In the context of a fishery management, the state variable x_n represents the total biomass and the growth of the resource is due to reproduction. To model these growth processes, we consider a stochastic version of the Beverton-Holt model:

$$x_{n+1} = f(s_n, w_n) = (1-m)s_n + w_n \frac{r_0 s_n}{1 + s_n/M}, \quad (15)$$

where $s_n = x_n - h_n$ is the escapement of fishing in year n. According to Equation (15), the stock size at the beginning of the next breeding season is given by the sum of two terms: the fraction of the population that survives natural mortality and the new recruitment. To account for variable factors and uncertainty (such as weather conditions or the temperature of the water) that affect the growth rate, we introduce stochasticity into the system in the form of seasonal shocks w_n that influence the new recruitment part. We will (a priori) assume that w_n are independent identically distributed *uniform* random variable with a finite support $I_w = [1 - 0.11, 1 + 0.06]$. The values of the parameters of Equation (15) fitted to the Halibut population in Area 3A are taken from [Ermon *et al.*, 2010] and reported in Table 1.

Parameter	Value
q	$9.07979\ 10^{-7}$
b	2.55465
p	$4,300,000\$$ / (10^6 pounds)
K	$5,000,000\$$
c	$200,000\$$ / 1000 skate soaks
α	$1/(1+0.05)$
m	0.15
M	$196.3923\ 10^6$ pounds
r_0	0.543365

Table 1: Parameter values for area 3A.

For the resource economics aspects of the model, we assume that there are two categories of costs involved: *fixed costs* (such as vessel repairs costs, license and insurance fees) and *variable costs* (such as fuel and wages). While the first component is independent of the size of the harvest, the second one depends on the effort involved. The sum of all the fixed costs will be denoted with K, and these costs will be incurred every year in which a harvest is undertaken.

The model for variable costs will capture the fact that harvesting is cheaper when the stock size is abundant. In particular, we assume that a harvest of size h that brings the stock size from x to $x-h$ results in variable costs given by

$$c\int_{x-h}^{x}\frac{1}{qy^b}\,dy$$

for some q and b. The function $c/(qx^b)$ is precisely the marginal harvesting cost function $g(x)$ introduced earlier. Furthermore, we assume that there is a fixed constant selling price p independent of the size of the harvest and we assume a fixed discount factor $\alpha = 1/(1+0.05)$. The values of the parameters are estimated in [Ermon *et al.*, 2010] and are reported in Table 1.

5.2 Optimal risk-sensitive policy

We compute the optimal risk-sensitive policy by solving Bellman Equations (7) using a Dynamic Programming approach for a discretized problem. In order to solve the problem, we need to discretize both the state space and the control space. In the experiments presented in this section we used a discretization step $\Delta = 0.25 \times 10^6$ pounds. Moreover, we also discretize the support I_w of the random variables involved, thus obtaining a completely discrete MDP that we can solve.

We computed the optimal risk-sensitive policy (R-S) for the Pacific Halibut fishery in Area 3A for several values of γ and for a management length $N = 33$ years. As predicted by Theorem 2, the optimal policy is of $S-s$ type. In particular, given a fixed *risk-sensitiveness* γ, an optimal policy is characterized by tuples S_n, s_n for $n = 1,\ldots,33$; at year n, the optimal policy prescribes to harvest the stock down to S_n if and only if the current stock is larger than s_n. This type of management policy involves *periodic closures* of the fishery, where for several consecutive years no action is taken so that the population can recover. In particular, after it has been harvested down to S, it is optimal to wait for the stock to become larger than s before opening the fishery again.

Periodic closures of the fishery are also prescribed by the worst-case scenario policy considered in [Ermon *et al.*, 2010], but the thresholds to be used are different, and in the case of a risk-sensitive policy they are dependent on the risk-sensitiveness γ used. However, this approach is very different from the Constant Proportional Policy (CPP) that has been traditionally used to manage the Halibut fishery, that harvests a fixed fraction of the current stock level every year in order to maintain the population at at level perceived to be optimal (also known as a *constant escapement policy* [Clark, 1990])

As we show in Table 2, $S-s$ policies are superior to historical harvests and CPP policy in terms of total discounted revenue (the experiment is initialized with an initial stock size $x_1 = X_{1975} = 90.989$ million pounds), both assuming a worst-case realization of the randomness and assuming *uniform* samples from I_w. While the worst-case policy is indeed (by definition!) optimal for a worst-case realization, the risk sensitive approach is far more versatile. In fact by choosing a suitable value of γ, we can take a more balanced attitude toward risk, giving us a much broader spectrum of possible alternatives, that range from close to worst-case (large values of γ) to something close to risk-neutral optimization (small value of γ).

Policy	**Average (10^7\$)**	**Worst case (10^7\$)**
R-S $\gamma = 0.1$	113.662	90.175
R-S $\gamma = 0.5$	113.581	90.244
R-S $\gamma = 2.0$	113.475	90.401
Worst case	112.940	90.514
Historical	96.491	70.686
Average CPP	90.709	65.185

Table 2: Policy Comparison

6 Conclusions and future work

In this paper, we consider a general class of MDPs used to model renewable resources allocation problems and we prove the optimality of $S-s$ policies in a risk-sensitive optimization framework. Our proof is based on a generalization of concavity known as K-concavity. As part of our proof, we significantly generalize some fundamental results on the composition of traditional concave functions.

Our framework generalizes previous approaches such as the worst-case analysis, since it provides more balanced approaches toward risk. In particular, it allows a range of risk behaviors, from a worst-case approach (for large γ) to a risk-neutral approach (for small γ), as well as a broad spectrum of intermediate cases.

We apply our results to the Pacific Halibut fishery management problem, and find new evidence that a cyclic policy involving periodic closures of the fishery should be employed instead of the traditional constant escapement policies.

We are currently working towards an extension of the K-concavity concept to multidimensional spaces, in order to generalize our results on the optimality of S-s policies to multidimensional settings. This would allow us to capture interesting scenarios involving for example the interactions between multiple species. Another interesting research direction is to examine whether the (multidimensional) K-concavity concept arises in other traditional continuous state space or hybrid MDPs (e.g. in robotic applications) from the AI literature [Sutton and Barto, 1998], or can be used to efficiently compute approximate threshold-based policies to less structured scenarios.

7 Acknowledgments

This research is funded by NSF Expeditions in Computing grant 0832782.

References

[Bertsekas, 1995] D.P. Bertsekas. *Dynamic programming and optimal control*. Athena Scientific Belmont, MA, 1995.

[Boyd and Vandenberghe, 2004] S.P. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge Univ Pr, 2004.

[Clark, 1990] C.W. Clark. *Mathematical bioeconomics: the optimal management of renewable resources*. Wiley New York, 1990.

[Clark, 2007] W.C. Clark. Sustainability Science: A room of its own. *PNAS*, 104:1737–1738, 2007.

[Conrad, 1999] J.M. Conrad. *Resource economics*. Cambridge Univ Pr, 1999.

[Ermon *et al.*, 2010] S. Ermon, J. Conrad, C. Gomes, and B. Selman. Playing games against nature: optimal policies for renewable resource allocation. *Proc. of The 26th Conference on Uncertainty in Artificial Intelligence*, 2010.

[FAO, 2005] FAO. Review of the state of world marine fishery resources. *FAO Fisheries Technical Paper*, 2005.

[Gomes, 2009] C. Gomes. Computational Sustainability Computational Methods for a Sustainable Environment,Economy, and Society. *The Bridge, National Academy of Engineering*, 39(4), 2009.

[Horwood, 1996] JW Horwood. Risk-sensitive optimal harvesting and control of biological populations. *Mathematical Medicine and Biology*, 13(1):35, 1996.

[Howard and Matheson, 1972] R.A. Howard and J.E. Matheson. Risk-sensitive Markov decision processes. *Management Science*, 18(7):356–369, 1972.

[Liu and Koenig, 2006] Yaxin Liu and Sven Koenig. Probabilistic planning with nonlinear utility functions. In *ICAPS-2006*, pages 410–413, 2006.

[Marcus *et al.*, 1997] S.I. Marcus, E. Fernández-Gaucherand, D. Hernández-Hernandez, S. Coraluppi, and P. Fard. Risk sensitive Markov decision processes. *Systems and Control in the 21st Century*, 29, 1997.

[Scarf, 1960] H. Scarf. The Optimality of (S, s) Policies in the Dynamic Inventory Problem. *Stanford mathematical studies in the social sciences*, page 195, 1960.

[Sutton and Barto, 1998] R.S. Sutton and A.G. Barto. *Reinforcement learning: An introduction*. The MIT press, 1998.

[Von Neumann *et al.*, 2007] J. Von Neumann, O. Morgenstern, A. Rubinstein, and H.W. Kuhn. *Theory of games and economic behavior*. Princeton Univ Pr, 2007.