JOTA manuscript No.
(will be inserted by the editor)

Optimal Stopping with a Probabilistic Constraint

Aaron Zeff Palmer · Alexander Vladimirsky

Received: date / Accepted: date

Abstract We present an efficient method for solving optimal stopping problems with a probabilistic constraint. The goal is to optimize the expected cumulative cost, but constrained by an upper bound on the probability that the cost exceeds a specified threshold. This probabilistic constraint causes optimal policies to be time-dependent and randomized, however, we show that an optimal policy can always be selected with "piecewise-monotonic" time-dependence and "nearly-deterministic" randomization. We prove these properties using the Bellman optimality equations for a Lagrangian relaxation of the original problem. We present an algorithm that exploits these properties for computational efficiency. Its performance and the structure of optimal policies are illustrated on two numerical examples.

Keywords Stochastic Optimal Control · Stopping-Times · Dynamic Programming · Chance Constraint

Mathematics Subject Classification (2000) 49L20 65K15 60G40

1 Introduction

Controlled stochastic processes arise in a wide variety of practical applications. It is frequently useful to consider different objective/utility functions

Communicated by Kyriakos G. Vamvoudakis.

✉ Aaron Zeff Palmer
University of British Columbia
Vancouver, BC, Canada
azp@math.ubc.ca

Alexander Vladimirsky
Cornell University
Ithaca, NY, USA
vladimirsky@cornell.edu

and optimize them with respect to the control policies. Many well established techniques exist to optimize either the expected total cost/reward [1], the probability of the desired outcome [2,3], the value-at-risk [4], or the expectation of a nonlinear utility function (e.g., leading to "risk-sensitive" controls) [1]. In many applications practitioners desire to optimize the expected performance among the control policies satisfying some hard constraint on the worst-case-scenario of the accrued cost; e.g., see [5] for examples in optimal routing on stochastic networks. However, such approaches are not suitable for applications where "the worst case scenario" is undefined. E.g., there is no way to guarantee that a particle undergoing (controlled) Brownian motion will reach the target before any specific deadline.

Our goal is to develop methods for optimizing the expected cost of feedback policies but under a *probabilistic constraint* on a specific undesirable outcome. When minimizing expected cost, the dynamic programming principle states that an optimal policy also minimizes the remaining expected cost from any node reached under that policy. Unfortunately, a probabilistic constraint destroys this property since we are constraining the probability of an event over an aggregate of trials. However, allowing for randomized policies leads to the existence of a Lagrange multiplier such that the dynamic programming equations hold for a Lagrangian relaxation of the original problem [6]. Specifically, the optimal policies minimize the expected cost penalized by the probabilistically constrained value with a Lagrange multiplier. The Bellman equations for the penalized problem lead to techniques to both analyze and compute optimal policies.

To illustrate this general approach, we focus on a particularly simple example in discrete time and space: an optimal stopping problem for a random walk on a graph (formulated in §2). Its simplicity allows us to emphasize the analytic properties of optimal policies (see §3) circumventing many technical difficulties present in more realistic applications in financial engineering or robotic path planning. We exploit the structure of the problem to introduce efficient algorithms with rigorous algorithmic analysis in §4, which are then tested on discretizations of 1D continuous optimal stopping problems in §5. We conclude by listing several directions for future work in §6.

Relation to prior work. The study of probabilistic constraints in optimal control goes back to at least the work of White [7], where he considers ergodic problems with constraints on the frequency of visiting parts of the state space. His work is based on the probability distributions over the set of deterministic policies (where the choice among them is made at the beginning). He demonstrates how deterministic policies can be used to determine the Lagrange multiplier – the approach similar to our Proposition 3.1 and Algorithm 4.1 of §4. Using linear programming, White shows that an optimal policy can be selected as a randomized choice between two deterministic policies.

More recently, two approaches to probabilistically constrained stochastic optimal control problems were considered by Pfeiffer [8]. One approach uses a dynamic programming principle and deterministic policies on a state space that is extended to include the constrained value. The second is a Lagrangian

relaxation approach similar to ours, which Pfeiffer realizes as a Legendre transform of the value function from the first approach. Pfeiffer finds that the Lagrangian relaxation provides the more efficient computational method; see [8] for detailed algorithmic/numeric analysis. The state space extension increases the dimension of the problem and introduces a new approximation of the set of possible constrained values, both of which are computationally undesirable.

In contrast with these prior papers, we focus on the randomized feedback (Markov) stopping-policies and the probabilistic constraint on the total cumulative cost. We define a subclass of "piecewise-monotonic *nearly-deterministic*" stopping-policies, and we show that it contains the optimal solution to the original problem. However, the presence of 'degenerate points' introduces computational challenges in computing that solution directly. Instead, we start by producing a pair of piecewise-monotonic *deterministic* policies, one feasible and the other super-optimal (see Algorithm 4.1), which are then efficiently mixed by Algorithms 4.2 and 4.3 to produce a nearly-deterministic output. Both of these deterministic policies could also be computed by the more general approaches in [7] and [8], but the mixing procedure is significantly different since White's and Pfeiffer's resulting policies are non-Markovian and the decision may be randomized at multiple points. In Theorem 4.1 we prove the convergence of our algorithm to the optimal stopping-policy in finitely many steps.

2 Stochastic Optimal Stopping Problem Formulation

The domain is a finite undirected graph with vertices X. Elements of $X_0 \subset X$ are designated as target nodes, and the optimal stopping problem is posed on $X_1 = X \backslash X_0$. We use $N(x) \subset X \backslash \{x\}$ to denote the set of vertices adjacent to $x \in X_1$. For simplicity, we will assume that $N(x)$ is non-empty and each $x \in X_1$ is path-connected to X_0. Supposing that a transition probability function $p : X \to [0, 1]$ is known, we consider a stopped random walk Ξ_t on X with (random) stopping-time $\tau \in \mathbb{N}$, time parameter $t \in \{0, \ldots, \tau\}$, and individual transition probabilities:

$$P\big(\Xi_{t+1} = y \mid \Xi_t = x\big) = \begin{cases} p(x)/|N(x)|, & \text{if } y \in N(x), \\ 1 - p(x), & \text{if } y = x, \\ 0, & \text{otherwise.} \end{cases}$$

We assume that $p(x) = 0$ for $x \in X_0$ and $p(x) > 0$ for $x \in X_1$. An initial distribution is prescribed on X_1 such that $P(\Xi_0 = x) = \Phi_0(x)$, for Φ_0 a non-negative function on X_1 that sums to 1. Each step prior to termination costs $k > 0$, and the decision to terminate at $x \in X$ costs $\psi(x)$. We assume that $\psi(x) > 0$ for $x \in X_1$ and $\psi(x) = 0$ for $x \in X_0$. If the process terminates at time τ, the total incurred cost is

$$\Upsilon = k\tau + \psi(\Xi_\tau). \tag{1}$$

We encode the termination decision as a randomized feedback stopping-policy, of the class $\mathcal{A}^R = \{A : X_1 \times \mathbb{N} \to [0,1]\}$, where $A(x,t)$ is the probability the process terminates given $\Xi_t = x$. Throughout the paper, the term policy, used without additional qualifiers, will refer to elements of $\mathcal{A}^R$. We assume the process always terminates prior to or upon entering X_0 for the first time as any other decision always increases the cost. We sometimes refer to $A(x,t)$ for $x \in X_0$, which is always 1. We now consider the expected cost and probabilistic constraint to be functions of the policy, although we do not make this explicit in the notation. The precise calculation of these quantities from $A(x,t)$ is in §3.2. Based on our assumptions that x is path-connected to X_0 and $p(x) > 0$ for each $x \in X_1$, the case $\tau = \infty$ occurs with probability zero. Similarly, it is not hard to show that $E[\tau] < \infty$ and thus $E[\Upsilon] < \infty$ for any policy.

We focus on a probabilistically constrained optimal stopping (PCOS) problem with constant non-negative parameters π and ϵ:

PCOS *Given $P(\Xi_0 = x) = \Phi_0(x)$, find $A^* \in \mathcal{A}^R$ that minimizes the expected cost, $E[\Upsilon]$, subject to the probabilistic constraint, $P(\Upsilon > \pi) \leq \epsilon$.*

We will refer to the expected cost of an optimal policy A^*, as E^* or the *the value* of PCOS and to the corresponding constrained value as P^*. The optimal policy depends not only on π and ϵ, but also on Φ_0. If the constraint is satisfied for a given policy, we say that this policy is feasible.

We briefly remark on a couple important policy subclasses and the special case of PCOS that does not include a constraint. The deterministic feedback stopping-policies are $\mathcal{A}^D = \{A : X_1 \times \mathbb{N} \to \{0,1\}\}$, and the process terminates at time t if $A(\Xi_t, t) = 1$. The stationary deterministic feedback policies, $\mathcal{A}^S$, are the policies in $\mathcal{A}^D$ that do not depend on time. The unconstrained problem ($\epsilon \geq 1$) over the set $\mathcal{A}^R$ has an optimal policy $\overline{A}^* \in \mathcal{A}^S$, and $\overline{A}^*$ also does not depend on Φ_0. We demonstrate how to determine $\overline{A}^*$ in §3.1.

3 Optimality Criteria

3.1 Unconstrained Problem

As a preliminary step to solving PCOS, we consider the unconstrained problem to minimize the expected cost. An optimal policy may be determined from the optimal cost-to-go function, which is defined by $U(x) = \inf_{A \in \mathcal{A}^R}\{E[\Upsilon \mid \Xi_0 = x]\}$ for $x \in X_1$, and $U(x) = 0$ for $x \in X_0$. The dynamic programming principle implies that U solves the Bellman equations for $x \in X_1$:

$$U(x) = \min\big\{\psi(x), M[U](x) + k\big\}; \tag{2}$$

where the difference operator, $M : \mathbb{R}^{|X|} \to \mathbb{R}^{|X_1|}$, is defined for functions on X and evaluated at a point $x \in X_1$ as

$$M[U](x) = \big(1 - p(x)\big)U(x) + \sum_{\xi \in N(x)} \frac{p(x)}{|N(x)|} U(\xi). \tag{3}$$

The unconstrained problem can be reduced to a simple form of a stochastic shortest path problem (Chapter 3.4 of [9]). The latter has a unique solution that can be found by value iterations, which is covered in [10] and Chapters 2 and 3.4 of [9]. The optimal policy at $x \in X_1$ is $\overline{A}^*(x) = 0$ (diffuse) if $U(x) < \psi(x)$ and $\overline{A}^*(x) = 1$ (terminate) if $M[U](x) + k > \psi(x)$. In the *degenerate* case that $M[U](x) + k = \psi(x)$, either choice is optimal.

3.2 Constrained Optimality Equations

For PCOS with $0 < \epsilon < 1$, the optimal policies are generally neither stationary (in $\mathcal{A}^S$) nor even deterministic (in $\mathcal{A}^D$). We let $T_1 = \lfloor \pi/k \rfloor$ and $T_0(x) = \lfloor (\pi - \psi(x))/k \rfloor$. Terminating a process at (x,t) for $t \le T_0(x)$ satisfies $\Upsilon \le \pi$, whereas terminating at the same position and $t > T_0(x)$ does not.

Observation 3.1 *To determine if there exists a feasible policy for PCOS, we consider the policy, A^m, that minimizes the constraint. It is given simply by $A^m(x,t) = 1$ if $t \le T_0(x)$ and $A^m(x,t) = 0$ otherwise. We define the minimal constrained value to be $P^m = P(\Upsilon > \pi)$ for this policy. If $P^m > \epsilon$, then there is no feasible policy for PCOS.*

The calculation of $A \mapsto E[\Upsilon]$ and $A \mapsto P(\Upsilon > \pi)$, in equations (4), (5) *and* (6)*, shows that both maps are continuous. The existence of minimizers over $\mathcal{A}^R$ can be obtained by reducing PCOS to a finite time horizon version, for which $\mathcal{A}^R$ is compact. Indeed, $\Upsilon \le \pi$ definitely fails whenever the stopping-time τ exceeds $T_1 = \lfloor \pi/k \rfloor$. Thus, any feasible $A \in \mathcal{A}^R$ will remain feasible and will not increase in expected cost if we set $A(x,t) = \overline{A}^*(x)$ for all $t \ge T_1$.*

The observation above suggests a reformulation of the problem with finite time horizon. We let the cost equal $k\tau + \psi(\Xi_\tau)$ as before if $\tau \le T_1$, and otherwise we take the cost to be $kT_1 + U(\Xi_{T_1})$. Recall that $U(x)$ is the value function of the unconstrained problem, or, equivalently, the expected cost of the policy given by $A(x,t) = \overline{A}^*(x)$ for all t.

We introduce three new functions, Φ, R and Z, in order to capture the dependence of $E[\Upsilon]$ and $P(\Upsilon > \pi)$ on $A(x,t)$. These functions depend on the policy, although this is not made explicit by the notation. We define $\Phi(x,t) = P(\Xi_t = x)$ to be the probability of finding the process at position x and time t prior to termination. At the initial time, $\Phi(x,0) = \Phi_0(x)$ is given, and for $x \in X$ and $0 < t \le T_1$

$$\begin{aligned}\Phi(x,t) =& \big(1 - A(x,t-1)\big)\big(1 - p(x)\big)\Phi(x,t-1) \\ &+ \sum_{\xi \in N(x)} \big(1 - A(\xi,t-1)\big)\frac{p(\xi)}{|N(\xi)|}\Phi(\xi,t-1).\end{aligned} \tag{4}$$

Let $Z(x,t) = E[\Upsilon - kt \mid \Xi_t = x]$ be the expected cost-to-go, and $R(x,t) = P(\Upsilon > \pi \mid \Xi_t = x)$ be the conditional constrained value. The expected cost

and constrained value are recovered by the inner products:

$$E[\Upsilon] = \sum_{\xi \in X_1} \Phi_0(\xi) Z(\xi, 0), \tag{5}$$

$$P(\Upsilon > \pi) = \sum_{\xi \in X_1} \Phi_0(\xi) R(\xi, 0). \tag{6}$$

The functions Z and R satisfy backward Kolmogorov equations that are adjoint to the evolution of Φ. At time T_1, $\Upsilon > \pi$ with probability one so the terminal conditions are $R(x, T_1) = 1$ and $Z(x, T_1) = U(x)$ for all $x \in X_1$. If $\Xi_t \in X_0$ for $t \leq T_1$ then $\Upsilon = kt \leq \pi$ so $R(x, t) = 0$ and $Z(x, t) = 0$ for all $x \in X_0$. The backwards evolutions at $x \in X_1$ and $0 \leq t < T_1$ are given by

$$R(x,t) = \big(1 - A(x,t)\big) M[R(\cdot, t+1)](x) + A(x,t)\chi(x,t), \tag{7}$$

$$Z(x,t) = \big(1 - A(x,t)\big)\big(M[Z(\cdot, t+1)](x) + k\big) + A(x,t)\psi(x), \tag{8}$$

where $\chi(x, t)$ encodes whether $\Upsilon \leq \pi$ for termination with $\Xi_t = x$;

$$\chi(x,t) = \begin{cases} 1, & kt + \psi(x) > \pi, \\ 0, & kt + \psi(x) \leq \pi. \end{cases} \tag{9}$$

From the definition of $T_0(x)$, $\chi(x, t)$ is also the indicator function of the set where $t > T_0(x)$, when termination causes failure of the constraint.

With the definitions of Z and R, and the evolution equations (7) and (8), the following relationships hold for $0 \leq t < T_1$:

$$\begin{aligned} E[\Upsilon] = &\sum_{s=0}^{t-1} \sum_{\xi \in X_1} \Phi(\xi, s)\Big(\big(1 - A(\xi, s)\big)k + A(\xi, s)\psi(\xi)\Big) \\ &+ \sum_{\xi \in X_1} \Phi(\xi, t)\Big(\big(1 - A(\xi, t)\big)\big(M[Z(\cdot, t+1)](\xi) + k\big) + A(\xi, t)\psi(\xi)\Big), \end{aligned} \tag{10}$$

$$\begin{aligned} P(\Upsilon > \pi) = &\sum_{s=0}^{t-1} \sum_{\xi \in X_1} \Phi(\xi, s) A(\xi, s) \chi(\xi, s) \\ &+ \sum_{\xi \in X_1} \Phi(\xi, t)\Big(\big(1 - A(\xi, t)\big) M[R(\cdot, t+1)](\xi) + A(\xi, t)\chi(\xi, t)\Big). \end{aligned} \tag{11}$$

In (10) and (11), we have isolated the dependence of $E[\Upsilon]$ and $P(\Upsilon > \pi)$ on $A(x, t)$ for fixed (x, t).

We will apply the KKT conditions with the constraints $A(x, t) \in [0, 1]$ and $P(\Upsilon > \pi) \leq \epsilon$. We first check the Mangasarian-Fromowitz constraint qualification condition. If we assume that the minimal constrained value satisfies $P^m < \epsilon$, it is sufficient to show that for every feasible policy, $A \in \mathcal{A}^R$, there exists a variation, B, such that for all sufficiently small $\delta > 0$: $A(x, t) + \delta B(x, t) \in]0, 1[$ for all (x, t) and, if $P(\Upsilon > \pi) = \epsilon$, $\frac{d}{d\delta} P(\Upsilon > \pi) < 0$. The variation $\frac{d}{d\delta} P(\Upsilon > \pi)$ can be computed directly from (11). If $P(\Upsilon > \pi) = \epsilon$ then there is some (x', t')

where $A(x',t') \neq A^m(x',t')$ and $\Phi_0(x',t')(\chi(x',t') - M[R(\cdot,t'+1)](x')) \neq 0$. We will define $B(x',t') = A^m(x',t') - A(x',t')$. At all other points we will define $B(x,t) = b$ where $A(x,t) = 0$, $B(x,t) = -b$ where $A(x,t) = 1$, and $B(x,t) = 0$ where $A(x,t) \in (0,1)$. This verifies the constraint qualification condition because $\frac{d}{d\delta}P(\Upsilon > \pi) < 0$ for sufficiently small $b > 0$.

For optimal A^*, the KKT optimality conditions provide the existence of multipliers for individual constraints:

- the Lagrange multiplier $\lambda^* \geq 0$ corresponding the probabilistic constraint,
- $\gamma^+(x,t) \geq 0$ corresponding to the constraints $A^*(x,t) \leq 1$, and
- $\gamma^-(x,t) \leq 0$ corresponding to $A^*(x,t) \geq 0$.

With these multipliers the following equality holds (using linearity of M) for each (x,t),

$$\begin{aligned} 0 =& \gamma^+(x,t) + \gamma^-(x,t) \\ &+ \Phi(x,t)\big(- M[Z(\cdot,t+1) + \lambda^* R(\cdot,t+1)](x) - k + \psi(x) + \lambda^* \chi(x,t)\big). \end{aligned} \tag{12}$$

Moreover, the complementary slackness principles are satisfied. If $P(\Upsilon > \pi) < \epsilon$ then $\lambda^* = 0$. If $A^*(x,t) < 1$ then $\gamma^+(x,t) = 0$, and if $A^*(x,t) > 0$ then $\gamma^-(x,t) = 0$. We will interpret these conditions as a dynamic programming principle.

We define $V^*(x,t) = Z(x,t) + \lambda^* R(x,t)$ and define the Hamiltonian, given $v \in \mathbb{R}^{|X|}$ and $a \in [0,1]$, to be

$$H\big(v,x,t,\lambda,a\big) = \big(1-a\big)\big(M[v](x) + k\big) + a\big(\psi(x) + \lambda\chi(x,t)\big). \tag{13}$$

We consider four cases:
1) If $\Phi(x,t) = 0$, then the expected cost and constraint are independent of the choice of policy. Otherwise:
2) If $M\left[V^*(\cdot,t+1)\right](x) + k > \psi(x) + \lambda^*\chi(x,t)$ then using $\gamma^- \leq 0$, (12) implies that $\gamma^+(x,t) > 0$. Complementary slackness implies that $A^*(x,t) = 1$.
3) If $M\left[V^*(\cdot,t+1)\right](x) + k < \psi(x) + \lambda^*\chi(x,t)$ then $\gamma^-(x,t) < 0$ and $A^*(x,t) = 0$.
4) If $M\left[V^*(\cdot,t+1)\right](x) + k = \psi(x) + \lambda^*\chi(x,t)$ then the Hamiltonian does not depend on a at (x,t). We denote the set of such *degenerate points* by

$$D^* = \big\{(x,t) \;:\; M\left[V^*(\cdot,t+1)\right](x) + k = \psi(x) + \lambda^*\chi(x,t)\big\}. \tag{14}$$

In all cases, the Hamiltonian achieves its minimum for $a \in [0,1]$ at $A^*(x,t)$.

The function V^* satisfies $V^*(x,T_1) = U(x) + \lambda^*$ for $x \in X_1$, $V^*(x,t) = 0$ for $x \in X_0$ and $0 \leq t \leq T_1$, and the backward evolution is given by

$$\begin{aligned} V^*(x,t) &= H(V^*(\cdot,t+1), x, t, \lambda^*, A^*(x,t)) \\ &= \min_{a\in[0,1]} H(V^*(\cdot,t+1), x, t, \lambda^*, a) \\ &= \min\big\{\psi(x) + \lambda^*\chi(x,t), M[V^*(\cdot,t+1)](x) + k\big\} \end{aligned} \tag{15}$$

for $x \in X_1$ and $0 \leq t < T_1$; where the last equality follows from the three cases analyzed above. Remarkably, the quantity $V^*(x,t)$ is the same for any optimal

policy and can be reinterpreted as the optimal cost-to-go of the following λ-penalized problem with $\lambda = \lambda^*$: Given $\lambda \geq 0$, find $A \in \mathcal{A}^R$ that minimizes $E[\Upsilon] + \lambda P(\Upsilon > \pi)$.

We have shown the following proposition:

Proposition 3.1 *Suppose that $P^m < \epsilon$. There exists an optimal Lagrange multiplier, $\lambda^* \geq 0$, such that if $A^* \in \mathcal{A}^R$ is optimal, i.e. a minimizer of PCOS, then*

- *for V^* determined by (15), $A^*(x,t)$ is a minimizer of the Hamiltonian defined in (13), with $v = V^*(\cdot, t+1)$, for each (x,t) where $\Phi(x,t) > 0$,*
- *and for the probabilistic constraint corresponding to A^*, the complementary slackness holds that $\lambda^* P^* = \lambda^* \epsilon$,*

Equations (15) are the optimality equations for the λ^*-penalized problem, which gives V^* the alternate interpretation of

$$V^*(x,t) = \inf_{A \in \mathcal{A}^R} \big\{ E[\Upsilon - kt \mid \Xi_t = x] + \lambda^* P(\Upsilon > \pi \mid \Xi_t = x) \big\}.$$

Another useful equation relating V^*, λ^* and the optimal expected cost is

$$\sum_{\xi \in X_1} \Phi_0(\xi) V^*(\xi, 0) = E^* + \lambda^* \epsilon. \tag{16}$$

Equation (16) follows from Equations 5 and 6, and $\lambda^* P^* = \lambda^* \epsilon$. This allows us to determine E^* from V^* and λ^*, avoiding the calculation of Z.

The difficulty remains that P^* depends on the policy through (7), and λ^* is determined implicitly from the constraint. By considering the family of λ-penalized problems with different λ, we can solve the constrained problem by determining the value of λ^* such that either $P^* = \epsilon$ or $\lambda^* = 0$. The solution V^* to (15) determines the policy, except at the degenerate points, D^*, where the Hamiltonian is minimized for all $a \in [0,1]$. Due to the presence of degenerate points, not every policy that is optimal for the λ^*-penalized problem is feasible or optimal with the constraint. We find in Theorem 4.1 that while such degenerate points occur generically, the optimal policy only needs to be randomized, $A^*(x,t) \in]0,1[$, for at most one pair $(x,t) \in X_1 \times \{0, \ldots, T_1\}$.

3.3 Linear Programming Approach

We remark on an alternative approach to formulate PCOS and prove Proposition 3.1 that has complementing advantages. The optimality conditions of the following linear program do not require a constraint qualification (thus covering the case $\epsilon = 0$) and are *sufficient* for the optimality of A (provided it is feasible, and A, λ and V satisfy the conditions of Proposition 3.1).

We consider the variables $\hat{\Phi}(x,t) = (1 - A(x,t))\Phi(x,t)$ and $\tilde{\Phi}(x,t) = A(x,t)\Phi(x,t)$. Clearly, $\hat{\Phi}(x,t) + \tilde{\Phi}(x,t) = \Phi(x,t)$, and if $\Phi(x,t) > 0$ then the correspondence between $(\hat{\Phi}, \tilde{\Phi}) \in \mathbb{R}^+ \times \mathbb{R}^+$ and $(\Phi, A) \in \mathbb{R}^+ \times [0,1]$ is

one-to-one. (In the degenerate case $\Phi(x,t) = 0$, the expected cost and the probabilistic constraint are independent of $A(x,t)$.)

Equation (4) becomes a linear equation of $\hat{\Phi}$ and $\tilde{\Phi}$, without additional occurrence of A, for $t \in \{1,\dots,T_1\}$ and $x \in X$,

$$\hat{\Phi}(x,t) + \tilde{\Phi}(x,t) = \big(1-p(x)\big)\hat{\Phi}(x,t-1) + \sum_{\xi\in N(x)} \frac{p(\xi)}{|N(\xi)|}\hat{\Phi}(\xi,t-1), \tag{17}$$

where we consider $\hat{\Phi}$ defined only on $X_0 \times \{0,\dots,T_1-1\}$, and the convention that $A(x,t)=1$ yielding $\Phi(x,t)=\tilde{\Phi}(x,t)$ is used when $x \in X_0$ or $t = T_1$. The initial condition is prescribed for $x \in X_1$ by

$$\hat{\Phi}(x,0) + \tilde{\Phi}(x,0) = \Phi_0(x). \tag{18}$$

Similarly, $E[\Upsilon]$ and $P(\Upsilon > \pi)$ can be expressed as linear functions of $\hat{\Phi}$ and $\tilde{\Phi}$ from (10) and (11),

$$E\big[\Upsilon\big] = \sum_{\xi\in X_1}\sum_{t=0}^{T_1-1}\Big[k\hat{\Phi}(\xi,t) + \psi(\xi)\tilde{\Phi}(\xi,t)\Big] + U(\xi)\Big(\hat{\Phi}(\xi,T_1) + \tilde{\Phi}(\xi,T_1)\Big), \tag{19}$$

$$-P\big(\Upsilon > \pi\big) = -\sum_{\xi\in X_1}\sum_{t=1}^{T_1-1}\chi(\xi,t)\tilde{\Phi}(\xi,t) + \Big(\hat{\Phi}(\xi,T_1) + \tilde{\Phi}(\xi,T_1)\Big) \geq -\epsilon. \tag{20}$$

We can now express PCOS as a linear program to minimize (19) for the variables $\hat{\Phi}(x,t) \geq 0$, where $x \in X_1$ and $t \in \{0,\dots,T_1-1\}$, and $\tilde{\Phi}(x,t) \geq 0$ for $x \in X$ and $t \in \{0,\dots,T_1\}$, with constraints (17), (18), and (20).

The dual variable $\sigma \geq 0$ corresponds to (20), and the dual variables $W(x,t)$ correspond to (17) for $t > 0$ and correspond to (18) for $t = 0$. The dual linear program is to maximize

$$-\sigma\epsilon + \sum_{\xi\in X_1}\Phi_0(\xi)W(\xi,0)$$

subject to

$$\begin{aligned}
&W(x,t) \leq U(x) + \sigma, && x \in X_1,\ t = T_1,\\
&W(x,t) \leq 0, && x \in X_0,\ t \in \{0,\dots,T_1\},\\
&W(x,t) \leq M\big[W(\cdot,t+1)\big](x) + k, && x \in X_1,\ t \in \{0,\dots,T_1-1\},\\
&W(x,t) \leq \psi(x) + \sigma\chi(x,t), && x \in X_1, t \in \{0,\dots,T_1-1\}.
\end{aligned}$$

It is easy to see that (λ^*, V^*) of Proposition 3.1 form the optimal (σ, W) for this dual linear program. This equivalence shows in particular that the value of PCOS is convex with respect to ϵ. While our formulation in terms of A and Φ loses convexity with respect to the policies A, we gain a causal dependency that allows us to prove monotonicity properties in time as well as develop an efficient numerical algorithm.

3.4 Qualitative Analysis of Optimal Policies

Before presenting our algorithms that take advantage of the structure of optimal policies, we review the qualitative behavior of optimal policies. Any (stationary feedback) policy in $\mathcal{A}^S$ can be equivalently described by specifying its "termination set" $\overline{\Sigma}$; i.e., all nodes in X_1 where that policy prescribes an immediate termination. For policies in $\mathcal{A}^D$ the description is similar except that the set $\Sigma(t)$ is now time-dependent. Even for the randomized policies in $\mathcal{A}^R$ that we consider, it is useful to study the set $\Sigma(t) \subset X_1$ where a policy prescribes termination with probability one.

Suppose $\overline{\Sigma}^* \subset X_1$ is an optimal "termination set" for the unconstrained problem. For PCOS, we can assume that $\Sigma^*(t) = \overline{\Sigma}^*$ for $t \geq T_1$. Prior to T_1, the probabilistic constraint might create an incentive to change the expectation-optimal behavior encoded in $\overline{\Sigma}^*$. Figure 3.1 illustrates this for two examples (described in detail in §5). We particularly highlight two regions in $X_1 \times \mathbb{N}$ where $\Sigma^*(t)$ and $\overline{\Sigma}^*$ are different:

- When $T_0(x) < t \leq T_1$, we have $\Sigma^*(t) \subset \overline{\Sigma}^*$. Region `I` in Figure 3.1A represents the nodes, for which it is now optimal to diffuse, even though the unconstrained optimal policy would terminate. Termination at this stage would cause the cost to exceed π, while there is still a chance to finish with cost less than π by diffusing.
- When $t \leq T_0(x)$, we have $\overline{\Sigma}^* \subset \Sigma^*(t)$. Region `II` represents the nodes, for which it is optimal to terminate even though the unconstrained optimal policy would continue to diffuse. Up until $T_0(x)$ immediate termination is guaranteed to make the total cost lower than π, making termination a more attractive option for some nodes.
- Despite a "discontinuous" change in the termination set at $t = T_0(x)$, Figure 3.1A shows a certain "piecewise-monotonicity." If either $0 \leq r \leq s \leq T_0(x)$ or $T_0(x) < r \leq s \leq T_1$, then $x \in \Sigma^*(r)$ implies that $x \in \Sigma^*(s)$.

Before proving the last property rigorously in Lemma 3.1, we define the subset $\mathcal{A}^P \subset \mathcal{A}^R$ of "piecewise-monotonic" policies.

Definition 3.1 We say that a policy $A \in \mathcal{A}^R$ is *piecewise-monotonic*, $A \in \mathcal{A}^P$, if there are switching-times $S_0 : X_1 \to \{0, \ldots, T_0(x) + 1\}$, $S_1 : X_1 \to \{T_0(x)+1, \ldots, T_1+1\}$, and $A_0, A_1 : X_1 \to [0, 1]$ such that for $t \in \{0, \ldots, T_0(x)\}$ and $x \in X_1$:

$$A(x,t) = \begin{cases} 0, & t < S_0(0), \\ A_0(x), & t = S_0(x), \\ 1, & t > S_0(x), \end{cases}$$

and, for $t \in \{T_0(x) + 1, \ldots, T_1\}$ and $x \in X_1$:

$$A(x,t) = \begin{cases} 0, & t < S_1(0), \\ A_1(x), & t = S_1(x), \\ 1, & t > S_1(x). \end{cases}$$

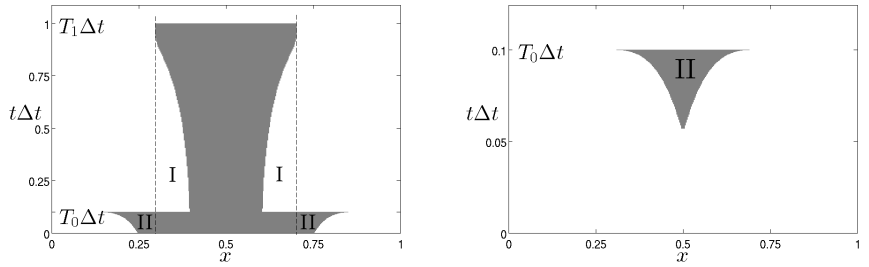


Fig. 3.1: The optimal termination set $\Sigma^*(t) = \{(x,t) : A^*(x,t) = 1\}$ for Example 5.2 on the left and Example 5.1 on the right. The vertical dashed lines indicate the boundaries of $\overline{\Sigma}^*$. In subfigure (B), the parameters values are such that $\overline{\Sigma}^* = \emptyset$.

If $A(x,t) \in \{0,1\}$ for each $t \in \{0,\dots,T_0(x)\}$, then we choose $S_0(x)$ as large as possible so that $A_0(x) = 1$, and the same for $S_1(x)$ and $A_1(x)$. We note that $S_0(x) = T_0(x) + 1$ or $S_1 = T_1 + 1$ correspond to the policy that diffuses for $t \in \{0,\dots,T_0(x)\}$ or $t \in \{T_0(x)+1,\dots,T_1\}$ respectively.

We also say that a policy is *nearly-deterministic*, $A \in \mathcal{A}^N \subset \mathcal{A}^R$, if $A(x,t) \in \{0,1\}$ for all but one point (x,t).

There are optimal policies for the λ-penalized problem that are piecewise-monotonic and deterministic. This structure is used to compute an optimal policy to PCOS of class $\mathcal{A}^P \cap \mathcal{A}^N$ in §4.

Lemma 3.1 *There exists a mapping $\lambda \mapsto A^\lambda \in \mathcal{A}^P \cap \mathcal{A}^D$ for $\lambda \geq 0$ such that:*

- *For all $\lambda \geq 0$, $A^\lambda \in \mathcal{A}^P \cap \mathcal{A}^D$ and A^λ minimizes $E[\Upsilon] + \lambda P(\Upsilon > \pi)$. We define $\{S_0^\lambda, S_1^\lambda\}$ to be the switching-times for A^λ as in Definition 3.1.*
- *For all $0 \leq \lambda_1 < \lambda_2$ and $x \in X_1$, $S_0^{\lambda_1}(x) \geq S_0^{\lambda_2}(x)$ and $S_1^{\lambda_1}(x) \leq S_1^{\lambda_2}(x)$.*

Proof Suppose that V is the solution to (15) for the given value of λ. First, we show by induction that $V(x,t)$ is non-decreasing in t for each $x \in X$, and

$$V(x,t) \geq \min\big\{\psi(x) + \lambda\chi(x,t), M[V(\cdot,t)](x) + k\big\}. \tag{21}$$

Since $U(x) + \lambda = \min\big\{\psi(x) + \lambda, M[U(\cdot) + \lambda](x) + k\big\}$, we may extend (15) for later times by defining $V(x,t) = \lambda + U(x)$ for all $x \in X$ and $t > T_1$. This makes (21) hold with equality for times later than T_1. Now we suppose that $V(x,t+1) \leq V(x,t+2)$ and (21) holds for $V(x,t+1)$. Then

$$\begin{aligned} V(x,t) &= \min\big\{\psi(x) + \lambda\chi(x,t), M[V(\cdot,t+1)](x) + k\big\} \\ &\leq \min\big\{\psi(x) + \lambda\chi(x,t+1), M[V(\cdot,t+1)](x) + k\big\} \leq V(x,t+1). \end{aligned}$$

The relation (21) holds at time t because M is monotone (the coefficients are non-negative) so $M[V(\cdot,t+1)](x) \geq M[V(\cdot,t)](x)$.

We define $A^\lambda \in \mathcal{A}^P \cap \mathcal{A}^D$ as

$$A^\lambda(x,t) = \begin{cases} 0, & M[V(\cdot,t+1)](x) + k < \psi(x) + \lambda\chi(x,t), \\ 1-\chi(t), & M[V(\cdot,t+1)](x) + k = \psi(x) + \lambda\chi(x,t), \\ 1, & M[V(\cdot,t+1)](x) + k > \psi(x) + \lambda\chi(x,t). \end{cases} \tag{22}$$

This policy is piecewise-monotone because $\chi(x,t)$ is piecewise-constant in time and $M[V(\cdot,t+1)](x)$ is monotonically-nondecreasing. A^λ minimizes $E[\Upsilon] + \lambda P(\Upsilon > \pi)$ because it minimizes the Hamiltonian of (13) at each point. In the case of a degenerate points, we have chosen to minimize the constrained value – this will imply that A^λ is feasible for PCOS if λ is an optimal Lagrange multiplier. We also note that $A^\lambda(x,T_1)$ equals $\overline{A}^*(x)$, an optimal policy for the unconstrained problem, and is the same for all λ.

Now suppose that $0 \le \lambda_1 < \lambda_2$ and V^1, V^2 are the corresponding solutions to (15). We set $W = V^2 - V^1$. By considering the four possibilities of the maximum in (15), we find that for all (x,t),

$$W(x,t) \ge \min\{M[W(\cdot,t+1)](x), (\lambda_2-\lambda_1)\chi(x,t)\}, \tag{23}$$

$$W(x,t) \le \max\{M[W(\cdot,t+1)](x), (\lambda_2-\lambda_1)\chi(x,t)\}. \tag{24}$$

Suppose that $W(y,s) \ge 0$ whenever $y \in X_1$ and $t < s \le T_1$. Then since the coefficients within M are non-negative, $M[W(\cdot,t+1)](x) \ge 0$, and (23) implies that $W(x,t) \ge 0$. When $t \le T_0(x)$ and $M[V^1(\cdot,t+1)](x) + k > \psi(x)$ (it is optimal to terminate), then it must be the case that $M[V^2(\cdot,t+1)](x) + k > \psi(x)$, hence $S_0^{\lambda_1}(x) \ge S_0^{\lambda_2}(x)$.

Equation (24) similarly implies that $W(x,t) \le \lambda_2 - \lambda_1$. For $t > T_0(x)$, if $M[V^2(\cdot,t+1)](x) + k > \psi(x) + \lambda_2$ then

$$\begin{aligned} M[V^1(\cdot,t+1)](x) + k \ge & M[V^2(\cdot,t+1)](x) + k + \lambda_1 - \lambda_2 \\ > & \psi(x) + \lambda_1, \end{aligned}$$

which implies $S_1^{\lambda_1}(x) \le S_1^{\lambda_2}(x)$. □

4 Solution Algorithms

Our computational approach is detailed in Algorithms 4.1, 4.2 and 4.3. We start with a brief discussion of their respective goals and the relationship to the theoretical results from §3. We let $\Pi = \{\psi, k, \epsilon, \pi, X_0, X_1, p, N, \Phi_0\}$ be the collection of problem parameters, with an additional algorithmic parameter Δ, which effects the performance but is not a part of the problem statement.

Our goal in Algorithm 4.1 is to find a pair of Lagrange multipliers, $\{\lambda^f, \lambda^s\}$, and a corresponding pair of policies, $\{A^f, A^s\}$, such that A^f is feasible, A^s is super-optimal, and $|\lambda^f - \lambda^s| < \Delta$. The expected cost and constrained probability pairs associated with these policies will be denoted as (E^f, P^f) and (E^s, P^s) respectively. Lemma 3.1 describes a constructive approach for implementing the map $\lambda \mapsto A^\lambda \in \mathcal{A}^P \cap \mathcal{A}^D$. Here we rely on this map, choosing

$A^f = A^{\lambda^f}$ and $A^s = A^{\lambda^s}$. Theorem 4.1 shows that A^λ is feasible whenever $\lambda \geq \lambda^*$, the optimal Lagrange multiplier, and A^λ is super-optimal whenever $\lambda < \lambda^*$. We define $\lambda = \frac{1}{2}(\lambda^f + \lambda^s)$ and compute the corresponding A^λ. The algorithm then checks whether this policy is feasible, which determines whether λ should replace the current λ^f or λ^s, producing a narrower interval straddling the optimal λ^*. The bisection continues until the width falls below the prescribed threshold Δ.

In order for Algorithm 4.1 to be successful, it must be initialized with a value of λ^f for which the corresponding policy is feasible. Recall that the policy which minimizes the constrained value is given in Observation 3.1 with $P(\Upsilon > \pi) = P^m$ and $E[\Upsilon] = E^m$, and we let $E^0 = \sum_{\xi \in X_1} \Phi_0(\xi) U(\xi)$ be the expected cost of the unconstrained problem. Thus, if we initialize $\lambda^f = (E^m - E^0)/(\epsilon - P^m)$ as in [8], the corresponding policy A^f is feasible because the constrained value satisfies

$$P^f = \frac{(E^f + \lambda^f P^f) - E^f}{\lambda^f} \leq \frac{(E^m + \lambda^f P^m) - E^0}{\lambda^f} = \epsilon.$$

One simple approach for combining A^f and A^s comes from the linear programming interpretation in §3.3. The interpolation of $(1-\gamma)\tilde{\Phi}^f + \gamma\tilde{\Phi}^s$ and $(1-\gamma)\hat{\Phi}^f + \gamma\hat{\Phi}^s$ would result in a policy

$$A^\gamma = \frac{(1-\gamma)A^f\Phi^f + \gamma A^s\Phi^s}{\big((1-\gamma)\Phi^f + \gamma\Phi^s\big)}. \tag{25}$$

To ensure its feasibility we could then solve

$$(1-\gamma)P^f + \gamma P^s = 0.02$$

for γ, however, this policy would be randomized at each point where A^f and A^s differ and generally would not inherit the piecewise-monotonic structure.

In contrast, the goal of Algorithms 4.2 and 4.3 is to carefully blend A^f and A^s to produce a feasible nearly-deterministic policy $A^\sharp$, whose value will be better than E^f. This improved policy is in fact optimal if Δ is sufficiently small, cf. Theorem 4.1. We focus on the set of "nearly-degenerate" points $\tilde{D} \subset X \times \{0, \ldots, T_1\}$ where A^f and A^s differ, and change from A^f to A^s as long as the policy remains feasible. Due to the piecewise-monotonic structure shown in Lemma 3.1 (i.e., $S_0^f \leq S_0^s$, but $S_1^s \leq S_1^f$), we move forward in time when changing points in $\tilde{D}_0 = \{(x,t) \in \tilde{D} \ : \ t \leq T_0(x)\}$ (Algorithm 4.2) and backward in time for points in $\tilde{D}_1 = \{(x,t) \in \tilde{D} \ : \ t > T_0(x)\}$ (Algorithm 4.3).

Additional implementation details:

1. All policies are stored in piecewise-monotonic form, and we refer to the switching-times of A^f as $\{S_0^f, S_1^f\}$ and to those of A^s as $\{S_0^s, S_1^s\}$. The earlier switching-times of the current policy, S_0, are increased in Algorithm 4.2, and then the stopping-times of S_1 are decreased in Algorithm 4.3.

Algorithm 4.1: Solve PCOS

Input: Δ, Π
Output: $A^\sharp$, $P^\sharp$, $E^\sharp$

Compute P^m and E^m from A^m as defined in Observation 3.1;
// If $\min_{x\in X_1} T_0(x) \geq 0$ then $P^m = 0$ and $E^m = \sum_{\xi\in X_1} \Phi_0(\xi)\psi(\xi)$
Solve the unconstrained problem by value iterations [9] to obtain U;
$E^0 = \sum_{\xi\in X_1} \Phi_0(\xi)U(\xi)$;
$\lambda = 0$; $\lambda^s = 0$; $\lambda^f = (E^m - E^0)/(\epsilon - P^m)$;
repeat
  Solve (7) and (15) with λ to determine V and R;
  $A = A^\lambda$, determined from (22) given V;
  $P = \sum_{\xi\in X_1} \Phi_0(\xi)R(\xi, 0)$;
  $E = -\lambda P + \sum_{\xi\in X_1} \Phi_0(\xi)V(\xi, 0)$;
  if $P \leq \epsilon$
    $[\lambda^f, A^f, P^f, E^f] = [\lambda, A, P, E]$;
  else
    $[\lambda^s, A^s, P^s, E^s] = [\lambda, A, P, E]$;
  $\lambda = \frac{1}{2}(\lambda^s + \lambda^f)$;
until $\lambda^f - \lambda^s < \Delta$;
if $P^f < \epsilon$ **and** $\lambda^f > 0$
  $[A^\flat, P^\flat, E^\flat]$ = Algorithm 4.2 $(A^f, A^s, P^f, E^f, \Pi)$;
  if $P^\flat < \epsilon$
    return $[A^\sharp, P^\sharp, E^\sharp]$ = Algorithm 4.3 $(A^\flat, A^s, P^\flat, E^\flat, \Pi)$;
  else
    return $[A^\sharp, P^\sharp, E^\sharp] = [A^\flat, P^\flat, E^\flat]$;
else
  return $[A^\sharp, P^\sharp, E^\sharp] = [A^f, P^f, E^f]$;

2. The current policy, A, of Algorithms 4.2 and 4.3 is feasible and deterministic until it is possible to solve for $P(\Upsilon > \pi) = \epsilon$ with a randomized termination probability, in which case the resulting nearly-deterministic policy is labelled $A^\sharp$. If all the points have been updated by Algorithm 2, the resulting policy $A^\flat$ equals A^s for $t \leq T_0(x)$ and equals A^f for $t > T_0(x)$.
3. The dependences of $P(\Upsilon > \pi)$ and $E[\Upsilon]$ on $A(x,t)$ are isolated in (11) and (10), and require $M[R(\cdot, t+1)](x)$, $M[Z(\cdot, t+1)](x)$ and $\Phi(x,t)$. In Algorithm 4.2, we compute $\Phi(x,t)$ from the values of $\Phi(x,t-1)$, and since we follow A^f for the remaining time (see Figure 4.1), we use the values of $M[R^f(\cdot, t+1)](x)$ and $M[Z^f(\cdot, t+1)](x)$ on $\tilde{D}_0$. In Algorithm 4.3, we compute $R(x,t)$ and $Z(x,t)$ using $R(\cdot, t+1)$ and $Z(\cdot, t+1)$, and we use the values of $\Phi(x,t)$ on $\tilde{D}_1$ that were computed in Algorithm 4.2.
4. Suppose that $P(\Upsilon > \pi) = P$ and $E[\Upsilon] = E$ with corresponding R and Z values, and that we change the value of the policy from $A(x,t) = A$ to $A(x,t) = A_n$. Then the new constrained value is

$$P_n = P + \big(A_n - A\big)\Phi(x,t)\big(\chi(x,t) - M[R(\cdot, t+1)](x)\big)\big), \tag{26}$$

and the new value of the expected cost is

$$E_n = E + \Big(A - A_n\big)\Phi(x,t)\Big(M[Z(\cdot, t+1)](x) + k - \psi(x)\Big). \tag{27}$$

We note that in Algorithm 4.2, when $A^f(x,t) \neq A^s(x,t)$, we will always have that $A = 1$ and $\chi(x,t) = 0$, yielding $\chi(x,t) - M[R(\cdot, t+1)] \leq 0$. On the other hand, in Algorithm 4.3 we will have $A = 0$, $\chi(x,t) = 1$ and $\chi(x,t) - M[R(\cdot,t+1)] \geq 0$. In either case, the choice of A_n that minimizes the expected cost while maintaining feasibility is

$$A_n = \min\Big\{\chi(x,t), \frac{\epsilon - P}{\Phi(x,t)\big(\chi(x,t) - M[R(\cdot,t+1)](x)\big)} + A\Big\}. \tag{28}$$

If $\Phi(x,t)(\chi(x,t) - M[R(\cdot,t+1)](x)) = 0$ then $A_n = \chi(x,t)$. However, we only use (28) under the conditions that it maintains piecewise-monotonicity, i.e. $S_0(x) = t$ if $t \leq T_0(x)$ or $S_1(x) = t+1$ if $t > T_0(x)$, and it does not increase the expected cost, i.e. $M[Z(\cdot,t+1)](x) + k - \psi(x) \leq (\geq)\ 0$ if $t \leq T_0(x)$ $(t > T_0(x))$.

5. While we cannot tell a priori if the selected Δ is small enough to guarantee that $A^\sharp$ is optimal, this is easy to check after the fact; see a brief discussion after Theorem 4.1. Although our implementation does not rely on this idea, it could be used to avoid specifying Δ and iterate until the full convergence.

Algorithm 4.2: Resolve degeneracies forward

Input: A^f, A^s, P^f, E^f, Π
Output: $A^\flat$, $P^\flat$, $E^\flat$

$\tilde{D}_0 = \{(x,t) \mid S_0^f(x) \leq t < S_0^s(x)\}$;
$\{S_0, S_1, A_0, A_1\} = \{S_0^f, S_1^f, A_0^f, A_1^f\}$;
Compute R^f and Z^f from $t = T_1$ to $t = 0$;
Store $M[R^f(\cdot,t+1)](x)$ and $M[Z^f(\cdot,t+1)](x)$ for all $(x,t) \in \tilde{D}_0$;
$P = P^f$; $E = E^f$; $\Phi_n = \Phi_0$;
for $t = 0:1:T_1$ **do**
  $\Phi_c = \Phi_n$;
  for $x \in X_1$ **do**
    if $t > 0$
      Update $\Phi_n(x)$ by (4) using $\Phi_c(\cdot)$ and $A(\cdot, t-1)$;
        // $\Phi_n(x) = \Phi(x,t)$ and $\Phi_c(x) = \Phi(x,t-1)$
    if $(x,t) \in \tilde{D}_0$ **and** $S_0(x) = t$
      if $\Phi_n(x) > 0$
        if $M[Z^f(\cdot,t+1)](x) + k \leq \psi(x)$
          Update $A_0(x)$, P and E from (26-28);
          if $P = \epsilon$
            return $[A^\flat, P^\flat, E^\flat] = [A, P, E]$;
          else
            $S_0(x) = t+1$; $A_0(x) = 1$;
        else
          $S_0(x) = t+1$; $A_0(x) = 1$;
return $[A^\flat, P^\flat, E^\flat] = [A, P, E]$;

Algorithm 4.3: Resolve degeneracies backward

Input: $A^\flat$, A^s, $P^\flat$, $E^\flat$, Π
Output: $A^\sharp$, $P^\sharp$, $E^\sharp$
$\tilde{D}_1 = \{(x,t) \mid S_1^s(x) \le t < S_1^\flat(x)\}$;
$\{S_0, S_1, A_0, A_1\} = \{S_0^\flat, S_1^\flat, A_0^\flat, A_1^\flat\}$;
Compute Φ from $t = 0$ to $t = T_1$; // Same Φ as Algorithm 4.2
Store $\Phi(x,t)$ for all $(x,t) \in \tilde{D}_1$;
$P = P^\flat$; $E = E^\flat$; $R_n = 1$; $Z_n = U$;
for $t = T_1 : -1 : 0$ **do**
  $R_c = R_n$; $Z_c = Z_n$;
  for $x \in X_1$ **do**
    if $(x,t) \in \tilde{D}_1$ **and** $S_1(x) = t+1$
      if $\Phi(x,t) > 0$
        if $M[Z(\cdot, t+1)](x) + k \ge \psi(x)$
          $S_1(x) = t$; $A_1(x) = 0$;
          Update $A_1(x)$, P and E from (26-28);
          if $P = \epsilon$
            return $[A^\sharp, P^\sharp, E^\sharp] = [A, P, E]$;
        else
          $S_1(x) = t$; $A_1(x) = 1$;
    if $t < T_1$
      Update $R_n(x)$ by (7) using $R_c(\cdot)$ and $A(\cdot, t+1)$;
        // $R_n(x) = R(x,t)$ and $R_c(x) = R(x,t+1)$
      Update $Z_n(x)$ by (8) using $R_c(\cdot)$ and $A(\cdot, t+1)$;
        // $Z_n(x) = Z(x,t)$ and $Z_c(x) = Z(x,t+1)$
return $[A^\sharp, P^\sharp, E^\sharp] = [A, P, E]$;

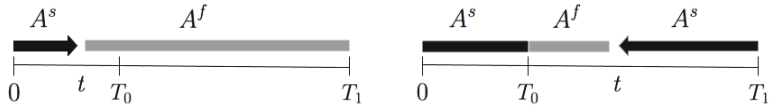


Fig. 4.1: The current policy for fixed $x \in X_1$ in Algorithm 4.2 is drawn schematically for $t \le T_0(x)$ on the left and for Algorithm 4.3 and $t > T_0(x)$ on the right. The Algorithms proceed in the direction indicated by the arrows.

4.1 Algorithm Analysis

The number of iterative steps of Algorithm 4.1 is $\lceil -\log_2 \Delta + \log_2 \lambda^f \rceil$ for the initial value of λ^f. For each iterative step, the solution of equations (7) and (15) occurs in one pass through space and time, i.e. of complexity $O(|X_1|T_1)$. More notably, the values are only stored for two time slices, so the memory required is $O(|X_1|)$. Of course, we also require a solution to the unconstrained problem. The value iterations will converge linearly to U, i.e. having complexity $O(|X_1||\log\kappa|)$, where $\kappa > 0$ is an error threshold. Alternatively, policy iterations can be used, which will often find the exact solution in a small number of steps, but each of them requires solving a linear system of size $|X_1|$.

Algorithms 4.2 and 4.3 work in a single pass through the space-time points so have complexity of $O(|X_1|T_1)$ (with possibly an additional pass to compute $M[R^f(\cdot,t+1)](x)$ and $M[Z^f(\cdot,t+1)](x)$ or $\Phi(x,t)$). We only store the value of Φ, R and Z in two times slices and on $\tilde{D}$. The piecewise-monotonic policy only requires $\{S_0(x), A_0(x), S_1(x), A_1(x)\}$ for each x. Thus, the memory requirement is $O(|X_1| + |\tilde{D}|)$.

The following Theorem summarizes the properties of Algorithms 4.1-4.3.

Theorem 4.1 *Let λ^* be an optimal Lagrange multiplier for PCOS.*

1. *If $\lambda^* = 0$ then there is an optimal deterministic policy, $A^* \in \mathcal{A}^P \cap \mathcal{A}^D$, which is also optimal for the unconstrained problem.*
2. *If $\lambda^f \geq \lambda^* > \lambda^s \geq 0$ then the policy A^f is feasible, and the policy A^s is super-optimal.*
3. *There exists $\Delta > 0$ (dependent on λ^*) such that if $\lambda^f - \lambda^s \leq \Delta$ then $\tilde{D} \subset D^*$ and $A^s = A^f = A^*$ outside of D^*.*
4. *For any $\Delta > 0$, Algorithm 4.1 outputs a feasible policy $A^\sharp \in \mathcal{A}^P \cap \mathcal{A}^N$. Algorithms 4.2 and 4.3 result in $E^\sharp \leq E^\flat \leq E^f$. If $\Delta > 0$ is sufficiently small then $A^\sharp$ is optimal.*

Proof 1. Any minimizer of the unconstrained problem has by definition expected cost $E^0 \leq E^*$. So, if it is also feasible, it must be optimal for PCOS. In Lemma 3.1 we choose to minimize P as a tie-breaker when the policy is not uniquely determined (see (22)), which ensures $A^\lambda \in \mathcal{A}^P \cap \mathcal{A}^D$ with $\lambda = 0$ is a feasible minimizer of the unconstrained problem.

2. We now show that if $\lambda^f \geq \lambda^*$ then A^f is feasible. By construction, A^f is a minimizer of $E[\Upsilon] + \lambda^f P(\Upsilon > \pi)$. If $\lambda^f = \lambda^*$ then A^f is feasible by the choice of tie-breaker. Assume now that $\lambda^f > \lambda^*$. Since A^f and A^* minimize the respective λ^f and λ^*-penalized problems, it follows that

$$\begin{aligned} E^f + \lambda^f P^f \leq & E^* + \lambda^f P^*, \\ E^f + \lambda^* P^f \geq & E^* + \lambda^* P^*. \end{aligned} \tag{29}$$

   Then subtracting the equations we have

$$(\lambda^f - \lambda^*)P^f \leq (\lambda^f - \lambda^*)\epsilon.$$

   Similarly, for the super-optimal policy, we consider the expected cost, E^s, and constraint, P^s. Then using the same argument, but multiplying the second line by λ^s/λ^*, we arrive at

$$\left(1 - \frac{\lambda^s}{\lambda^*}\right) E^s \leq \left(1 - \frac{\lambda^s}{\lambda^*}\right) E^*.$$

3. Next we show that if $\lambda^f - \lambda^s$ is small enough, the policies A^f and A^s do not differ from A^* outside of D^*. We let V^α be the solution of (15) with λ^α. By finiteness of the domain $X_1 \times \{0, \ldots, T_1\}$, there exists $\delta > 0$ such that if $(x,t) \notin D^*$ then

$$|M[V^*(\cdot, t+1)](x) + k - \psi(x) - \lambda^* \chi(x,t)| \geq \delta. \tag{30}$$

In Lemma 3.1 we found that $|V^\alpha(x,t) - V^\beta(x,t)| \le |\lambda^\alpha - \lambda^\beta|$. Suppose that $M[V^*(\cdot,t+1)](x) + k < \psi(x) + \lambda^*\chi(x,t)$ and it is optimal to diffuse at (x,t). Then

$$\begin{aligned} M[V^\alpha(\cdot,t+1)](x) + k \le & M[V^*(\cdot,t+1)](x) + k + |\lambda^\alpha - \lambda^*| \\ \le & \psi(x) + \lambda^*\chi(x,t) + |\lambda^\alpha - \lambda^*| - \delta \\ \le & \psi(x) + \lambda^\alpha\chi(x,t) + 2|\lambda^* - \lambda^\alpha| - \delta. \end{aligned} \tag{31}$$

Thus if $2|\lambda^\alpha - \lambda^*| < \delta$ then

$$M[V^\alpha(\cdot,t+1)](x) + k < \psi(x) + \lambda^\alpha\chi(x,t),$$

and A^α agrees with the optimal policy at (x,t). In the case that $\psi(x) + \lambda^*\chi(x,t) < M[V^*(\cdot,t+1)](x) + k$,

$$\begin{aligned} \psi(x) + \lambda^\alpha\chi(x,t) \le & \psi(x) + \lambda^*\chi(x,t) + |\lambda^\alpha - \lambda^*| \\ \le & M[V^*(\cdot,t+1)](x) + k + |\lambda^\alpha - \lambda^*| - \delta \\ \le & M[V^\alpha(\cdot,t+1)](x) + k + 2|\lambda^\alpha - \lambda^*| - \delta. \end{aligned} \tag{32}$$

If we select $\Delta > 0$ such that $2\Delta < \delta$, then A^f and A^s both agree with A^* for $(x,t) \notin D^*$, and in particular $\tilde{D} \subset D^*$.

4. We use A^f and A^s to construct an optimal policy $A^\sharp \in \mathcal{A}^P \cap \mathcal{A}^N$ assuming Δ is sufficiently small, as detailed in Algorithms 4.2 and 4.3. If $\lambda^f = 0$ then part 1 implies that $A^\sharp = A^f$ is optimal. If instead $P^f = \epsilon$, the same follows from part 2.

   We assume that $P < \epsilon$ is the value of $P(\Upsilon > \pi)$ for the current policy of Algorithm 4.2. We update the policy when $t \le T_0(x)$, $(x,t) \in \tilde{D}$, $M[Z^f(\cdot,t+1)](x)+k \le \psi(x)$ (so that the update does not increase the cost), and $S_0(x) = t$ (so that the update does not break piecewise-monotonicity). In this case we set $A(x,t)$ from (28), which maximizes $P(\Upsilon > \pi)$ subject to $A(x,t) \in [0,1]$ and $P(\Upsilon > \pi) \le \epsilon$. This update increases the switching-time, $S_0(x)$, maintains the feasibility and the piecewise-monotonic structure of A, and does not increase $E[\Upsilon]$. The constraint P updates from (26), leading to two possible outcomes: if $P = \epsilon$ then we have constructed our desired $A^\sharp = A^\flat = A \in \mathcal{A}^P \cap \mathcal{A}^N$, otherwise $P < \epsilon$ and A is still piecewise-monotonic and deterministic. We will need to show that in the case that $\tilde{D} \subset D^*$, the conditions $M[Z^f(\cdot,t+1)](x) + k \le \psi(x)$ and $S_0(x) = t$ are always satisfied. The switching-time begins with $S_0(x) = S_0^f(x)$ so that if $(x,t) \in \tilde{D}$ and $t \le T_0(x)$ then $t \ge S_0^f(x)$ and $(x, S_0^f(x)) \in \tilde{D}$. Since the update either increases $S_0(x)$ or terminates the algorithm, we will only need to check that $M[Z^f(\cdot,t+1)](x) + k \le \psi(x)$ for $(x,t) \in \tilde{D}$ with $t \le T_0(x)$.

   Supposing that $P^\flat < \epsilon$ still holds after Algorithm 4.2, we now work backwards through the points $(x,t) \in \tilde{D}$ when $t > T_0(x)$ in Algorithm 4.3; see Figure 4.1 for the structure of the policy. If $M[Z(\cdot,t+1)](x) + k \ge \psi(x)$ and $S_1(x) = t+1$, then we update $A(x,t)$ and P at degenerate points as described in (28) and (26). Again there are two cases: if $P = \epsilon$ then we are finished with $A^\sharp = A \in \mathcal{A}^P \cap \mathcal{A}^N$, otherwise $P < \epsilon$ and

A remains deterministic and feasible. We will also need to check that if $\tilde{D} \subset D^*$ then $M[Z(\cdot, t+1)](x) + k \geq \psi(x)$ and $S_1(x) = t+1$ are always satisfied. Since we begin with $S_1(x) = S_1^f(x) \geq S_1^s(x)$, if $(x,t) \in \tilde{D}$ and $t > T_0(x)$ then $t < S_1^f(x)$ and $(x, S_0^f(x) - 1) \in \tilde{D}$. The update either decreases $S_1(x)$ or terminates the algorithm, so we only need to check that $M[Z(\cdot, t+1)](x) + k \geq \psi(x)$ for $(x,t) \in \tilde{D}$ with $t > T_0(\mathrm{x})$.

The process described above, and detailed in Algorithms 4.2 and 4.3, constructs a policy $A^\sharp \in \mathcal{A}^P \cap \mathcal{A}^D$ with $P^\sharp \leq \epsilon$ and $E^\sharp \leq E^f$. When $\lambda^f - \lambda^s \leq \Delta$ is sufficiently small, $A^\sharp$ only differs from an optimal policy on the degenerate set D^* by part 3. For any policy A that differs from an optimal policy only on D^*, the corresponding Z and R satisfy $Z(x,t) + \lambda^* R(x,t) = V^*(x,t)$ for all (x,t) and

$$M[Z(\cdot, t+1)](x) + \lambda^* M[R(\cdot, t+1)](x) + k = \psi(x) + \lambda^* \chi(x,t) \qquad (33)$$

for $(x,t) \in D^*$. Then it follows that $M[Z^f(\cdot, t+1)](x) + k \leq \psi(x)$ for $(x,t) \in \tilde{D}$ with $t \leq T_0(x)$, and that $M[Z(\cdot, t+1)](x) + k \geq \psi(x)$ for Z corresponding to the current policy of Algorithm 4.3 and $(x,t) \in \tilde{D}$ with $t > T_0(x)$. Thus for Δ sufficiently small as in part 3, Algorithms 4.2 and 4.3 update the policy for each $(x,t) \in \tilde{D}$. They must terminate with $P^\sharp = \epsilon$ because otherwise, by the end of Algorithm 3, $A^\sharp$ would agree with A^s and would thus satisfy $P^\sharp > \epsilon$. From Proposition 3.1, $E^\sharp + \lambda^* P^\sharp = E^* + \lambda^* \epsilon$, so the termination with $P^\sharp = \epsilon$ ensures that $E^\sharp = E^*$. □

We briefly comment on how to check if $A^\sharp$ is optimal, i.e. whether Δ was sufficiently small. By Proposition 3.1, if $P^\sharp < \epsilon$ and $\lambda^f > 0$ then $A^\sharp$ is not optimal. Recall that any pair (λ, V) that solves (15) is feasible for the dual linear program of §3.3. The duality principle implies that $E^* + \lambda\epsilon \geq \sum_{\xi \in X_1} \Phi_0(\xi) V(\xi, 0)$, which becomes an equality (16), for optimal (λ^*, V^*). In the case that $P^\sharp = \epsilon$, if we define

$$\lambda^\sharp = \frac{E^f + \lambda^f P^f - E^\sharp}{\epsilon}$$

then part 3 of Theorem 4.1 implies $\lambda^\sharp = \lambda^*$ for sufficiently small Δ. For $V^\sharp$ solving (15) with $\lambda^\sharp$, if $E^\sharp + \lambda^\sharp \epsilon = \sum_{\xi \in X_1} \Phi_0(\xi) V^\sharp(\xi, 0)$ then $A^\sharp$ is optimal.

5 Examples

We present two examples corresponding to a discretization of a continuous one-dimensional problem. The continuous domain is $[0,1]$, and the process is a Brownian motion scaled by $\sqrt{2d}$. The target set is the boundary $\{0,1\}$. Cost is accrued at a rate of $\hat{k}$ and the early termination penalty is a constant $\overline{\psi} > 0$. Without the probabilistic constraint, the expected cost can be minimized by computing the value function $u : [0,1] \to \mathbb{R}$, which is a viscosity solution of a

quasi-variational inequality [11], [12]:

$$\begin{aligned} u(x) =& 0,\ x \in \{0,1\}, \\ \max\left\{u(x) - \overline{\psi}, -d\Delta u(x) + \hat{k}\right\} =& 0,\ x \in]0,1[. \end{aligned} \tag{34}$$

Here, we discretize the spatial domain by $X = \{x_i = i\Delta x\}_{i=0}^{2n}$ with the separation $\Delta x = 1/(2n)$, and the target set $X_0 = \{x_0 = 0,\ x_{2n} = 1\}$. Each interior node ($|X_1| = 2n-1$) is adjacent to its two neighbors on the interval. Given a time-step $\Delta t > 0$, the consistent discretized probability to transition to a neighboring node is $d\Delta t/\Delta x^2$ so $p(x) = 2d\Delta t/\Delta x^2$. The CFL condition (i.e. $p(x) \leq 1$) yields $\Delta t \leq \Delta x^2/(2d)$. Once we have chosen Δt, we set $k = \hat{k}\Delta t$. With the constraint parameters π and ϵ, we now have a discrete problem of the form introduced in §2. We consider two different initial conditions, a point-mass in the center with $\Phi_0(x_n) = 1$, or the discretization of a uniform distribution with $\Phi_0(x) = 1/(2n-1)$ for each $x \in X_1$. The C++ code used to generate the numerical data is available on GitHub [13].

Remark 5.1 The function V, which solves (15), is always symmetric,

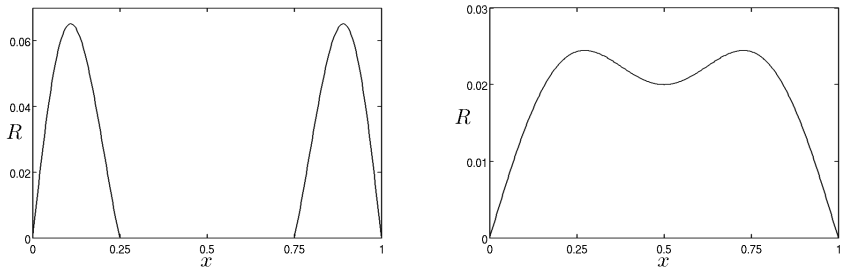
$$V(x,t) = V(1-x,t)\ \forall\ (x,t) \in X \times \mathbb{N}, \tag{35}$$

so V could be determined on X from the values on nodes $\{0,\ldots,x_n\}$. Despite the computational gains from this reduction, we do not pursue it here, solving equations on the entire X_1 to highlight the generality of our approach.

Example 5.1 We use the parameters $d = 0.25$, $\hat{k} = 1$, $\pi = 1$, $\overline{\psi} = 0.9$, $\epsilon = 0.02$, $\Phi_0(x_n) = 1$, $n = 200$, and $\Delta t = 1/100{,}000$. Thus $T_1 = \lfloor \pi/k \rfloor = \lfloor \pi/(\hat{k}\Delta_t) \rfloor = 100{,}000$ and $T_0 = \lfloor (\pi - \overline{\psi})/k \rfloor = 10{,}000$. The policy to always diffuse is optimal without the constraint with expected cost $E[\Upsilon] \approx 0.5$ but fails the constraint with constrained value $P(\Upsilon > \pi) \approx 0.1080$. For PCOS, we use $\Delta = 10^{-6}$ in Algorithm 4.1 and it terminates with $\lambda^f \approx 4.2441$. The corresponding constrained values satisfy $P^s > 0.02 > P^f$ with $P^s - P^f \leq 10^{-7}$. The expected cost is $E^f \approx 0.7842$ and $E^f - E^s \leq 10^{-7}$. The termination set of A^f is shown in Figure 3.1B. As explained in §3.4, in this example there is no incentive to trigger an early termination for $t > T_0$. The switching-times for A^f and A^s agree everywhere except for $x \in \{x_{183}, x_{217}\}$ where $S_0^f(x) = 7814$ and $S_0^s(x) = 7815$. Algorithm 4.2 finds the nearly-deterministic optimal policy $A^\sharp = A^\flat = A^*$ with $A^\sharp(x_{183}, 7814) \approx 0.4572$ and $A^\sharp(x_{217}, 7814) = 1$.

The conditional constrained value $R(x,0)$ using $A^\sharp$ is shown in Figure 5.1B. As required, $P(\Upsilon > \pi) = \sum_{\xi \in X_1} \Phi_0(\xi) R(\xi,0) = R(x_n,0) = 0.02$ holds up to machine precision. However, $R(x,0) > 0.02$ on a large part of the domain. This counterintuitive property is a result of using the optimal $A^\sharp$. ($R(x,0)$ would certainly be monotone increasing on $\{0,\ldots,x_n\}$ if we used the unconstrained optimal policy to always diffuse instead.)

In Figure 5.2 we plot the dependence of λ^f on ϵ for small values of n ($n = 5$ and $n = 25$ with $\Delta t = 1/(50n^2)$). Since we use $\Delta = 10^{-6}$, this can can be also viewed as graphs of λ^*. The Lagrange multiplier, $\lambda^*(\epsilon)$, jumps



(A) (B)

Fig. 5.1: The conditional constrained value, $R(x,0)$, for Examples 5.2 (A) and 5.1 (B).

discontinuously to 0 when the policy to always diffuse becomes feasible. The discontinuity occurs because $U(x) < \overline{\psi}$ so penalizing the constrained value by $\lambda > 0$ may not be enough incentive to terminate. We also plot the values of E^f and E^s showing how the benefit of randomization becomes smaller as n increases. The optimal randomized policy will attain the expected cost that linearly interpolates the points at which E^f and E^s agree. These points can be seen as corners of the rectangles traced by E^f and E^s in Figure 5.2. The difference $E^f - E^\sharp$ is as high as 0.997 for $n = 5$ and $\epsilon = 0.0703$, but decreases with n, e.g. for $n = 25$ the difference $E^f - E^\sharp$ never exceeds 0.0190 regardless of ϵ.

Example 5.2 All the parameters are the same as in Example 5.1 except for $d = 0.05$, $\Delta t = 1/20{,}000$, and a uniform initial distribution, $\Phi_0(x) = 1/(2n-1)$ for all $x \in X_1$. Due to the smaller diffusive constant, the CFL condition requires only 1/5 as many time steps. In our case we have $T_1 = 20{,}000$ and $T_0 = 2{,}000$. The unconstrained optimal policy $\overline{A}^*$ terminates for $x \in [0.3, 0.7]$, yielding $P(\Upsilon > \pi) \approx 0.1421$ and $E[\Upsilon] \approx 0.7218$. Again using $\Delta = 10^{-6}$, we find $\lambda^f \approx 0.7605$, the expected cost is $E^f \approx 0.7434$ and $E^f - E^s \leq 10^{-7}$, and again we have $P^s > 0.02 > P^f$ and $P^s - P^f \leq 10^{-7}$. The termination set is non-empty for all times; see Figure 3.1A. The policies A^f and A^s agree for all $t > T_0$, and the switching-times, S_0^f and S_0^s, only differ for $x \in \{x_{96}, x_{304}\}$ where $S_0^f(x) = 421$ and $S_0^s(x) = 422$. The nearly-deterministic optimal policy $A^\flat$ from Algorithm 4.2 is actually optimal, i.e. $A^\flat = A^\sharp = A^*$, with $A^\sharp(x_{96}, 421) = 0$ and $A^\sharp(x_{304}, 421) \approx 0.8820$. $R(x,0)$ is plotted in Figure 5.1A.

6 Conclusions

We have studied a prototypical stochastic optimal stopping problem with a probabilistic constraint, and found that it can be solved using dynamic programming with a Lagrange multiplier appearing as an additional parameter. It is easy to determine the optimal value of the Lagrange multiplier due to

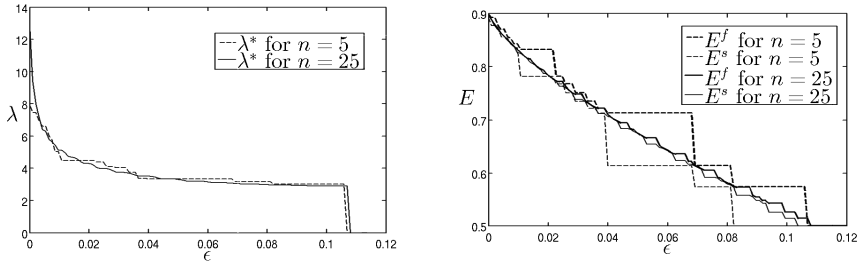


Fig. 5.2: On the left $\lambda^*(\epsilon)$ is plotted with parameters and initial distribution from Example 5.1 and different discretizations of the continuous problem. On the right we show the feasible and super-optimal values E^f and E^s for the same problems.

a monotonic relationship with the constraint. The optimal policies are time-dependent, depend on the initial distribution, and require randomization. However, we prove there are optimal policies that are nearly-deterministic with a piecewise-monotonic structure, which allows for efficient computation.

A few generalizations of this problem present interesting questions. Dependence of transition probabilities on additional control variables will result in a more general stochastic shortest path problem (SSP) with more complicated optimality equations; however, the arguments in Proposition 3.1 will still apply. We therefore expect to find "nearly-deterministic" optimal policies, but not the structural property of "piecewise-monotonicity." Whether there are more general assumptions on the transition probabilities that lead to computationally useful properties of optimal policies is an interesting question for further research. Another non-trivial extension is to allow for inhomogeneous or random running costs $k(x,t)$. The usual approach is to expand the state space to keep track of the accumulated cost as an additional dimension. The obvious computational drawbacks make it attractive to search for a subclass of problems or alternate solution techniques, where the increase in dimensionality can be avoided.

Multiple probabilistic constraints (e.g., $P(\Upsilon > \pi_1) \leq \epsilon_1$ **and** $P(\Upsilon > \pi_2) \leq \epsilon_2$) can be handled similarly [7,8], although our notion of "piecewise-monotonic" and "nearly-deterministic" policies would have to be generalized. We have focused on the problem of minimizing the expected cost, but our approach might also be applicable with other objective functions, e.g. "risk-sensitive" controls [1].

Finally, a continuous version of this problem provides interesting exercises in stochastic analysis and variational inequalities. A part of the difficulty is that randomized stopping-policies in feedback form are not as natural in the continuous setting. Instead, the analysis will have to focus on trajectory-dependent randomized stopping-times [14] or a linear programming formulation analogous to that of §3.3. For state-constraints in general controlled drift-diffusion processes, a natural approach is the "stochastic maximum principle" [15]. But

the probabilistic constraint violates its technical assumptions, so some modification of that theory would be required.

Acknowledgements Aaron Zeff Palmer was supported in part as NSF GRFP Fellow 2011122749. Alexander Vladimirsky was supported in part by the NSF grants DMS-1016150 and DMS-1738010. We would like to thank the associate editor and the reviewers for their carefully reading and suggestions that helped us greatly improve this paper.

References

1. Fleming, W.H., Soner, H.M.: Controlled Markov processes and viscosity solutions. Applications of mathematics. Springer-Verlag, New York (1993)
2. Fan, Y., Nie, Y.: Optimal routing for maximizing the travel time reliability. Networks and Spatial Economics **6**(3), 333–344 (2006)
3. Browne, S.: Optimal investment policies for a firm with a random risk process: Exponential utility and minimizing the probability of ruin. Mathematics of Operations Research **20**(4), 937–958 (1995). DOI 10.1287/moor.20.4.937
4. Rockafellar, R.T., Uryasev, S.: Optimization of conditional value-at-risk. Journal of Risk **2**, 21–41 (2000)
5. Ermon, S., Gomes, C., Selman, B., Vladimirsky, A.: Probabilistic planning with nonlinear utility functions and worst-case guarantees. In: Proceedings of the 11th International AAMAS Conference - Vol. 2, pp. 965–972 (2012)
6. Bertsekas, D.P.: Nonlinear programming. Athena scientific Belmont (1999)
7. White, D.: Dynamic programming and probabilistic constraints. Operations Research **22**(3) (1974)
8. Pfeiffer, L.: Two approaches to stochastic optimal control problems with a final-time expectation constraint. Applied Mathematics and Optimization (2016)
9. Bertsekas, D.P.: Dynamic Programming and Optimal Control, Vol. II, 3rd edn. Athena Scientific (2007)
10. Bertsekas, D.P., Tsitsiklis, J.N.: An analysis of stochastic shortest path problems. Mathematics of Operations Research **16**, 580–595 (1991)
11. Crandall, M.G.: Viscosity Solutions and Applications: Lectures given at the 2nd Session of the Centro Internazionale Matematico Estivo (C.I.M.E.) held in Montecatini Terme, Italy, June 12–20, 1995, chap. Viscosity solutions: A primer, pp. 1–43. Springer Berlin Heidelberg, Berlin, Heidelberg (1997)
12. Krylov, N.V., Balakrishnan, A.V.: Controlled diffusion processes / N. V. Krylov ; translated by A. B. Aries ; [editor, A. V. Balakrishnan]. Springer-Verlag New York (1980)
13. Palmer, A.Z.: a C++ implementation of algorithms for PCOS problem. https://github.com/AaronZPalmer/PCOS (2017)
14. Baxter, J., Chacon, R.: Compactness of stopping times. Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete **40**(3), 169–181 (1977)
15. Karoui, N.E., Peng, S., Quenez, M.C.: A dynamic maximum principle for the optimization of recursive utilities under constraints. The Annals of Applied Probability (2001)