New relative value iteration algorithms for the ergodic risk-sensitive control of Markov chains

SUMITH REDDY ANUGU*, GUODONG PANG*, AND NICOLA SASSONE*

Abstract. In this paper, we propose new Jacobi-like and Gauss-Seidel-like relative value iteration (RVI) algorithms for the ergodic risk-sensitive control (ERSC) problem of a controlled discrete-time Markov chain (DTMC) on discrete state space and establish their convergence. With finite state space, whenever the DTMC is irreducible and recurrent under every stationary Markov control policy, we prove that the iterates of our algorithms converge geometrically. The proof requires establishing a local contraction property, and also a local Lipschitz continuity property of the fixed point of the risk-sensitive Bellman-like operator. To tackle the challenges associated with the multiplicative nature of the operator, we employ the well-known entropy variational formula, which re-casts the operator and its fixed point equation in an additive form, albeit with an additional optimization problem over a corresponding set of probability vectors. In the case of countable state space, we propose an implementing procedure by properly truncating the problem to have finite state space. Under either uniform stability or a near-monotonicity condition, we show that as the truncation size grows indefinitely, the iterate values of our algorithms converge to the value function and the optimal ERSC cost. Finally, we implement our algorithms on two examples to demonstrate the performances of our proposed RVI algorithms: maximizing the exit rate from a finite domain (on a graph) and a single-server queue of finite capacity in the case of finite state space, and also a single-server queue of infinite capacity with abandonment in the case of countable state space.

1. Introduction

We consider the ergodic risk-sensitive control (ERSC) problem of a controlled discrete-time Markov chain (DTMC) $X = \{X_t\}_{t=0}^{\infty}$ on a discrete state space with an associated transition probability $p(i, j, u) = \mathbb{P}(X_1 = j | X_0 = i, u)$, where our objective is to minimize the following criterion for a risk-sensitivity parameter $\delta > 0$:

$$\limsup_{T\to\infty} \frac{1}{\delta T} \log \mathbb{E}\left[e^{\delta \sum_{t=0}^{T-1} c(X_t, U_t)} | X_0 = i \right] \tag{1.1}$$

over U which lies in an appropriate class of control policies. One distinctive feature of this problem is that it concerns with all higher moments of the total accumulated cost $\delta \sum_{t=0}^{T-1} c(X_t, U_t)$, due to the exponential in (1.1). This is in contrast to the average cost control problem, where the minimization criterion (over an appropriate class of controls) is given by

$$\limsup_{T\to\infty} \frac{1}{T} \mathbb{E}\left[\sum_{t=0}^{T-1} c(X_t, U_t) | X_0 = i \right]. \tag{1.2}$$

In other words, the average cost problem only concerns with the first moment (the expectation) of the accumulated cost. Hence, the average cost control problem is also commonly referred to as the risk-neutral problem - this terminology is further aided by the fact that the sequence of optimal

*Department of Computational Applied Mathematics and Operations Research, George R. Brown School of Engineering and Computing, Rice University, Houston, TX 77005

E-mail addresses: sa167, gdpang, ngs6@rice.edu.

Date: February 12, 2025.

Key words and phrases. Ergodic risk-sensitive control problem, Markov chains, Jacobi-like and Gauss-Seidel-like relative value iteration (RVI) algorithms, convergence analysis.

values of the ERSC problem converges to the optimal value of the average cost problem, as $\delta \to 0$. Given its ability to incorporate risk-sensitivity into decision-making, risk-sensitive control with an exponential utility function has been widely applied across various domains, including inventory control [17], revenue management [4, 24], portfolio optimization [10, 25], and multi-armed bandit problems [22].

The primary objective of the ERSC problem is to find and characterize optimal control policies. It is well-known that in the case of countable state space (and also in the case of finite state space), under a suitable uniform stability assumption on the controlled DTMC in relation to the growth of running cost or an appropriate near-monotonicity assumption where the running cost penalizes the unstable behavior of the controlled DTMC, there exist optimal policies that are stationary (in the case of finite state space, it suffices for the DTMC to be irreducible and recurrent for all stationary Markov policies); see [19, Theorem 3.1]) in the case of finite state space and [13, Theorems 2.5 and 2.17] in the case of countable state space. Furthermore, these optimal stationary Markov policies can be completely characterized by the multiplicative Bellman equation (or simply the Bellman equation). More precisely, if (V^*, λ^*) is the solution pair of the Bellman equation (where λ^* turns out to be the optimal ERSC cost and V^* is referred to as the value function), then a stationary Markov policy is optimal if and only if it is a minimizer of the Bellman equation given below (for precise statements, see Theorems 2.1 and 5.2 for the case of finite state space and countable state space, respectively):

$$V^*(i) = \min_u \Big(e^{\delta(c(i,u)-\lambda^*)} \sum_j V^*(j) p(i,j,u) \Big), \tag{1.3}$$

where i varies over the state space and the minimization is over an appropriate control set that possibly depends on i. Unfortunately, computing the minimizers of the Bellman equation requires the knowledge of its solution pair (V^*, λ^*) which is not readily available. Therefore, it is a common practice to turn towards numerical algorithms to compute an optimal stationary Markov policy. These numerical algorithms typically fall into two categories: (i) relative value iteration (RVI) algorithms: these algorithms recursively evaluate the value function so that it approaches the true value function, and (ii) policy iteration algorithms: these algorithms recursively evaluate policies and improve them (see [12, Section 6.1] for a historical review of policy iteration algorithms for the ERSC problem).

Our contributions are two new RVI algorithms (introduced below) for the ERSC problem in the case of finite state space (given by $\{1, \dots, n\}$, for simplicity with n being chosen as the reference state below). The first algorithm takes the following form (see Algorithm 1): for an initialization (V^0, λ^0), $k \geq 1$ and step-size parameters $\{\gamma_k\}_{k\geq 0}$,

$$\begin{aligned} V^{k+1}(i) &\doteq \min_u \Big[e^{\delta(c(i,u)-\lambda^k)} \Big(p(i,n,u) + \sum_{j=1}^{n-1} V^k(j) p(i,j,u) \Big) \Big], \quad 1 \leq i \leq n, \\ \lambda^{k+1} &\doteq \lambda^k + \frac{1}{\delta} \gamma_k \log V^{k+1}(n)\,. \end{aligned} \tag{1.4}$$

The second algorithm takes the following form (see Algorithm 2): for an initialization (V^0, λ^0), $k \geq 1$, $2 \leq i \leq n$, and step-size parameters $\{\gamma_k\}_{k\geq 0}$,

$$\begin{aligned} V^{k+1}(1) &\doteq \min_u \Big[e^{\delta(c(1,u)-\lambda^k)} \Big(p(1,n,u) + \sum_{j=1}^{n-1} V^k(j) p(1,j,u) \Big) \Big], \\ V^{k+1}(i) &\doteq \min_u \Big[e^{\delta(c(i,u)-\lambda^k)} \Big(p(i,n,u) + \sum_{j=1}^{i-1} V^{k+1}(j) p(i,j,u) + \sum_{j=i}^{n-1} V^k(j) p(i,j,u) \Big) \Big], \\ \lambda^{k+1} &\doteq \lambda^k + \frac{1}{\delta} \gamma_k \log V^{k+1}(n)\,. \end{aligned} \tag{1.5}$$

Note that (1.4) and (1.5) resemble the well-known Jacobi and Gauss-Seidel iterations, respectively, that are used in solving strictly diagonally dominant linear system of equations. Due to this resemblance, we refer to them as the Jacobi-like RVI and Gauss-Seidel-like RVI algorithms, respectively, in the rest of this paper. These algorithms can be regarded as the risk-sensitive analogs of the algorithms introduced in [6]. For the sake of discussion below, let (1.4) and (1.5) be represented by $V^{k+1} = \widetilde{F}(V^k, \lambda^k)$ and $V^{k+1} = \widetilde{G}(V^k, \lambda^k)$, for some appropriate operators $\widetilde{F}$ and $\widetilde{G}$, respectively.

We first show that under the assumptions of irreducibility and recurrence of the DTMC and an appropriate choice of γ_k, the above algorithms converge, at a geometric rate to (V^*, λ^*), the solution pair satisfying (1.3). The analysis hinges heavily on the analysis of the fixed points equations associated with the algorithms. However, for reasons that will become apparent below, we consider operators F and G such that $F(h, \lambda) = \log \widetilde{F}(e^h, \lambda)$ and $G(h, \lambda) = \log \widetilde{G}(e^h, \lambda)$, respectively. In particular, we analyze two properties: (i) the local contraction property of F (and G), and (ii) the local bi-Lipschitz continuity (in terms of λ) of the fixed points (denoted by h_λ) of $F(\cdot, \lambda)$ and $G(\cdot, \lambda)$. Due to the multiplicative nature of the problem, we obtain properties that are 'local', which is in contrast to the risk-neutral case where the analogous properties are satisfied globally. There is another representation of h_λ (or equivalently, V_λ) that is key in obtaining the aforementioned continuity of $\lambda \mapsto h_\lambda$: $h_\lambda(i)$ is the minimum of

$$\log \mathbb{E}\Big[e^{\sum_{t=0}^{\tau_n - 1} \left(\delta c(X_t, v(X_t)) - \lambda\right)} | X_0 = i\Big], \tag{1.6}$$

over all stationary Markov policies v. Here, τ_n is the first return time to state n. This representation also sheds light on the reason behind the term $\log V^{k+1}(n)$ *via.* the following result (see Proposition 2.1): $h_\lambda(n) = 0$ if and only if $\lambda = \lambda^*$, and hence $h_{\lambda^*} = e^{V^*}$ (or, $V_{\lambda^*} = V^*$).

We overcome the challenges arising from the multiplicative structure of the operators $\widetilde{F}$ and $\widetilde{G}$, by a repeated application of the well-known entropy variational formula: for a probability measure P and an appropriately bounded function f on a space $\mathcal{X}$, we have

$$\log \int_{\mathcal{X}} e^{f(x)} P(dx) = \sup_Q \Big[\int_{\mathcal{X}} f(x) Q(dx) - R(Q||P) \Big]$$

with $R(Q||P)$ being the relative entropy of Q with respect to P. This re-casts $F(h, \lambda)$ and $G(h, \lambda)$ into a form that is linear in h - this entails the analysis of F and G (and thereby, h_λ), rather than $\widetilde{F}$ and $\widetilde{G}$ (and thereby V_λ).

From here, the proofs of the local contraction and the local bi-Lipschitz continuity involve choosing an appropriate control u and an appropriate measure Q, and bounding the resulting operator. Subsequently, the proof of the local contraction of F closely resembles the arguments in [33]. However, the local contraction of operator G is much more involved compared to that of F, owing to the iterative nature in the definition of operator G; see Section 4.2.

The proof of bi-Lipschitz continuity of $\lambda \mapsto h_\lambda(n)$ is more subtle. For $\lambda' > \lambda$ in a suitable domain, to prove the upper bound of $h_\lambda(n) - h_{\lambda'}(n)$, we consider a stationary Markov policy that attains the infimum in $h_{\lambda'}(n)$ and then choose a probability measure Q that attains the supremum in $h_\lambda(n)$. From here, we obtain a new Markov chain associated with Q. Under these choices of stationary Markov policy and probability measure Q, $h_{\lambda'}(n)$ satisfies a sub-solution-like equation and $h_\lambda(n)$ satisfies a super-solution-like equation. An application of Dynkin's formula to both $h_\lambda(n)$ and $h_{\lambda'}(n)$, and bounding the expectation of the return times (to state n, for the new Markov chain), give us the desired upper bound. Similarly, the lower bound of $h_\lambda(n) - h_{\lambda'}(n)$ is proved.

Although the Jacobi-like and Gauss-Seidel-like RVI algorithms are inspired by [6] in the risk-neutral case, the necessary properties - the local contraction properties of $F(\cdot, \lambda)$ and $G(\cdot, \lambda)$), and local bi-Lipschitz continuity of $\lambda \mapsto h_\lambda(n)$, are more difficult to prove than in the risk-neutral case in [6]. In particular, the risk-neutral analog of bi-Lipschitz continuity is an immediate consequence of the additivity of the risk-neutral cost (in terms of the running cost) and of the uniform recurrence

of the Markov chain, whereas in our case, bi-Lipschitz continuity is not at all obvious due to the multiplicative structure within the expectation and requires a novel approach.

Another contribution of this paper is designing RVI algorithms in the case of countable state space that are implementable in practice. The naive extensions of Jacobi-like and Gauss-Seidel-like algorithms are not implementable in practice (due to the requirement of infinitely many computations) and it is not clear if the techniques used in the proof of convergence in the case of finite state space can be extended to countable state space. Therefore, we introduce a new ERSC problem associated with a new Markov chain on a truncated (finite) set and a modified running cost. We show that the optimal ERSC cost of this family of truncated problems converges to the optimal ERSC cost of our original ERSC problem (on the countable state space) as the truncation size grows indefinitely. This is shown by relating the truncated ERSC problems to another closely related family of Dirichlet problems (with the Dirichlet eigenvalue being the optimal ERSC cost of the truncated problem; see, e.g., [13]).

Finally, we demonstrate the performance of our algorithms and compare them with the existing RVI algorithm by implementing them on two models. The first one is a controlled DTMC on a finite connected graph, in particular, the objective is to maximize the exit-rate from a fixed connected sub-graph given that it starts from this sub-graph. This problem can be re-cast into the ERSC problem defined *via.* (1.1), for an appropriate running cost. We provide numerical evidence that our algorithms outperform the existing RVI algorithm, whenever the graph is neither complete nor a tree. The second one is a service-effort control problem for a discrete-time single-server queue in two cases: finite capacity, and infinite capacity with abandonment. In the finite capacity case, we perform sensitivity analysis to study the effect of the sensitivity parameter and the size of the state space on the performance of our algorithms. At a high sensitivity parameter, we find that the Gauss-Siedel-like algorithm performs well in comparison to the other algorithms. However, the performances of our algorithms (and the existing RVI algorithm) deteriorate when the sensitivity is low or the state space is large. In the case of infinite capacity, we implement our algorithms on the aforementioned truncated versions, and find that as the truncation size grows large, the iterates converge, verifying that our truncation procedure indeed approximates the original ERSC problem.

1.1. **Literature review.** The relative value iteration (RVI) algorithm for the ergodic control of Markov chains with finite state space and control set was first introduced in the seminal work by White [36]. Since then, there has been an extensive body of work studying RVI algorithms for ergodic control of Markov chains under various assumptions and settings; see [7, Section 4.5] for a full account of these developments. One important development is by Bertsekas [6], which introduced the Jacobi and Gauss-Seidel versions of the RVI algorithm, inspired by the stochastic shortest path problem (SSP). Under an irreducibility condition on the controlled DTMC, the author uses the contraction property of the Bellman operator in conjunction with the structure of the SSP to prove that the iterates of both versions converge to the value function and the optimal cost at a geometric rate.

In the context of ERSC problems, to our knowledge, the first work discussing the RVI algorithm appeared in [8], where the authors study a multiplicative analogue of the risk-neutral RVI algorithm studied in [36] for finite-state controlled DTMCs. In addition to assuming irreducibility of the controlled DTMC, the authors assume that the probability of a self-loop is bounded away from 0 for all states and controls. This allowed the authors to establish a contraction property of the Bellman operator with respect to a span semi-norm, which is then used to show that the RVI iterates converge at a geometric rate. This version of the RVI algorithm has been subsequently studied by several authors under various setups and conditions in [1, 16, 20]. The strict positivity condition on the probability of a self-loop is removed in [20], where the authors study a transformed version of the original process, under the general assumption that a solution to the multiplicative Bellman equation exists. Borkar and Meyn [16] establish the convergence of this RVI algorithm in the

countable setting under the conditions of aperiodicity and irreducibility, and the near-monotonicity of the running cost, by employing multiplicative ergodic theorems. The extension of this work to compact Polish state space is carried out in [1]. We note that the RVI algorithm for the ERSC problem has seen applications across various domains, including portfolio optimization [9, 11], manufacturing systems [21], and stochastic differential games [3]. Additionally, this RVI algorithm is the foundation for the risk-sensitive Q-learning algorithms proposed in [5, 15].

Lastly, we review the literature on approximating countable state controlled DTMCs using finite-state models. To our knowledge, this is first discussed in [37]. Since then, truncation procedures with provable guarantees have been studied for both the long-run average cost problem and discounted problem. We refer the reader to [28, 32] for a history of these developments and an overview of the truncation procedure under general assumptions in the risk-neutral context. Truncation procedures in the context of risk-sensitive control have also been studied. See for example, [34] for the finite horizon problem and [27] for the discounted cost problem. Prior to our work, to our knowledge, truncation procedures for the ERSC problem are investigated in [26, 35]. The novelty of our work lies in the relating our truncation procedure to the well-studied Dirichlet problem associated with the original ERSC problem. See Section 5.3.1 for a more elaborate discussion.

1.2. **Organization of the paper.** We introduce the necessary notation in the next subsection. In Section 2, we introduce the ERSC problem in the case of finite state space, propose our new Jacobi-like and Gauss-Seidel-like RVI algorithms, and state our main result on their convergence. In Section 3, we establish the local contraction property and the local bi-Lipschitz continuity property of the operator F. Section 4 contains the proof of our main result. In Section 5, we propose analogous algorithms in the case of countable state space and prove their convergence. In Section 6, we provide the numerical examples to illustrate the performances of our proposed algorithms. Finally, in the Appendix, we collect the proofs of two key supporting results.

1.3. **Notation.** We use $(\Omega, \mathcal{F}, \mathbb{P})$ to denote the underlying abstract probability space with $\mathbb{E}$ as the associated expectation. The set of nonnegative real numbers (integers) is denoted by $\mathbb{R}_+$ ($\mathbb{Z}_+$), $\mathbb{N}$ stands for the set of natural numbers, and $\mathbb{1}_A(\cdot)$ denotes the indicator function corresponding to set A. Let $\mathbb{R}^k$ denote the k-dimensional Euclidean space, $e \doteq (1, 1, \dots, 1)^\top \in \mathbb{R}^k$ and $\|\cdot\|_\infty$ denote the usual supremum norm. For a Polish space $\mathcal{X}$, let $\mathcal{P}(\mathcal{X})$ denote the space of Borel probability measures on $\mathcal{X}$. For two probability measures $P, Q \in \mathcal{P}(\mathcal{X})$, we say $Q \ll P$ if Q is absolutely continuous with respect to P.

2. New RVI algorithms in the case of finite state space and their convergence

2.1. **ERSC problem on a finite space.** Consider a controlled DTMC $X = \{X_t\}_{t=0}^\infty$ which takes values in a finite-state space $S = \{1, \dots, n\}$. Let $\mathbb{U}(i)$ be a compact metric space for every $i \in S$. At each instant, if the current state is i, then, for a chosen control $u \in \mathbb{U}(i)$, the next transition to state j occurs with probability $p(i, j, u)$. Namely, $\{p(i, j, u) : i, j \in S, \text{ and } u \in \mathbb{U}(i)\}$ is the associated controlled transition probability of X. Let $S_{\mathbb{U}} \doteq \{(i, u) : i \in S \text{ and } u \in \mathbb{U}(i)\}$ denote the set of all allowed state-action pairs.

A control policy (sometimes referred to simply as a policy) is a sequence of controls for the DTMC X at each time instant, which we denote by $U = \{U_0, U_1, \dots\}$ where $(X_t, U_t) \in S_{\mathbb{U}}$ for all $t \in \mathbb{Z}_+$. We say a control U is admissible if for every $t \in \mathbb{Z}_+$, U_t is measurable with respect to the σ–algebra generated by $\{X_0, X_1, \dots, X_t\}$. Let $\mathfrak{U}$ denote the set of all admissible control policies. A policy U is called Markov if $U_t = v_t(X_t)$, where $v_t(\cdot)$ is such that $\big(X_t, v_t(X_t)\big) \in S_{\mathbb{U}}$ for $t \in \mathbb{Z}_+$ and it is called stationary Markov if $v_t(\cdot)$ is independent of t. With slight abuse of notation, we refer to such a stationary Markov policy by v and let $\mathfrak{U}_{\mathrm{SM}}$ denote the space of stationary Markov control polices. Given an admissible control policy U and an initial state i, the corresponding cost functional for the ergodic risk-sensitive control (ERSC) problem, with a risk-sensitive parameter

$\delta > 0$, is given by

$$\mathcal{E}_i^\delta(U) \doteq \limsup_{T\to\infty} \frac{1}{\delta T} \log \mathbb{E}_i^U \left[e^{\delta \sum_{t=0}^{T-1} c(X_t, U_t)} \right], \tag{2.1}$$

where $c : S_{\mathbb{U}} \to \mathbb{R}_+$ is the running cost and we write $\mathbb{E}_i^U$ to emphasize that the underlying control policy is U and $X_0 = i$. Whenever the control policy v lies in $\mathfrak{U}_{\mathrm{SM}}$, we write $\mathbb{E}_i^v$.

The objective of the ERSC problem is to minimize the cost in (2.1) over all the admissible controls U and initial conditions $1 \le i \le n$, *i.e.*, to find

$$\lambda^{*,\delta} \doteq \min_{1\le i\le n} \inf_{U\in\mathfrak{U}} \mathcal{E}_i^\delta(U) \tag{2.2}$$

and also to characterize optimal stationary Markov control policies v (if they exist), *i.e.*, a stationary Markov policy v satisfying $\mathcal{E}_i^\delta(v) = \lambda^{*,\delta}$, for $1 \le i \le n$.

Remark 2.1. It is well known that as $\delta \to 0$, $\lambda^{*,\delta}$ converges to $\lambda^{*,0}$ the optimal cost of the long-run average (ergodic) cost control problem, given by

$$\lambda^{*,0} \doteq \min_{1\le i\le n} \inf_{U\in\mathfrak{U}} \mathcal{E}_i^0(U), \quad \text{where} \quad \mathcal{E}_i^0(U) \doteq \limsup_{T\to\infty} \frac{1}{T} \mathbb{E}_i^U \left[\sum_{t=0}^{T-1} c(X_t, U_t) \right].$$

For this reason, we often refer to this problem as the risk-neutral ergodic control problem.

We now state the conditions on the controlled transition kernel $p(i, j, u)$ and the process X under which we later prove the convergence of the RVI algorithms.

Assumption 2.1. The following conditions hold:

(1) For any $1 \le i, j \le n$, the functions $u \mapsto c(i, u)$ and $u \mapsto p(i, j, u)$ are continuous on $\mathbb{U}(i)$.
(2) The DTMC X is irreducible under all stationary Markov policies.

The following theorem (which is [19, Theorem 3.1]) provides the Bellman optimality criterion which completely characterizes the optimal stationary Markov controls.

Theorem 2.1. *Under Assumption 2.1, there exists a function $V^{*,\delta} : S \to \mathbb{R}_+$ that is unique up to a multiplicative constant and satisfies*

$$e^{\delta\lambda^{*,\delta}} V^{*,\delta}(i) = \min_{u\in\mathbb{U}(i)} \left[e^{\delta c(i,u)} \sum_{j=1}^n V^{*,\delta}(j) p(i, j, u) \right], \quad \text{for } 1 \le i \le n\,. \tag{2.3}$$

Moreover, $v \in \mathfrak{U}_{\mathrm{SM}}$ is optimal if and only if it satisfies

$$\min_{u\in\mathbb{U}(i)} \left[e^{\delta c(i,u)} \sum_{j=1}^n V^{*,\delta}(j) p(i, j, u) \right] = e^{\delta c(i,v(i))} \sum_{j=1}^n V^{*,\delta}(j) p(i, j, v(i)), \quad \text{for } 1 \le i \le n\,. \tag{2.4}$$

In what follows, we often suppress the dependence of $(V^{*,\delta}, \lambda^{*,\delta})$ on δ and simply write (V^*, λ^*).

2.2. **Design of RVI algorithms.** It is evident from (2.4) that to characterize an optimal stationary Markov policy, we require the knowledge of the value function V^* which in turn, requires the knowledge of λ^*. Therefore, it is of significant interest to numerically compute the pair (V^*, λ^*). As discussed in the Section 1.1, it is well known that the pair (V^*, λ^*) can be computed from a variety of numerical algorithms. Many of them fall into two categories: relative value iteration (RVI) and policy iteration. As mentioned already, we confine ourselves to investigating RVI type algorithms in this paper. Most of the earlier works on RVI algorithms for the ERSC problem study a version from [16], which we refer to as Algorithm 0, which generates iterates designed to find a fixed point of the Bellman equation from Theorem 2.1. In [16, Theorem 4.5], it is shown that as $k \to \infty$, $(V^k, \lambda^k) \to (V^*, \lambda^*)$, under the assumption that the controlled DTMC is irreducible, aperiodic, and recurrent for every stationary Markov control policy. However, the associated rate of convergence is not investigated.

Algorithm 0 Existing RVI Algorithm

(i) Initialize with $k = 0$ and $V^0 : S \to \mathbb{R}^+$.

(ii) Update: for $1 \le i \le n$,

$$\lambda^{k+1} = \min_{u \in \mathbb{U}(n)} \frac{1}{\delta} \log \left[e^{\delta c(n,u)} \sum_{j=1}^{n} V^k(j) p(n,j,u) \right],$$

$$V^{k+1}(i) = \min_{u \in \mathbb{U}(i)} \left[e^{\delta(c(i,u) - \lambda^{k+1})} \sum_{j=1}^{n} V^k(j) p(i,j,u) \right].$$

(iii) Set $k = k + 1$.

(iv) Repeat Steps (ii) and (iii).

The design of our RVI algorithms is inspired from the following observation: we rewrite (2.3) as

$$V^*(i) = \min_{u \in \mathbb{U}(i)} \left[e^{\delta(c(i,u) - \bar{\lambda}^*)} \Big(p(i,n,u) + \sum_{j=1}^{n-1} \frac{V^*(j)}{V^*(n)} p(i,j,u) \Big) \right], \quad \text{for } 1 \le i \le n, \tag{2.5}$$

$$\bar{\lambda}^* \doteq \lambda^* - \frac{1}{\delta} \log V^*(n). \tag{2.6}$$

Here and in what follows, state n is chosen as the reference state. We re-arrange the above displays with $\bar{V}^* \doteq \big(V^*(n)\big)^{-1} V^*$ as

$$V^*(i) = \min_{u \in \mathbb{U}(i)} \left[e^{\delta(c(i,u) - \bar{\lambda}^*)} \Big(p(i,n,u) + \sum_{j=1}^{n-1} \bar{V}^*(j) p(i,j,u) \Big) \right], \quad \text{for } 1 \le i \le n,$$

$$\lambda^* = \bar{\lambda}^* + \frac{1}{\delta} \log V^*(n).$$

From here, we see the clear evidence of a Jacobi-type iteration procedure with iterates (V^k, λ^k) if $(\bar{V}^*, \bar{\lambda}^*)$ is replaced by (V^k, λ^k) and (V^*, λ^*) is replaced by (V^{k+1}, λ^{k+1}), for $k \ge 0$ with (V^0, λ^0) chosen as the initial condition. We note two key observations regarding the resulting iteration procedure: (i) the condition for $\lambda^{k+1} = \lambda^k$, *i.e.*, the fixed point of the iteration procedure, is that $V^{k+1}(n) = 1$, and (ii) (V^*, λ^*) is a fixed point. We also note that these facts remain true if we replace $\delta^{-1} \log V^{k+1}(n)$ by $\gamma_k \delta^{-1} \log V^{k+1}(n)$, for any family of positive real numbers $\{\gamma_k\}_{k \in \mathbb{Z}_+}$. From the above discussion, we have the following RVI algorithm:

Algorithm 1 Jacobi-like RVI algorithm

(i) Initialize $k = 0$, $V^0 : S \to \mathbb{R}^+$ and $\lambda^0 \in \mathbb{R}_+$.

(ii) Choose step-size $0 < \gamma_k < 1$ and update: for $1 \le i \le n$,

$$\begin{aligned} V^{k+1}(i) &= \min_{u \in \mathbb{U}(i)} \left[e^{\delta(c(i,u) - \lambda^k)} \Big(p(i,n,u) + \sum_{j=1}^{n-1} V^k(j) p(i,j,u) \Big) \right], \\ \lambda^{k+1} &= \lambda^k + \frac{1}{\delta} \gamma_k \log V^{k+1}(n). \end{aligned} \tag{2.7}$$

(iii) Set $k = k + 1$.

(iv) Repeat Steps (ii) and (iii).

Replacing (2.7) with its Gauss-Seidel version gives us the algorithm below.

Algorithm 2 Gauss-Seidel-like RVI algorithm

(i) Initialize $k=0$, $V^0 : S \to \mathbb{R}^+$ and $\lambda^0 \in \mathbb{R}_+$.

(ii) Choose step-size $0 < \gamma_k < 1$ and update: for $2 \le i \le n$,

$$V^{k+1}(1) = \min_{u\in\mathbb{U}(1)} \left[e^{\delta(c(1,u)-\lambda^k)} \Big(p(1,n,u) + \sum_{j=1}^{n-1} V^k(j) p(1,j,u) \Big) \right],$$

$$V^{k+1}(i) = \min_{u\in\mathbb{U}(i)} \left[e^{\delta(c(i,u)-\lambda^k)} \Big(p(i,n,u) + \sum_{j=1}^{i-1} V^{k+1}(j) p(i,j,u) + \sum_{j=i}^{n-1} V^k(j) p(i,j,u) \Big) \right],$$

$$\lambda^{k+1} = \lambda^k + \frac{1}{\delta}\gamma_k \log V^{k+1}(n)\,.$$

(iii) Set $k = k+1$.

(iv) Repeat Steps (ii) and (iii).

Remark 2.2. The reason behind the inclusion of $\{\gamma_k\}_{k\in\mathbb{Z}_+}$ in the designs of Algorithms 1 and 2 is two-fold: (i) we will see in Theorem 2.2 that the inclusion of an appropriate family of $\{\gamma_k\}_{k\in\mathbb{Z}_+}$ ensures the geometric convergence of these algorithms, and (ii) while numerically implementing these algorithms, the inclusion of an appropriate choice of $\{\gamma_k\}_{k\in\mathbb{Z}_+}$ can aid us in achieving faster convergence; see Section 6 for more elaborate discussions.

To investigate the convergence of Algorithms 1 and 2, it is evident that the risk-sensitive Bellman-like operators $\widetilde{F}, \widetilde{G} : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$ defined by

$$\widetilde{F}_i(V,\lambda) \doteq \min_{u\in\mathbb{U}(i)} \left[e^{(\delta c(i,u)-\lambda)} \Big(p(i,n,u) + \sum_{j=1}^{n-1} V(j) p(i,j,u) \Big) \right], \quad \text{for } 1 \le i \le n,$$

$$\widetilde{G}_i(V,\lambda) \doteq \begin{cases} \min_{u\in\mathbb{U}(1)} \left[e^{(\delta c(1,u)-\lambda)} \Big(p(1,n,u) + \sum_{j=1}^{n-1} V(j) p(1,j,u) \Big) \right], & \text{for } i = 1, \\ \min_{u\in\mathbb{U}(i)} \left[e^{(\delta c(i,u)-\lambda)} \Big(p(i,n,u) + \sum_{j=1}^{i-1} \widetilde{G}_j(V,\lambda) p(i,j,u) + \sum_{j=i}^{n-1} V(j) p(i,j,u) \Big) \right], & \text{for } i \ne 1, \end{cases}$$

play a fundamental role. In terms of these operators, Algorithm 1 can be expressed as

$$V^{k+1} = \widetilde{F}(V^k, \lambda^k),$$
$$\lambda^{k+1} = \lambda^k + \frac{1}{\delta}\gamma_k \log V^{k+1}(n),$$

and Algorithm 2 can be expressed as

$$V^{k+1} = \widetilde{G}(V^k, \lambda^k),$$
$$\lambda^{k+1} = \lambda^k + \frac{1}{\delta}\gamma_k \log V^{k+1}(n)\,.$$

Observe that for every $\lambda \in \mathbb{R}$, the fixed point equations of $\widetilde{F}(\cdot,\lambda)$ and $\widetilde{G}(\cdot,\lambda)$ are the same. The following result provides an alternative characterization of the fixed points, whenever they exist - we only discuss the case of $\widetilde{F}(\cdot,\lambda)$, as the same discussion applies to $\widetilde{G}(\cdot,\lambda)$. However, it is not a priori clear if for every $\lambda \in \mathbb{R}$, there exists a fixed point for $\widetilde{F}(\cdot,\lambda)$; see Remark 2.5. To that end, define

$$\Lambda \doteq \Big\{ \lambda \in \mathbb{R} : \text{ a fixed point of } \widetilde{F}(\cdot,\lambda) \text{ exists and is finite} \Big\}. \tag{2.8}$$

Proposition 2.1. *For $\lambda \in \Lambda$, under Assumption 2.1, there exists a unique function $V_\lambda : S \to \mathbb{R}$ such that $V_\lambda = \widetilde{F}(V_\lambda, \lambda)$. Moreover, the following hold.*

(i) $V_\lambda(n) = 1$ if and only if $\lambda = \lambda^*$, where λ^* is as defined in (2.2).

(ii) An alternative characterization of V_λ holds: for $1 \le i \le n$,

$$V_\lambda(i) \doteq \inf_{v \in \mathfrak{U}_{\mathrm{SM}}} \mathbb{E}_i^v \left[e^{\sum_{t=0}^{\tau_n - 1} \left(\delta c(X_t, v(X_t)) - \lambda \right)} \right]. \tag{2.9}$$

Here, $\tau_n \doteq \min\{t \ge 1 : X_t = n\}$ with i-dependence suppressed.

Proof. Part (i) follows from the fact that $V_{\lambda^*}(n) = 1$ and that the map $\lambda \mapsto V_\lambda$ is injective, and Part (ii) follows from [19, Theorem 5.1]. □

A few remarks are now in order.

Remark 2.3. Observe that from Theorem 2.1 and the above proposition, whenever $V_\lambda(n) = 1$ (or equivalently, $\lambda = \lambda^*$), we have $V_\lambda = V^*$, where V^* is given by Theorem 2.1.

Remark 2.4. From Proposition 2.1, it is clear that V_λ can be interpreted as the infimum of the exponential cost

$$\mathbb{E}_i^v \left[e^{\sum_{t=0}^{\tau_n - 1} \left(\delta c(X_t, v(X_t)) - \lambda \right)} \right],$$

over all $v \in \mathfrak{U}_{\mathrm{SM}}$, incurred while the DTMC X starts at state i and enters state n, for the first time. In other words, this means that $v^* \in \mathfrak{U}_{\mathrm{SM}}$ for which the infimum in (2.9) is attained (which exists by Assumption 2.1) corresponds to the exponentially stochastic shortest path from state i to state n, while varying over all $v \in \mathfrak{U}_{\mathrm{SM}}$. The analogous connection between the risk-neutral stochastic shortest path problem and risk-neutral RVI algorithms was thoroughly investigated in [6]. This problem was also studied in [19], although the authors referred to it as the "auxiliary expected-total cost" problem.

Remark 2.5. From the alternative characterization in Proposition 2.1, due to the exponential nature of the cost, it is more apparent that V_λ may not exist for every λ, *i.e.*, $V_\lambda(i)$ may be infinite for some $1 \le i \le n$ and $\lambda \in \mathbb{R}$. This illustrates more clearly the reason for defining Λ in (2.8).

We remark that much of the analysis that follows is for operator F, as many of these techniques can be implemented for operator G as well, in conjunction with Lemma 4.1 and Proposition 4.3 from Section 4.2.

From Theorem 2.1, we make the following trivial, but important observation: for that we set $h^* = \log V^*$ and take the logarithm on both sides of (2.3) and (2.4). It is clear that (2.3) and (2.4), respectively become

$$h^*(i) = \min_{u \in \mathbb{U}(i)} \left[\delta c(i,u) + \log \Big(\sum_{j=1}^n e^{h^*(j)} p(i,j,u) \Big) \right] - \delta \lambda^*, \tag{2.10}$$

$$\min_{u \in \mathbb{U}(i)} \left[\delta c(i,u) + \log \Big(\sum_{j=1}^n e^{h^*(j)} p(i,j,u) \Big) \right] = \delta c(i, v(i)) + \log \Big(\sum_{j=1}^n e^{h^*(j)} p(i,j,v(i)) \Big), \tag{2.11}$$

for $1 \le i \le n$. From the above, it is clear that $v \in \mathfrak{U}_{\mathrm{SM}}$ satisfies (2.4) if and only if it satisfies (2.11). Hence, to characterize the optimal stationary Markov policies associated with our ERSC problem, we can equivalently work with (2.10) and (2.11). For this reason, instead of investigating operators $\widetilde{F}$ and $\widetilde{G}$, we investigate the operators $F, G : \mathbb{R}^n \to \mathbb{R}^n$, also referred to as risk-sensitive Bellman operators, given by

$$F_i(h, \lambda) \doteq \min_{u \in \mathbb{U}(i)} \left[\delta c(i,u) + \log \Big(p(i,n,u) + \sum_{j=1}^{n-1} e^{h(j)} p(i,j,u) \Big) \right] - \lambda, \quad \text{for } 1 \le i \le n, \tag{2.12}$$

$$G_i(h,\lambda) \doteq$$
$$\begin{cases} \min_{u\in\mathbb{U}(1)} \left[\delta c(1,u) + \log\left(p(1,n,u) + \sum_{j=1}^{n-1} e^{h(j)} p(1,j,u)\right)\right] - \lambda, & \text{for } i = 1, \\ \min_{u\in\mathbb{U}(i)} \left[\delta c(i,u) + \log\left(p(i,n,u) + \sum_{j=1}^{i-1} e^{G_j(h,\lambda)} p(i,j,u) + \sum_{j=i}^{n-1} e^{h(j)} p(i,j,u)\right)\right] - \lambda, & \text{for } i \neq 1\,. \end{cases} \tag{2.13}$$

Observe that for every $(V,\lambda) \in \mathbb{R}^n \times \mathbb{R}$, $F(\log V, \lambda) = \log \widetilde{F}(V,\lambda)$ and $G(\log V, \lambda) = \log \widetilde{G}(V,\lambda)$. In terms of operators F and G with $h^k \doteq \log V^k$, Algorithm 1 can be equivalently expressed as

$$\begin{aligned} h^{k+1} &= F(h^k, \lambda^k), \\ \lambda^{k+1} &= \lambda^k + \frac{1}{\delta}\gamma_k h^{k+1}(n), \end{aligned} \tag{2.14}$$

and Algorithm 2 can be expressed as

$$\begin{aligned} h^{k+1} &= G(h^k, \lambda^k), \\ \lambda^{k+1} &= \lambda^k + \frac{1}{\delta}\gamma_k h^{k+1}(n)\,. \end{aligned} \tag{2.15}$$

2.3. **Main result.** We are now in a position to state our first main result which concerns the convergences of Algorithms 1 and 2. Before we proceed to do this, we introduce a weighted supremum norm that turns out to be fundamental in proving these convergences. We follow the construction from [33, Pg. 293]. First, we consider a finite partition $\{S_k\}_{1\le k\le l}$ of the state space $S = \{1,\dots,n\}$ defined as follows:

$$S_1 \doteq \{n\}, \quad S_k \doteq \left\{ i \in S : i \notin \cup_{r\le k-1} S_r \text{ and } \min_{u\in\mathbb{U}(i)} \max_{j\in S_1\cup\cdots\cup S_{k-1}} p(i,j,u) > 0 \right\}.$$

The non-emptiness of S_k, for $k > 1$, is due to the irreducibility of the DTMC X in Assumption 2.1. For any state i, let $1 \le k(i) \le l$ be such that $i \in S_{k(i)}$. We now define the relevant quantities which will be used to construct our weighted supremum norm. First, we define η to be the *smallest, strictly positive* transition probability. In other words,

$$\eta \doteq \inf\left\{p(i,j,u) :\ p(i,j,u) > 0 \ \text{ for } 1 \le i,j \le n,\ u \in \mathbb{U}(i)\right\}. \tag{2.16}$$

The continuity of $u \mapsto p(i,j,u)$ (for every $i,j \in S$) from Assumption 2.1 implies that $\eta \in (0,1)$. Using η and the partition $\{S_k\}_{1\le k\le l}$ defined above, we define weights $\{w_i^m\}_{1\le i\le n}$ as follows:

$$w_i^m \doteq 1 - (\eta e^{-2m})^{2k(i)}, \ \text{ for } i < n \text{ with } w_n^m = 1, \tag{2.17}$$

$$\beta_m \doteq \frac{1-(\eta e^{-2m})^{2l-1}}{1-(\eta e^{-2m})^{2l}}, \tag{2.18}$$

for every $m > 0$. Since $\eta \in (0,1)$, it follows that $\beta_m \in (0,1)$. The weighted supremum norm is then defined as follows: for $x \in \mathbb{R}^n$,

$$\|x\|_m \doteq \max_{1\le i\le n} \frac{|x_i|}{w_i^m}\,. \tag{2.19}$$

From the definition of w_i^m in (2.17), it is clear that $w_i^m > w_1^m > w_1^0$. This means that for $x \in \mathbb{R}^n$,

$$\|x\|_\infty \le \|x\|_m \le \frac{1}{w_1^0}\|x\|_\infty\,. \tag{2.20}$$

Let

$$\underline{c} \doteq \min_{(i,u)\in S_{\mathbb{U}}} c(i,u) \quad \text{and} \quad \bar{c} \doteq \max_{(i,u)\in S_{\mathbb{U}}} c(i,u)\,. \tag{2.21}$$

From the definition of λ^* in (2.2), it is clear that $\underline{c} \le \lambda^* \le \bar{c}$. Hence, we simply choose the initial condition of Algorithms 1 and 2, (V^0,λ^0) to be such that $\underline{c} \le \lambda^0 \le \bar{c}$.

Theorem 2.2. *Suppose that $\{(V^k, \lambda^k)\}_{k\in\mathbb{N}}$ is the sequence of iterate pairs of either Algorithm 1 or Algorithm 2 and $h^k \doteq \log V^k$. Under Assumption 2.1 and the condition that $\lambda^* \le \lambda^0 \le \bar{c}$, there exist constants $m > 0$ (depending on (h^0, λ^0)), $0 < \underline{\gamma} < \overline{\gamma}$ and positive functions $c_h(\gamma, m)$, $c_\lambda(\gamma, m)$ and $L(m)$ with the following property: for $k \ge 1$, whenever $\gamma_k \in [\underline{\gamma}, \overline{\gamma}]$, we have $0 < c_h(\gamma_k, m), c_\lambda(\gamma_k, m) < 1$, and*

$$\begin{aligned} \|h^{k+1} - h^*\|_m &\le c_h(\gamma_k, m)\|h^k - h^*\|_m + L(m)|\lambda^k - \lambda^*|, \\ |\lambda^{k+1} - \lambda^*| &\le c_\lambda(\gamma_k, m)|\lambda^k - \lambda^*|\,. \end{aligned} \tag{2.22}$$

In particular, there exists some constant $\varrho = \varrho(m) > 1$ (depending on m such that $\varrho(m) \to 1$ as $m \to \infty$) such that

$$\lim_{k\to\infty} \varrho^k \big(\|V^k - V^*\|_m + |\lambda^k - \lambda^*|\big) = 0\,.$$

Remark 2.6. The condition $\lambda^0 \ge \lambda^*$ can be easily satisfied - simply choose a $v \in \mathfrak{U}_{\mathrm{SM}}$, $1 \le i \le n$ and compute $\mathcal{E}_i^\delta(v)$. Then, by definition of λ^*, $\lambda^0 \doteq \mathcal{E}_i^\delta(v) \ge \lambda^*$.

We end this section with the sketch of the proof of Theorem 2.2 and defer the detailed proof to Section 4 and the Electronic Companion. Recall that Algorithms 1 and 2, in terms of $h^k = \log V^k$, are expressed according to (2.14) and (2.15), respectively. The main properties given below (which are proved in Section 3) form the building blocks of the proof:

(P1) Local contraction property of $F(\cdot, \lambda)$ (see Theorem 3.1) and $G(\cdot, \lambda)$ (see Proposition 4.3) under the weighted supremum norm defined in (2.19).

(P2) Lipschitz continuity of $F(h, \cdot)$ and $G(h, \cdot)$: for $F(h, \cdot)$, we obtain this continuity globally. However, due to the iterative nature in the definition of $G(h, \lambda)$, we only obtain this continuity locally; see Proposition 4.3.

(P3) Local bi-Lipschitz continuity of the map $\lambda \mapsto h_\lambda(n)$, whenever $F(\cdot, \lambda)$ has a unique fixed point h_λ: more precisely, we show that both $\lambda \mapsto h_\lambda$ and its inverse are locally Lipschitz in Lemma 3.2.

In the sketch below, we mainly discuss the case where $\{(V^k, \lambda^k)\}_{k\in\mathbb{N}}$ is a sequence of iterates of Algorithm 1 as the sketch of the proof for Algorithm 2 follows similar arguments from hereon. To illustrate the main idea behind the proof, we first discuss the local one-iteration analysis below, *i.e.*, given that (h^k, λ^k) is close to $(h_{\lambda^*}, \lambda^*)$ for some $k \in \mathbb{N}$, does (h^{k+l}, λ^{k+l}) converge to $(h_{\lambda^*}, \lambda^*)$, as $l \to \infty$? Suppose for an initialization of Algorithm 1 and $k \in \mathbb{N}$, we have $\lambda^k \simeq \lambda^*$, $\lambda^k \le \lambda^*$ and $h^k \simeq h_{\lambda^*}$ ($x \simeq y$ means that x is very close to y). Then, Property (P2) gives us $h^{k+1} = F(h^k, \lambda^k) \simeq F(h^k, \lambda^*)$. Since $F(\cdot, \lambda^*)$ is a local contraction and h_{λ^*} is the unique fixed point (see Property (P1)), $\|h^{k+1} - h_{\lambda^*}\|$ is strictly smaller than $\|h^k - h_{\lambda^*}\|$. Turning to $\lambda^{k+1} - \lambda^*$, from the first part of Property (P3), we have $h_{\lambda^k} \simeq h_{\lambda^*} \simeq h^k$ which means that the continuity of $F(\cdot, \lambda^k)$ (in fact, it is a local contraction; Property (P1)) gives us $h^{k+1} = F(h^k, \lambda^k) \simeq F(h_{\lambda^k}, \lambda^k) \simeq h_{\lambda^k}$. From the second part of Property (P3), monotonicity of $\lambda \mapsto h_\lambda(n)$ and the fact that $h_{\lambda^*}(n) = 0$, we have $h^{k+1}(n) \simeq h_{\lambda^k}(n) \gtrsim \underline{C}(\lambda^* - \lambda^k)$ (here, $x \gtrsim y$ means that $y - x$ less than a small positive number), for some constant $\underline{C} > 0$. On the other hand, from the first part of Property (P3), we also have $h^{k+1}(n) \lesssim \overline{C}(\lambda^* - \lambda^k)$, for some constant $\overline{C} > 0$ (here, $x \lesssim y$ means that $x - y$ is less than a small positive number). Combining this with the recursion for λ^k and choosing γ_k from an appropriate closed interval, it will be shown to ensure that $|\lambda^{k+1} - \lambda^*|$ is strictly smaller than $\lambda^* - \lambda^k$.

Similarly, we can argue that if $\lambda^k \simeq \lambda^*$, $\lambda^k \ge \lambda^*$ and $h^k \simeq h_{\lambda^*}$, then $\|h^{k+1} - h_{\lambda^*}\|$ is strictly smaller than $\|h^k - h_{\lambda^*}\|$ and $|\lambda^{k+1} - \lambda^*|$ is strictly smaller than $\lambda^k - \lambda^*$. Iterating the above argument indefinitely establishes the local convergence. It turns out that the precise argument is robust enough to relax the above assumption of $\lambda^k \simeq \lambda^*$ and $h^k \simeq h_{\lambda^*}$, and to be used in the proof of convergence of Algorithm 1.

Due to the 'local' nature of the above properties, it is essential to ensure that the constants involved in the above arguments do not deteriorate in the subsequent iterations. This is a very subtle issue owing to the fact that in general, $\mathbb{R} \setminus \Lambda \neq \emptyset$, which is addressed in Proposition 4.1.

3. Properties of the risk-sensitive Bellman operator $F(\cdot, \lambda)$

In this section, we establish the following key results regarding the risk-sensitive Bellman operator $F(\cdot, \lambda)$:

(i) For every $\lambda \in \mathbb{R}$, we establish a local contraction property of operator $F(\cdot, \lambda)$. This, in turn, will help us in establishing the existence and uniqueness of a $\mathbb{R}^n$–valued vector h_λ such that $h_\lambda = F(h_\lambda, \lambda)$.

(ii) For every $\lambda \in \Lambda$ for Λ in (2.8), we establish the local Lipschitz continuity of the map $\lambda \mapsto h_\lambda$ and also that of its inverse.

The existing approaches in proving (local) contraction properties involve showing that the multiplicative Bellman operator is a (local) contraction in the span-norm, and then normalizing the operator output at state n to be 1 (see [8]). However, these methods do not extend to our case because $F(h, \lambda)$ is not fixed. The techniques in [6] are amenable for adaptation to our case and as mentioned already, this is the reason for investigating F instead of $\widetilde{F}$. However, this is not straightforward due to the logarithm term in the definition of $F(h, \lambda)$ in (2.12). This can be handled using the entropy variational formula below.

3.1. Entropy variational formula.

In this section, we first present the well-known entropy variational formula as given in [23, Proposition 1.4.2] and then apply it to the operator $F(h, \lambda)$ which will help us re-write the logarithm term in the definition of F in a much more amenable form.

Proposition 3.1. *Let $(\mathcal{X}, \mathcal{G})$ be a Borel measurable space and $f : \mathcal{X} \to \mathbb{R}$ be a bounded and measurable functional. Then for a probability measure P on $\mathcal{X}$,*

$$\log \int_{\mathcal{X}} e^{f(x)} P(dx) = \sup_{Q \in \mathcal{P}(\mathcal{X})} \Big[\int_{\mathcal{X}} f(x) Q(dx) - R(Q \,||\, P) \Big], \tag{3.1}$$

where $R(Q||P)$ denotes the relative entropy of Q with respect to P, i.e.,

$$R(Q||P) \doteq \begin{cases} \int_{\mathcal{X}} \log \frac{dQ}{dP}(x) Q(dx), & \text{if } Q \ll P, \\ \infty, & \text{otherwise.} \end{cases}$$

Furthermore, the supremum in (3.1) is uniquely attained at the probability measure Q^* defined by

$$Q^*(dx) \doteq \frac{e^{f(x)}}{\int_{\mathcal{X}} e^{f(y)} P(dy)} P(dx)\,. \tag{3.2}$$

Applying Proposition 3.1 to the sum (which plays the role of integral in our case) in the definition of $F(\cdot, \cdot)$ in (2.12) gives us the following corollary.

Corollary 3.1. *For any $h \in \mathbb{R}^n$, $\lambda \in \mathbb{R}$ and $1 \leq i \leq n$, we have*

$$F_i(h, \lambda) = \min_{u \in \mathbb{U}(i)} \Big[\delta c(i,u) + \sup_{q(i,\cdot,u) \in \mathcal{P}(S)} \Big(\sum_{j=1}^{n-1} h(j) q(i,j,u) - R\big(q(i,\cdot,u) || \, p(i,\cdot,u)\big) \Big) \Big] - \lambda\,. \tag{3.3}$$

Here, $R\big(q(i,\cdot,u) || \, p(i,\cdot,u)\big) = \sum_{j=1}^{n} q(i,j,u) \log \frac{q(i,j,u)}{p(i,j,u)}$.

Remark 3.1. Observe that the supremum in (3.3) can be equivalently written as

$$\sup_{q \in \mathcal{P}(S)} \Big(\sum_{j=1}^{n-1} h(j) q(j) - R\big(q(\cdot) || p(i,\cdot,u)\big) \Big)\,. \tag{3.4}$$

However, we choose to write as the supremum over $q(i,\cdot,u) \in \mathcal{P}(S)$ (in (3.3) and in what follows) to aid the reader in keeping track of the underlying dependency of the optimal probability measure $q \in \mathcal{P}(S)$ on (i,u).

Proof. To begin with, for any $h \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$, define $\tilde{h} \in \mathbb{R}^n$ as $\tilde{h}(j) \doteq h(j)$ for $1 \le j \le n-1$ and 0 for $j = n$. Then, the expression defining $F(h,\lambda)$ in (2.12) becomes

$$F_i(h,\lambda) = \min_{u \in \mathbb{U}(i)} \Big[\delta c(i,u) + \log \mathbb{E}_i^u [e^{\tilde{h}(X_1)}] \Big] - \lambda, \quad \text{for } 1 \le i \le n\,. \tag{3.5}$$

Applying Proposition 3.1 to $\log \mathbb{E}_i^u \left[e^{\tilde{h}(X_1)} \right]$ gives us

$$\begin{aligned} \log \mathbb{E}_i^u \Big[e^{\tilde{h}(X_1)} \Big] &= \sup_{q(i,\cdot,u) \in \mathcal{P}(\mathcal{S})} \Big[\sum_{j=1}^{n} \tilde{h}(j) q(i,j,u) - R\big(q(i,\cdot,u) \,||\, p(i,\cdot,u)\big) \Big] \\ &= \sup_{q(i,\cdot,u) \in \mathcal{P}(\mathcal{S})} \Big[\sum_{j=1}^{n-1} h(j) q(i,j,u) - R\big(q(i,\cdot,u) \,||\, p(i,\cdot,u)\big) \Big]\,. \end{aligned}$$

Plugging this back into (3.5), we obtain the desired result. □

3.2. **Local contraction of the operator $F(\cdot,\lambda)$.**

Theorem 3.1. *For $m > 0$ and $\lambda \in \mathbb{R}$, we have*

$$\|F(h_1,\lambda) - F(h_2,\lambda)\|_m \le \beta_m \|h_1 - h_2\|_m, \tag{3.6}$$

whenever $\|h_1\|_\infty$, $\|h_2\|_\infty \le m$. Here, $\beta_m \in (0,1)$ is as defined in (2.18).

Remark 3.2. Observe that the above theorem implies that $F(\cdot,\lambda)$ is only a local contraction as β_m depends on m and $\beta_m \to 1$, as $m \to \infty$ (this follows from (2.18)).

Proof. Fix $m > 0$, $\lambda \in \mathbb{R}$ and $h_1, h_2 \in \mathbb{R}^n$ such that $\|h_1\|_\infty, \|h_2\|_\infty \le m$. From Corollary 3.1, for $1 \le i \le n$, we have

$$F_i(h_1,\lambda) = \min_{u \in \mathbb{U}(i)} \Big[\delta c(i,u) + \sup_{q(i,\cdot,u) \in \mathcal{P}(S)} \Big(\sum_{j=1}^{n-1} h_1(j) q(i,j,u) - R\big(q(i,\cdot,u) \,||\, p(i,\cdot,u)\big) \Big) \Big] - \lambda, \tag{3.7}$$

$$F_i(h_2,\lambda) = \min_{u \in \mathbb{U}(i)} \Big[\delta c(i,u) + \sup_{q(i,\cdot,u) \in \mathcal{P}(S)} \Big(\sum_{j=1}^{n-1} h_2(j) q(i,j,u) - R\big(q(i,\cdot,u) \,||\, p(i,\cdot,u)\big) \Big) \Big] - \lambda\,. \tag{3.8}$$

Next, we choose $v_2 \in \mathfrak{U}_{\mathrm{SM}}$ such that it is a minimizer for (3.8). This gives us

$$\begin{aligned} &F_i(h_1,\lambda) - F_i(h_2,\lambda) \\ &\quad \le \sup_{q(i,\cdot,v_2(i)) \in \mathcal{P}(S)} \Big(\sum_{j=1}^{n-1} h_1(j) q(i,j,v_2(i)) - \sum_{j=1}^{n} q(i,j,v_2(i)) \log \Big(\frac{q(i,j,v_2(i))}{p(i,j,v_2(i))} \Big) \Big) \\ &\qquad - \sup_{q(i,\cdot,v_2(i)) \in \mathcal{P}(S)} \Big(\sum_{j=1}^{n-1} h_2(j) q(i,j,v_2(i)) - \sum_{j=1}^{n} q(i,j,v_2(i)) \log \Big(\frac{q(i,j,v_2(i))}{p(i,j,v_2(i))} \Big) \Big)\,. \end{aligned} \tag{3.9}$$

To arrive at the inequality, we use the fact that $v_2 \in \mathfrak{U}_{\mathrm{SM}}$ is sub-optimal for the minimization in (3.7). From Proposition 3.1, we know that

$$\sup_{q(i,\cdot,v_2(i)) \in \mathcal{P}(S)} \Big(\sum_{j=1}^{n-1} h_1(j) q(i,j,v_2(i)) - \sum_{j=1}^{n} q(i,j,v_2(i)) \log \Big(\frac{q(i,j,v_2(i))}{p(i,j,v_2(i))} \Big) \Big)$$

has a unique maximizer $q^* = q^*(i,\cdot,v_2(i))$ given by

$$q^*(i,j,v_2(i)) = \frac{e^{\tilde{h}_1(j)}p(i,j,v_2(i))}{\sum_{k=1}^n e^{\tilde{h}_1(k)}p(i,k,v_2(i))} \geq e^{-2m}p(i,j,v_2(i))\,. \tag{3.10}$$

To get the above inequality, we use $\|h_1\|_\infty \leq m$. Here, $\tilde{h}_1 \in \mathbb{R}^n$ is defined as $\tilde{h}_1(j) \doteq h_1(j)$, for $1 \leq j \leq n-1$ and 0, for $j = n$. Plugging this into (3.9) and using the fact that $q^*(i,\cdot,v_2(i))$ is sub-optimal for

$$\sup_{q(i,\cdot,v_2(i))\in\mathcal{P}(S)} \Big(\sum_{j=1}^{n-1} h_2(j)q(i,j,v_2(i)) - \sum_{j=1}^{n} q(i,j,v_2(i)) \log\Big(\frac{q(i,j,v_2(i))}{p(i,j,v_2(i))}\Big)\Big),$$

we get

$$F_i(h_1,\lambda) - F_i(h_2,\lambda) \leq \sum_{j=1}^{n-1}(h_1(j)-h_2(j))q^*(i,j,v_2(i))\,. \tag{3.11}$$

Following the arguments from the proof in [33, Lemma 3], we can conclude that $q^*(i,\cdot,v_2(i))$ satisfies $\sum_{j=1}^{n-1} w_j^m q^*(i,j,v_2(i)) \leq \beta_m w_i^m$, for $1 \leq i \leq n$. Recall w^m and β_m from (2.17) and (2.18), respectively. Now for the right hand side of (3.11), we have

$$\begin{aligned}\sum_{j=1}^{n-1} \frac{(h_1(j)-h_2(j))}{w_j^m} w_j^m q^*(i,j,v_2(i)) &\leq \beta_m w_i^m \max_{1\leq j\leq n}\left\{\frac{|h_1(j)-h_2(j)|}{w_j^m}\right\} \\ &= \beta_m w_i^m \|h_1-h_2\|_m.\end{aligned} \tag{3.12}$$

Hence we obtain $F_i(h_1,\lambda) - F_i(h_2,\lambda) \leq \beta_m w_i^m \|h_1-h_2\|_m$. Interchanging the roles of h_1 and h_2, we obtain $F_i(h_2,\lambda) - F_i(h_1,\lambda) \leq \beta_m w_i^m \|h_1-h_2\|_m$. From the above two displays, we have

$$\frac{|F_i(h_1,\lambda) - F_i(h_2,\lambda)|}{w_i^m} \leq \beta_m \|h_1-h_2\|_m, \quad \text{for } 1 \leq i \leq n,$$

which in turn, using the definition of $\|\cdot\|_m$ in (2.19) gives us the desired result. □

3.3. **Analysis of the fixed point of $F(\cdot,\lambda)$.** In this section, we consider the existence of fixed points of $F(\cdot,\lambda)$ and also discuss their properties, in terms of λ. We remark that the existence of these fixed points does not follow from Theorem 3.1 as the theorem only gives us the local contraction. Recall $\tau_n = \min\{t \geq 1 : X_t = n\}$, and define $\underline{\tau} \doteq \min_{1\leq i\leq n} \inf_{v\in\mathfrak{U}_{\rm SM}} \mathbb{E}_i^v[\tau_n]$ and $\overline{\tau} \doteq \max_{1\leq i\leq n} \sup_{v\in\mathfrak{U}_{\rm SM}} \mathbb{E}_i^v[\tau_n]$. It is well-known that under Assumption 2.1, $\underline{\tau} \leq \overline{\tau} < \infty$ (see [19, Theorem 4.1]). The following lemma gives us the lower bound of h_λ, in terms of λ, whenever h_λ is finite. Recall $\underline{c}$ from (2.21).

Lemma 3.1. *For any $\lambda \in \Lambda$ and $1 \leq i \leq n$, $h_\lambda(i) \geq \big(\delta\underline{c} - \lambda\big)\underline{\tau}$.*

Proof. From Proposition 2.1(ii), we know that $h_\lambda(i) = \inf_{v\in\mathfrak{U}_{\rm SM}} \log \mathbb{E}_i^v\Big[e^{\sum_{t=0}^{\tau_n-1}\left(\delta c(X_t,v(X_t))-\lambda\right)}\Big]$. Using Jensen's inequality, we have

$$h_\lambda(i) \geq \inf_{v\in\mathfrak{U}_{\rm SM}} \mathbb{E}_i^v\Big[\sum_{t=0}^{\tau_n-1}\big(\delta c(X_t,v(X_t))-\lambda\big)\Big] \geq (\delta\underline{c}-\lambda)\min_{1\leq i\leq n}\inf_{v\in\mathfrak{U}_{\rm SM}} \mathbb{E}_i^v[\tau_n]\,.$$

From the definition of $\underline{\tau}$, the lemma is proved. □

Next we investigate the local bi-Lipschitz continuity of the map $\lambda \mapsto h_\lambda$, whenever $\lambda \in \Lambda$. To do this, we further define

$$\Lambda_m \doteq \big\{\lambda \in \mathbb{R} : \max_{1\leq i\leq n} h_\lambda(i) \leq m\big\} \subset \Lambda\,. \tag{3.13}$$

Lemma 3.2. *Suppose $\lambda, \lambda' \in \Lambda$ and $\lambda > \lambda'$ and let $m' \doteq \max_{1\le i\le n} h_{\lambda'}(i)$ and $m \doteq \max_{1\le i\le n} h_{\lambda}(i)$. Then, we have*

$$\widetilde{N}_*(m,\lambda)(\lambda-\lambda') \le h_{\lambda'}(i) - h_\lambda(i) \le \widetilde{N}^*(m',\lambda')(\lambda-\lambda'), \tag{3.14}$$

for $1 \le i \le n$. Here,

$$\begin{aligned} \widetilde{N}^*(m',\lambda') &\doteq 1 + \max_{1\le i\le n} \sup_{v\in\mathfrak{U}_{\mathrm{SM}}} \mathbb{E}_i^v\Big[(\tau_n-1)e^{\left(m'-(\delta\underline{c}-\lambda')\underline{\tau}\right)\tau_n}\Big], \\ \widetilde{N}_*(m,\lambda) &\doteq 1 + \min_{1\le i\le n} \inf_{v\in\mathfrak{U}_{\mathrm{SM}}} \mathbb{E}_i^v\Big[(\tau_n-1)e^{\left((\delta\underline{c}-\lambda)\underline{\tau}-m\right)\tau_n}\Big]. \end{aligned} \tag{3.15}$$

Proof. We first show that

$$h_{\lambda'}(i) - h_\lambda(i) \le \widetilde{N}^*(m,\lambda')(\lambda-\lambda'), \qquad \text{for } 1\le i\le n\,. \tag{3.16}$$

From Corollary 3.1, we have

$$h_\lambda(i) = \min_{u\in\mathbb{U}(i)}\Big[\delta c(i,u) + \sup_{q(i,\cdot,u)\in\mathcal{P}(S)}\Big(\sum_{j=1}^n \tilde{h}_\lambda(j)q(i,j,u) - \sum_{j=1}^n q(i,j,u)\log\Big(\frac{q(i,j,u)}{p(i,j,u)}\Big)\Big)\Big] - \lambda,$$

where $\tilde{h}_\lambda(j) \doteq h_\lambda(j)$ for $1\le j\le n-1$ and 0 for $j=n$. Now choose $v\in\mathfrak{U}_{\mathrm{SM}}$ such that it is a minimizer of the above equation. For this choice of $v\in\mathfrak{U}_{\mathrm{SM}}$, we have

$$h_{\lambda'}(i) \le \delta c(i,v(i)) + \sup_{q(i,\cdot,v(i))\in\mathcal{P}(S)}\Big(\sum_{j=1}^n \tilde{h}_{\lambda'}(j)q(i,j,v(i)) - \sum_{j=1}^n q(i,j,v(i))\log\Big(\frac{q(i,j,v(i))}{p(i,j,v(i))}\Big)\Big) - \lambda'\,. \tag{3.17}$$

Here, $\tilde{h}_{\lambda'}$ is defined similarly as above. Next, choose $q^*(i,\cdot,v(i))\in\mathcal{P}(S)$ such that it achieves the supremum in the above display. Note that $q^*(i,\cdot,v(i))$ is given by

$$q^*(i,j,v(i)) = \frac{e^{\tilde{h}_{\lambda'}(j)}p(i,j,v(i))}{\sum_{j=1}^n e^{\tilde{h}_{\lambda'}(j)}p(i,j,v(i))} \le e^{m'-(\delta\underline{c}-\lambda')\underline{\tau}}p(i,j,v(i))\,. \tag{3.18}$$

In the above, we use the definition of m' from the hypothesis of the lemma and the bound from Lemma 3.1. This gives us

$$h_\lambda(i) \ge \delta c(i,v(i)) + \sum_{j=1}^n \tilde{h}_\lambda(j)q^*(i,j,v(i)) - \sum_{j=1}^n q^*(i,j,v(i))\log\Big(\frac{q^*(i.j,v(i))}{p(i,j,v(i))}\Big) - \lambda\,. \tag{3.19}$$

It is easy to deduce that $q^* = \big(q^*(i,j,v(i))\big)_{i,j\in S}$ is a Markov transition probability. Let $X^* = \{X_t^*\}_{t=0}^\infty$ denote the associated DTMC on S and τ_n^* denote the corresponding first return time to state n starting from i (we have suppressed the dependence on i). Finally, to keep the expressions below concise, we write the expectation associated with q^* as $\mathbb{E}^*$ and denote the law of X^* by $\mathbb{Q}^*$. If $X_0^* = i$, then we write $\mathbb{E}_i^*$ to emphasize this.

For $T>0$, we apply Dynkin's formula to $h_\lambda(X_t^*)$ and $h_{\lambda'}(X_t^*)$ (up to the stopping time $(\tau_n^*-1)\wedge T$), where $h_\lambda(\cdot)$ and $h_{\lambda'}(\cdot)$ satisfy (3.17) and (3.19), respectively. Fixing $i\ne n$, from Dynkin's formula, we obtain

$$h_{\lambda'}(i) = \tilde{h}_{\lambda'}(i) = \mathbb{E}_i^*\big[\tilde{h}_{\lambda'}(X^*_{\tau_n^*\wedge T})\big] + \mathbb{E}_i^*\left[\sum_{t=0}^{(\tau_n^*-1)\wedge T} -\mathcal{A}^*\tilde{h}_{\lambda'}(X_t^*)\right], \tag{3.20}$$

where $\mathcal{A}^*$ is the one step generator for the DTMC X^*, *i.e.*, $\mathcal{A}^*f(i) \doteq \mathbb{E}_i^*[f(X_1^*)] - f(i)$ for any $f: S\to\mathbb{R}$. Note that we can replace $h_{\lambda'}(x)$ with $\tilde{h}_{\lambda'}(x)$ inside the sum of the generators. The same procedure holds if we take $i=n$. From the definition of $\mathcal{A}^*\tilde{h}_{\lambda'}(x)$ and the inequality in (3.19), we have that

$$\mathcal{A}^*\tilde{h}_{\lambda'}(x) \ge -\delta c(x,v(x)) + \lambda' + R\big(q^*(x,\cdot,v(x))\|\,p(x,\cdot,v(x))\big)\,. \tag{3.21}$$

Combining (3.20) and (3.21) gives us

$$h_{\lambda'}(i) \le \mathbb{E}_i^*\left[\sum_{t=0}^{(\tau_n^*-1)\wedge T}\Big(\delta c(X_t^*, v(X_t^*)) - R\big(q^*(X_t^*,\cdot,v(X_t^*))\|\, p(X_t^*,\cdot,v(X_t^*))\big) - \lambda'\Big) + \tilde{h}_{\lambda'}(X^*_{\tau_n^*\wedge T})\right].$$

An analogous approach can be used to establish

$$h_{\lambda}(i) \ge \mathbb{E}_i^*\left[\sum_{t=0}^{(\tau_n^*-1)\wedge T}\big(\delta c(X_t^*, v(X_t^*)) - R\big(q^*(X_t^*,\cdot,v(X_t^*))\|\, p(X_t^*,\cdot,v(X_t^*))\big) - \lambda\big) + \tilde{h}_{\lambda}(X^*_{\tau_n^*\wedge T})\right].$$

From the above, we clearly have

$$h_{\lambda'}(i) - h_{\lambda}(i) \le \mathbb{E}_i^*\Big[\tau_n^*\wedge T\Big](\lambda-\lambda') + \mathbb{E}_i^*\Big[\tilde{h}_{\lambda'}(X^*_{\tau_n^*\wedge T}) - \tilde{h}_{\lambda}(X^*_{\tau_n^*\wedge T})\Big]. \tag{3.22}$$

From the monotone convergence theorem, we have $\mathbb{E}_i^*[\tau_n^*\wedge T]\uparrow\mathbb{E}_i^*[\tau_n^*]$, as $T\uparrow\infty$. Owing to this fact, we take $T\uparrow\infty$ in (3.22). This gives us

$$h_{\lambda'}(i) - h_{\lambda}(i) \le \mathbb{E}_i^*\big[\tau_n^*\big](\lambda-\lambda') + \limsup_{T\to\infty}\mathbb{E}_i^*\Big[|\tilde{h}_{\lambda'}(X^*_{\tau_n^*\wedge T})| + |\tilde{h}_{\lambda}(X^*_{\tau_n^*\wedge T})|\Big].$$

The term with lim sup above goes to zero as $T\uparrow\infty$ as $\tilde{h}_\lambda(n)$ and $\tilde{h}_{\lambda'}(n)$ are zero. To summarize, until now, we have shown that

$$h_{\lambda'}(i) - h_{\lambda}(i) \le \mathbb{E}_i^*\big[\tau_n^*\big](\lambda-\lambda'). \tag{3.23}$$

Below we obtain a simpler upper bound for $\mathbb{E}_i^*\big[\tau_n^*\big]$:

$$\mathbb{E}_i^*\big[\tau_n^*\big] = 1 + \sum_{N=1}^{\infty}(N-1)\mathbb{Q}^*(\tau_n^*=N) \le 1 + \sum_{N=1}^{\infty}(N-1)e^{\left(m'-(\delta\underline{c}-\lambda)\underline{\tau}\right)N}\mathbb{P}(\tau_n=N).$$

The above inequality can be proved as follows. The probability $\mathbb{Q}^*(\tau_n^*=N)$ involves exactly N transitions. On the other hand, using (3.18), we know that the probabilities of each of the intermediate transitions (under $\mathbb{Q}^*$) are bounded from above by $e^{m'-(\delta\underline{c}-\lambda')\underline{\tau}}$ times the probability of that transition under $\mathbb{P}$ (the original measure). In other words, we have

$$\mathbb{Q}^*(\tau_n^*=N) \le e^{\left(m'-(\delta\underline{c}-\lambda')\underline{\tau}\right)N}\mathbb{P}(\tau_n=N).$$

Now it is clear that the following holds

$$\sum_{N=1}^{\infty}(N-1)e^{m'N-(\delta\underline{c}-\lambda')\underline{\tau}N}\mathbb{P}(\tau_n=N) = \mathbb{E}_i^v\big[(\tau_n-1)e^{\left(m'-(\delta\underline{c}-\lambda')\underline{\tau}\right)\tau_n}\big].$$

This in turn implies that

$$\begin{aligned}\mathbb{E}_i^*\big[\tau_n^*\big] &\le 1 + \mathbb{E}_i^v\big[(\tau_n-1)e^{\left(m'-(\delta\underline{c}-\lambda')\underline{\tau}\right)\tau_n}\big]\\ &\le 1 + \max_{1\le i\le n}\sup_{v\in\mathfrak{U}_{\rm SM}}\mathbb{E}_i^v\big[(\tau_n-1)e^{\left(m'-(\delta\underline{c}-\lambda')\underline{\tau}\right)\tau_n}\big] = \widetilde{N}^*(m',\lambda').\end{aligned} \tag{3.24}$$

Combining (3.23) and (3.24) completes the proof of (3.16).

We now move on to show that

$$h_{\lambda'}(i) - h_{\lambda}(i) \ge \widetilde{N}_*(m,\lambda')(\lambda-\lambda'), \qquad \text{for } 1\le i\le n. \tag{3.25}$$

The proof follows along the same lines as the proof of (3.16). Hence we omit the details and only provide the expression analogous to (3.23). To do this, let us define a transition probability (analogous to (3.18))

$$q_*(i,j,v'(i)) = \frac{e^{\tilde{h}_\lambda(j)}p(i,j,v'(i))}{\sum_{j=1}^n e^{\tilde{h}_\lambda(j)}p(i,j,v'(i))} \ge e^{(\delta\underline{c}-\lambda)\underline{\tau}-m}p(i,j,v'(i)).$$

To get the inequality, we use the definition of m from the hypothesis of the lemma and the bound from Lemma 3.1. In the above, we choose $v' \in \mathfrak{U}_{\mathrm{SM}}$ such that

$$h_{\lambda'}(i) = \delta c(i, v'(i)) - \lambda' + \sup_{q(i,\cdot,v'(i)) \in \mathcal{P}(S)} \left(\sum_{j=1}^{n} \tilde{h}_{\lambda'}(j) q(i,j,v'(i)) - \sum_{j=1}^{n} q(i,j,v'(i)) \log \left(\frac{q(i,j,v'(i))}{p(i,j,v'(i))} \right) \right).$$

Let $X_* = \{X_{*,t}\}_{t=0}^{\infty}$ denote the associated DTMC on S and $\tau_{*,n}$ denote the corresponding first return time to state n starting from i (we have suppressed the dependence of i). Finally, to keep the expressions below concise, we write the expectation associated with q_* as $\mathbb{E}_*$ and denote the law of X_* by $\mathbb{Q}_*$. If $X_{*,0} = i$, then we write $\mathbb{E}_{*,i}$ to emphasize this.

Using the above construction, the expression analogous to (3.23) is given by

$$h_{\lambda'}(i) - h_{\lambda}(i) \geq \mathbb{E}_{*,i}\big[\tau_{*,n}\big](\lambda - \lambda'). \tag{3.26}$$

Using the fact that $q_*(i,j,v'(i)) \geq e^{(\underline{c}-\lambda)\underline{\tau}-m} p(i,j,v'(i))$, for $1 \leq i,j \leq n$ and again the arguments analogous to those in the proof of (3.24), we obtain

$$\mathbb{E}_{*,i}\big[\tau_{*,n}\big] \geq 1 + \min_{1\leq i\leq n} \inf_{v' \in \mathfrak{U}_{\mathrm{SM}}} \mathbb{E}_i^{v'}\left[(\tau_n - 1) e^{\left((\delta\underline{c}-\lambda)\underline{\tau}-m\right)\tau_n}\right] = \widetilde{N}_*(m,\lambda).$$

This completes the proof of the lemma. □

Remark 3.3. In the above proof, it may appear that we could have stopped after (3.23) while proving (3.16) as it gives us the desired upper bound of $h_{\lambda'}(i) - h_{\lambda}(i)$ in terms of $\lambda - \lambda'$. But it is still undesirable in the sense that it is in terms of $\mathbb{E}_i^*\big[\tau_n^*\big]$ which is difficult to compute. This is the reason behind us proceeding to obtain a much desirable upper bound on $\mathbb{E}_i^*\big[\tau_n^*\big]$ that is more readily computable as it is in terms of the original measure $\mathbb{P}$. A similar remark applies to the proof of (3.25).

Lemma 3.2 gives us the following important proposition. Recall Λ_m from (3.13).

Proposition 3.2. *The following hold:*

(i) For $\lambda, \lambda' \in \Lambda$ such that $\lambda > \lambda'$, we have $h_{\lambda}(i) < h_{\lambda'}(i)$, for every $1 \leq i \leq n$.

(ii) There exists $\lambda_c \in \mathbb{R}$ such that the set Λ defined in (2.8) can be written as $\Lambda = (\lambda_c, \infty)$. Moreover, it follows that for $1 \leq i \leq n$,

$$\lim_{\lambda \uparrow \infty} h_{\lambda}(i) = -\infty, \quad \lim_{\lambda \downarrow \lambda_c} h_{\lambda}(i) = \infty.$$

(iii) Let $m^* \doteq \max_{1\leq i\leq n} h^*(i)$. Then, for $m > m^*$ we have

$$\left(\lambda^* - \frac{m - m^*}{\widetilde{N}_*(m,\lambda^*)}, \infty\right) \subset \Lambda_m.$$

Proof. Part (i) follows immediately from the application of Proposition 2.1(ii) and the fact that $h_{\lambda} = \log V_{\lambda}$ and $h_{\lambda'} = \log V_{\lambda'}$.

To prove part (ii), we apply Lemmas 3.1 and 3.2 to get

$$(\delta c - \lambda')\underline{\tau} \leq h_{\lambda'}(i) \leq h_{\lambda}(i) - \widetilde{N}_*(m,\lambda)(\lambda' - \lambda), \quad \text{for } 1 \leq i \leq n,$$

where $m = \max_{1\leq i\leq n} h_{\lambda}(i)$ and $\widetilde{N}^*(m,\lambda)$ is defined in (3.15). From here, we can conclude that for any $\lambda' > \lambda$, $h_{\lambda'}$ exists. In other words, Λ is of the form (λ_c, ∞) for some $\lambda_c \in \mathbb{R}$. The rest of the proof for part (ii) follows trivially from here.

To prove part (iii), we apply Lemma 3.2 with λ^* and $\lambda \in \Lambda$ such that $\lambda^* \geq \lambda \geq \lambda^* - \frac{m-m^*}{\widetilde{N}^*(m,\lambda^*)}$. This gives us

$$h_{\lambda}(i) - h^*(i) \leq \widetilde{N}^*(m,\lambda)(\lambda^* - \lambda) \leq \frac{N^*(m,\lambda)(m - m^*)}{N(m,\lambda^*)}.$$

Using the fact that $\widetilde{N}^*(m,\lambda) \le \widetilde{N}^*(m,\lambda^*)$ (which follows from the definition) and the definition of m^*, we obtain that $h_\lambda(i) \le m - m^* + h^*(i) \le m$. Taking the maximum over $1 \le i \le n$ and from the definition of Λ_m, we conclude the claim in part (iii). □

4. Proof of Theorem 2.2

We split the proof into two cases: $\{(V^k,\lambda^k)\}_{k\in\mathbb{N}}$ is the sequence of iterate pairs of either Algorithm 1 or Algorithm 2. To begin with, recall that for $\lambda \in \Lambda$, h_λ satisfies $h_\lambda = F(h_\lambda,\lambda)$ (or $h_\lambda = G(h_\lambda,\lambda)$, as F and G inherit identical fixed point equations). We fix (h^0,λ^0) such that $\lambda^* \le \lambda^0 \le \bar{c}$, according to the hypothesis of Theorem 2.2; see also Remark 2.6. Since h^* exists, from Proposition 3.2(i), we can also infer that h_{λ^0} exists, *i.e.*, the operator $F(\cdot,\lambda^0)$ has a unique fixed point denoted by h_{λ^0}.

4.1. Proof of (2.22) in the case of Algorithm 1.

Set

$$\begin{aligned} m &= \frac{1}{w_0^1}\Big\{\big(\|h^0 - h^*\|_\infty + \|h^*\|_\infty + |\lambda^* - \lambda^0|\big) \vee \max_{1\le i\le n} h_{\lambda^0}(i)\Big\}, \\ N_*(m) &\doteq \widetilde{N}_*(m,\bar{c}), \qquad N^*(m) \doteq \widetilde{N}^*(m,\bar{c}), \\ \alpha_0 &\doteq \big(\|h^0 - h_{\lambda^0}\|_m\big) \vee \big(N_*(m)(\lambda^0 - \lambda^*)\big)\,. \end{aligned} \tag{4.1}$$

We begin by proving the following result concerning the one-iteration behavior of Algorithm 1.

Proposition 4.1. *Suppose for any $k \in \mathbb{Z}_+$, $0 < \alpha \le \alpha_0$ and the following holds:*

$$\begin{aligned} \|h^k - h_{\lambda^k}\|_m \le \alpha \quad \textit{and} \quad |\lambda^k - \lambda^*| \le \frac{\alpha}{N_*(m)}, \\ \|h^k\|_\infty \le m \quad \textit{and} \quad \lambda^k \in \Lambda_m\,. \end{aligned} \tag{4.2}$$

Then, whenever $|\lambda^{k+1} - \lambda^*| \le N_*(m)^{-1}\alpha$ and

$$\gamma_k \le \Big(\frac{m - m^*}{N_*(m)}\Big)\Big(\frac{1}{\frac{\alpha_0}{N_*(m)} + \bar{c} - \underline{c} + \varkappa(m)}\Big), \tag{4.3}$$

we have $\|h^{k+1}\|_\infty \le m$ and $\lambda^{k+1} \in \Lambda_m$. Here, $\varkappa(m) \doteq \max_{u\in\mathbb{U}(n)}\{(1-p(n,n,u))m - p(n,n,u)\}$.

Proof. To prove that $\|h^{k+1}\|_\infty \le m$, using (2.20), we get

$$\begin{aligned} \|h^{k+1}\|_\infty \le \|h^{k+1}\|_m &\le \|F(h^k,\lambda^*) - F(h^*,\lambda^*)\|_m + \|h^*\|_m + |\lambda^k - \lambda^*| \\ &\le \beta_m\|h^k - h^*\|_m + \|h^*\|_m + |\lambda^0 - \lambda^*| \\ &\le \frac{1}{w_1^0}\big(\|h^k - h^*\|_\infty + \|h^*\|_\infty\big) + |\lambda^0 - \lambda^*|\,. \end{aligned}$$

In the above, to get the first line, we use the definition of h^{k+1}, the fact that h^* satisfies $h^* = F(h^*,\lambda^*)$ and the triangle inequality; the second and third lines are consequences of Theorem 3.1, the fact that $\beta_m < 1$ and $|\lambda^k - \lambda^0| \le |\lambda^0 - \lambda^*|$ (which is clear from the hypothesis), and (2.20). From the definition of m, we can infer that $\|h^{k+1}\|_\infty \le m$.

To prove that $\lambda^{k+1} \in \Lambda_m$, we consider two cases:

Case (a): $\lambda^k \le \lambda^*$. Since $|\lambda^{k+1} - \lambda^*| \le N_*(m)^{-1}\alpha$ from the hypothesis, it is clear that $\lambda^k \le \lambda^{k+1}$ and from Proposition 3.2(i), we know that $h_{\lambda^{k+1}}$ exists and $h_{\lambda^{k+1}}(i) \le h_{\lambda^k}(i) \le m$, for $1 \le i \le n$. This proves that $\lambda^{k+1} \in \Lambda_m$.

Case (b): $\lambda^k > \lambda^*$. We further split the proof into two sub-cases: (b1) $h^{k+1}(n) > 0$ and (b2) $h^{k+1}(n) \le 0$. Under Case (b1), the proof is similar to that of proof in Case (a) as $\lambda^k < \lambda^{k+1}$.

For Case (b2), we obtain a lower bound on $h^{k+1}(n)$ as follows: let $u^* \in \mathbb{U}(n)$ be such that

$$\begin{aligned} h^{k+1}(n) &= c(n,u^*) + \log\Big(p(n,n,u^*) + \sum_{j=1}^{n-1} e^{h^k(j)} p(n,j,u^*)\Big) - \lambda^k \\ &\geq \underline{c} + p(n,n,u^*) + \sum_{j=1}^{n-1} h^k(j) p(n,j,u^*) - \lambda^k \\ &\geq \underline{c} + p(n,n,u^*) - (1-p(n,n,u^*))m - \lambda^k \\ &\geq \underline{c} - \lambda^k - \lambda^* + \lambda^* - \varkappa(m) \\ &\geq -\frac{\alpha}{N_*(m)} - \overline{c} + \underline{c} - \varkappa(m)\,. \end{aligned}$$

In the above, to get the first inequality, we use Jensen's inequality; to get the second inequality, we use the fact that $\|h^k\|_\infty \leq m$ and to get the fourth inequality, we use the second inequality in (4.2) and the fact that $\underline{c} \leq \lambda^* \leq \overline{c}$. We then obtain

$$\lambda^{k+1} - \lambda^* = \lambda^k - \lambda^* + \gamma_k h^{k+1}(n) \geq \gamma_k\Big(-\frac{\alpha}{N_*(m)} - \overline{c} + \underline{c} - \varkappa(m)\Big)\,.$$

From Proposition 3.2(iii), the fact that $\alpha \leq \alpha_0$ and the hypothesis of the lemma, $\lambda^{k+1} \in \Lambda_m$. This completes the proof. □

Proposition 4.2. *Suppose for any $k \in \mathbb{Z}_+$ and $0 < \alpha \leq \alpha_0$, (4.2) holds. Then, we have*

$$\|h^{k+1} - h_{\lambda^{k+1}}\|_m \leq \widetilde{c}_h \alpha, \quad \text{and} \quad |\lambda^{k+1} - \lambda^*| \leq \widetilde{c}_\lambda \frac{\alpha}{N_*(m)}, \tag{4.4}$$

for some constants $\widetilde{c}_h = \widetilde{c}_h(\gamma_k, m)$, $\widetilde{c}_\lambda = \widetilde{c}_\lambda(\gamma_k, m)$, such that $\widetilde{c}_h, \widetilde{c}_\lambda \in (0,1)$, whenever $\gamma_k \in (0, \widetilde{\gamma}]$, for some $\widetilde{\gamma} = \widetilde{\gamma}(m) > 0$.

The proof of Proposition 4.2 follows very closely the arguments used in the risk-neutral setting (see [6, Proposition 2.1]) with the only difference being that the constants involved are dependent on m. We however, provide the proof for the sake of completeness in Appendix A.

We now show that Propositions 4.1 and 4.2 together imply the first and second inequalities in (2.22): from Proposition 4.1, we know that $\|h^k\|_\infty \leq m$, $\lambda^k \in \Lambda_m$, for every $k \in \mathbb{Z}_+$. Hence, we obtain

$$\begin{aligned} \|h^{k+1} - h^*\|_m &\leq \|h^{k+1} - h_{\lambda^{k+1}}\|_m + \|h_{\lambda^{k+1}} - h^*\|_m \\ &\leq \widetilde{c}_h(\gamma_k, m)\|h^k - h^*\|_m + \widetilde{c}_h(\gamma_k, m)\|h_{\lambda^k} - h^*\|_m + \|h_{\lambda^{k+1}} - h^*\|_m \\ &\leq \widetilde{c}_h(\gamma_k, m)\|h^k - h^*\|_m + \frac{\big(\widetilde{c}_h(\gamma_k, m) + \widetilde{c}_\lambda(\gamma_k, m)\big)N^*(m)}{w_1^0}|\lambda^k - \lambda^*|\,. \end{aligned}$$

In the above, to get the second inequality we use (4.4) and the triangle inequality; to get the third inequality, we apply Lemma 3.2 to the pairs (h_{λ^k}, h^*) and $(h_{\lambda^{k+1}}, h^*)$ and use (4.4).

Therefore, the first and second inequalities of (2.22) are implied by setting $\bar{\gamma} = \min\{\widehat{\gamma}, \widetilde{\gamma}\}$, $c_h(\gamma, m) = \widetilde{c}_h(\gamma, m)$, $c_\lambda(\gamma, m) = \widetilde{c}_\lambda(\gamma, m)$ and

$$L(m) = \frac{\big(\widetilde{c}_h(\gamma_k, m) + \widetilde{c}_\lambda(\gamma_k, m)\big)N^*(m)}{w_1^0}\,.$$

4.2. **Proof of (2.22) in the case of Algorithm 2.** In this section, we suppose that $\{(V^k, \lambda^k)\}_{k \in \mathbb{N}}$ is the sequence of iterate pairs of Algorithm 2. Again, we provide the proof in terms of $h^k = \log V^k$. Recall operator G defined in (2.13). Also, recall that in terms of operator G and (h^k, λ^k), Algorithm 2 can be equivalently expressed as in (2.15).

We first show that $G(\cdot,\lambda)$ satisfies a local contraction property for any fixed $\lambda\in\Lambda$, by using the local contraction property of the operator $F(\cdot,\lambda)$ represented according to (3.5) and an argument which parallels the argument in the proof of Theorem 3.1. Before we proceed, we give a lemma, which provides a bound on $G_i(h,\lambda)-\lambda$ in terms $\|h\|_\infty$, for $1\le i\le n$. Since η, defined in (2.16), is strictly positive, we have $\eta n\le\sum_{j=1}^n p(i,j,u)=1$. This means that

$$0<\bar\eta\doteq 1-(n-1)\eta<1 \tag{4.5}$$

and recall $\bar c$ from (2.21).

Lemma 4.1. *For $1\le i\le n$, $\lambda\in\mathbb{R}$ and $m>0$, let $h\in\mathbb{R}^n$ be such that $\|h\|_\infty\le m$. Then, we have the following:*

$$-\frac{1}{1-\bar\eta}\le\frac{G_i(h,\lambda)}{\bar c+\|h\|_\infty+|\lambda|}\le\frac{1-\chi(m+|\lambda|)^{n-1}}{1-\chi(m+|\lambda|)}\,. \tag{4.6}$$

Here, for $l\ge 0$,

$$\chi(l)\doteq\max_{(i,u)\in S_{\mathbb{U}}}\frac{(1-p(i,n,u))e^{\bar c+l}}{p(i,n,u)+(1-p(i,n,u))e^{\bar c+l}}\,.$$

Remark 4.1. Observe that $\chi(l)<1$ for every $l>0$ and $\lim_{l\to\infty}\chi(l)=1$.

Proof. Fix $\lambda\in\mathbb{R}$ and $h\in\mathbb{R}^n$ satisfying $\|h\|_\infty\le m$. Now for $1\le i\le n$, choose $v\in\mathfrak{U}_{\rm SM}$ such that $v(i)$ achieves the minimum in the definition of $G_i(h,\lambda)$. For $1\le i\le n$, we apply Proposition 3.1 for the following choices: $\mathcal{X}=S$, $P=p(i,\cdot,v(i))$ and $f(\cdot)=\bar G^i(\cdot)$ with $\bar G^i:\mathbb{R}^n\to\mathbb{R}$ defined as

$$\bar G^i(j)=\begin{cases}G_j(h,\lambda), & j\le i-1,\\ h(j), & i\le j\le n-1,\\ 0, & j=n,\end{cases}$$

for $1<i\le n$ and $\bar G^1(\cdot)=h(\cdot)$. This gives us

$$\begin{aligned}G_1(h,\lambda)&=c(1,v(1))+\sup_{q(1,\cdot,v(1))\in\mathcal{P}(S)}\Big(\sum_{j=1}^{n-1}h(j)q(1,j,v(1))-\sum_{j=1}^{n}q(1,j,v(1))\log\Big(\frac{q(1,j,v(1))}{p(1,j,v(1))}\Big)\Big)-\lambda,\\ G_i(h,\lambda)&=c(i,v(i))+\sup_{q(i,\cdot,v(i))\in\mathcal{P}(S)}\Big(\sum_{j=1}^{i-1}G_j(h,\lambda)q(i,j,v(i))+\sum_{j=i}^{n-1}h(j)q(i,j,v(i))\\ &\qquad-\sum_{j=1}^{n}q(i,j,v(i))\log\Big(\frac{q(i,j,v(i))}{p(i,j,v(i))}\Big)\Big)-\lambda,\end{aligned}$$

for $2\le i\le n$. It is clear from choosing $q(i,j,v(i))=p(i,j,v(i))$ that we get

$$\begin{aligned}G_1(h,\lambda)&\ge c(1,v(1))+\sum_{j=1}^{n-1}h(j)p(1,j,v(1))-|\lambda|,\\ G_i(h,\lambda)&\ge c(i,v(i))+\sum_{j=1}^{i-1}G_j(h,\lambda)p(i,j,v(i))+\sum_{j=i}^{n-1}h(j)p(i,j,v(i))-|\lambda|,\end{aligned} \tag{4.7}$$

for $2\le i\le n$. From the first inequality in (4.7), it is clear that

$$G_1(h,\lambda)\ge-\bar c-\|h\|_\infty-|\lambda|\,. \tag{4.8}$$

Now let us compute the lower bound on $G_2(h,\lambda)$. From the second inequality in (4.7) and (4.8), we get

$$\begin{aligned} G_2(h,\lambda) &\geq c(2,v(2)) + G_1(h,\lambda)p(2,1,v(2)) + \sum_{j=2}^{n-1} h(j)p(2,j,v(2)) - |\lambda| \\ &\geq -(\bar{c} + \|h\|_\infty + |\lambda|) - (\bar{c} + \|h\|_\infty + |\lambda|)p(2,1,v(2)) \\ &\geq -(\bar{c} + \|h\|_\infty + |\lambda|) - (\bar{c} + \|h\|_\infty + |\lambda|)\bar{\eta}\,. \end{aligned}$$

Recall $\bar{\eta}$ from (4.5). In the above, to arrive at the second line, we use the definition of $\bar{c}$, $\|h\|_\infty$ and the fact that $\sum_{j=2}^{n-1} p(2,j,v(2)) < 1$. Similarly, for any $i \geq 2$, we obtain

$$G_i(h,\lambda) \geq -(\bar{c} + \|h\|_\infty + |\lambda|)\big(1 + \bar{\eta} + \bar{\eta}^2 + \ldots + \bar{\eta}^{i-1}\big) \geq -\frac{\bar{c} + m + |\lambda|}{1 - \bar{\eta}}\,. \tag{4.9}$$

To get the second inequality, we use the fact that $\bar{\eta} < 1$; see (4.5). Therefore, (4.8) and (4.9) together give us the desired lower bound in (4.6).

We next move on to the upper bound in (4.6). It is easy to see that

$$G_1(h,\lambda) \leq \bar{c} + \|h\|_\infty + |\lambda|\,. \tag{4.10}$$

Now we consider the case $i = 2$. Again, with the same $v = v(\cdot)$ as above, we have

$$\begin{aligned} G_2(h,\lambda) &= c(2,v(2)) + \log\Big(p(2,n,v(2)) + e^{G_1(h,\lambda)}p(2,1,v(2)) + \sum_{j=2}^{n-1} e^{h(j)}p(2,j,v(2))\Big) - \lambda \\ &\leq \bar{c} + |\lambda| + \log\Big(p(2,n,v(2)) + e^{G_1(h,\lambda)}p(2,1,v(2)) + \sum_{j=2}^{n-1} e^{h(j)}p(2,j,v(2))\Big) \\ &\leq \bar{c} + |\lambda| + \log\Big(p(2,n,v(2)) + e^{G_1(h,\lambda)\vee\|h\|_\infty}\sum_{j=1}^{n-1} p(2,j,v(2))\Big) \\ &\leq \bar{c} + |\lambda| + \log\Big(p(2,n,v(2)) + e^{G_1(h,\lambda)\vee\|h\|_\infty}\big(1 - p(2,n,v(2))\big)\Big)\,. \end{aligned}$$

In the above, the first inequality is obtained from the definition of $\bar{c}$; the third inequality is obtained from the fact that $\sum_{j=1}^{n} p(2,j,v(2)) = 1$. Now for $\rho > 0$, consider the following function $f(x) \doteq \log\big(\rho + (1-\rho)e^x\big)$. It is clear that for $x \in \mathbb{R}$,

$$f'(x) = \frac{(1-\rho)e^x}{\rho + (1-\rho)e^x} \leq 1\,.$$

From here, using the mean value theorem and the fact that $f(0) = 0$, we have $f(x) \leq f'(x)x$, for $x \geq 0$. Using this, (4.11) and the definition of $\chi(\cdot)$ from the hypothesis of the lemma, we have

$$\begin{aligned} G_2(h,\lambda) &\leq \bar{c} + |\lambda| + \chi(m + |\lambda|)\big(\bar{c} + \|h\|_\infty + |\lambda|\big) \vee \|h\|_\infty \\ &\leq \bar{c} + \chi(m + |\lambda|)\|h\|_\infty + |\lambda| + \chi(m + |\lambda|)\big(\bar{c} + \|h\|_\infty + |\lambda|\big) \\ &\leq \bar{c} + \|h\|_\infty + |\lambda| + \chi(m + |\lambda|)\big(\bar{c} + \|h\|_\infty + |\lambda|\big)\,. \end{aligned}$$

In the above, to get the first inequality, we use the fact that $f(x) \leq f'(x)x$, for the function f defined above, the bound on G_1 from (4.10) and the definition of $\chi(\cdot)$; to get the second inequality, we use the fact that $a \vee b \leq a + b$, for non-negative real numbers a and b; to get the final inequality, we use the fact that $\chi(m + |\lambda|) \leq 1$.

Following this argument similarly for $2 < i \leq n$, we obtain the desired upper bound in (4.6). This completes the proof of the lemma. □

Just like in the case of Algorithm 1, local Lipschitz continuity of $G(\cdot,\cdot)$ is crucial for the proof of (2.22). However, due to the iterative structure of the definition of $G(h,\lambda)$, the proof is more involved than in the case of $F(h,\lambda)$. For example, local Lipschitz continuity of $G(h,\cdot)$ is also not at all immediate, unlike the local Lipschitz continuity of $F(h,\cdot)$, which directly follows from its definition.

Proposition 4.3. *Fix $\lambda_1,\lambda_2\in\Lambda$ and $l>0$. Let $h_1,h_2\in\mathbb{R}^n$ be such that $\|h_1\|_\infty,\|h_2\|_\infty\le l$ and*

$$m \doteq \big(\bar{c}+l+|\lambda_1|\vee|\lambda_2|\big)\max\Big\{\frac{1}{1-\bar{\eta}},\frac{1-\chi(l+|\lambda_1|\vee|\lambda_2|)^{n-1}}{1-\chi(l+|\lambda_1|\vee|\lambda_2|)}\Big\},$$

with $\chi(\cdot)$ as defined in Lemma 4.1. Then, we have

$$\|G(h_1,\lambda_1)-G(h_2,\lambda_2)\|_m \le \beta_m\|h_1-h_2\|_m+\Delta^m|\lambda_1-\lambda_2|, \tag{4.11}$$

where $\Delta^m\doteq\max_{1\le i\le n}\Delta_i^m$, and $\{\Delta_i^m\}_{1\le i\le n}$ is defined recursively as follows:

$$\Delta_1^m=\frac{1}{w_1^m},\quad \Delta_i^m=\frac{1+\max_{1\le j\le i-1}\Delta_j^m}{w_i^m},\quad 2\le i\le n\,. \tag{4.12}$$

The proof of this result is given in Appendix B.

4.2.1. *Completing the proof of (2.22) in the case of Algorithm 2.* As mentioned earlier, the proof of Theorem 2.2 for Algorithm 2 follows from the same arguments as those used in the proof of Theorem 2.2 for Algorithm 1. Hence, we only discuss the differences and omit the proof, with the main difference being the explicit value of the constant $\widetilde{c}_h(\gamma,m)$ (appearing in Lemma A.5). Therefore, to avoid repetition of the arguments for Algorithm 1, we simply provide the explicit value of the analogous constant in this case. Also, to avoid introducing new notation we simply use the existing notation. Set

$$m=\frac{1}{w_0^1}\Big\{\big(\|h^0-h^*\|_\infty+\|h^*\|_\infty+\widetilde{\Delta}|\lambda^*-\lambda^0|\big)\vee\max_{1\le i\le n}h_{\lambda^0}(i)\Big\}, \tag{4.13}$$

$$\widetilde{c}_h(\gamma,m)=\beta_m+\gamma\Big(\Delta^m+\frac{\beta_m N^*(m)}{w_1^0}\Big)\Big(\beta_m+\frac{N^*(m)}{N_*(m)}\Big)\,.$$

Here, $\widetilde{\Delta}\doteq\sup_{m>0}\Delta^m$ which can be shown to be finite, from the definition of $\{w_i^m\}_{1\le i\le n}$ in (2.17).

4.3. **Completing the proof of Theorem 2.2.** From Sections 4.1 and 4.2, the first part of Theorem 2.2 follows. To see how the second part of the theorem follows from the first part, we observe that the first part implies the following:

$$\|h^k-h^*\|_m\le C_0e^{-C_1k},$$

for some constants $C_0=C_0(m),C_1=C_1(m)>0$ with m to be chosen accordingly *i.e.*, m is given by (4.1) for Algorithm 1 and by (4.13) for Algorithm 2. Therefore, in terms of V^k and V^*, this gives us

$$\|\mathcal{V}^k\|_m\le C_0e^{-C_1k}\text{ with }\mathcal{V}^k(i)\doteq\log\frac{V^k(i)}{V^*(i)},\text{ for }1\le i\le n\,.$$

From the definition of $\|\cdot\|_m$ in (2.19), supposing that for a large k, $V^k(i)\ge V^*(i)$, then $\frac{V^k(i)}{V^*(i)}\le e^{w_i^mC_0e^{-C_1k}}$ which in turn means that $V^k(i)-V^*(i)\le V^*(i)(e^{w_i^mC_0e^{-C_1k}}-1)\le C_2w_i^mC_0V^*(i)e^{-C_1k}$, for some $C_2=C_2(m)>0$. Similarly, supposing that $V^k(i)\le V^*(i)$ gives $V^*(i)-V^k(i)\le C_3w_i^mC_0V^*(i)e^{-C_1k}$, for some $C_3=C_3(m)>0$. To summarize, we have argued that $\|V^k-V^*\|_m\le\max\{C_2,C_3\}C_0e^{-C_1k}$. From here, the second part of the theorem follows.

5. RVI algorithms in the case of countable state space and their convergence

In this section, we consider the ERSC problem of a DTMC taking values in a countable state space. We state the important existing results pertaining to the characterizations of the optimal cost and optimal stationary Markov controls. Since the state space is now infinite, for numerical implementation, it is desirable to consider a truncated version (which includes an appropriate transition probability, an appropriate running cost and thereby an associated ERSC problem) of our original problem; see [28, 30, 32] for extensive works in this direction. The truncated version is an approximation of our original problem in the sense that the ERSC optimal cost associated with the truncated version is close to the optimal ERSC cost of our original problem, whenever, the truncation size is large. Using their associated stationary optimal controls, one can then construct nearly optimal controls of our original ERSC problem. This is achieved by using Algorithms 0, 1 and 2 for the truncated ERSC problem. The truncated version we discuss is motivated from the associated family of Dirichlet problems. We end this section with a discussion on an analogous truncation procedure in the risk-neutral case.

5.1. ERSC problem in the countable state space: assumptions and existing results.

To avoid introducing more notation, we use the same notation as in the previous sections. As earlier, $X = \{X_t\}_{t=0}^{\infty}$ is a controlled DTMC which now takes values in $\mathbb{N}$. At each time, if the current state is i, a control u is chosen from the compact set $\mathbb{U}(i)$, and the next state of the chain is j with probability $p(i,j,u)$. Let $\mathbb{N}_{\mathbb{U}} \doteq \{(i,u) : i \in \mathbb{N} \text{ and } u \in \mathbb{U}(i)\}$ represent the set of all allowed state-action pairs. We also let $U = \{U_0, U_1, \dots\}$ denote the control policy, with $(X_t, U_t) \in \mathbb{N}_{\mathbb{U}}$ for all $t \in \mathbb{Z}_+$. Following similar definitions as in Section 2, $\mathfrak{U}^{\infty}$ and $\mathfrak{U}_{\mathrm{SM}}^{\infty}$ denote the set of admissible controls and the set of stationary Markov controls, respectively.

For a running cost $c : \mathbb{N}_{\mathbb{U}} \to \mathbb{R}_+$, the corresponding cost functional for the ERSC problem is

$$\mathcal{E}_i^{\delta}(U) \doteq \limsup_{T\to\infty} \frac{1}{\delta T} \log \mathbb{E}_i^U \left[e^{\delta \sum_{t=0}^{T-1} c(X_t, U_t)} \right], \tag{5.1}$$

for an admissible control policy U and an initial state i, with $\delta > 0$ being the risk-sensitivity parameter. The optimal ERSC cost is given by $\lambda^{*,\delta} \doteq \inf_{i\in\mathbb{N}} \inf_{U\in\mathfrak{U}^{\infty}} \mathcal{E}_i^{\delta}(U)$. We also consider the optimal ERSC cost over stationary Markov controls given by $\lambda_m^{*,\delta} \doteq \inf_{i\in\mathbb{N}} \inf_{v\in\mathfrak{U}_{\mathrm{SM}}^{\infty}} \mathcal{E}_i^{\delta}(v)$. As earlier, we are interested in finding an optimal stationary Markov control policy v, *i.e.*, $v \in \mathfrak{U}_{\mathrm{SM}}^{\infty}$ satisfying $\lambda^{*,\delta} = \mathcal{E}_i^{\delta}(v),\ \forall i \in \mathbb{N}$.

In the literature, just as in the risk-neutral cost control problem (see [2] for the relevant literature), the ERSC problem is well-explored (see [12, 18] for surveys in this context) under two different conditions: uniform stability or near-monotonicity, which we state below.

Assumption 5.1 (Uniform stability)**.** For any $i \in \mathbb{N}$ and $f : \mathbb{N} \to \mathbb{R}$, the functions $u \mapsto c(i,u)$ and $u \mapsto \mathbb{E}_i^u[f]$ are continuous on $\mathbb{U}(i)$. Furthermore, one of the following conditions hold.

(i) There exist an inf-compact function $\mathcal{V} : \mathbb{N} \to [1,\infty)$, constants $C, \vartheta > 0$, and a finite set $\mathcal{K} \subset \mathbb{N}$ such that $\delta \sup_{(i,u)\in\mathbb{N}_{\mathbb{U}}} c(i,u) < \vartheta$ and

$$\mathbb{E}_i^u\big[\mathcal{V}(X_1)\big] \le C\mathbb{1}_{\mathcal{K}}(i) + e^{-\vartheta}\mathcal{V}(i), \quad \text{for } (i,u) \in \mathbb{N}_{\mathbb{U}}.$$

(ii) There exist inf-compact functions $\mathcal{V} : \mathbb{N} \to [1,\infty)$, $l : \mathbb{N} \to \mathbb{R}_+$, a constant $C > 0$ and a finite set $\mathcal{K} \subset \mathbb{N}$ such that $l(\cdot) - \delta \max_{(\cdot,u)\in\mathbb{N}_{\mathbb{U}}} c(\cdot,u)$ is inf-compact and

$$\mathbb{E}_i^u\big[\mathcal{V}(X_1)\big] \le C\mathbb{1}_{\mathcal{K}}(i) + e^{-l(i)}\mathcal{V}(i), \quad \text{for } (i,u) \in \mathbb{N}_{\mathbb{U}}.$$

Remark 5.1. Unlike in the finite case (see Assumption 2.1), recurrence and irreducibility of the controlled Markov chain is no longer sufficient for our analysis and hence, we consider a more quantitative assumption above. To be more elaborate, the above assumption implies that the

DTMC X is uniformly exponentially ergodic and the following holds: for every finite set $B \subset \mathbb{N}$,

$$\mathbb{E}_i^v[e^{\sum_{t=0}^{\tau_B-1}\left(l(X_t)-C\right)}\mathcal{V}(X_{\tau_B})] \leq \mathcal{V}(i), \quad \text{for } i \notin B\,.$$

Here, τ_B is the first hitting time of set B by the DTMC X. This follows from arguments similar to those in the proof in [13, Lemma 2.3(ii)].

Assumption 5.2 (Near-monotonicity)**.** For any $i \in \mathbb{N}$ and $f : \mathbb{N} \to \mathbb{R}$, the functions $u \mapsto c(i,u)$ and $u \mapsto \mathbb{E}_i^u[f]$ are continuous on $\mathbb{U}(i)$. Furthermore, the following conditions hold:

(i) There exists $\bar{v} \in \mathfrak{U}_{\mathrm{SM}}^{\infty}$ such that $\mathcal{E}_i^{\delta}(\bar{v}) < \infty$, which means that $\lambda_m^{*,\delta} < \infty$.

(ii) The function $\inf_{(\cdot,u)\in\mathbb{N}_{\mathbb{U}}} c(\cdot,u)$ is near-monotone relative to $\lambda_m^{*,\delta}$, *i.e.*,

$$\liminf_{n\to\infty} \inf_{l\geq n} \inf_{u\in\mathbb{U}(l)} c(l,u) > \lambda^{*,\delta}\,.$$

Remark 5.2. The above assumption ensures that we penalize the unstable behavior of X. This assumption is similar to its counterpart in the risk-neutral setting. However, it is not clear if this assumption is sufficient to completely characterize the optimal stationary Markov controls in the ERSC setting; see Theorem 5.1(ii). This contrasts with the analogous risk-neutral case where a complete characterization of optimal stationary Markov controls is guaranteed.

Under the setup above, we have the following existing result (see [13, Theorems 2.1 and 2.3]) that characterizes the optimal ERSC cost and optimal stationary Markov controls.

Theorem 5.1. *The following hold:*

(i) Suppose Assumption 5.1 holds. Then, there exists a function $V^{*,\delta} : \mathbb{N} \to \mathbb{R}_+$ that is unique up to a multiplicative constant and satisfies

$$e^{\delta\lambda_m^{*,\delta}}V^{*,\delta}(i) = \min_{(i,u)\in\mathbb{N}_{\mathbb{U}}}\left[e^{\delta c(i,u)}\mathbb{E}_i^u\left[V^{*,\delta}(X_1)\right]\right], \quad \text{for } i \in \mathbb{N}\,. \tag{5.2}$$

Moreover, $\lambda^{*,\delta} = \lambda_m^{*,\delta}$ and $v^* \in \mathfrak{U}_{\mathrm{SM}}^{\infty}$ is an optimal stationary Markov control if and only if

$$e^{\delta c(i,v^*(i))}\mathbb{E}_i^{v^*(i)}\left[V^{*,\delta}(X_1)\right] = \min_{(i,u)\in\mathbb{N}_{\mathbb{U}}}\left[e^{\delta c(i,u)}\mathbb{E}_i^u\left[V^{*,\delta}(X_1)\right]\right], \quad \text{for } i \in \mathbb{N}\,. \tag{5.3}$$

(ii) Suppose Assumption 5.2 holds and for every $v \in \mathfrak{U}_{\mathrm{SM}}^{\infty}$, DTMC X is positive recurrent. Then, there exists a function $V^{*,\delta} : \mathbb{N} \to \mathbb{R}_+$ that satisfies (5.2) with '$\geq$' instead of '$=$'. Moreover, $\lambda^{*,\delta} = \lambda_m^{*,\delta}$ and $v^* \in \mathfrak{U}_{\mathrm{SM}}^{\infty}$ is an optimal stationary Markov control if it satisfies (5.3).

Remark 5.3. In comparison with Theorem 5.1(i), Theorem 5.1(ii) is weaker in three ways: under the hypothesis of Theorem 5.1(ii), (a) (5.2) just holds with '$\geq$' instead of '$=$', (b) the uniqueness of the function $V^{*,\delta}$ is unclear and (c) the optimal stationary Markov controls are given as the minimizers of (5.3) while the converse is not shown to hold.

Remark 5.4. In Theorem 5.1(ii), the assumption of positive recurrence for every $v \in \mathfrak{U}_{\mathrm{SM}}^{\infty}$ (albeit possibly non-uniform in v; for example, the Lyapunov function may depend on v) is stronger than what is needed for the statement of Theorem 5.1(ii) to hold. In fact, the authors in [13] prove that the statement of Theorem 5.1(ii) holds under a much weaker assumption, which allows for the DTMC to be even transient for some stationary Markov control. However, the need for positive recurrence in the hypothesis of Theorem 5.1(ii) stems from the following: we are interested in designing RVI algorithms for a DTMC that is constructed on a truncated set, and analyzing their behavior as the number of iterations and/or the size of the truncated set grows indefinitely. This forces us to restrict ourselves to the case where the DTMC X is positive recurrent for every $v \in \mathfrak{U}_{\mathrm{SM}}$. An analogous assumption was also considered in [31, Proposition 4.1] in the context of the risk-neutral problem.

Just as in Section 2, we are again interested in finding optimal stationary Markov controls which satisfy (5.3) when either Assumption 5.1 or 5.2 holds and the DTMC X is positive recurrent for every $v \in \mathfrak{U}^{\infty}_{\mathrm{SM}}$, from Theorem 5.1. Since (5.3) requires the knowledge of $V^{*,\delta}$, we are consequently, interested in computing $V^{*,\delta}$. We take the approach using truncated ERSC problems with finite state space and applying Algorithms 0–2 to them. We remark that the reader may notice that one can naively extend Algorithm 1 where S is replaced by $\mathbb{N}$ and $n \in S$ is replaced by some $\bar{n} \in \mathbb{N}$. But this algorithm can neither be implemented in practice (due to the infiniteness of the state space) nor is it clear if the techniques from Sections 3 and 4 can be adapted in proving the convergence of Algorithm 1 (because the weighted norm $\|\cdot\|_m$ relies on the finiteness of the state space). Although the convergence techniques used in [16] can be applied to a naive extension of Algorithm 0, the problem of implementation still persists.

5.2. **A family of truncated ERSC problems.** We define the truncated ERSC problem on a finite state space as follows. Fix $K \in \mathbb{N}$ and set $S_K \doteq \{1, 2, \ldots, K\}$. Let $X^K = \{X^K_t\}_{t=0}^{\infty}$ be the S_K–valued controlled DTMC associated with transition probability $p_K(\cdot,\cdot,\cdot)$ given by

$$p_K(i,j,u) \doteq \frac{p(i,j,u)}{\sum_{k=1}^{K} p(i,k,u)}, \qquad \text{for } i,j \in S_K \text{ such that } (i,u) \in \mathbb{N}_{\mathbb{U}}\,.$$

Let c_K be the running cost given by

$$c_K(i,u) \doteq c(i,u) + R_K(i,u), \tag{5.4}$$

where $R_K(i,u) \doteq \delta^{-1}\log\big(\sum_{j=1}^{K} p(i,j,u)\big)$, $i \in S_K$ and $u \in \mathbb{U}(i)$. Here, we suppose that the denominator $\sum_{k=1}^{K} p(i,k,u)$ is non-zero for every $i \in S_K$ and $u \in \mathbb{U}(i)$. Let $\mathfrak{U}^K$ and $\mathfrak{U}^K_{\mathrm{SM}}$ denote the sets of admissible controls and stationary Markov control policies in this case.

Remark 5.5. Before we proceed further, it is important to understand a crucial aspect of the above construction in regards to $p_K(i,\cdot,u)$. For any given K, the function $p_K(\cdot,\cdot,\cdot)$ is a transition probability on S_K as long as $\sum_{k=1}^{K} p(i,k,u)$ is non-zero for every $i \in S_K$ and $u \in \mathbb{U}(i)$. However, for a general transition probability $p(\cdot,\cdot,\cdot)$ and a given K, this is not necessarily the case.

In the following, we prove the existence of a sequence of truncation parameters $\{K_l\}_{l\in\mathbb{N}}$ along which the failure discussed in Remark 5.5 does not occur.

Lemma 5.1. *Suppose the DTMC X is positive recurrent for every $v \in \mathfrak{U}^{\infty}_{\mathrm{SM}}$. Then there exists an increasing sequence $\{K_l\}_{l\in\mathbb{N}}$ (independent of $v \in \mathfrak{U}^{\infty}_{\mathrm{SM}}$) such that $\sum_{k=1}^{K_l} p(i,k,u) > 0$, for every $i \in S_{K_l}$ and $u \in \mathbb{U}(i)$.*

Proof. We prove this lemma by contradiction. Suppose for any sequence $\{K_l\}_{l\in\mathbb{N}}$, there exists $i_l \leq K_l$ that is the largest among $i \in S_{K_l}$ such that for some $v_l \in \mathfrak{U}_{\mathrm{SM}}$, $\sum_{k=1}^{K_l} p(i,k,v_l(i)) = 0$. Clearly, $\sum_{k=1}^{i_l} p(i_l,k,v_l(i_l)) = 0$ as $i_l \leq K_l$ and $p(i,j,v_l(i)) \geq 0$. Therefore, without loss in generality, we can take $i_l = K_l$.

Since X is positive recurrent for every $v \in \mathfrak{U}_{\mathrm{SM}}$ (and in particular, for $v_l \in \mathfrak{U}_{\mathrm{SM}}$), we know that there exists an inf-compact function $\mathcal{W} : \mathbb{N} \to \mathbb{R}_+$ and a positive constant η such that

$$\mathbb{E}^{v_l}_i\big[\mathcal{W}(X_1)\big] - \mathcal{W}(i) \leq -\eta, \tag{5.5}$$

whenever i is outside a large finite set. We now compute the left hand side of the above inequality for $i = i_l = K_l$ with large enough l, such that $\mathcal{W}(K_l) = \min_{i \in S^c_{K_l}} \mathcal{W}(i)$ and obtain

$$\begin{aligned}\sum_{k=1}^{\infty} p(K_l, k, v_l(i))\mathcal{W}(k) - \mathcal{W}(i) &= \sum_{k=K_l+1}^{\infty} p(K_l, k, v_l(i))\mathcal{W}(k) - \mathcal{W}(i) \\ &\geq \mathcal{W}(K_l)\Big(\sum_{k=K_l+1}^{\infty} p(K_l, k, v_l(K_l)) - 1\Big) \\ &= 0\,.\end{aligned}$$

In the above, to obtain the first line, we use the fact that $\sum_{k=1}^{K_l} p(K_l, k, v_l(i)) = 0$; the second line follows from the choice of l and the third line follows from the fact that $\sum_{k=K_l+1}^{\infty} p(K_l, k, v_l(i)) = 1$.

Therefore, combining (5.5) and (5.6) gives us the contradiction and proves the lemma. □

The above lemma gives the existence of the desired sequence $\{K_l\}_{l\in\mathbb{N}}$, but an explicit construction of such a sequence is often a straightforward exercise as shown in the following lemma.

Lemma 5.2. *Suppose the transition probability p is such that $p(i, j, u) > 0$, for some $j \leq i$ and every $u \in \mathbb{U}(i)$. Then, the sequence $\{K_l\}_{l\in\mathbb{N}}$ in Lemma 5.1 can be given by $K_l = l + 1$, $l \in \mathbb{N}$.*

Proof. The claim follows as $\sum_{j=1}^{i} p(i, j, u) > 0$, for every $i \in \mathbb{N}$ and $u \in \mathbb{U}(i)$. □

Remark 5.6. This result is applicable in many practical examples. In the service-effort control problem of single-server queues with or without abandonment, it is easy to check that the hypothesis of the lemma is satisfied, since with positive service rate, we have $p(i, i-1, u) > 0$, for any $u \in \mathbb{U}(i)$.

Let $\mathcal{K}$ denote a sequence of $\mathbb{N}$ obtained from Lemma 5.1. For $K \in \mathcal{K}$, consider the following ERSC problem (which is referred to as the truncated problem, from hereon) associated with X^K and the running cost c_K:

$$\lambda^{*,\delta}_K \doteq \min_{i\in S_K} \inf_{U^K \in \mathfrak{U}^K} \limsup_{T\to\infty} \frac{1}{\delta T} \log \mathbb{E}^U_i \Big[e^{\delta \sum_{t=0}^{T-1} c_K(X^K_t, U^K_t)} \Big]\,. \tag{5.6}$$

It is clear that from the above setup, we have the following result that is similar to Theorem 2.1.

Theorem 5.2. *Suppose either Assumption 5.1 holds or Assumption 5.2 holds and DTMC X is positive recurrent for every $v \in \mathfrak{U}^\infty_{\mathrm{SM}}$. Then, for every $K \in \mathcal{K}$, there exists a function $V^{*,\delta}_K : S_K \to \mathbb{R}_+$ that is unique up to a multiplicative constant and satisfies*

$$e^{\delta\lambda^{*,\delta}_K} V^{*,\delta}_K(i) = \min_{u\in\mathbb{U}(i)} \Big[e^{\delta c_K(i,u)} \sum_{j=1}^{K} V^{*,\delta}_K(j) p_K(i, j, u) \Big], \quad \text{for } 1 \leq i \leq K\,. \tag{5.7}$$

Moreover, $v \in \mathfrak{U}^K_{\mathrm{SM}}$ is optimal if and only if it satisfies

$$\min_{u\in\mathbb{U}(i)} \Big[e^{\delta c_K(i,u)} \sum_{j=1}^{K} V^{*,\delta}_K(j) p_K(i, j, u) \Big] = e^{\delta c_K(i,v(i))} \sum_{j=1}^{K} V^{*,\delta}_K(j) p_K(i, j, v(i)), \quad \text{for } 1 \leq i \leq K\,. \tag{5.8}$$

5.3. **RVI algorithms for the truncated ERSC problem.** We are now in a position to apply Algorithms 0, 1 and 2 to the ERSC problem defined in (5.6) (for the running cost c_K and the DTMC X^K on S_K) with one change, *viz.,* we choose state 1 as the reference state instead of K (recalling state n was chosen in (2.5) as the reference state in Section 2.2). In the rest of this section, Algorithms 0, 1 and 2, when applied to the ERSC problem defined in (5.6), are represented as Algorithms 0(K), 1(K) and 2(K), respectively. In terms of the truncation parameter K, Theorem 2.2 can be expressed as follows.

Theorem 5.3. *For every $K \in \mathcal{K}$, suppose that $\{(V_K^k, \lambda_K^k)\}_{k\in\mathbb{N}}$ is the sequence of iterate pairs of either Algorithm 1(K) or Algorithm 2(K). Suppose the hypothesis of Theorem 5.2 holds. Then, there exist constants $\underline{\gamma}_K < \overline{\gamma}_K$, depending on K and (V_K^0, λ_K^0), such that for $k \geq 0$, whenever $\gamma_{k,K} \in [\underline{\gamma}_K, \overline{\gamma}_K]$,*

$$V_K^k \to V_K^{*,\delta} \quad \text{and} \quad \lambda_K^k \to \lambda_K^{*,\delta} \tag{5.9}$$

geometrically in k with a possibly K–dependent rate.

Remark 5.7. If $\{(V_K^k, \lambda_K^k)\}_{k\in\mathbb{N}}$ is the sequence of iterates of Algorithm 0(K), from the work of [16], one can only infer the convergence in (5.9) but not the rate of convergence.

From here, all that remains to show is that $\lambda_K^{*,\delta} \to \lambda^{*,\delta}$ as $K \to \infty$. To prove this, we introduce the notion of a Dirichlet eigenpair and state an associated crucial existing result. Define $\mathfrak{B}_K = \{f : \mathbb{N} \to \mathbb{R}_+ : f(i) = 0 \text{ for } i \in S_K^c\}$. We say a pair $(V_{D,K}^\delta, \lambda_{D,K}^\delta) \in \mathfrak{B}_K \times \mathbb{R}_+$ is a Dirichlet eigenpair on S_K, if $V_{D,K}^\delta$ is not identically zero and the pair satisfies

$$V_{D,K}^\delta(i) = \min_{(i,u)\in\mathbb{N}_\mathbb{U}} \left(e^{\delta\left(c(i,u) - \lambda_{D,K}^\delta\right)} \mathbb{E}_i^u \left[V_{D,K}^\delta(X_1) \right] \right), \quad \text{for } 1 \leq i \leq K\,. \tag{5.10}$$

Lemma 5.3. *For every $K \in \mathbb{N}$, suppose $(V_{D,K}^\delta, \lambda_{D,K}^\delta)$ is a Dirichlet eigenpair on S_K. Then, as $K \to \infty$, we have the following:*

$$\lambda_{D,K}^\delta \to \lambda^{*,\delta} \quad \text{and} \quad V_{D,K}^\delta \to V^{*,\delta}, \quad \text{as } K \to \infty,$$

in the topology of pointwise convergence. Recall that $(V^{*,\delta}, \lambda^{*,\delta})$ is the pair obtained from Theorem 5.1.

Proof. The result follows directly from the combination of Lemmas 2.8, 2.10 and 2.15 of [13]. □

The following theorem proves that $(V_K^{*,\delta}, \lambda_K^{*,\delta})$ is indeed an approximation of the pair $(V^{*,\delta}, \lambda^{*,\delta})$ in Theorem 5.1.

Theorem 5.4. *For $K \in \mathcal{K}$, define the function $\widetilde{V}_K^{*,\delta} \in \mathfrak{B}_K$ as*

$$\widetilde{V}_K^{*,\delta}(i) \doteq \begin{cases} V_K^{*,\delta}(i), & \text{if } i \in S_K, \\ 0, & \text{otherwise}\,. \end{cases}$$

Suppose the hypothesis of Theorem 5.2 holds. Then,

$$\lambda_K^{*,\delta} \to \lambda^{*,\delta} \quad \text{and} \quad \widetilde{V}_K^{*,\delta} \to V^{*,\delta}, \quad \text{as } K \to \infty,$$

in the topology of pointwise convergence.

Proof. From Lemma 5.3, it suffices to show that for every $K \in \mathcal{K}$, with $\widetilde{V}_K^{*,\delta}$ as defined in the hypothesis of the theorem, $(\widetilde{V}_K^{*,\delta}, \lambda_K^{*,\delta})$ is a Dirichlet eigenpair. Since $\widetilde{V}_K^{*,\delta}$ is the value function associated with $\lambda_K^{*,\delta}$, we know that for $i \in S_K$, it satisfies

$$\begin{aligned} V_K^{*,\delta}(i) &= \min_{(i,u)\in\mathbb{N}_\mathbb{U}} \left(e^{\delta\left(c_K(i,u) - \lambda_K^{*,\delta}\right)} \mathbb{E}_i^u \left[V_K^{*,\delta}(X_1^K) \right] \right) \\ &= \min_{(i,u)\in\mathbb{N}_\mathbb{U}} \left(e^{\delta\left(c(i,u) - \lambda_K^{*,\delta}\right)} \left(\sum_{j=1}^K V_K^{*,\delta}(j) p(i,j,u) \right) \right). \end{aligned} \tag{5.11}$$

The second line above is obtained from the definition of $c_K(\cdot,\cdot)$ and $p_K(\cdot,\cdot,\cdot)$. Rewriting the second equation in (5.11) in terms of $\widetilde{V}_K^{*,\delta}$, we get

$$\widetilde{V}_K^{*,\delta}(i) = \min_{(i,u)\in\mathbb{N}_\mathbb{U}} \left(e^{\delta\left(c(i,u) - \lambda_K^{*,\delta}\right)} \mathbb{E}_i^u \left[\widetilde{V}_K^{*,\delta}(X_1) \right] \right).$$

Then, by the fact that $\widetilde{V}_K^{*,\delta} \in \mathfrak{B}_K$, we can conclude that $(\widetilde{V}_K^{*,\delta}, \lambda_K^{*,\delta})$ is a Dirichlet eigenpair. This completes the proof. □

Remark 5.8. Theorem 5.3 provides us with a geometric rate of convergence of (V_K^k, λ_K^k) in k for every $K \in \mathcal{K}$. However, in general, we cannot quantify the rate of convergence in K of $(\widetilde{V}_K^{*,\delta}, \lambda_K^{*,\delta})$ or equivalently of $(V_K^{*,\delta}, \lambda_K^{*,\delta})$ in Theorem 5.4. Consequently, this implies that the rate of convergence in

$$\lim_{\mathcal{K}\ni K\to\infty} \lim_{k\to\infty} |V_K^k(i) - V^{*,\delta}(i)| = 0 \quad \text{and} \quad \lim_{\mathcal{K}\ni K\to\infty} \lim_{k\to\infty} |\lambda_K^k - \lambda^{*,\delta}| = 0$$

is not quantifiable using the current techniques of this paper.

Remark 5.9 (An alternative to the running cost in the truncated ERSC problem). In the definition of the truncated ERSC problem, the running cost c_K, defined in (5.4), clearly converges to c, as $K \to \infty$. Therefore, it is natural to consider using c, instead of c_K in the definition of the truncated ERSC problem. However, the reason behind using c_K instead of c is more of a technical reason than a practical one. More precisely, for the algorithms designed with c, it is not clear if one can show a result that is analogous to Theorem 5.4. On the other hand, for numerical implementation, one may equivalently use the algorithms with c, instead of c_K. Our numerical experiments suggest that one may equivalently use the algorithms with c instead of c_K.

5.3.1. *Discussion on the truncation approach.* The truncation approaches in the context of Markov chains in countable state space have been thoroughly investigated for both uncontrolled (see [30, Ch. 7]) and controlled cases (see [28, 32] and the references therein for the risk-neutral control problem). Here, we only discuss the truncation approaches developed in the context of risk-sensitive control. In [34], a truncation scheme is designed to prove certain quantitative bounds (in terms of the truncation size) on the difference between the value functions for the truncated version and the original problem, in the context of finite horizon risk-sensitive control. The problem of infinite horizon discounted risk-sensitive control is investigated in [27], where the authors make use of the truncation approach as an intermediate step in proving the well-posedness of the Bellman equation and the optimality criterion. See also [26, 35] for the ERSC of continuous-time Markov chains with countable state space. In terms of relevance to the current paper, these works come the closest. Our truncation approach differs fundamentally from those in [26, 35] in terms of its origins. In our case, the construction of the truncated version begins with the well-known construction of the transition probability on the truncated space. The novelty of our construction lies in the definition of c_K in (5.4) and the ERSC problem in (5.6) which are all motivated from the family of Dirichlet problems associated with our DTMC X on $\mathbb{N}$ and the running cost $c(\cdot,\cdot)$. This has helped us to invoke a few existing results from [13] which can be concisely stated as Lemma 5.3.

5.3.2. *Analogous RVI algorithms in the risk-neutral case.* Observe that the construction of the truncated DTMC in Section 5.2 is independent of the type of control problems under study. This naturally raises the question of applicability of the our construction of the truncated DTMC to the limiting case of $\delta \downarrow 0$, *i.e.*, the risk-neutral ergodic control problem. The answer to this question is affirmative as will be argued below.

Adopting the existing notation, the risk-neutral ergodic control problem is to minimize

$$\mathcal{E}_i^0(U) \doteq \limsup_{T\to\infty} \frac{1}{T} \mathbb{E}_i^U \Big[\sum_{t=0}^{T-1} c(X_t, U_t) \Big]$$

over all $U \in \mathfrak{U}^\infty$. The optimal ergodic cost is given by $\lambda^{*,0} \doteq \inf_{i\in\mathbb{N}} \inf_{U\in\mathfrak{U}^\infty} \mathcal{E}_i^0(U)$. In this case, we can relax Assumptions 5.1 and 5.2 to their risk-neutral analogs (details are omitted). First, fix $K \in \mathcal{K}$. Then, the risk-neutral analog of Algorithm 0(K) is given in [6, Eqs. (9) and (10)], of Algorithm 1(K) is given in [6, Eqs. (7) and (8)] and of Algorithm 2(K) is given in [6, Eqs. (43) and

(44)]. We remind the reader that the algorithms given in [6, Eqs. (7) and (8)] and [6, Eqs. (43) and (44)] involve a 'step-size' parameter that can depend on k (the iteration number) and K (the truncation size) which we denote by $\gamma^0_{k,K}$. We then have the results below which can be considered as the risk-neutral analog of the combination of all the results from Sections 5.1 and 5.2.

(i) [14, Theorems 4.1 and 6.1]: These results give us the well-posedness of the Bellman equation, *i.e.,* existence of the value function $V^{*,0}$ that is unique up to an additive constant and the optimality criterion for stationary Markov controls, associated with the original risk-neutral cost control problem on $\mathbb{N}$.

(ii) [6, Propositions 1 and 2] and [29, Section 8.5.2]: Suppose for $K \in \mathcal{K}$, $\{(V_K^{k,0}, \lambda_K^{k,0})\}_{k\in\mathbb{N}}$ is the sequence of iterate pairs of the risk-neutral analogs of either Algorithm 0(K) or Algorithm 1(K) or Algorithm 2(K). Then, these results show the geometric convergence of $(V_K^{k,0}, \lambda_K^{k,0})$.

(iii) [31, Propositions 4.1 and 4.2]: These results help us prove that $V_K^0 \doteq \lim_{k\to\infty} V_K^{k,0}$ and $\lambda_K^0 \doteq \lim_{k\to\infty} \lambda_K^{k,0}$ converge to $V^{*,0}$ and $\lambda^{*,0}$, respectively, as $K \to \infty$.

6. Numerical Implementations

In this section, we numerically illustrate the performances of Algorithms 0–2 in two examples. From Theorem 2.2, we know that an approximate choice of γ_k, for every k (in particular, not arbitrary), is sufficient to ensure the convergence of Algorithms 1 and 2. Our numerical experiments suggests that a judicious choice of γ_k, for every k, is also necessary for these algorithms to perform well. For instance, if the γ_k's are too large, Theorem 2.2 does not guarantee the convergence of Algorithms 1 and 2. However, if the γ_k's are too small, then the iterates generated by the algorithm will update quite slowly, which may lead to a slower convergence. The contingency on the size of the step-size parameters is analogous to the average cost algorithms in [6, Pg. 746], and following the author's approach there, we make a choice to use a γ_k which starts at a value closer to 1 and slowly vanishes based on the path of the algorithm. The different forms of γ_k as well as the corresponding constants for each example are chosen by experimentation. This ensures that the step-sizes do not decrease too quickly, and only decrease after each algorithm has made enough progress in the sense that the signs of $h^k(n)$ oscillate about 0.

6.1. **Example: Maximizing the exit-rate from a finite domain.** Consider a controlled DTMC on a finite set that is irreducible for all stationary policies. Our goal is to find a control policy that maximizes the rate of exit of the DTMC from a fixed subset, if the DTMC starts inside that subset. For the sake of concreteness, we fix the finite set to be a weighted connected graph with n nodes (we denote the set of nodes by $S = \{1, 2, \ldots, n\}$) and the control set $\mathbb{U}$ to be a finite set. The controlled DTMC $X = \{X_t\}_{t=0}^\infty$ is now defined through the transition probability $\{p(i, j, u) : i, j \in S, u \in \mathbb{U}\}$ which is in turn assigned through the weight function $w : S \times S \times \mathbb{U} \to [0, \infty)$ (with $w(i, j, \cdot) = 0$ only if the nodes i and j are not connected) as follows:

$$p(i, j, u) = \frac{w(i, j, u)}{\sum_{k\in S} w(i, k, u)}\,. \tag{6.1}$$

The graph connectivity implies that the above $p(i, j, u)$ is a valid transition probability and that the DTMC X is irreducible under any stationary Markov control policy. Let us fix a connected sub-graph S_0 of S and for a control $U = \{U_t\}_{t=0}^\infty$, denote by $\hat{\tau}(U)$, the first exit time from the set S_0, *i.e.*, $\hat{\tau}(U) = \inf\{t \geq 0 : X_t \notin S_0\}$. As mentioned above, we are interested in the following problem:

$$\lambda^* = \max_{i\in S_0} \sup_{U\in\mathfrak{U}} \liminf_{T\to\infty} -\frac{1}{T} \log \mathbb{P}(\hat{\tau}(U) > T | X_0 = i)\,. \tag{6.2}$$

Here, $\mathfrak{U}$ is the class of admissible controls that are $\mathbb{U}$–valued. Note the minus sign in the above display - this is included because the exit-rate, by convention, is non-negative and the quantity $\frac{1}{T}\log\mathbb{P}(\hat{\tau}(U) > T) \leq 0$ as the exit time $\hat{\tau}(U)$ is almost surely finite.

We are now in a position to re-cast this problem as an ERSC problem involving the DTMC X and an appropriate running cost. To do this, we first define a controlled DTMC that is restricted to S_0. This can be done as follows: define $\overline{p}_0(i,u) \doteq \sum_{j\in S_0} p(i,j,u)$ and for $i,j \in S_0$, $q(i,j,u) \doteq \frac{p(i,j,u)}{\overline{p}_0(i,u)}$. The above expression is well-defined due to the connectedness of S_0, *i.e.*, $\overline{p}_0(i,\cdot) > 0$ for all $i \in S$. Hence, $q(i,j,u)$ is also a transition probability. Let the associated controlled DTMC be denoted by $Y = \{Y_t\}_{t=0}^{\infty}$. From the definition of $q(\cdot,\cdot,\cdot)$, it is clear that Y is restricted to S_0 and is irreducible. From here, for any $U \in \mathfrak{U}$, we can verify that

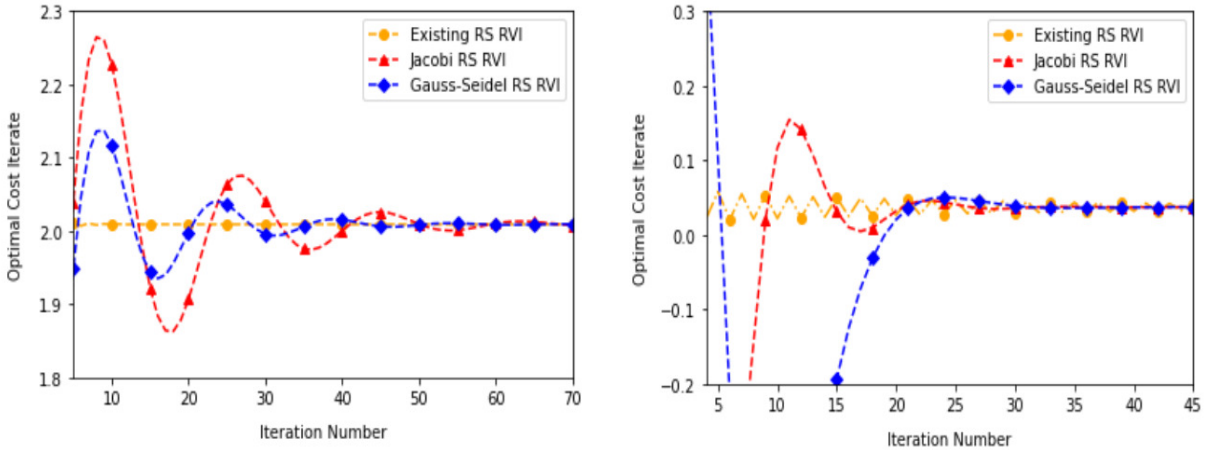
$$\mathbb{P}(\hat{\tau}(U) > T | X_0 = i) = \mathbb{E}_i^U\left[e^{\sum_{t=0}^{T-1}\log\overline{p}_0(Y_t,U_t)}\right].$$

This, in addition to interchanging the lim sup of a negative quantity with $-\liminf$, reduces (6.2) to

$$\lambda^* = -\min_{i\in S_0}\inf_{U\in\mathfrak{U}}\limsup_{T\to\infty}\frac{1}{T}\log\mathbb{E}_i^U\left[e^{\sum_{t=0}^{T-1}c(Y_t,U_t)}\right],$$

which resembles the ERSC problem defined in (2.2) with the running cost $c(i,u) = \log\overline{p}_0(i,u)$.

For the above problem to be completely defined, it is only remains to make a particular choice of the weight function. In our numerical implementation, we consider two particular choices of weight functions which correspond to the case of a complete graph and the case of a 'sparse' graph (that is neither complete nor a tree). In both the cases, we fix (i) the number of nodes $n = 20$ that are randomly chosen inside a disc of radius $R = 10$, (ii) the nodes are connected (depending on the case), and (iii) we set $\mathbb{U} = \{1, 1.3, 1.6, 1.9\}$ and $S_0 = \{1,\dots,m\}$ with $m = 5$. The weight function $w : S\times S\times\mathbb{U} \to [0,\infty)$ as $w(i,j,u) = \frac{d(i,j)}{u}$ if nodes i,j are connected and $w(i,j,u) = 0$ otherwise. Here, $d(i,j)$ is the Euclidean distance between nodes i and j. Regarding the control u as the speed with which the controller travels across the edge connecting i and j, the weight function $w(i,j,u)$ represents the time taken to travel across the edge connecting nodes i and j.



(a) **(b)**

Figure 1. Plot of $-\lambda^k$ *vs* the iteration number k for graphs with two different connectivities and parameters $n = 20$, $R = 10$, $\mathbb{U} = \{1, 1.3, 1.6, 1.9\}$, and $m = 5$. On the left, the graph is complete and the iterates of Algorithms 0, 1 and 2 fall within the tolerance of $\varepsilon = 1\times10^{-4}$, after 18, 104 and 88 iterations, respectively. On the right, the graph is neither complete nor a tree and the iterates of Algorithms 0, 1 and 2 fall within the tolerance of $\varepsilon = 1\times10^{-4}$, after 319, 66 and 68 iterations, respectively.

In the implementation of Algorithms 1 and 2, we set step-sizes $\gamma_k \propto \hat{k}^{-1}$, where $\hat{k}$ counts the number of iterations that the iterates $h^k(n)$ sequentially change signs and are larger than a fixed threshold of $\theta = 0.5$. Figure 1a corresponds to the case where the underlying graph is complete with $\gamma_k = 0.1\hat{k}^{-1}$ and $\gamma_k = 0.08\hat{k}^{-1}$ for Algorithms 1 and 2, respectively. Figure 1b corresponds to the case where the graph is neither complete nor a tree with $\gamma_k = 0.25\hat{k}^{-1}$ and $\gamma_k = 0.15\hat{k}^{-1}$ for Algorithms 1 and 2, respectively. We observe that for a connected graph, the less the number of edges (in the graph), the better the performance of our Jacobi-like and Gauss-Seidel-like algorithms, in comparison to the performance of the existing algorithm (Algorithm 0). This observation is further substantiated by another numerical implementation in the case where the graph is a tree (for which the DTMC is well-known to be periodic with period 2): Algorithm 0 fails to converge (even after a large number of iterations), while the Jacobi-like and Gauss-Seidel-like algorithms converge after a small number of iterations.

6.2. **Example: Service-effort control of a single-server queue.** We consider a service rate control problem for a discrete-time single-server queue. We let Q_t denote the number of customers in the system at time t in the case of (i) finite capacity and (ii) infinite capacity. Let $A = \{A_t\}_{t=0}^{\infty}$ be the arrival process, with A_t being the number of customers who enter the system at time t. We assume that A_t is a sequence of i.i.d. random variables taking values on a finite set - we make precise the choice of this finite set in both cases later. Every customer that enters the system will either wait in a queue or get served by a single server, if available, in the first-come first-served discipline. At time t, a customer in service is successfully served with a probability s, which will be appropriately chosen and hence, will be treated as the control. Suppose that s lies in a compact subset $\mathbb{U}$ of $(0,1)$ and denote by $\mathfrak{U}$ the set of admissible service-rate control policies that are $\mathbb{U}$–valued. The Bernoulli process associated with the customers serviced is denoted by $\{S_t\}_{t=0}^{\infty}$. The objective is to find a control policy which achieves the following:

$$\inf_i \inf_{\{S_t\}\in\mathfrak{U}} \limsup_{T\to\infty} \frac{1}{\delta T}\mathbb{E}\big[e^{\delta\sum_{t=0}^{T-1} c(Q_t,S_t)}|Q_0 = i\big],$$

where $c(i,s) \doteq r(i) + g(s)$ with $r(i)$ being the congestion cost when i customers are present in the system and $g(s)$ is the energy cost when service effort is s, and $\delta > 0$ is the sensitivity parameter. We assume that both r and g are increasing and g is convex.

6.2.1. *The case of finite capacity.* Suppose that the system has a maximum capacity of n, so that any new arrivals when there are n customers in the system are rejected. Suppose that A_t is a sequence of i.i.d. Bernoulli random variables taking values $\{0,1\}$, with $\mathbb{P}(A_t = 1) = \alpha \in (0,1)$. The dynamics of Q_t can be expressed as follows: $Q_{t+1} = \big[Q_t + A_t\mathbb{1}_{Q_t<n} - S_t\big]^+$. For a fixed $s \in \mathbb{U}$, Q_t is a controlled DTMC with the transition probability $p(i,j,s)$ (depending the choice of $s \in \mathbb{U}$) given by the following: for $1 < i < n$,

$$p(i,j,s) = \begin{cases} \alpha(1-s), & j = i+1, \\ (1-\alpha)(1-s) + \alpha\, s, & j = i, \\ (1-\alpha)s, & j = i-1, \end{cases}$$

$$p(0,1,s) = \alpha, \quad p(0,0,s) = 1-\alpha, \quad p(n,n-1,s) = s \quad \text{and} \quad p(n,n,s) = 1-s\,.$$

Figure 2 illustrates the convergence of the RVI algorithms. The first trend is how the Jacobi iterates seem to frequently oscillate in a damped manner. This trend also occurs for the Gauss-Seidel iterates, but within the first 30 iterations. We next perform a sensitivity analysis to understand the effect that the size of the state space n and the sensitivity parameter δ have on the convergence of the discussed RVI algorithms (again we use $\gamma_k = 0.95^{\hat{k}}$). The results are shown in Table 1. We can infer the following for this example: (i) among our three algorithms, the higher the risk sensitivity, the better the performance of the Gauss-Seidel iteration algorithm, and (ii) as one would expect,

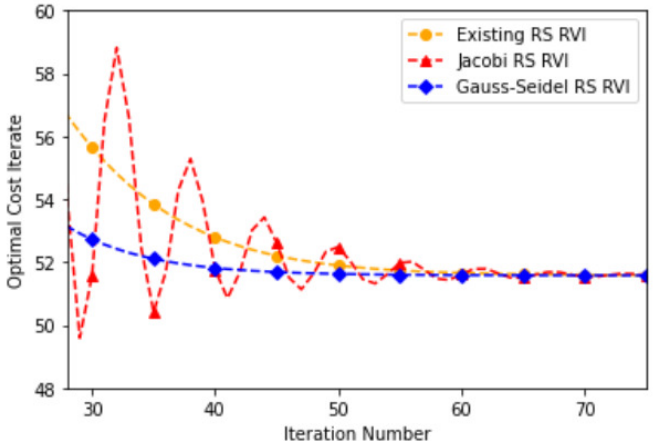


Figure 2. Plot of λ^k *vs* the iteration number k for the single server queue with a maximum capacity $n = 20$, $\alpha = 0.4$, $\mathbb{U} = \{0.1, 0.25, 0.4, 0.5, 0.75, 0.9\}$, $c(i,s) = 5(i-1)^+ + 0.25s^2$, $\gamma_k = 0.95^{\hat{k}}$ (with updating thresholds $\theta = 0.75, 0.85$ for Algorithms 1 and 2, respectively), and $\delta = 1 \times 10^{-2}$. Algorithms 0, 1, and 2 fall within the tolerance of $\varepsilon = 1 \times 10^{-4}$, after 120, 119, and 105 iterations, respectively.

n	δ	Existing ERSC RVI	**Jacobi** ERSC RVI	**Gauss-Seidel** ERSC RVI
20	5×10^{-2}	48	66	45
40	5×10^{-2}	68	86	62
60	5×10^{-2}	88	104	82
20	1×10^{-2}	120	119	105
40	1×10^{-2}	144	144	130
60	1×10^{-2}	166	164	150
20	1×10^{-3}	71	83	177
40	1×10^{-3}	147	147	405
60	1×10^{-3}	276	277	738

Table 1. Number of iterations required for the difference of successive cost iterates produced by each algorithm to be smaller than a tolerance of 10^{-4}.

the larger the state space, the slower the convergence of all three algorithms; this effect is more pronounced when the risk sensitivity is lower.

6.2.2. *The case of infinite capacity.* In this section, we consider the case without a capacity constraint and allow abandonment to ensure Assumption 5.1 is satisfied. In particular, we consider the following queueing dynamics: $Q_{t+1} = \big[\lfloor (1-\theta)Q_t \rfloor + A_t - S_t\big]^+$, where as before, $\{A_t\}_{t=0}^{\infty}$ denotes the arrival process and each A_t is i.i.d. and supported in a finite subset of $\mathbb{N}$, $\{S_t\}_{t=0}^{\infty}$ denotes the controllable $\{0,1\}$-valued service process of the single server and $\theta \in (0,1)$ indicates the fraction of customers that abandons in each period. Note that this is a simplified model with abandonment, and is similar to [13, Example 2.3]. The ERSC problem is formulated in the same way as above, with the congestion cost $r(i) = 3(i-1)^+$ and the energy cost $g(s) = 1.1s^2$.

In Figure 3, we implement Algorithms 0(K)–2(K) for a range of truncation sizes. For every truncation size in the range, we run Algorithms 0(K)–2(K) until the absolute value of the difference of the cost iterates and the Euclidean norm of the difference of the value function iterates are smaller than a tolerance of 1×10^{-6} (we refer to the final cost iterates as λ_K^*). We use step-sizes $\gamma_k^K = 0.9^{\hat{k}}$ and $\gamma_k^K = \left(\frac{K}{70}\right)^{\hat{k}}$ with updating thresholds $\theta = 0.5, 0.4$ for Algorithms 1(K) and 2(K), respectively.

As the truncation size K becomes larger, λ_K^* can be seen to converge and hence numerically verifying Theorem 5.4 for this specific problem.

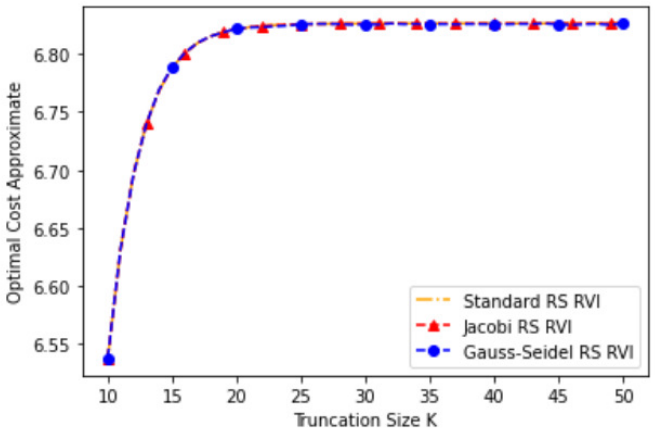


Figure 3. Plot of λ_K^* *v.s.* the truncation parameter K, where A_t is distributed on $\{0,1,2\}$ with $\mathbb{P}(A_1 = 0) = 0.2$, $\mathbb{P}(A_1 = 1) = \mathbb{P}(A_1 = 2) = 0.4$, $\theta = 0.001$, $s \in \mathbb{U} = \{0.1, 0.2, 0.3, 0.4\}$, $c(i,s) = 3(i-1)^+ + 1.1s^2$, and the sensitivity parameter $\delta = 1 \times 10^{-4}$.

Acknowledgement

This work is funded by the NSF Grant DMS 2216765.

References

[1] A. Arapostathis and V. S. Borkar. On the relative value iteration with a risk-sensitive criterion. *Banach Center Publications*, 122:9–24, 2020.

[2] A. Arapostathis, V. S. Borkar, E. Fernández-Gaucherand, M. K. Ghosh, and S. I. Marcus. Discrete-time controlled Markov processes with average cost criterion: A survey. *SIAM Journal on Control and Optimization*, 31(2):282–344, 1993.

[3] A. Arapostathis, V. S. Borkar, and K. S. Kumar. Relative value iteration for stochastic differential games. In V. Křivan and G. Zaccour, editors, *Advances in Dynamic Games: Theory, Applications, and Numerical Methods*, Annals of the International Society of Dynamic Games 13, pages 3–27. Springer International Publishing, 2013.

[4] C. Barz and K. Waldmann. Risk-sensitive capacity control in revenue management. *Mathematical Methods of Operations Research*, 65:565–579, 2007.

[5] A. Basu, T. Bhattacharyya, and V. S. Borkar. A learning algorithm for risk-sensitive cost. *Mathematics of Operations Research*, 33(4):880–898, 2008.

[6] D. P. Bertsekas. A new value iteration method for the average cost dynamic programming problem. *SIAM Journal on Control and Optimization*, 36(2):742–759, 1998.

[7] D. P. Bertsekas. *Dynamic programming and optimal control*. Athena Scientific, 2005.

[8] T. Bielecki, D. Hernandez-Hernandez, and S. Pliska. Value iteration for controlled Markov chains with risk sensitive cost criterion. *Proceedings of the 38th IEEE Conference on Decision and Control*, 1:126–130, 1999.

[9] T. Bielecki, D. Hernández-Hernández, and S. Pliska. Risk sensitive control of finite state Markov chains in discrete time, with applications to portfolio management. *Mathematical Methods of Operations Research*, 50:167–188, 1999.

[10] T. Bielecki and S. Pliska. Risk-sensitive dynamic asset management. *Applied Mathematics and Optimization*, 39:337–360, 1999.

[11] T. Bielecki, S. Pliska, and S.-J. Sheu. Risk sensitive portfolio management with Cox–Ingersoll–Ross interest rates: the HJB equation. *SIAM Journal on Control and Optimization*, 44(5):1811–1843, 2005.
[12] A. Biswas and V. S. Borkar. Ergodic risk-sensitive control—a survey. *Annual Reviews in Control*, 55:118–141, 2023.
[13] A. Biswas and S. Pradhan. Ergodic risk-sensitive control of Markov processes on countable state space revisited. *ESAIM: Control, Optimisation and Calculus of Variations*, 28, 2021.
[14] V. S. Borkar. Control of Markov chains with long-run average cost criterion: The dynamic programming equations. *SIAM Journal on Control and Optimization*, 27(3):642–657, 1989.
[15] V. S. Borkar. Q-learning for risk-sensitive control. *Mathematics of Operations Research*, 27(2):294–311, 2002.
[16] V. S. Borkar and S. P. Meyn. Risk-sensitive optimal control for Markov decision processes with monotone cost. *Mathematics of Operations Research*, 27(1):192–209, 2002.
[17] M. Bouakiz and M. J. Sobel. Inventory control with an exponential utility criterion. *Operations Research*, 40(3):603–608, 1992.
[18] N. Bäuerle and A. Jaśkiewicz. Markov decision processes with risk-sensitive criteria: An overview. *Mathematical Methods of Operations Research*, 99:141–178, 2024.
[19] R. Cavazos-Cadena and E. Fernández-Gaucherand. Risk-sensitive optimal control in communicating average Markov decision chains. In M. Dror, P. L'Ecuyer, and F. Szidarovszky, editors, *Modeling Uncertainty: An Examination of Stochastic Theory, Methods, and Applications*, International Series in Operations Research & Management Science 46, pages 515–553. Springer US, 2002.
[20] R. Cavazos-Cadena and R. Montes-de Oca. The value iteration algorithm in risk-sensitive average Markov decision chains with finite state space. *Mathematics of Operations Research*, 28(4):752–776, 2003.
[21] S. P. Coraluppi and S. I. Marcus. Risk-sensitive and minimax control of discrete-time, finite-state Markov decision processes. *Automatica*, 35(2):301–309, 1999.
[22] E. V. Denardo, H. Park, and U. G. Rothblum. Risk-sensitive and risk-neutral multiarmed bandits. *Mathematics of Operations Research*, 32(2):374–294, 2007.
[23] P. Dupuis and R. S. Ellis. *A weak convergence approach to the theory of large deviations*. John Wiley & Sons, 1997.
[24] Y. Feng and B. Xiao. A risk-sensitive model for managing perishable products. *Operations Research*, 56(5):1305–1311, 2008.
[25] W. H. Fleming and S. Sheu. Optimal long term growth rate of expected utility of wealth. *The Annals of Applied Probability*, 9(3):871–903, 1999.
[26] X. Guo and Y. Huang. Risk-sensitive average continuous-time Markov decision processes with unbounded transition and cost rates. *Journal of Applied Probability*, 58(2):523–550, 2021.
[27] X. Guo and Z.-W. Liao. Risk-sensitive discounted continuous-time Markov decision processes with unbounded rates. *SIAM Journal on Control and Optimization*, 57(6):3857–3883, 2019.
[28] O. O. Hernández-Lerma. *Adaptive Markov control processes*. Applied Mathematical Sciences. Springer New York, 1989.
[29] M. L. Puterman. *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons, 1994.
[30] E. Seneta. *Non-negative Matrices and Markov Chains*. Springer Series in Statistics. Springer New York, 1981.
[31] L. I. Sennott. The computation of average optimal policies in denumerable state Markov decision chains. *Advances in Applied Probability*, 29(1):114–137, 1997.
[32] L. I. Sennott. *Stochastic dynamic programming and the control of queueing systems*. John Wiley & Sons, 1999.

[33] P. Tseng. Solving H-horizon, stationary Markov decision problems in time proportional to $\log(H)$. *Operations Research Letters*, 9(5):287–297, 1990.

[34] Q. Wei. Continuous-time Markov decision processes with risk-sensitive finite-horizon cost criterion. *Mathematical Methods of Operations Research*, 84(3):461–487, 2016.

[35] Q. Wei and X. Chen. Risk-sensitive average continuous-time Markov decision processes with unbounded rates. *Optimization*, 68(4):773–800, 2019.

[36] D. White. Dynamic programming, Markov chains, and the method of successive approximations. *Journal of Mathematical Analysis and Applications*, 6(3):373–376, 1963.

[37] W. Whitt. Approximations of dynamic programs, I. *Mathematics of Operations Research*, 3(3):231–243, 1978.

Appendix A. Proof of Proposition 4.2

Recall that for $k \geq 1$, the iterates (h^k, λ^k) are given recursively by (2.14). The proof in this case is divided into several lemmas whose proofs hinge heavily on the use of Theorem 3.1 and Lemma 3.2. We begin by proving that $|\lambda^{k+1} - \lambda^*| \leq \widetilde{c}_\lambda N_*^{-1}(m)\alpha$. The proof of this estimate is split into two cases depending on whether $\lambda^k \leq \lambda^*$ or $\lambda^k > \lambda^*$. We only provide the proof of the estimate of $|\lambda^{k+1} - \lambda^*|$ in detail in the case where $\lambda^k \leq \lambda^*$. For the case where $\lambda^k > \lambda^*$, we highlight the key differences in the arguments involved at the end of the proof of Lemma A.4.

Case: $\lambda^k \leq \lambda^*$. The idea is to use Proposition 3.2 (i)-(ii) to choose intermediate λ values such that $|\lambda - \lambda^*| \propto \alpha$ which is similar to (4.4) and then, on a case-by-case basis, show that we can choose a step-size γ_k so that λ^{k+1} is close enough to the intermediate λ values. The approach is summarized as follows:

(i) Using the explicit bounds for the difference between any two scalars in Λ from Lemma 3.2 as well as the bijectivity of the function $\lambda \mapsto h_\lambda(n)$ from Proposition 3.2 (i)-(ii), we choose $\overline{\lambda}, \underline{\lambda} \in \Lambda_m$ so that $h_{\overline{\lambda}}(n) = \alpha\beta_m$, $h_{\underline{\lambda}}(n) = \alpha$. We then define $\widehat{\lambda}$ as the midpoint of $\overline{\lambda}$ and $\underline{\lambda}$.

(ii) We next analyze $|\lambda^* - \lambda^{k+1}|$ on a case-by-case basis depending on whether our newly defined $\widehat{\lambda}$ is larger or smaller than λ^k, where both cases are handled through comparisons with the intermediate values $\widehat{\lambda}, \underline{\lambda}$, and $\overline{\lambda}$. In each case, we show there exists a function $c(\gamma, m)$ that is strictly less than 1 whenever γ is less than a certain threshold value. For the step size γ_k less than this threshold, we then have $|\lambda^* - \lambda^{k+1}| \leq c(\gamma_k, m)N_*^{-1}(m)\alpha$.

We now give a lemma which provides the existence of the aforementioned $\overline{\lambda}$ and $\underline{\lambda}$.

Lemma A.1. *For any $\alpha > 0$, there exist $m > 0$ large enough and unique $\overline{\lambda}, \underline{\lambda} \in \Lambda_m$ such that $\lambda^* > \overline{\lambda} > \underline{\lambda}$ and*

$$h_{\overline{\lambda}}(n) = \alpha\beta_m \quad \text{and} \quad h_{\underline{\lambda}}(n) = \alpha\,.$$

This is straightforward consequence of Proposition 3.2(ii) and hence, we omit the proof. In what follows, we refer to $\overline{\lambda}$ and $\underline{\lambda}$ as those defined in the above lemma. Also, let $\widehat{\lambda} \doteq \frac{1}{2}(\overline{\lambda} + \underline{\lambda})$. We immediately have $\lambda^* > \overline{\lambda} > \widehat{\lambda} > \underline{\lambda}$, from Proposition 3.2(i). For a more quantitative comparison between the intermediate λ values, we provide the following lemma.

Lemma A.2. *The following inequalities hold among λ^*, $\overline{\lambda}$, $\underline{\lambda}$ and $\widehat{\lambda}$.*

$$\begin{aligned}
\frac{(1-\beta_m)\alpha}{N^*(m)} \leq \overline{\lambda} - \underline{\lambda} \leq \frac{(1-\beta_m)\alpha}{N_*(m)},\\
\frac{\alpha\beta_m}{N^*(m)} \leq \lambda^* - \overline{\lambda} \leq \frac{\alpha\beta_m}{N_*(m)},\\
\frac{\alpha}{N^*(m)} \leq \lambda^* - \underline{\lambda} \leq \frac{\alpha}{N_*(m)},\\
\frac{(1+\beta_m)\alpha}{2N^*(m)} \leq \lambda^* - \widehat{\lambda} \leq \frac{(1+\beta_m)\alpha}{2N_*(m)},\\
\frac{(1-\beta_m)\alpha}{2N^*(m)} \leq \overline{\lambda} - \widehat{\lambda} \leq \frac{(1-\beta_m)\alpha}{2N_*(m)}\,.
\end{aligned} \tag{A.1}$$

Proof. Using Lemma 3.2, we have

$$\begin{aligned}
&0 < N_*(m)(\overline{\lambda} - \underline{\lambda}) \leq h_{\underline{\lambda}}(n) - h_{\overline{\lambda}}(n) \leq N^*(m)(\overline{\lambda} - \underline{\lambda})\\
&\implies N_*(m)(\overline{\lambda} - \underline{\lambda}) \leq (1-\beta_m)\alpha \leq N^*(m)(\overline{\lambda} - \underline{\lambda})\,.
\end{aligned}$$

by our choice of $\underline{\lambda}$ and $\overline{\lambda}$. This gives us the first inequality in (A.1).

Similarly, using the fact that $h_{\lambda^*}(n) = 0$, we obtain the second and third inequalities in (A.1). Lastly, the fourth inequality in (A.1) is obtained from the definition of $\widehat{\lambda}$, and the fifth inequality is a result of subtraction of the second inequality in (A.1) from the fourth. □

Lemma A.3. *Suppose* (4.2) *holds for some* $\alpha > 0$, $\lambda^k \le \widehat{\lambda}$ *and* $\gamma_k \le \frac{1}{N^*(m)}$. *Then,*

$$|\lambda^{k+1} - \lambda^*| \le c_1(\gamma_k, m)\frac{\alpha}{N_*(m)},$$

where

$$c_1(\gamma, m) \doteq \max\left\{1 - \gamma\frac{(1-\beta_m)N_*(m)^2}{2N^*(m)},\ \beta_m\gamma N_*(m)\right\}.$$

Proof. Since the map $\lambda \mapsto h_\lambda(n)$ is monotonically decreasing, Lemma 3.2 and our choice of $\overline{\lambda}$ give us

$$h_{\lambda^k}(n) \ge h_{\widehat{\lambda}}(n) \ge h_{\overline{\lambda}}(n) + N_*(m)(\overline{\lambda} - \widehat{\lambda}) \ge \alpha\beta_m + \frac{\alpha(1-\beta_m)N_*(m)}{2N^*(m)}. \tag{A.2}$$

From the definition of h^{k+1}, the norm $\|\cdot\|_m$ defined in (2.19), Theorem 3.1 and the first inequality in (4.2), we obtain

$$|h^{k+1}(n) - h_{\lambda^k}(n)| \le \|F(h^k, \lambda^k) - F(h_{\lambda^k}, \lambda^k)\|_m \le \beta_m\|h^k - h_{\lambda^k}\|_m \le \alpha\beta_m. \tag{A.3}$$

Hence, $h^{k+1}(n) \ge h_{\lambda^k}(n) - \alpha\beta_m$. Combining this with (A.2) shows that $h^{k+1}(n) \ge \frac{\alpha(1-\beta_m)N_*(m)}{2N^*(m)}$, which enables us to bound $\lambda^* - \lambda^{k+1}$ from above as follows:

$$\begin{aligned}
\lambda^* - \lambda^{k+1} &= \lambda^* - \lambda^k - \gamma_k h^{k+1}(n) \\
&\le \lambda^* - \lambda^k - \gamma_k\frac{\alpha(1-\beta_m)N_*(m)}{2N^*(m)} \\
&\le \frac{\alpha}{N_*(m)} - \gamma_k\frac{\alpha(1-\beta_m)N_*(m)}{2N^*(m)} \\
&= \frac{\alpha}{N_*(m)}\left(1 - \gamma_k\frac{(1-\beta_m)N_*(m)^2}{2N^*(m)}\right).
\end{aligned} \tag{A.4}$$

Next, we turn to bound $\lambda^* - \lambda^{k+1}$ from below. To begin with, from (A.3), we obtain

$$h^{k+1}(n) \le h_{\lambda^k}(n) + \alpha\beta_m.$$

From the above display, Lemma 3.2 and the fact that $\lambda^k \le \lambda^*$, we immediately get

$$h^{k+1}(n) \le N^*(m)(\lambda^* - \lambda^k) + \alpha\beta_m. \tag{A.5}$$

Now observe that

$$\begin{aligned}
\lambda^* - \lambda^{k+1} &= \lambda^* - \lambda^k - \gamma_k h^{k+1}(n) \\
&\ge \lambda^* - \lambda^k - \gamma_k\left(N^*(m)(\lambda^* - \lambda^k) + \alpha\beta_m\right) \\
&= (1 - \gamma_k N^*(m))(\lambda^* - \lambda^k) - \alpha\beta_m\gamma_k \\
&\ge -\alpha\beta_m\gamma_k.
\end{aligned} \tag{A.6}$$

In the above, to get the first line we use (2.14); to get the second line, we use (A.5); to get the fourth line, we use the fact that $\gamma_k \le \frac{1}{N^*(m)}$. Therefore, the fourth lines in (A.4) and (A.6) imply that

$$|\lambda^* - \lambda^{k+1}| \le c_1(\gamma_k, m)\frac{\alpha}{N_*(m)}.$$

This completes the proof. □

Lemma A.4. *Suppose (4.2) holds for some $\alpha > 0$, $\widehat{\lambda} < \lambda^k \le \lambda^*$ and*

$$\gamma_k \le \min\Big\{\frac{1+\beta_m}{\beta_m N_*(m) + N^*(m)},\ \frac{1-\beta_m}{2\beta_m N^*(m)}\Big\}. \tag{A.7}$$

Then

$$|\lambda^{k+1} - \lambda^*| \le c_2(m)\frac{\alpha}{N_*(m)}, \quad \text{where} \quad c_2(m) \doteq \frac{(1+\beta_m)}{2}. \tag{A.8}$$

Proof. The proof of this lemma is divided into two cases depending on the sign of $h^{k+1}(n)$, *i.e.*, (a) $h^{k+1}(n) \ge 0$, or (b) $h^{k+1}(n) < 0$.

Case (a): From the definition of λ^{k+1}, we have $\lambda^{k+1} \ge \lambda^k$. Using the bounds on the difference between $\widehat{\lambda}$ and λ^* in the fourth inequality of (A.1) as well as the hypothesis that $\lambda^k > \widehat{\lambda}$, we get

$$\lambda^* \le \widehat{\lambda} + \frac{(1+\beta_m)\alpha}{2N_*(m)} \le \lambda^k + \frac{(1+\beta_m)\alpha}{2N_*(m)} \le \lambda^{k+1} + \frac{(1+\beta_m)\alpha}{2N_*(m)}. \tag{A.9}$$

We next bound $h^{k+1}(n)$ from above:

$$\begin{aligned} 0 \le h^{k+1}(n) &\le |h^{k+1}(n) - h_{\lambda^k}(n)| + |h_{\lambda^k}(n)| \\ &\le \|h^{k+1} - h_{\lambda^k}\|_m + N^*(m)|\lambda^k - \lambda^*| \\ &\le \|F(h^k, \lambda^k) - F(h_{\lambda^k}, \lambda^k)\|_m + N^*(m)|\lambda^k - \lambda^*| \\ &\le \beta_m \|h^k - h_{\lambda^k}\|_m + N^*(m)|\lambda^k - \lambda^*| \\ &\le \beta_m\alpha + \alpha\frac{N^*(m)}{N_*(m)}. \end{aligned} \tag{A.10}$$

This, in conjunction with $\lambda^k \le \lambda^*$, implies that

$$\lambda^{k+1} = \lambda^k + \gamma_k h^{k+1}(n) \le \lambda^* + \gamma_k\Big(\alpha\beta_m + \alpha\frac{N^*(m)}{N_*(m)}\Big). \tag{A.11}$$

Hence,

$$\lambda^{k+1} - \lambda^* \le \gamma_k\alpha\Big(\beta_m + \frac{N^*(m)}{N_*(m)}\Big) \le \frac{(1+\beta_m)\alpha}{N_*(m)}.$$

To get the second inequality above, we use the fact that γ_k satisfies (A.7). Additionally, using (A.9), we see that

$$\lambda^* - \lambda^{k+1} \le \frac{(1+\beta_m)\alpha}{2N_*(m)}.$$

As a result, we obtain (A.8). This completes the proof for Case (a).

Case (b): The assumption that $\|h^k - h_{\lambda^k}\|_m \le \alpha$ and the bound on $|h^{k+1}(n) - h_{\lambda^k}(n)|$ from (A.3) give us

$$h_{\lambda^k}(n) \le h^{k+1}(n) + \alpha\beta_m \le \alpha\beta_m, \tag{A.12}$$

where the last inequality holds because $h^{k+1}(n) < 0$ by assumption. However, recall that $\overline{\lambda}$ was chosen so that $h_{\overline{\lambda}}(n) = \alpha\beta_m$. Since $\lambda \mapsto h_\lambda(n)$ is monotonically decreasing, it follows that $\lambda^k \ge \overline{\lambda}$. Since $\lambda^k \le \lambda^*$ and $\lambda \mapsto h_\lambda(n)$ is monotonically decreasing, we get $0 = h_{\lambda^*}(n) \le h_{\lambda^k}(n)$. Hence, we get $0 \le h_{\lambda^k}(n) \le \alpha\beta_m$. Combining this with (A.12) and re-arranging the resulting inequality, we get

$$-\alpha\beta_m \le h^{k+1}(n) \le 0. \tag{A.13}$$

We use this to construct the desired upper bound for $\lambda^* - \lambda^{k+1}$. This is sufficient because $\lambda^{k+1} = \lambda^k + \gamma_k h^{k+1}(n) < \lambda^k$ as $h^{k+1}(n) < 0$, and $\lambda^k \le \lambda^*$ by assumption. From (A.13) and (A.7), it follows that

$$|\gamma_k h^{k+1}(n)| \le \frac{(1-\beta_m)\alpha}{2N^*(m)}.$$

Combining this with our bound on the difference of $\overline{\lambda}$ and $\widehat{\lambda}$ in the fifth inequality of (A.1), we have

$$|\gamma_k h^{k+1}(n)| \le \overline{\lambda} - \widehat{\lambda} \le \lambda^k - \widehat{\lambda},$$

where $\lambda^k \ge \overline{\lambda}$ was shown in the paragraph following (A.12). Following the definition of λ^{k+1}, this leads to $\lambda^{k+1} = \lambda^k + \gamma_k h^{k+1}(n) \ge \widehat{\lambda}$. As a result, we can apply the bound on $\lambda^* - \widehat{\lambda}$ from the fourth inequality in (A.1) to obtain

$$\lambda^* - \lambda^{k+1} \le \lambda^* - \widehat{\lambda} \le \frac{(1+\beta_m)\alpha}{2N^*(m)} \le c_2(m)\frac{\alpha}{N_*(m)}.$$

This completes the proof for Case (b) as well as the proof for the lemma. □

Case: $\lambda^k > \lambda^*$. The proof in this case involves similar arguments as those used in the case where $\lambda^k \le \lambda^*$. However, we only discuss one of the main differences. We restrict ourselves to using the same notation as in the previous case. In this case, we define our intermediate values $\overline{\lambda}$ and $\underline{\lambda}$ in a similar manner to Lemmas A.1 and A.2 but with the following modification: $\overline{\lambda}$ and $\underline{\lambda}$ are the unique real numbers such that $\underline{\lambda}$, $\overline{\lambda} \in \Lambda_m$ satisfying $h_{\underline{\lambda}}(n) = -\alpha\beta_m$ and $h_{\overline{\lambda}}(n) = -\alpha$. In this case, we again define $\widehat{\lambda} \doteq \frac{1}{2}(\overline{\lambda} + \underline{\lambda})$. From these definitions, it follows that $\lambda^* < \underline{\lambda} < \widehat{\lambda} < \overline{\lambda}$, and these values satisfy the same inequalities as those given in Lemma A.2 (but with $\lambda^* - \overline{\lambda}$, $\lambda^* - \underline{\lambda}$ and $\lambda^* - \widehat{\lambda}$ replaced by $\overline{\lambda} - \lambda^*$, $\underline{\lambda} - \lambda^*$ and $\widehat{\lambda} - \lambda^*$, respectively).

From here, the rest of the proof is split into two cases depending on whether $\lambda^k \in (\lambda^*, \widehat{\lambda})$ or not. In the case where $\lambda^k \notin (\lambda^*, \widehat{\lambda})$, we obtain a result analogous to Lemma A.3 and in the case where $\lambda^k \in (\lambda^*, \widehat{\lambda})$, we obtain a result analogous to Lemma A.4. The proofs of these results follow closely the main arguments of the proofs of Lemmas A.3 and A.4, with minor changes. Hence, we omit them.

From Lemmas A.3 and A.4, and their analogs for the case where $\lambda^k > \lambda^*$, whenever

$$\gamma_k \le \min\Big\{\frac{1}{N^*(m)},\ \frac{1+\beta_m}{\beta_m N_*(m) + N^*(m)},\ \frac{1-\beta_m}{2\beta_m N^*(m)}\Big\},$$

we have $|\lambda^{k+1} - \lambda^*| \le \widetilde{c}_\lambda(\gamma_k, m)\frac{\alpha}{N_*(m)}$, where $\widetilde{c}_\lambda(\gamma, m) \doteq \max\{c_1(\gamma, m), c_2(m)\}$. Moreover, if

$$\gamma \le \widehat{\gamma} \doteq \min\Big\{\frac{1}{\beta_m N_*(m)},\ \frac{2N^*(m)}{(1-\beta_m)N_*(m)^2},\ \Big(\frac{m-m^*}{N_*(m)}\Big)\Big(\frac{1}{\frac{\alpha_0}{N_*(m)} + \overline{c} - \underline{c} + \varkappa(m)}\Big)\Big\},$$

then $\widetilde{c}_\lambda(\gamma, m) < 1$ and from Proposition 4.1, we know that $\lambda^{k+1} \in \Lambda$ and in fact $\lambda^{k+1} \in \Lambda_m$. In particular, $h_{\lambda^{k+1}}$ exists. In the following lemma, we estimate $\|h^{k+1} - h_{\lambda^k}\|_m$ in terms of α.

Lemma A.5. *Suppose (4.2) holds for some $\alpha > 0$. Then,*

$$\|h^{k+1} - h_{\lambda^{k+1}}\|_m \le \widetilde{c}_h(\gamma_k, m)\alpha, \tag{A.14}$$

where

$$\widetilde{c}_h(\gamma, m) \doteq \beta_m + \gamma\big(\|e\|_m + \frac{\beta_m N^*(m)}{w_1^0}\big)\Big(\beta_m + \frac{N^*(m)}{N_*(m)}\Big).$$

Recall w_1^0 from (2.17), β_m from (2.18) and $N_*(m)$ is as defined in the second line of (4.1).

Proof. From (2.14), we have

$$\begin{aligned}\|h^{k+1} - h_{\lambda^{k+1}}\|_m &\le \|(\lambda^{k+1} - \lambda^k)e\|_m + \|F(h^k, \lambda^{k+1}) - F(h_{\lambda^{k+1}}, \lambda^{k+1})\|_m \\ &\le |\lambda^{k+1} - \lambda^k|\|e\|_m + \beta_m\|h_{\lambda^k} - h_{\lambda^{k+1}}\|_m + \beta_m\|h^k - h_{\lambda^k}\|_m.\end{aligned} \tag{A.15}$$

In the above, we apply Theorem 3.1 and the triangle inequality. We now estimate the first two terms. Consider $|\lambda^{k+1}-\lambda^k|$. Using the definition of λ^{k+1} from (2.14), we see that

$$\begin{aligned}|\lambda^{k+1}-\lambda^k| = \gamma_k|h^{k+1}(n)| &\le \gamma_k\Big(|h^{k+1}(n)-h_{\lambda^k}(n)|+|h_{\lambda^k}(n)|\Big)\\ &\le \gamma_k\Big(\|F(h^k,\lambda^k)-F(h_{\lambda^k},\lambda^k)\|_m + N^*(m)|\lambda^k-\lambda^*|\Big)\\ &\le \gamma_k\Big(\beta_m\|h^k-h_{\lambda^k}\|_m + N^*(m)|\lambda^k-\lambda^*|\Big).\end{aligned}$$

To arrive at the second line, we use the fact that $\|x_n\|\le\|x\|_m$, for $x\in\mathbb{R}^n$, Lemma 3.2 for $h_{\lambda^k}(n)$, and $h^*(n)$, and the definition of h^{k+1} and h_{λ^k}. To get the last line, we apply Theorem 3.1.

From (2.20), Lemma 3.2 and the fourth inequality in (A.10), we have

$$\|h_{\lambda^k}-h_{\lambda^{k+1}}\|_m \le \frac{N^*(m)}{w_1^0}|\lambda^{k+1}-\lambda^k| \le \frac{\gamma_k N^*(m)}{w_1^0}\Big(\beta_m\|h^k-h_{\lambda^k}\|_m + N^*(m)|\lambda^k-\lambda^*|\Big).$$

Combining the inequalities in (A.15) and (A.10) with the above inequality, we obtain

$$\begin{aligned}\|h^{k+1}-h_{\lambda^{k+1}}\|_m &\le \beta_m\|h^k-h_{\lambda^k}\|_m + \gamma_k\|e\|_m\Big(\beta_m\|h^k-h_{\lambda^k}\|_m + N^*(m)|\lambda^k-\lambda^*|\Big)\\ &\quad + \gamma_k\beta_m N^*(m)\Big(\beta_m\|h^k-h_{\lambda^k}\|_m + N^*(m)|\lambda^k-\lambda^*|\Big)\\ &\le \beta_m\alpha + \gamma_k\|e\|_m\Big(\beta_m\alpha + N^*(m)\frac{\alpha}{N_*(m)}\Big) + \frac{\gamma_k N^*(m)}{w_1^0}\Big(\beta_m\alpha + N^*(m)\frac{\alpha}{N_*(m)}\Big)\\ &= \widetilde{c}_h(\gamma_k,m)\alpha\,.\end{aligned}$$

In the second inequality, we use the first two inequalities in (4.2). This proves the first part of (A.14). This completes the proof of the lemma. □

From the above analysis, and the definitions of $\widetilde{c}_h$ and $\widetilde{c}_\lambda$, it is easy to see that whenever

$$\gamma_k \le \widetilde{\gamma} \doteq \min\left\{\frac{1}{\beta_m N_*(m)}, \frac{2N^*(m)}{(1-\beta_m)N_*(m)^2}, \frac{(1-\beta_m)N_*(m)}{\big(\|e\|_m+\beta_m N^*(m)\big)\big(\beta_m N_*(m)+N^*(m)\big)}, \widehat{\gamma}\right\},$$

we have $\widetilde{c}_h(\gamma_k,m)<1$ and $\widetilde{c}_\lambda(\gamma_k,m)<1$. This completes the proof of Proposition 4.2. □

Appendix B. Proof of Proposition 4.3

We proceed by using the contraction property of $F(\cdot,\lambda)$ (obtained in Theorem 3.1) and inductively showing that

$$\frac{|G_i(h_1,\lambda_1)-G_i(h_2,\lambda_2)|}{w_i^m} \le \beta_m\|h_1-h_2\|_m + \Delta_i^m|\lambda_1-\lambda_2| \tag{B.1}$$

for $1\le i\le n$. From here, we use the definition of $\|\cdot\|_m$ in (2.19) to obtain (4.11).

Initialization step: We prove (B.1) for $i=1$. Observe that $G_1(h_1,\lambda_1)=F_1(h_1,\lambda_1)$. Furthermore, the local contraction property of F (Theorem 3.1) implies that

$$\frac{|F_1(h_1,\lambda_1)-F_1(h_2,\lambda_1)|}{w_1^m} \le \beta_m\|h_1-h_2\|_m\,. \tag{B.2}$$

From the definition of $G_1(\cdot,\cdot)$, we have

$$G_1(h_1,\lambda_1)-G_1(h_2,\lambda_1) = F_1(h_1,\lambda_1)-F_1(h_2,\lambda_1),$$

and using (B.2), we get

$$\frac{|G_1(h_1,\lambda_1)-G_1(h_2,\lambda_1)|}{w_1^m} \le \beta_m \|h_1-h_2\|_m\,.$$

Using the definitions of $G_1(\cdot,\cdot)$ and Δ_1^m, and the above display, it is now clear that

$$\begin{aligned}\frac{G_1(h_1,\lambda_1)}{w_1^m} &\le \frac{G_1(h_2,\lambda_1)}{w_1^m} + \beta_m\|h_1-h_2\|_m \\ &\le \frac{G_1(h_2,\lambda_2)}{w_1^m} + \beta_m\|h_1-h_2\|_m + \Delta_1^m|\lambda_1-\lambda_2|\,.\end{aligned}$$

Interchanging the roles of (h_1,λ_1) and (h_2,λ_2), we obtain

$$\frac{G_1(h_2,\lambda_2)}{w_1^m} \le \frac{G_1(h_1,\lambda_1)}{w_1^m} + \beta_m\|h_1-h_2\|_m + \Delta_1^m|\lambda_1-\lambda_2|\,.$$

Combining the above two displays proves (B.1) for $i=1$.

Induction step: Suppose that (B.1) holds for $1\le i\le r-1$, where $r<n$. We show that it also holds for r. Let $v^*\in\mathfrak{U}_{\rm SM}$ be a minimizing control policy for the right hand side of $G_r(h_2,\lambda_1)$. Using an argument analogous to that in the proof of Theorem 3.1, it follows that

$$\begin{aligned}&\frac{G_r(h_1,\lambda_1)-G_r(h_2,\lambda_1)}{w_r^m} \\ &= \frac{1}{w_r^m}\min_{u\in\mathbb{U}(r)}\Big[c(r,u)+\log\Big(p(r,n,u)+\sum_{j=1}^{r-1}e^{G_j(h_1,\lambda_1)}p(r,j,u)+\sum_{j=r}^{n-1}e^{h_1(j)}p(r,j,u)\Big)\Big] \\ &\quad -\frac{1}{w_r^m}\min_{u\in\mathbb{U}(r)}\Big[c(r,u)+\log\Big(p(r,n,u)+\sum_{j=1}^{r-1}e^{G_j(h_2,\lambda_1)}p(r,j,u)+\sum_{j=r}^{n-1}e^{h_2(j)}p(r,j,u)\Big)\Big] \\ &\le \frac{1}{w_r^m}\log\Big(p(r,n,v^*(r))+\sum_{j=1}^{r-1}e^{G_j(h_1,\lambda_1)}p(r,j,v^*(r))+\sum_{j=r}^{n-1}e^{h_1(j)}p(r,j,v^*(r))\Big) \\ &\quad -\frac{1}{w_r^m}\log\Big(p(r,n,v^*(r))+\sum_{j=1}^{r-1}e^{G_j(h_2,\lambda_1)}p(r,j,v^*(r))+\sum_{j=r}^{n-1}e^{h_2(j)}p(r,j,v^*(r))\Big).\end{aligned}$$

From here, following similar calculations as those leading up to (3.11), we have

$$\begin{aligned}&\frac{G_r(h_1,\lambda_1)-G_r(h_2,\lambda_1)}{w_r^m} \\ &\le \frac{1}{w_r^m}\Big(\sum_{j=1}^{r-1}[G_j(h_1,\lambda_1)-G_j(h_2,\lambda_1)]q^*(r,j,v^*(r))+\sum_{j=r}^{n-1}[h_1(j)-h_2(j)]q^*(r,j,v^*(r))\Big),\end{aligned} \tag{B.3}$$

with

$$q^*(r,j,v^*(r)) = \frac{e^{\tilde h_1(j)}p(r,j,v^*(r))}{\sum_{k=1}^n e^{\tilde h_1(k)}p(r,k,v^*(r))}\,.$$

Here, $\tilde h_1:\mathbb{R}^n\to\mathbb{R}^n$, as follows

$$\tilde h_1(j) = \begin{cases} G_j(h_1,\,\lambda_1), & j\le r-1,\\ h_1(j), & r\le j\le n-1,\\ 0, & j=n\,.\end{cases}$$

Similarly, $\tilde h_2$ is defined. From the hypothesis of the theorem and Lemma 4.1, we have $\|\widetilde{h}_1\|_\infty\le m$. Therefore, $q^*(r,\cdot,v^*(r))\ge e^{-2m}p(r,\cdot,v^*(r))$.

Now following similar calculations as those leading up to (3.12), we get

$$\frac{G_r(h_1,\lambda_1)-G_r(h_2,\lambda_1)}{w_r^m} \le \beta_m \max_{1\le i\le n} \frac{|\tilde{h}_1(i)-\tilde{h}_2(i)|}{w_i^m} \le \beta_m \|h_1-h_2\|_m\,.$$

Interchanging the roles of h_1 and h_2, we get

$$\frac{G_r(h_2,\lambda_1)-G_r(h_1,\lambda_1)}{w_r^m} \le \beta_m \|h_1-h_2\|_m\,.$$

From the above two displays, we have

$$\frac{|G_r(h_1,\lambda_1)-G_r(h_2,\lambda_1)|}{w_r^m} \le \beta_m \|h_1-h_2\|_m\,.$$

Recall that the induction hypothesis implies

$$\frac{|G_i(h_2,\lambda_1)-G_i(h_2,\lambda_2)|}{w_i^m} \le \Delta_i^m|\lambda_1-\lambda_2|, \quad 1\le i\le r-1\,. \tag{B.4}$$

Using this, we next obtain a bound on $\frac{|G_r(h_2,\lambda_1)-G_r(h_2,\lambda_2)|}{w_r^m}$. To do this, we again follow similar arguments as those leading up to (B.3) and obtain

$$\begin{aligned}
\frac{G_r(h_2,\lambda_1)-G_r(h_2,\lambda_2)}{w_r^m} &\le \frac{1}{w_r^m}\Big(\sum_{j=1}^{r-1}[G_j(h_2,\lambda_1)-G_j(h_2,\lambda_2)]\widehat{q}(r,j,v^*(r))+(\lambda_2-\lambda_1)\Big)\\
&\le \frac{1}{w_r^m}\Big(\sum_{j=1}^{r-1}|G_j(h_2,\lambda_1)-G_j(h_2,\lambda_2)|\widehat{q}(r,j,v^*(r))+|\lambda_2-\lambda_1|\Big)\\
&\le \frac{1}{w_r^m}\Big(\sum_{j=1}^{r-1}\Delta_i^m|\lambda_1-\lambda_2|\widehat{q}(r,j,v^*(r))+|\lambda_2-\lambda_1|\Big)\\
&\le \Delta_r^m|\lambda_2-\lambda_1|\,.
\end{aligned}$$

In the above, $\widehat{q}(\cdot,\cdot,\cdot)$ is some transition probability (unlike earlier, the exact form of $\widehat{q}(\cdot,\cdot,\cdot)$ is irrelevant for our purpose). To get the third line, use (B.4) and to get the fourth line, we use the definition of Δ_r^m and the fact that $\sum_{j=1}^{r-1}\widehat{q}(r,j,v^*(r))\le 1$. Interchanging the roles of λ_1 and λ_2, we obtain

$$\frac{G_r(h_2,\lambda_1)-G_r(h_2,\lambda_2)}{w_r^m} \le \Delta_r^m|\lambda_2-\lambda_1|,$$

which gives us

$$\frac{|G_r(h_2,\lambda_1)-G_r(h_2,\lambda_2)|}{w_r^m} \le \Delta_r^m|\lambda_2-\lambda_1|\,.$$

From here, it follows that

$$\begin{aligned}
\frac{G_r(h_1,\lambda_1)}{w_r^m} &\le \frac{G_r(h_2,\lambda_1)}{w_r^m}+\beta_m\|h_1-h_2\|_m\\
&\le \frac{G_r(h_2,\lambda_2)}{w_r^m}+\frac{|G_r(h_2,\lambda_1)-G_r(h_2,\lambda_2)|}{w_r^m}\\
&\quad +\beta_m\|h_1-h_2\|_m\\
&\le \frac{G_r(h_2,\lambda_2)}{w_r^m}+\beta_m\|h_1-h_2\|_m+\Delta_r^m|\lambda_1-\lambda_2|\,.
\end{aligned} \tag{B.5}$$

Again, interchanging the roles of (h_1, λ_1) and (h_2, λ_2), we obtain a bound similar to the one in (B.5), which, combined with (B.5), gives us (B.1) for $i = r$. This completes the proof of the proposition. □