Efficient schedules for the problem of optimal node visitation in acyclic stochastic digraphs

Theologos Bountourelis and Spyros Reveliotis
School of Industrial & Systems Engineering
Georgia Institute of Technology
{tbountou,spyros}@isye.gatech.edu

Abstract

Given a stochastic, acyclic, connected digraph with a single source node and a control agent that repetitively traverses this graph, each time starting from the source node, we want to define a control policy that will enable this agent to visit each of the graph terminal nodes a prespecified number of times, while minimizing the expected number of the graph traversals. We formulate this problem as a specially structured Discrete Time Markov Decision Process and we develop a series of computationally efficient and asymptotically optimal policies, by building upon the special structure of the problem and the relevant theory of suboptimal control.

1 Introduction

The problem addressed in this work can be stated as follows: Given a stochastic, acyclic, connected digraph with a single source node and a control agent that repetitively traverses this graph, each time starting from the source node, we want to define a control policy that will enable this agent to visit each of the graph terminal nodes a prespecified number of times, while minimizing the expected number of the graph traversals. A practical motivation for this problem has been the work presented in [13], where a learning agent must compute on-line an optimal policy for a task that evolves episodically over a state space that is stochastic and acyclic, and it has a single source state that defines the task initial state. As established in [13], the agent can obtain an $\epsilon$-optimal policy with probability at least $1 - \delta$, by sampling the various actions available at each state a certain number of times\textsuperscript{1} and selecting the action that results to the highest sample mean. Furthermore, this sampling must be performed on a layer

\textsuperscript{1}that depends on the graph structure and the performance parameters $\epsilon$ and $\delta$
by layer basis, starting from the terminal states and proceeding towards the initial state of the underlying state space. Higher-level states that have covered all the required sampling and have their actions selected are declared “fully explored” and abandon the layer of “actively explored” states. On the other hand, lower-level states join the layer of “actively explored” states when all their immediate successors become fully explored. It is clear that, in this setting, expeditious learning translates to the completion of all the required sampling in a minimum number of episodes. However, this minimization can be defined only in an expected sense, since the stochasticity of the environment implies that the agent might fail to reach any of the actively explored states during some episodes, under any policy. Another potential application context for the problem considered in this work is provided by various experimental setups where the subject must be studied in a number of states that are obtained from an initial state through some sequential treatment with probabilistic outcomes at the various stages. Assuming that the performed treatment has a destructive effect on the subject, one would like to obtain the required measurements while minimizing the number of subjects utilized in the experiment.

From a methodological standpoint, the aforementioned problem falls in the broader category of stochastic scheduling problems [11, 12]. As indicated in [11], most stochastic scheduling problems are notoriously hard to solve optimally, and one has to compromise for solutions that are suboptimal but computationally tractable. In particular, the last few years have seen the emergence of a number of works that seek to provide suboptimal solutions to various stochastic scheduling problems by exploiting some “relaxed” version of the original problem. Furthermore, in many cases, this line of analysis also provides guaranteed bounds for the potential suboptimality of the derived policies; c.f., for instance, the works of [4, 6] and the references provided therein.

Our results follow the spirit of these broader developments. Hence, in the first part of the paper, we provide a formal characterization of the considered problem and we show that it abstracts to a specially structured “stochastic shortest path” (SSP) problem [3]. However, the solution of this SSP formulation through standard approaches based on Dynamic Programming is of non-polynomial complexity with respect to the underlying problem size, and therefore, in the rest of the work we develop a series of suboptimal policies that seek to trade off operational efficiency for computational tractability. Some important traits of these policies are that (a) they are asymptotically optimal, with the ratio of their performance to the performance of the optimal policy converging to unity as the node visitation requirements grow uniformly to infinity, and (b) collectively they establish a broad range of options for the effective and

---

Footnote:

1We also identify significant special structure that guarantees stronger convergence results for the proposed policies.
systematic resolution of the aforementioned trade-off between efficiency and computational
delay and tractability. The development of these policies is based on (i) the pertinent
exploitation of a continuous – or “fluid” – relaxation of the problem towards the characterization
of an efficient randomized policy, and (ii) the ability to derive a closed-form expression for the
performance of this randomized policy, which further enables (iii) the optimization of the policy
parameters, and (iv) its embedding to adaptive control schemes that can lead to even more
enhanced performance. Our results regarding item (i) above are similar in spirit to the results
of [4, 6] concerning the computation of near-optimal policies for the job shop scheduling problem,
but the underlying analysis is substantially different. The results regarding item (ii) are based
on our ability to represent the dynamics generated by the considered randomized policy as a
Generalized Semi-Markov3 Scheme (GSMS) [10]. The results on item (iii) employ standard
techniques borrowed from non-linear optimization [2], and those on item (iv) are building on
notions borrowed from adaptive control and “rollout” algorithms [1, 3].

The rest of the paper is organized as follows: Section 2 provides a formal characterization of
the considered problem and its abstraction to a specially structured SSP. Section 3 introduces
the aforementioned suboptimal policies, establishes their properties, including their asymptotic
optimality, and investigates their relevant dominance. Subsequently, Section 4 complements the
results developed in Section 3 through a number of computational experiments that demonstrate
and validate them, but also offer additional practical insights. Finally, Section 5 concludes the
paper and suggests directions for the extension of the presented results.

2 Problem description and its MDP formulation

2.1 A formal description of the considered problem

An instance of the problem considered in this work is completely defined by a quadruple \( \mathcal{E} = (X, \mathcal{A}, \mathcal{P}, \mathcal{N}) \), where

- \( X \) is a finite set of nodes, that is partitioned into a sequence of “layers”, \( X^0, X^1, \ldots, X^L \). \( X^0 = \{x^0\} \) defines the source or root node, while nodes \( x \in X^L \) are the terminal or leaf nodes.

- \( \mathcal{A} \) is a set function defined on \( X \), that maps each \( x \in X \) to the finite, non-empty set \( \mathcal{A}(x) \), comprising all the decisions / actions that can be executed by the control agent at node \( x \). It is further assumed that for \( x \neq x' \), \( \mathcal{A}(x) \cap \mathcal{A}(x') = \emptyset \).

\(^3\) actually, Markovian
• $\mathcal{P}$ is the transition function, defined on $\bigcup_{x \in X} \mathcal{A}(x)$, that associates with every action $a$ in this set a discrete probability distribution $p(\cdot; a)$. The support sets, $\mathcal{S}(a)$, of the distributions $p(\cdot; a)$ are subsets of the set $X$ that satisfy the following property: For any given action $a \in \mathcal{A}(x)$ with $x \in X^i$ for some $i = 0, \ldots, L - 1$, $\mathcal{S}(a) \subseteq \bigcup_{j=i+1}^L X^j$; for $a \in \mathcal{A}(x)$ with $x \in X^L$, $\mathcal{S}(a) = X^0$. In words, the previous assumption implies that the control agent traverses the considered graph in an iterative manner, where each iteration is an acyclic traversal that starts from the root node and ends at a leaf node $x \in X^L$.

Furthermore, it is assumed that for every $x \in X$, there exists at least one action sequence $\xi(x) = a^{(0)}a^{(1)}\ldots a^{(k(x))}$ such that (i) $a^{(0)} \in \mathcal{A}(x^0)$, (ii) $\forall i = 1, \ldots, k(x)$, $a^{(i)} \in \mathcal{A}(x^{(i)})$ with $p(x^{(i)}; a^{(i-1)}) > 0$, and (iii) $p(x; a^{k(x)}) > 0$; we shall refer to this action sequence as an action path from node $x^0$ to node $x$.

• $\mathcal{N}$ is the visitation requirement vector, that associates with each node $x \in X^L$ a visitation requirement $\mathcal{N}_x \in \mathbb{Z}^+ \cup \{0\}$. The support $||\mathcal{N}||$ of $\mathcal{N}$ is defined by the nodes $x \in X^L$ with $\mathcal{N}_x > 0$; we shall refer to nodes $x \in ||\mathcal{N}||$ as the problem target nodes.

• Finally, we define the instance size $|\mathcal{E}| \equiv |X| + |\bigcup_{x \in X} \mathcal{A}(x)| + ||\mathcal{N}||$, where application of the operator $|$ on a set returns the cardinality of this set, while application on a vector returns its $l_1$ norm.

In the subsequent discussion we shall also employ the variable vector $\mathcal{N}^c$ to denote the vector of the remaining visitation requirements. The control agent starts from the initial node $x^0$ at period $t = 0$, sets $\mathcal{N}^c := \mathcal{N}$, and at every consecutive period $t = 1, 2, 3, \ldots$, it (i) observes its current position, $x$, on the graph, and the vector of the remaining node visitation requirements, $\mathcal{N}^c$, (ii) selects an action $a \in \mathcal{A}(x)$ and commands its execution, and (iii) upon reaching one of the terminal nodes, $x \in X^L$, updates $\mathcal{N}_x^c$ to $(\mathcal{N}_x^c - 1)^+$, and subsequently, resets itself back to the initial node $x^0$, in order to start another traversal. The entire operation terminates when all the node visitation requirements have been satisfied, i.e., $\mathcal{N}^c$ has been reduced to zero. Our intention is to determine an action selection scheme – or, a policy – $\pi$, that maps each tuple $(x, \mathcal{N}^c)$ to an action $\pi(x, \mathcal{N}^c) \in \mathcal{A}(x)$ in a way that minimizes the expected number of graph traversals until $\mathcal{N}^c = 0$.

Example 1: Figure 1 depicts a problem instance $\mathcal{E}$ where $X$ is partitioned into layers $X^0 = \{x^0\}$ and $X^1 = \{x^1, x^2\}$. The decisions associated with each node are $\mathcal{A}(x^0) = \{\alpha^1, \alpha^2\}$, $\mathcal{A}(x^1) = \{\alpha^3\}$, and $\mathcal{A}(x^2) = \{\alpha^4\}$. The corresponding transition probabilities are $p(x^1; \alpha^1) = 0.5$, $p(x^2; \alpha^1) = 0.5$, $p(x^1; \alpha^2) = 0.3$, $p(x^2; \alpha^2) = 0.7$, and $p(x^0; \alpha^3) = p(x^0; \alpha^4) = 1$. Finally, the visitation requirement vector is defined by $\mathcal{N}_{x^1} = 2$, $\mathcal{N}_{x^2} = 1$. 

4
2.2 The induced stochastic shortest path problem

The problem defined in Section 2.1 can be further abstracted to a Discrete Time Markov Decision Process (DT-MDP), \( \mathcal{M} = (S, A, t, c) \), where

- \( S \) is the finite set of states, identified with the tuples \( (x, N^c) \), where \( x \in X \) and \( N^c \in \prod_{x \in X^L} \{0, \ldots, N_x^L\} \).

- \( A \) is a set function defined on \( S \) that maps each state \( s \in S \) to the finite, non-empty set \( A(s) \), comprising all the decisions / actions that are feasible in \( s \). More specifically, for \( s = (x, N^c) \), \( A(s) \) coincides with \( A(x) \) as specified in the definition of \( \mathcal{E} \).

- \( t : S \times \bigcup_{s \in S} A(s) \times S \rightarrow [0, 1] \) is the MDP state transition function, i.e., a partial function defined on all triplets \( (s, a, s') \) with \( a \in A(s) \), and with \( t(s, a, s') \) being the probability to reach state \( s' \) from state \( s \) on decision \( a \). More specifically, for \( s = (x, N^c) \), \( a \in A(s) \), \( s' = (x', N'^c) \),

\[
t(s, a, s') = \begin{cases} 
  p(x'; a), & \text{if } x \in X^l, l \in \{0, \ldots, L - 1\}, x' \in \bigcup_{k=l+1}^L X^k, N'^c = N^c; \\
  1, & \text{if } x \in X^L, x' = x^0, N'^c = (N^c_x - 1)^+, N'^c_y = N^c_y, \forall y \in X^L/\{x\}; \\
  0, & \text{otherwise.} 
\end{cases}
\]

(1)

- \( c : S \rightarrow \{0, 1\} \) is the cost function, where for \( s = (x, N^c) \),

\[
c(s) = \begin{cases} 
  1, & \text{if } x \in X^L \text{ and } N^c \neq 0; \\
  0, & \text{otherwise.} 
\end{cases}
\]

(2)

Notice that the cost function defined by Equation 2 assigns a unit cost to every resetting transition that takes the control agent from a leaf node back to the root node, but only when

Figure 1: An example problem instance
there is at least one leaf node with a positive requirement. Hence, the set of states $s = (x, \mathcal{N}^c)$ with $\mathcal{N}^c = 0$ constitute a closed class which is also cost-free, i.e., once the process enters this class of states it will remain in it, and there will be no further cost accumulation. For the purposes of the subsequent development, we shall represent this entire class of states with a single aggregate state, $s^T$, which we shall refer to as the problem terminal state; clearly, $s^T$ is absorbing and cost-free under any policy $\pi$. Furthermore, the MDP state set $S$ will be redefined to $S \equiv \{(x, \mathcal{N}^c)|\mathcal{N}^c \neq 0\} \cup \{s^T\}$, and the action, state transition and cost functions, $A$, $t$ and $c$, will also be appropriately redefined to reflect the above aggregation. In particular, for the terminal state $s^T$, we define $A(s^T) = \{a^T\}$ with $t(s^T, a^T, s^T) = 1; t(s^T, a^T, s) = 0, \forall s \in S \setminus \{s^T\}$, and $c(s^T) = 0$. The redefinition of the remaining elements of $A$, $t$ and $c$ is straightforward and the relevant details are left to the reader. Figure 2 exemplifies the above construct by depicting the state transition diagram for the MDP induced by the problem instance depicted in Figure 1.

In the above MDP modelling framework, we are particularly interested in a policy, $\pi^*$, that, starting from the initial state $s^0 \equiv (x^0, \mathcal{N})$, will drive the underlying process to the terminal state $s^T$. Figure 2: The State Transition Diagram for the stochastic shortest path problem induced by the problem instance depicted in Figure 1.
state \( s^T \) with the minimum expected total cost. Formally,

\[
\pi^* = \arg \min_{\pi \in \Pi} \mathbb{E}_{\pi} \left[ \sum_{t=0}^{\infty} c(s_t) \mid s_0 = s^0 \right]
\]

where \( \Pi \) denotes the entire set of policies and the expectation \( \mathbb{E}_{\pi} \left[ \cdot \right] \) is taken over all possible realizations under policy \( \pi \). This specification of \( \pi^* \) brings the considered MDP problem to a particular class of MDP problems known as stochastic shortest path (SSP) problems [3]. It is easy to see that, under the assumptions stated in Section 2.1, this SSP problem is well-defined, and therefore, according to [3]:

**Theorem 1** For the SSP formulation characterizing the problem considered in this work there exists a unique vector \( V^*(s), s \in S \), with \( V^*(s^T) = 0 \) and with its remaining components satisfying the Bellman equation

\[
\forall s \in S \setminus \{s^T\}, \quad V^*(s) = \min_{a \in A(s)} \left\{ c(s) + \sum_{s' \in S} t(s, a, s') \cdot V^*(s') \right\}
\]

Furthermore, the vector \( V^*(s) \) defines an optimal policy \( \pi^* \) by setting

\[
\forall s \in S \setminus \{s^T\}, \quad \pi^*(s) := \arg \min_{a \in A(s)} \left\{ c(s) + \sum_{s' \in S} t(s, a, s') \cdot V^*(s') \right\}
\]

The vector \( V^*(s) \) introduced in Theorem 1 is known as the optimal value function or the optimal cost-to-go vector for the considered SSP formulation. Each component of \( V^*(s) \) expresses the expected total cost of initiating the underlying process at state \( s \in S \) and subsequently following an optimal policy. In particular, the expected performance for an optimal policy \( \pi^* \) is characterized by \( V^* \equiv V^*(s^0) \). From a computational standpoint, \( V^*(s) \) can be obtained through a number of approaches coming from the broader area of Dynamic Programming (DP) [3]. Next, we focus on an approach that employs a linear programming (LP) formulation and it will be useful in the subsequent developments presented in this document.

**Theorem 2** [3] The optimal value vector \( V^*(s), s \in S \setminus \{s^T\} \), for the SSP formulation considered in this work is the optimal solution of the following linear program:

\[
\max \sum_{s \in S \setminus \{s^T\}} V(s)
\]
\[ s.t. \]
\[ \forall s \in S \setminus \{s^T\}, \; \forall a \in A(s), \]
\[ V(s) \leq c(s) + \sum_{s' \in S \setminus \{s^T\}} t(s, a, s') \cdot V(s') \]  

\[ \square \]

From a practical computational standpoint, the value of Theorems 1 and 2 in the determination of the optimal policy for any given problem instance, \( \mathcal{E} = (X, A, \mathcal{P}, \mathcal{N}) \), is severely limited by the fact that the size of the state space, \( S \), of the induced SSP problem grows exponentially to the number of the problem target nodes, \(|\mathcal{N}|\), since \(|S| = |X| \cdot \prod_{x \in X} (N_x + 1) - |X| + 1\). On the other hand, the monotonic decrease of \( N^c \), and the acyclic structure in the underlying state space that is implied by this effect, enable the incremental solution of the formulation of Theorem 2 through a series of subproblems that are defined on the subspaces obtained by fixing the value for the remaining visitation requirement vector \( N^c \). Clearly, each of these subproblems will be of polynomial complexity with respect to \(|\mathcal{E}|\). But the set of all possible values for \( N^c \) is an exponential function of \(|\mathcal{N}|\), and therefore, the complexity of the overall approach remains super-polynomial. Motivated by these observations, in the next section we develop a number of suboptimal policies for the considered problem that seek to trade off some operational efficiency for computational tractability. However, all of the presented policies maintain asymptotic optimality, in that the ratio of their expected value over the expected value of the optimal policy converges to unity as the node visitation requirements grow uniformly to infinity. Furthermore, when viewed from a collective standpoint, the proposed policies define a broad and systematic range of options for effecting the aforementioned trade-off between performance and computational expediency and tractability.

3 Suboptimal control policies

3.1 The class of simple randomized policies

It is clear from the concluding discussion of the previous section that the main reason for the non-polynomial complexity presented by the standard DP-based approaches when applied to the considered SSP problem, is the exponentially large number of the possible values of the vector \( N^c \) that constitutes part of the system state \( s = (x, N^c) \). This observation motivates the introduction and study of a class of policies that is defined only on the basis of the first component of the system state, i.e., the position \( x \in X \) of the acting agent. This idea is formalized by the concept of simple randomized policy as follows:
Definition 1 Given a problem instance $E$, the class of simple randomized policies, $\Pi^S$, is defined by the following two properties: (i) For any $\pi \in \Pi^S$ and $s = (x, N^c) \in S$, the action $\pi(s)$ is chosen according to a probability distribution $D^\pi(\cdot; s) = D^\pi(\cdot; x)$, i.e., this distribution depends only upon the first component of $s$. (ii) For $N_x > 0$, $x \in X^L$, $\pi$ connects $(x^0, N)$ and $(x, N)$ with a path of positive probability. □

The satisfaction of Assumption (ii) in Definition 1 is guaranteed by the existence of action paths from node $x^0$ to any node $x \in X$, that was presumed in the problem statement, and the policy randomization. The next proposition establishes that simple randomized policies are characterized uniquely by the action selection probabilities that they induce for any single traversal of the underlying graph, and it provides an interesting “flow” interpretation for these probabilities. The proof of this proposition is based on simple inductive arguments and it is provided for completeness in Appendix A.

Proposition 1 There is a bijection between the space of simple randomized policies $\Pi^S$ and the space $\mathcal{X}$ of vectors $\chi = \{\chi_\alpha | \alpha \in \mathcal{A}(x), x \in X \setminus X^L\}$ satisfying

$$\sum_{\alpha \in \mathcal{A}(x^0)} \chi_\alpha = 1$$

$$\sum_{\alpha : x \in S(\alpha)} \chi_\alpha \cdot p(x, \alpha) = \sum_{\alpha \in \mathcal{A}(x)} \chi_\alpha, \ \forall x \in X \setminus \{x^0, X^L\}$$

$$\sum_{\alpha : x \in S(\alpha)} \chi_\alpha \cdot p(x, a) > 0, \ \forall x \in X^L, N_x > 0.$$  

□

The variables $\chi_\alpha$, $\alpha \in \mathcal{A}(x)$, $x \in X \setminus X^L$, denote the probability of executing action $\alpha$ during any single traversal of the graph under policy $\pi$. The vector $\chi$ can also be interpreted as the “flow” pattern that would result in the considered graph if a unit flow was induced in the source node $x^0$ and subsequently it was distributed at the different nodes $x \in X \setminus X^L$ according to the proportions suggested by the distributions $D^\pi(\cdot, x)$.

Given a simple randomized policy $\pi$ and the corresponding vector $\chi^\pi \in \mathcal{X}$, we also define the vector $\rho^\pi \equiv \rho(\chi^\pi)$, of dimensionality $|X^L|$, with

$$\rho^\pi_x \equiv \sum_{\alpha : x \in S(\alpha)} \chi^\pi_\alpha \cdot p(x, a), \ x \in X^L$$

Clearly, $\rho^\pi_x$, $x \in X^L$, expresses the probability of reaching node $x \in X^L$ during a single traversal of the underlying graph $G$, under $\pi$. The following theorem gives an explicit characterization of the connection between the vector $\rho^\pi$ and the performance of a policy $\pi \in \Pi^S$. 

Theorem 3 Consider a problem instance $\mathcal{E} = (X, A, P, N)$, a simple randomized policy $\pi \in \Pi^S$ for it, and the corresponding probability vector $\rho^\pi = \rho(\chi^\pi)$. Then,

$$V^\pi = V(\rho^\pi) = E[\max_{j: N_j > 0} \left\{ \frac{1}{\rho^\pi_j} \sum_{i=1}^{N_j} \Xi^i_j \right\}]$$

(12)

where $\Xi^i_j$ are independent identically distributed exponential random variables with rate $\lambda = 1$.

Proof: Consider a continuous-time version of the problem where the process is guided by the simple randomized policy $\pi$ and a graph traversal is concluded at random times $Y_i$ generated by a Poisson process with rate $\lambda = 1$. Let $T_j$ denote the time until target leaf node $j$ has satisfied its visitation requirements, and $N$ denote the total number of graph traversals required until every visitation requirement is satisfied. Then it is easy to see that (i) $T_j$ is distributed according to a Gamma distribution with parameters $N_j$ and $\rho^\pi_j$, and (ii) the $T_j$’s are independent. Let $T = \max_{j: N_j > 0} \{T_j\}$. Then

$$E[\max_{j: N_j > 0} \{T_j\}] = E[T]$$

$$= E[\sum_{i=1}^{N} Y_i]$$

$$= E[E[\sum_{i=1}^{N} Y_i | N]]$$

$$= E[N \cdot E[Y_1]]$$

$$= E[N]$$

$$= V^\pi$$

(13)

Since $T_j$ is equal in distribution to $\frac{1}{\rho^\pi_j} \sum_{i=1}^{N_j} \Xi^i_j$, we have that

$$E[\max_{j: N_j > 0} \{T_j\}] = E[\max_{j: N_j > 0} \left\{ \frac{1}{\rho^\pi_j} \sum_{i=1}^{N_j} \Xi^i_j \right\}].$$

(14)

The result now follows by combining Equations 13 and 14. □

An immediate implication of Theorem 3 is that the performance, $V^\pi$, of a simple randomized policy $\pi$, can be evaluated through the numerical integration of a continuous function since, for $\rho^\pi = \rho(\chi^\pi)$,
\[ V^\pi = V(\rho^\pi) \]
\[ = E[\max_{j: N_j > 0} \{ \frac{1}{\rho_j^\pi} \sum_{i=1}^{N_j} \Xi_j^i \}] \]
\[ = \int_0^\infty P(\max_{j: N_j > 0} \{ \frac{1}{\rho_j^\pi} \sum_{i=1}^{N_j} \Xi_j^i \} > t) dt \]
\[ = \int_0^\infty (1 - \prod_{j: N_j > 0} P(\frac{1}{\rho_j^\pi} \sum_{i=1}^{N_j} \Xi_j^i \leq t)) dt \]
\[ = \int_0^\infty (1 - \prod_{j: N_j > 0} F_{\Xi_j^i}(\rho_j^\pi \cdot t)) dt \] (15)

where \( F_{\Xi_j^i}(t) \) is the cumulative distribution function of the Gamma(\( N_j \), 1) distribution. Another consequence of Equation 12 is the convexity of the function \( V(\rho^\pi) \) with respect to \( \rho^\pi \). This last property subsequently enables the effective and efficient solution of the optimization problem

\[ \min_{\pi \in \Pi^S} V^\pi \] (16)

which, under Theorem 3 and Proposition 1, can be alternatively stated as

\[ \min V(\rho) \] (17)

s.t. \( \rho = \rho(\chi), \; \chi \in \mathcal{X} \).

Indeed, the objective function of Formulation 17, \( V(\rho) \), is convex in \( \rho \) and continuously differentiable. Furthermore, the convexity of space \( \mathcal{X} \), as delineated by Equations 8-10, when combined with the linearity of \( \rho(\chi) \) with respect to \( \chi \), as revealed by Equation 11, imply that the space \( \{ \rho \mid \rho = \rho(\chi), \chi \in \mathcal{X} \} \) is also convex. Hence, the optimization problem defined by Equation 17 possesses a convex smooth structure and therefore it can be effectively addressed by standard solution techniques coming from the area of non-linear programming; we refer to [2] for the relevant details. In the following, we shall denote an optimal solution for the formulation of Equation 17 by \( \chi^{opt} \), and the corresponding simple randomized policy by \( \pi^{opt} \).

### 3.2 Asymptotically optimal simple randomized policies

In this section we establish that the simple randomized policy \( \pi^{opt} \), introduced in the previous section, is *asymptotically optimal*, with the ratio of its expected performance to \( V^* \) converging to unity, as the node visitation requirement vector, \( N \), grows uniformly to infinity. However, in order to establish this result, we need to introduce and analyze the performance of another
simple randomized policy that is obtained through a continuous – or “fluid” – relaxation of the original MDP problem. We shall refer to this policy as \( \pi^{rel} \), and as it will be revealed in the following, \( \pi^{rel} \) has its own merit as a suboptimal policy for the considered problem, since (i) it demonstrates the same asymptotically optimal performance with \( \pi^{opt} \), but (ii) it is computationally simpler to derive than the latter, and in addition, (iii) some numerical experimentation reported in Section 4 indicates that it provides the basis for one of the most efficient suboptimal policies for this problem. The definition of \( \pi^{rel} \) relies on the optimal solution of the following LP formulation, that will be called the “relaxing LP”:

\[
\min \sum_{a \in A(x^0)} \chi_a \quad \text{(18)}
\]

s.t.

\[
\sum_{a \in S(a)} p(x; a) \cdot \chi_a = \sum_{a \in A(x)} \chi_a, \quad \forall x \in X \setminus \{x^0\} \cup X^L \quad \text{(19)}
\]

\[
\sum_{a \in S(a)} p(x; a) \cdot \chi_a \geq N_x, \quad \forall x \in X^L \quad \text{(20)}
\]

\[
\chi_a \geq 0, \quad \forall a \in \bigcup_{x \in X \setminus X^L} \mathcal{A}(x) \quad \text{(21)}
\]

In the light of the flow-based interpretation of Equations 8–10, a natural interpretation of an optimal solution of the relaxing LP, \( \chi^* \), is that it constitutes a flow pattern that can satisfy the flow requirements for the terminal nodes \( x \in X^L \) expressed by the visitation requirement vector, \( N \), while minimizing the total amount of flow induced into the graph. Policy \( \pi^{rel} \) is the simple randomized policy induced by \( \chi^* \) according to Proposition 1. More specifically, given an optimal solution \( \{\chi_a^* | a \in \bigcup_{x \in X \setminus X^L} \mathcal{A}(x)\} \) of the LP defined by Equations 18-21, policy \( \pi^{rel} \) assigns to a state \( s = (x, N^c) \) with \( x \in X \setminus X^L \) and \( \sum_{a \in A(x)} \chi_a^* > 0 \), an action \( \pi(x, N^c) \in A(s) \) according to the probability distribution

\[
\text{Prob}(\pi^{rel}(x, N^c) = a) = \frac{\chi_a^*}{\sum_{a \in A(x)} \chi_a^*}, \quad a \in A(x). \quad \text{(22)}
\]

On the other hand, states \( s = (x, N^c) \) with \( x \in X \setminus X^L \) and \( \sum_{a \in A(x)} \chi_a^* = 0 \), are inaccessible under \( \pi^{rel} \), and the policy is indeterminate at them. Finally, for states \( s = (x, N^c), \ x \in X^L \), the policy executes the unique transition \( a \in A(s) \) with probability one. Clearly, the deployment and execution of the aforesaid policy \( \pi^{rel} \) is of polynomial complexity with respect to the problem size \( |E| \). Furthermore, another consequence of the above characterizations of

\footnote{Notice that a single problem instance, \( \mathcal{E} \), can have more than one instantiations of \( \pi^{rel} \) since, in general, there will be more than one optimal solutions, \( \chi^* \), for the corresponding relaxing LP.}
the relaxing LP and the policy \( \pi^{rel} \), is the following theorem, which is formally proven in Appendix A:

**Theorem 4** Given a problem instance \( \mathcal{E} = (X, \mathcal{A}, \mathcal{P}, \mathcal{N}) \), let \( V^*_r(\mathcal{N}) \) denote the optimal value of the relaxing LP, \( \chi^* \) denote an optimal solution of it, and \( \rho^{rel} = \rho(\chi^*) \). Then,

\[
V^*_r(\mathcal{N}) = \max_{j : N_j > 0} \left\{ \frac{N_j}{\rho^{rel}_j} \right\} \leq V^*.
\]  
(23)

□

Next, we proceed to establish the asymptotic optimality of \( \pi^{rel} \). For this, consider the problem sequence, \( \{\mathcal{E}(n)\} \), that is induced by a problem instance \( \mathcal{E} = (X, \mathcal{A}, \mathcal{P}, \mathcal{N}) \), through the scaling of the visitation requirement vector, \( \mathcal{N} \), by a factor \( n \in \mathbb{Z}^+ \). Also, let \( \{V^*_r(n)\} \) denote the sequence of the optimal objective values of the relaxing LP implied by the problem sequence \( \{\mathcal{E}(n)\} \), and \( \{V^*(n)\} \) denote the sequence of the corresponding optimal expected total costs. On the other hand, the perusal of the formulation of Equations 18–21 and of the first part of Equation 23 reveals that the policy \( \pi^{rel} \) remains invariant across the entire sequence \( \{\mathcal{E}(n)\} \). Hence, we also define \( \{V^\pi_r(n)\} \) as the sequence of the expected costs resulting by the application of the randomized policy \( \pi^{rel} \) to the problem instances \( \mathcal{E}(n) \). Then, we have:\footnote{We remind the reader that the notation \( f(n) = O(g(n)) \) implies that there exist positive constants \( c \) and \( n_0 \) such that \( 0 \leq f(n) \leq cg(n) \) for all \( n \geq n_0 \). Similarly, the notation \( f(n) = \Theta(g(n)) \) implies that there exist positive constants \( c_1, c_2 \), and \( n_0 \) such that \( 0 \leq c_1g(n) \leq f(n) \leq c_2g(n) \) for all \( n \geq n_0 \). [9]}

**Theorem 5**

\[
V^\pi_r(n) - V^*_r(n) = O(\sqrt{n}), \ n \in \mathbb{Z}^+.
\]  
(24)

**Proof:** Since \( V^*_r(n) \leq V^*(n), \ n \in \mathbb{Z}^+ \), it suffices to prove that

\[
V^\pi_r(n) - V^*_r(n) = O(\sqrt{n}).
\]  
(25)

Observe that

\[
V^\pi_r(n) - V^*_r(n) = E[\max_{j : N_j > 0} \left\{ \frac{1}{\rho^{rel}_j} \sum_{i=1}^{N_j} \xi_j \right\} - n \cdot \max_{x : N_x > 0} \left\{ \frac{N_x}{\rho^{rel}_x} \right\}]
\]

\[
\leq E[\max_{j : N_j > 0} \left\{ \frac{1}{\rho^{rel}_j} \sum_{i=1}^{N_j} \xi_j - \frac{nN_j}{\rho^{rel}_j} \right\}]
\]  
(26)

where the above equality results from Theorems 3 and 4 and the inequality is the result of the following property:

\[
\forall a_1, b_1 \in \mathbb{R}, i = 1, \ldots, n,
\]

\[
|\max\{a_1, a_2, \ldots, a_n\} - \max\{b_1, b_2, \ldots, b_n\}| \leq \max\{|a_1 - b_1|, |a_2 - b_2|, \ldots, |a_n - b_n|\}
\]
The application of the Central Limit Theorem [7] gives
\[
\frac{1}{\sqrt{n}} \cdot \left( \frac{1}{\rho^*_j} \sum_{i=1}^{nN_j} \Xi_j^i - \frac{nN_j}{\rho^*_j} \right) \Rightarrow N(0, N_j/(\rho^*_j)^2), \quad j : N_j > 0
\]  
(27)
where ‘\(\Rightarrow\)’ denotes convergence in distribution as \(n \to \infty\) and \(N(a, b)\) denotes the normal distribution with mean \(a\) and variance \(b\).

Also, observe that
\[
\frac{1}{n} \cdot E\left[ \frac{1}{\rho^*_j} \sum_{i=1}^{nN_j} \Xi_j^i - \frac{nN_j}{\rho^*_j} \right]^2 = \frac{N_j}{(\rho^*_j)^2}, \quad n \in \mathbb{Z}^+, \quad j : N_j > 0
\]  
(28)
which implies the uniform integrability [7] of \(\left\{ \frac{1}{\sqrt{n}} \left( \frac{1}{\rho^*_j} \sum_{i=1}^{nN_j} \Xi_j^i - \frac{nN_j}{\rho^*_j} \right) \right\}, \quad n \in \mathbb{Z}^+\)، for every \(j\) with \(N_j > 0\). But then, Equation 27, when combined with the independence of the \(\Xi_j^i\)’s and the Continuous Mapping Theorem [7], imply that
\[
\frac{1}{\sqrt{n}} E[\max_{j: N_j > 0} \left| \frac{1}{\rho^*_j} \sum_{i=1}^{nN_j} \Xi_j^i - \frac{nN_j}{\rho^*_j} \right|] \to E[\max_{j: N_j > 0} \left| N(0, N_j/(\rho^*_j)^2) \right|] \quad (29)
\]
as \(n \to \infty\). Finally, Equation 25 follows by combining Equation 29 with Equation 26. □

An immediate implication of Theorem 5 is the asymptotic optimality of the policy \(\pi^*_j\):

**Corollary 1**

\[
\frac{V^{\pi^*_j}(n)}{V^*(n)} \to 1 \quad \text{as} \quad n \to \infty
\]  
(30)

**Proof:** The combination of Theorems 4 and 5 implies that \(\lim_{n \to \infty} \frac{V^{\pi^*_j}(n)}{V^*(n)} \leq 1\), while the definition of \(V^*\) implies that \(V^{\pi^*_j}(n) \geq V^*(n), \quad \forall n \in \mathbb{Z}^+.\) □

Theorem 5 implies also the asymptotic optimality of the policy \(\pi^{opt}\), which was defined in Section 3.1. To obtain a formal statement of this result, let \(\{V^{\pi^{opt}}(n)\}\) denote the sequence of the expected costs that results from the application on the problem sequence \(\{E(n)\}\) of the corresponding randomized policies \(\pi^{opt}(n)\). Then, we have:

**Corollary 2**

\[
\frac{V^{\pi^{opt}}(n)}{V^*(n)} \to 1 \quad \text{as} \quad n \to \infty
\]  
(31)

**Proof:** Equation 31 is an immediate consequence of Corollary 1 when noticing that the definition of \(\pi^{opt}(n)\) implies that \(V^*(n) \leq V^{\pi^{opt}}(n) \leq V^{\pi^*_j}(n)\). □

Next we show that the bound implied by Equation 24 can be tight – i.e., that \(V^{\pi^*_j}(n) - V^*(n) = \Theta(\sqrt{n})\) – in certain cases, but there is also a significant problem sub-class for which the difference of Equation 24 converges to zero, as \(n \to \infty\). The first of these two results is established through the following example:
Example 2: Consider the very simple problem instance depicted in Figure 3, where the root node, $x^0$, is immediately connected to the two leaf nodes, $x^1$ and $x^2$, through two actions, $\alpha^1$ and $\alpha^2$, each leading to the corresponding leaf node with probability 1. Also, assume that the visitation requirement vector is $N = (1, 1)$. Then, it is clear that for any scaled requirement visitation vector $n \cdot N = (n, n)$, $V^*(n) = V^*_\rel(n) = 2n$. Furthermore, the problem symmetries imply that $\pi^{opt} = \pi^\rel$, with $\rho^\rel_i = \rho^{opt}_i = 0, i = 1, 2$. Finally, $V_{\pi^{opt}}(n) = V_{\pi^\rel}(n) = E[\max\{\frac{1}{0.5} \sum_{i=1}^n \Xi^1_i, \frac{1}{0.5} \sum_{i=1}^n \Xi^2_i\}]$, which gives

$$V_{\pi^{opt}}(n) - V^*(n) = V_{\pi^\rel}(n) - V^*(n) = V_{\pi^\rel}(n) - V^*_\rel(n) =$$

$$E[\max\{\frac{1}{0.5} \sum_{i=1}^n \Xi^1_i, \frac{1}{0.5} \sum_{i=1}^n \Xi^2_i\}] - 2n =$$

$$E[\max\{\frac{1}{0.5} \sum_{i=1}^n (\Xi^1_i - 1), \frac{1}{0.5} \sum_{i=1}^n (\Xi^2_i - 1)\}]$$

(32)

According to an argument similar to that provided in the proof of Theorem 5,

$$\frac{1}{\sqrt{n}} E[\max\{\frac{1}{0.5} \sum_{i=1}^n (\Xi^1_i - 1), \frac{1}{0.5} \sum_{i=1}^n (\Xi^2_i - 1)\}] \rightarrow E[\max\{N(0, 4), N(0, 4)\}]$$

(33)

as $n \rightarrow \infty$. But then, the $\Theta(\sqrt{n})$ nature of the quantities involved in the different parts of Equation 32 follows immediately from the fact that $E[\max\{N(0, 4), N(0, 4)\}] > 0$. □

Notice that in the previous example, $N_1/\rho^{rel}_1 = N_2/\rho^{rel}_2$. The equality of these ratios can be interpreted as an equality of the “difficulty” for the posed visitation requirements, and it turns out that it is a fundamental reason for the inability of the difference $V_{\pi^\rel}(n) - V^*(n)$ to converge to zero, as $n$ grows to infinity. A formal statement and a proof for this result is provided in Theorem 6 below. However, first we introduce a technical lemma that is needed in the proof of this theorem; the proof of this lemma is provided in the Appendix.

Lemma 1 Let $\{\Xi^i\}$ be a sequence of independent identically distributed exponential random variables with rate $\lambda = 1$. Then, for any $\rho \in \mathbb{R}\{0\}$ and $N \in \mathbb{Z}^+$, it holds that

$$E[\exp\{\frac{1}{\rho} \sum_{i=1}^{nN} \frac{\Xi^i - 1}{\sqrt{n}}\}] \rightarrow e^{N/2\rho^2}$$

(34)
as $n \to \infty$. □

**Theorem 6** Suppose that for a given problem instance $\mathcal{E} = (X, \mathcal{A}, \mathcal{P}, \mathcal{N})$, with $l \geq 2$ target leaf nodes, there exists a target leaf node $x^k$ such that, for any other target leaf node $x^j$, $\frac{N_j}{\rho_j} > \frac{N_k}{\rho_k}$. Then, as $n \to \infty$,

$$V_{rel}^\pi(n) - V_{rel}^*(n) \to 0 \quad (35)$$

**Proof:** Without loss of generality assume that $k = 1$. Then, the left part of Equation 35 can be re-written as follows:

$$V_{rel}^\pi(n) - V_{rel}^*(n) = E[\max_{j:N_j > 0} \left\{ \frac{n}{\rho_j} \sum_{i=1}^{nN_j} \Xi_j^i \right\}] - \frac{n}{\rho_1} \sum_{i=1}^{nN_1} \Xi_1^i - E[\frac{n}{\rho_1} \sum_{i=1}^{nN_1} \Xi_1^i]\]$$

$$= E[\max\{0, \frac{n}{\rho_2} \sum_{i=1}^{nN_2} \Xi_2^i - \frac{n}{\rho_1} \sum_{i=1}^{nN_1} \Xi_1^i, \ldots, \frac{n}{\rho_l} \sum_{i=1}^{nN_l} \Xi_l^i - \frac{n}{\rho_1} \sum_{i=1}^{nN_1} \Xi_1^i\}]$$

$$\leq E[(\frac{n}{\rho_2} \sum_{i=1}^{nN_2} \Xi_2^i - \frac{n}{\rho_1} \sum_{i=1}^{nN_1} \Xi_1^i)^+] + \ldots + E[(\frac{n}{\rho_l} \sum_{i=1}^{nN_l} \Xi_l^i - \frac{n}{\rho_1} \sum_{i=1}^{nN_1} \Xi_1^i)^+]$$

In order to prove the result of Theorem 6, it suffices to prove that, for all $j = 2, \ldots, l$,

$$E[\frac{n}{\rho_j} \sum_{i=1}^{nN_j} \Xi_j^i - \frac{n}{\rho_1} \sum_{i=1}^{nN_1} \Xi_1^i]^+ \to 0 \quad (36)$$

as $n \to \infty$.

Hence, consider an arbitrary $j \in \{2, \ldots, l\}$, and let $a_j = \frac{n}{\rho_j} - \frac{n}{\rho_1} > 0$. Then, by basic probability arguments and the Markov inequality, we get:
\[
E[(\frac{1}{\rho_{j}^{rel}} \sum_{i=1}^{nN_j} \Xi_j^i - \frac{1}{\rho_{1}^{rel}} \sum_{i=1}^{nN_1} \Xi_1^i)]^+ \\
= E[(\frac{1}{\rho_{j}^{rel}} \sum_{i=1}^{nN_j} (\Xi_j^i - 1) - \frac{1}{\rho_{1}^{rel}} \sum_{i=1}^{nN_1} (\Xi_1^i - 1) - n(\frac{N_1}{\rho_1^{rel}} - \frac{N_j}{\rho_j^{rel}}))^+] \\
= \sqrt{n} \cdot E[(\frac{1}{\rho_{j}^{rel}} \sum_{i=1}^{nN_j} \frac{\Xi_j^i - 1}{\sqrt{n}} - \frac{1}{\rho_{1}^{rel}} \sum_{i=1}^{nN_1} \frac{\Xi_1^i - 1}{\sqrt{n}} - \sqrt{n}(\frac{N_1}{\rho_1^{rel}} - \frac{N_j}{\rho_j^{rel}}))^+] \\
= \sqrt{n} \cdot \int_{a_j \sqrt{n}}^{\infty} P(\frac{1}{\rho_j^{rel}} \sum_{i=1}^{nN_j} \frac{\Xi_j^i - 1}{\sqrt{n}} - \frac{1}{\rho_1^{rel}} \sum_{i=1}^{nN_1} \frac{\Xi_1^i - 1}{\sqrt{n}} > t)dt \\
= \sqrt{n} \cdot \int_{a_j \sqrt{n}}^{\infty} e^{-t}E[\exp\{\frac{1}{\rho_j^{rel}} \sum_{i=1}^{nN_j} \frac{\Xi_j^i - 1}{\sqrt{n}} - \frac{1}{\rho_1^{rel}} \sum_{i=1}^{nN_1} \frac{\Xi_1^i - 1}{\sqrt{n}}\}]dt \\
\leq \sqrt{n} \cdot e^{-a_j \sqrt{n}} E[\exp\{\frac{1}{\rho_j^{rel}} \sum_{i=1}^{nN_j} \frac{\Xi_j^i - 1}{\sqrt{n}} - \frac{1}{\rho_1^{rel}} \sum_{i=1}^{nN_1} \frac{\Xi_1^i - 1}{\sqrt{n}}\}] \\
= \sqrt{n} \cdot e^{-a_j \sqrt{n}} \cdot E[\exp\{\frac{1}{\rho_j^{rel}} \sum_{i=1}^{nN_j} \frac{\Xi_j^i - 1}{\sqrt{n}}\}] \cdot E[\exp\{-\frac{1}{\rho_1^{rel}} \sum_{i=1}^{nN_1} \frac{\Xi_1^i - 1}{\sqrt{n}}\}] \\
= (37)
\]

The result of Equation 36 follows from Equation 37, when noticing that, according to Lemma 1,

\[
E[\exp\{\frac{1}{\rho_j^{rel}} \sum_{i=1}^{nN_j} \frac{\Xi_j^i - 1}{\sqrt{n}}\}] \rightarrow e^{N_j/2(\rho_{j}^{rel})^2} \\
E[\exp\{-\frac{1}{\rho_1^{rel}} \sum_{i=1}^{nN_1} \frac{\Xi_1^i - 1}{\sqrt{n}}\}] \rightarrow e^{N_1/2(\rho_{1}^{rel})^2}
\]
as \(n \rightarrow \infty\), while

\[
\sqrt{n} \cdot e^{-a_j \sqrt{n}} \rightarrow 0
\]

\(\square\)

Hence, under the condition of Theorem 6, the performance of all three policies, \(\pi^{rel}, \pi^{opt}\) and \(\pi^*\), converges to the lower bound \(V_{\pi^{rel}}(n)\), as the scaling factor \(n\) grows to infinity. Furthermore, Equation 37 indicates that this convergence will be quite fast, and its rate will be determined by the maximum difference \(\frac{N_k}{\rho_k^{rel}} - \frac{N_j}{\rho_j^{rel}}\) among all the target leaf nodes \(x_j^i\) with \(j \neq k\). An intuitive interpretation of this result is that, as this difference grows to larger values, the information contained in the optimal solution of the relaxing LP is adequate in order to strongly bias the system behavior towards the optimal policy. On the other hand, when the maximal ratio \(\frac{N_k}{\rho_k^{rel}}\)
is attained at more than one leaf nodes, both $\pi^{rel}$ and $\pi^{opt}$ will treat all these nodes as “equally difficult targets”. But due to the static nature of these policies, this impartiality can turn into a disadvantage in the later stages of the problem evolution, where the original ties have been resolved by the underlying randomness. In the next section we discuss how these problems can be alleviated, and the performance of the considered policies can be substantially improved, through some adaptive implementation mechanisms that enable the applied policy to revise its action selection scheme, and the resultant probability vector $\rho^{\pi}$, according to the information provided by the remaining requirement vector $N^{c}$.

### 3.3 Adaptive Policies

In order to derive the enhanced suboptimal policies sought in this section, it is pertinent to consider the partitioning of the state space $S$, of the SSP defined in Section 2.2, into the state subsets defined by a common remaining visitation requirement vector, $N^{c}$. Each of these subsets defines a notion of “macro-state” for the underlying process, while, as it was observed at the end of Section 2.2, the monotonic decrease of $N^{c}$ implies that the induced space of macro-states is traversed in an acyclic manner. More specifically, the process starts from the macro-state defined by $N^{c} = N$, and at every macro-transition, it proceeds to a macro-state where the corresponding vector $N^{c'}$ is obtained from $N^{c}$ by reducing one of its components by one unit. Next we show that this structure enables the specification of computationally efficient suboptimal policies that perform better than the policy $\pi^{opt}$ defined in Section 3.1. In the subsequent developments, we shall use the notation $\pi(N)$ to denote the instantiation of the policy $\pi$ on the problem instance $E = (X, A, P, N)$, and this notation will extend to any other element pertaining to the considered policy $\pi$.

The first improvement to $\pi^{opt}$ is easily obtained by an adaptive implementation of it, that recomputes the optimized vector $\chi^{opt}$ at every visited macro-state, by solving the corresponding optimization problem defined by Equation 17. We shall refer to the resulting policy as $\pi^{adopt}$. Next we establish that

**Proposition 2** $V_{\pi^{adopt}} \leq V_{\pi^{opt}}$

**Proof**: We prove this result by induction on $|N|$, i.e., the total number of visitation requirements. For $|N| = 1$, the process will visit only one macro-state before its termination, and therefore, $V_{\pi^{adopt}} = V_{\pi^{opt}}$. Next, we assume that the inequality of Proposition 2 holds for $|N| \leq n$, and we show that it will also hold for $|N| = n + 1$. To obtain this result, notice that the value...
function of any proper policy $\pi$ will satisfy the following recursion:

$$V^\pi(x^0, N) = \frac{1}{\sum_{x \in X^L \cdot N_x > 0} \rho_x} \cdot [1 + \sum_{x \in X^L \cdot N_x > 0} \rho_x^{\pi(N)} \cdot V^\pi(x^0, N - 1_x)]$$  \hspace{1cm} (38)$$

where (i) $\rho_x^{\pi(N)}$ denotes the probability of reaching node $x \in X^L$ in any single traversal of graph $G$ under policy $\pi$, while starting from state $(x^0, N)$ (c.f. Equation 11), and (ii) $1_x$ denotes the unit vector of dimensionality equal to $|X^L|$ and with its non-zero component corresponding to node $x$. Application of Equation 38 to $\pi^{\text{adopt}}$ gives that

$$V^{\pi^{\text{adopt}}}(x^0, N) = \frac{1}{\sum_{x \in X^L \cdot N_x > 0} \rho_x^{\pi^{\text{adopt}}(N)}} \cdot [1 + \sum_{x \in X^L \cdot N_x > 0} \rho_x^{\pi^{\text{adopt}}(N)} \cdot V^{\pi^{\text{adopt}}}(x^0, N - 1_x)]$$  \hspace{1cm} (39)$$

However, the definition of $\pi^{\text{adopt}}$ implies that $\rho_x^{\pi^{\text{adopt}}(N)} = \rho_x^{\pi^{\text{opt}}(N)}$ and $V^{\pi^{\text{adopt}}}(x^0, N - 1_x) = V^{\pi^{\text{adopt}}}(N-1_x)(x^0, N - 1_x)$, for all $x \in X^L$. Furthermore, $V^{\pi^{\text{adopt}}}(N-1_x)(x^0, N - 1_x) \leq V^{\pi^{\text{opt}}}(N-1_x)(x^0, N - 1_x)$, $\forall x \in X^L : N_x > 0$, where the first inequality results from the induction hypothesis and the second from the definition of $\pi^{\text{opt}}$. But then, Equation 39 implies that

$$V^{\pi^{\text{adopt}}}(N)(x^0, N) \leq \frac{1}{\sum_{x \in X^L \cdot N_x > 0} \rho_x^{\pi^{\text{opt}}(N)}} \cdot [1 + \sum_{x \in X^L \cdot N_x > 0} \rho_x^{\pi^{\text{opt}}(N)} \cdot V^{\pi^{\text{opt}}}(x^0, N - 1_x)] = V^{\pi^{\text{opt}}}(N)(x^0, N)$$  \hspace{1cm} (40)$$

\square

When combined with Corollary 2, Proposition 2 implies also the asymptotic optimality of policy $\pi^{\text{adopt}}$, in the sense of Corollaries 1 and 2. Next we define another class of policies that can outperform $\pi^{\text{opt}}$ and they constitute a customized implementation on the considered MDP problem of the “rollout” policies discussed in [1, 3]. Under this new regime, the policy to be applied at the macro-state defined by the visitation requirement vector $N^c$, is the “greedy” policy determined by Equation 5 while employing the value function $V(s)$, $s \in \{(x, N^c) | x \in \bigcup_{l=0}^{L} X^l\}$, that is obtained by restricting the LP of Theorem 2 to the considered macro-state and setting the value function of the “boundary” states $(x^0, N^c - 1_y)$, $y = 1, \ldots, |X^L| : N^c_y > 0$, equal to $V^{\pi^{\text{opt}}}(N^c-1_y)(x^0, N^c - 1_y)$. The solution of these LP’s and the determination of the corresponding local policies is performed every time that the process enters a new macro-state. The resulting policy is characterized as $\pi^{\text{roll}}$, and it holds that

**Proposition 3** $V^{\pi^{\text{roll}}} \leq V^{\pi^{\text{opt}}}$

*Proof:* Similar to the case of Proposition 2, we prove this result by induction on $|N|$. It is clear that for $|N| = 1$, $V^{\pi^{\text{roll}}} = V^*$, and therefore, Proposition 3 is true. Next suppose that
Proposition 3 holds true for $|\mathcal{N}| \leq n$. We shall show that it also holds true for $|\mathcal{N}| = n + 1$. The application of Equation 38 to policy $\pi^{roll}$ gives

$$V^{\pi^{roll}}(x^0, \mathcal{N}) = \frac{1}{\sum_{x \in X^L: N_y \geq 0} \rho_x^{\pi^{roll}(x)} \cdot [1 + \sum_{x \in X^L: N_y > 0} \rho_x^{\pi^{roll}(N)} \cdot V^{\pi^{roll}}(x^0, \mathcal{N} - 1_x)]} (41)$$

The definition of the policy $\pi^{roll}$ implies that $V^{\pi^{roll}}(x^0, \mathcal{N} - 1_x) = V^{\pi^{roll}}(N - 1_x)(x^0, \mathcal{N} - 1_x)$, which when combined with the induction hypothesis and Equation 41 imply that

$$V^{\pi^{roll}}(x^0, \mathcal{N}) \leq \frac{1}{\sum_{x \in X^L: N_y \geq 0} \rho_x^{\pi^{roll}(N)} \cdot [1 + \sum_{x \in X^L: N_y > 0} \rho_x^{\pi^{opt}(N)} \cdot V^{\pi^{opt}}(N - 1_x)(x^0, \mathcal{N} - 1_x)]} (42)$$

From the definition of the policy $\pi^{roll}$ we also have that

$$\frac{1}{\sum_{x \in X^L: N_y \geq 0} \rho_x^{\pi^{roll}(N)} \cdot [1 + \sum_{x \in X^L: N_y > 0} \rho_x^{\pi^{roll}(N)} \cdot V^{\pi^{opt}}(N - 1_x)(x^0, \mathcal{N} - 1_x)]} \leq \frac{1}{\sum_{x \in X^L: N_y \geq 0} \rho_x^{\pi^{opt}(N)} \cdot [1 + \sum_{x \in X^L: N_y > 0} \rho_x^{\pi^{opt}(N)} \cdot V^{\pi^{opt}}(N - 1_x)(x^0, \mathcal{N} - 1_x)]} (43)$$

while the definition of the policy $\pi^{opt}$ further implies that

$$\frac{1}{\sum_{x \in X^L: N_y \geq 0} \rho_x^{\pi^{opt}(N)} \cdot [1 + \sum_{x \in X^L: N_y > 0} \rho_x^{\pi^{opt}(N)} \cdot V^{\pi^{opt}}(N - 1_x)(x^0, \mathcal{N} - 1_x)]} \leq \frac{1}{\sum_{x \in X^L: N_y \geq 0} \rho_x^{\pi^{opt}(N)} \cdot [1 + \sum_{x \in X^L: N_y > 0} \rho_x^{\pi^{opt}(N)} \cdot V^{\pi^{opt}}(N)(x^0, \mathcal{N})]} (44)$$

But then, Proposition 3 follows immediately from Equations 42–44. □

Clearly, policy $\pi^{roll}$ is also asymptotically optimal in the sense of Corollaries 1 and 2. Furthermore, by employing $V^{\pi^{adroll}(\mathcal{N}_y - 1_y)}(x^0, \mathcal{N}_y - 1_y)$ instead of $V^{\pi^{opt}(\mathcal{N}_y - 1_y)}(x^0, \mathcal{N}_y - 1_y)$ as an estimate of the value function of the “boundary” states $(x^0, \mathcal{N}_y - 1_y)$, $y = 1, \ldots, |X^L|$ : $\mathcal{N}_y > 0$, and denoting the resulting rollout policy as $\pi^{adroll}$, we can also establish through arguments similar to those provided above that

**Proposition 4** $V^{\pi^{adroll}} \leq V^{\pi^{opt}}$

One complication regarding the implementation of policy $\pi^{adroll}$ compared to those of $\pi^{opt}$ and $\pi^{roll}$, is that the estimates $V^{\pi^{adroll}(\mathcal{N}_y - 1_y)}(x^0, \mathcal{N}_y - 1_y)$ are not available in closed form. However, in most practical cases they should be easily computed through simulation. Finally, one can also envision additional versions of $\pi^{roll}$ and $\pi^{adroll}$ where the LP that specifies the policy to be followed at any given macro-state is formulated over an extended state subset that includes the states of the considered macro-state plus the states of all the macro-states.
that can be reached from it in up to $k$ transitions. These policies are characterized as $k$-step rollout policies, and typically, they will outperform the corresponding policies resulting from single-step lookahead; we refer to [3] for some relevant discussion.

4 Computational Studies

In this section we present two examples that provide a concrete demonstration of the convergence results developed in Section 3, and also enable an assessment of the relative performance of the policies introduced in that section.

Example 3: This example pursues a detailed study of the problem instance depicted in Figure 4, in an effort to provide some additional insights on (i) the effects underlying the sub-optimality of the simple randomized policies $\pi^{rel}$ and $\pi^{opt}$; (ii) the role of the policy adaptation for mitigating this suboptimality, through the recalculation of the policy-defining vectors $\chi^{rel}$ and $\chi^{opt}$ at the visited macro-states; and (iii) the relative performance of the resulting policies $\pi^{adrel}$ and $\pi^{adopt}$ with respect to each other and the optimal policy. It should be obvious to the reader that, for the case depicted in Figure 4, the optimal policy is to choose action $\alpha^2$ until the visitation requirement of node $x^2$ has been satisfied. At this point, if the visitation requirement of node $x^1$ is still unmet, the policy switches to action $\alpha^1$ and satisfies this requirement in a single traversal. But any simple randomized policy, $\pi$, will fail to take advantage of the deterministic nature of action $\alpha^1$, as suggested above, since it must maintain a fixed vector $\rho^{\pi}$ at every visited macro-state. Hence, both $\pi^{rel}$ and $\pi^{opt}$ will apply a randomization over $\alpha^1$ and $\alpha^2$ that will maintain a significant positive probability for selecting action $\alpha^1$, in an effort to

\[\text{Figure 4: Example 3 – The considered problem instance}\]
Figure 5: Example 3 – The performance of the simple randomized policies obtained for different values of the selection probability, $\chi$, for action $\alpha^2$

Figure 6: Example 3 – The performance of the adaptive randomized policies obtained for different values of the selection probability, $\chi$, for action $\alpha^2$ in the initial macro-state
increase the accessibility of node \( x^1 \). In particular, \( \pi^{rel} \) will choose action \( \alpha^2 \) with a probability \( \chi^{rel} \) that balances the ratios \( N_i / \rho_i^{rel} \) for \( i = 1, 2 \); i.e.,

\[
\frac{1}{0.5 \chi^{rel}} = \frac{1}{0.01 \chi^{rel} + 1 - \chi^{rel}} \iff \chi^{rel} = 0.671
\]  

(45)

Furthermore, Equation 38 implies that the performance, \( V^{\pi^{rel}} \), of the resulting policy, can be evaluated by plugging the obtained value for \( \chi^{rel} \) into the following function:

\[
V^{\pi}(\chi) = \frac{1}{0.51 \chi + 1 - \chi} \left( 1 + \frac{0.5 \chi}{0.01 \chi + 1 - \chi} + \frac{0.01 \chi + 1 - \chi}{0.5 \chi} \right)
\]  

(46)

Thus, it is found that \( V^{\pi^{rel}} = 4.47 \). On the other hand, the \( \chi \) value that defines \( \pi^{opt} \) can be computed by solving the equation \( \frac{dV^{\pi}(\chi)}{d\chi} = 0 \) and picking the root that belongs in the interval \([0, 1]\). It turns out that \( \chi^{opt} = 0.61129 \) and \( V^{\pi^{opt}} = 4.37693 \). Finally, Figure 5 characterizes the performance of all simple randomized policies for the considered problem instance, by plotting \( V^{\pi}(\chi) \) for \( \chi \in [0.1, 0.9] \).

Policies \( \pi^{adrel} \) and \( \pi^{adopt} \) present enhanced performance with respect to their static counterparts, \( \pi^{rel} \) ad \( \pi^{opt} \), since they are able to optimize their decision in the second macro-state, on the basis of the remaining requirement vector \( \mathcal{N}^c \). However, they remain suboptimal since their decision in the initial macro-state is compromised by the aforementioned suboptimality of \( \pi^{rel} \) and \( \pi^{opt} \). A closed-form evaluation of these policies can be based again on Equation 38: The performance of the adaptive randomized policy that selects action \( \alpha^2 \) at the initial macro-state with probability \( \chi \), and in the next macro-state applies the optimal policy, is given by:

\[
V^{ad-\pi}(\chi) = \frac{1}{0.51 \chi + 1 - \chi} \left( 1 + 0.5 \chi + 2(0.01 \chi + 1 - \chi) \right)
\]  

(47)

Hence, from Equation 47 we obtain that \( V^{\pi^{adrel}} = 2.99 \), \( V^{\pi^{adopt}} = 2.99127 \) and \( V^{ad-\pi(1.0)} = 2.98039 \). Furthermore, Figure 6 plots the performance of all the adaptive randomized policies that are obtained by varying \( \chi \in [0.1] \) and validates our original suggestion that the optimal policy is obtained for \( \chi^* = 1.0 \).

We conclude the discussion of this example with two additional observations: First, it is interesting to notice the proximity of \( V^{\pi^{adrel}} \) and \( V^{\pi^{adopt}} \) to each other and to the value of the optimal policy, \( V^{ad-\pi(1.0)} \). Second, in this example it even holds that \( V^{\pi^{adrel}} < V^{\pi^{adopt}} \), as manifested by the values quoted above and by the strictly decreasing nature of \( V^{ad-\pi(\chi)} \). These two observations are indicative of our collective experience with the empirical performance of

---

7It is interesting to notice the proximity of the \( \chi^{rel} \) and \( \chi^{opt} \) values. This seems to be a more general effect for the considered problem, with \( \chi^{opt} \) and the resulting probability vector \( \rho^{opt} \) being minor “corrections” of \( \chi^{rel} \) and \( \rho^{rel} \). Furthermore, it can be shown that \( \rho^{opt}(n) \to \rho^{rel} \) as \( n \to \infty \).
Example 4: In this example we consider two problem instances defined by the stochastic graph of Figure 7 and the visitation requirement vectors $N = (3, 1, 1, 0, 0)$ and $N = (1, 2, 2, 2, 1)$. The solution of the corresponding relaxing LPs indicates that the problem instance defined by $N = (3, 1, 1, 0, 0)$ satisfies the conditions of Theorem 6, with the most difficult visitation requirement determined by the leaf node $x_4$. On the other hand, the problem instance defined by $N = (1, 2, 2, 2, 1)$ has a constant ratio $N_i/\rho_{rel}^i$ across all $i = 4, \ldots, 8$. Figures 8 and 9 report the performance of the policies $\pi_{rel}$, $\pi_{adrel}$ and $\pi_{roll}$ in each of these two cases, as the corresponding vector $N$ is scaled to increasingly larger values. The reported values for the policy $\pi_{rel}$ were obtained from the closed-form expression that characterizes the performance of a simple randomized policy $\pi$ as a function of the corresponding probability vector $\rho^\pi$, that was derived in Section 3. The performance of the policies $\pi_{adrel}$ and $\pi_{roll}$ was estimated through simulation. As expected from Theorem 6, in the case of the visitation requirement vector $N = (3, 1, 1, 0, 0)$, the performance of all three policies converges very fast to the lower bound $V_{rel}^*(n)$ – c.f. Figure 8. On the other hand, the ties of the ratios $N_i/\rho_{rel}^i$, $i = 4, \ldots, 8$, in the case of the visitation requirement vector $N = (1, 2, 2, 2, 1)$, result in the divergence of the performance of the considered policies from the lower bound $V_{rel}^*(n)$ – c.f. Figure 9. However, as expected, the distance of the performance of these policies from $V_{rel}^*(n)$ increases in a slow, sub-linear manner with respect to $n$, so that the corresponding ratios $V^\pi(n)/V_{rel}^*(n)$ decrease to one.
Figure 8: Example 4 – The performance of various simple and adaptive randomized policies compared to the lower bound $V^*_\text{rel}(n)$, for the basic visitation requirement vector $N = (3, 1, 1, 0, 0)$ and $n = 1, \ldots, 7$

Figure 9: Example 4 – The performance of various simple and adaptive randomized policies compared to the lower bound $V^*_\text{rel}(n)$, for the basic visitation requirement vector $N = (1, 2, 2, 2, 1)$ and $n = 1, \ldots, 15$
Finally, it is worth-noticing that $\pi^{adrel}$ outperforms again the other two policies, demonstrating a performance that is pretty close to the lower bound $V^\ast_{rel}(n)$.

5 Conclusions

This paper introduced the problem of the optimal node visitation in acyclic stochastic digraphs, and developed a number of suboptimal but computationally efficient policies for it that are expected to demonstrate very good performance, especially as the posed visitation requirements grow to larger values. The presented results are motivated by and are similar in spirit to some recent developments in stochastic scheduling theory and the suboptimal control of Markov Decision Processes. Future work will seek to (i) formally analyze the computational complexity of the considered problem; (ii) capitalize upon further insights and results from stochastic scheduling theory, like those presented in [5], in order to identify additional structure and properties for it; and (iii) extend the results and the policies developed herein to other problem variations, like in the case that each graph traversal might generate more than one threads and therefore result in the coverage of more than one visitation requirements.

A Proofs

A.1 Proof of Proposition 1

First we prove by induction that, given a simple randomized policy $\pi$, there is a unique vector $\chi^\pi$, $\alpha \in A(x)$, $x \notin X \cup X^L$, denotes the probability that action $\alpha$ will be executed during a single graph traversal under $\pi$, and this vector satisfies Constraints 8–10. Our induction runs on the number of layers, $L$, of the underlying acyclic graph. Hence, first consider a problem instance with $L = 1$ and assume two different simple randomized policies $\pi$ and $\pi'$ and the respective vectors $\chi^\pi$ and $\chi'^\pi$ defined by $\chi^\pi_a = D^\pi(\alpha; x^0)$ and $\chi'^\pi_a = D'^\pi(\alpha; x^0)$, $\alpha \in A(x^0)$.\footnote{We remind the reader that, according to the definitions provided in Section 2, $L = 1$ implies a two-layered graph $G$, where the first layer consists of the source node $x^0$, and the second layer consists of the terminal nodes.}

Since $\pi \neq \pi'$ and $L = 1$, there is an $\alpha \in A(x^0)$ such that $D^\pi(\alpha; x^0) \neq D'^\pi(\alpha; x^0)$, which further implies that $\chi^\pi \neq \chi'^\pi$. Next, assume that the hypothesis holds for all problem instances with $L \leq n$. We consider a problem instance with $L = n + 1$ and two different simple randomized policies $\pi, \pi'$. To proceed, first consider the two policies, $\pi, \pi'$, on the truncated acyclic graph consisting of the layers $X^0, \ldots, X^n$. According to our induction hypothesis, there exist vectors $\psi^\pi, \psi'^\pi$ such that for all $\alpha \in A(x), x \in X^l, 0 \leq l \leq n - 1$, the components $\psi^\pi_a, \psi'^\pi_a$ denote the probability that action $\alpha$ will be executed during a single traversal of the truncated graph.
under $\pi$ and $\pi'$ respectively. Define the vector $\chi^\pi$ where

$$
\chi^\pi_\alpha = \begin{cases} 
\psi^\pi_\alpha, & \text{if } \alpha \in \mathcal{A}(x), \ x \in X^l, \ 0 \leq l \leq n - 1 \\
(\sum_{\alpha': x \in S(\alpha')} \psi^\pi_\alpha \cdot p(x, \alpha')) \cdot D^\pi(\alpha; x), & \text{if } \alpha \in \mathcal{A}(x), \ x \in X^n
\end{cases}
$$

(48)

The vector $\chi^\pi'$ is defined accordingly. Clearly, $\chi^\pi_\alpha, \chi^\pi'$ denote the probability that action $\alpha$ will be executed during a single graph traversal under $\pi$ and $\pi'$ respectively. Now let $l^* = \min \{l | \alpha \in \mathcal{A}(x), x \in X^l, D^\pi(\alpha; x) \neq D^\pi'(\alpha; x), 0 \leq l \leq n \}$. In words, $l^*$ is the first graph layer where the two policies $\pi, \pi'$ disagree. If $l^* \leq n - 1$ then, according to the induction hypothesis, there is an $\alpha \in \mathcal{A}(x), x \in X^l, 0 \leq l \leq n - 1$, such that $\psi^\pi_\alpha \neq \psi^\pi'_\alpha$, which together with Equation 48 imply that $\chi^\pi \neq \chi^\pi'$. On the other hand, if $l^* = n$, there is an $\alpha \in \mathcal{A}(x), x \in X^L$, such that $D^\pi(\alpha; x) \neq D^\pi'(\alpha; x)$, whereas $\psi^\pi_\alpha = \psi^\pi'_\alpha$ for all $\alpha \in \mathcal{A}(x), x \in X^l, 0 \leq l \leq n - 1$, which when combined with Equation 48, imply again that $\chi^\pi \neq \chi^\pi'$. Hence for every simple randomized policy $\pi$, there is a unique vector $\chi^\pi$ such that $\chi^\pi_\alpha, \alpha \in \mathcal{A}(x), x \in X^L$, denotes the probability that action $\alpha$ will be executed during a single traversal of graph $\mathcal{G}$ under $\pi$. Clearly, $\chi^\pi$ should satisfy the balance conditions expressed by Equations 8-9. Furthermore, part (ii) of Definition 1 implies that every target leaf node, $x \in X^L$, of the underlying graph $\mathcal{G}$, is reachable under $\pi$, and therefore, Equation 10 is also satisfied by $\chi^\pi$. Hence, $\chi^\pi \in \chi$ and $\pi \to \chi^\pi$ is injective.

On the other hand, given a vector $\chi \in \chi$, we define the simple randomized policy $\pi$ that assigns to a state $s = (x, \mathcal{N}^c)$, with $x \in X \setminus X^L$ and $\sum_{\alpha \in \mathcal{A}(x)} \chi_\alpha > 0$, an action $\pi(x, \mathcal{N}^c) \in \mathcal{A}(s)$ according to the probability distribution

$$
D^\pi(\alpha; x) = \frac{\chi_\alpha}{\sum_{\alpha \in \mathcal{A}(x)} \chi_\alpha}, \ \alpha \in \mathcal{A}(x).
$$

(49)

Furthermore, for states $s = (x, \mathcal{N}^c)$ with $x \in X \setminus X^L$ and $\sum_{\alpha \in \mathcal{A}(x)} \chi_\alpha = 0$, the policy is indeterminate. Finally, for states $s = (x, \mathcal{N}^c), x \in X \setminus X^L$, the policy executes the unique transition $\alpha \in \mathcal{A}(s)$ with probability 1. Then, it can be shown, with a simple induction on the number of layers of graph $\mathcal{G}$, that $\chi_\alpha, \alpha \in \mathcal{A}(x), x \in X \setminus X^L$, denotes the probability that action $\alpha$ will be executed during a single graph traversal under $\pi$. Hence, for every terminal node $x \in X^L$, the underlying process guided by the randomized policy $\pi$, reaches $x$ with probability

$$
\rho_x = \sum_{\alpha: x \in S(\alpha)} \chi_\alpha \cdot p(x, a).
$$

(50)

When combined with Equation 10, this last equality implies that $\rho_x > 0$, for $\mathcal{N}_x > 0$, and establishes that $\pi$ belongs to the class of simple randomized policies. Thus, the mapping $\pi \to \chi^\pi$ is also surjective. □

27
A.2 Proof of Theorem 4

The validity of the equality part in Equation 23 is immediately obvious when realizing that $\rho^{rel}_j, j \in X^L$, denotes the amount of flow routed to node $j$ by the flow pattern corresponding to policy $\pi^{rel}$, for every unit of flow induced in the underlying graph (cf. the discussion after Prop. 1).

In order to prove the inequality of Equation 23, first notice that $V^*$ can also be computed by a variation of the LP formulation of Equations 6–7 where the original objective function has been substituted by $\max V(s^0)$; this substitution is legitimate since it is well-known in the relevant MDP theory that the SSP optimal value function $V^*(s), s \in S$, is the componentwise maximal vector that satisfies the constraint of Equation 7. Then, taking the dual of this new LP formulation, it suffices to show that (i) every feasible solution for this dual problem induces a feasible solution $\chi_a, a \in \bigcup_{x \in X \setminus X^L} A(x)$, for the relaxing LP, and (ii) the corresponding objective values are equal. The considered dual LP formulation is as follows [8]:

$$\min \sum_{s \in S \setminus \{s^T\}} \sum_{x \in X^L \setminus A(s)} q(s, a)$$ \quad (51)

s.t.

$$\forall s \in S \setminus \{s^T\},$$

$$\sum_{a \in A(s)} q(s, a) = 1_{\{s = s^0\}} + \sum_{s' \in S \setminus \{s^T\}} \sum_{a \in A(s')} t(s', a, s) \cdot q(s', a)$$ \quad (52)

$$\forall s \in S \setminus \{s^T\}, \forall a \in A(s),$$

$$q(s, a) \geq 0$$ \quad (53)

Let $q(s, a), s \in S \setminus \{s^T\}, a \in A(s)$, denote a feasible solution for this formulation, and define

$$\chi_a \equiv \sum_{s \in S \setminus \{s^T\} \setminus A(s)} q(s, a), \forall a \in \bigcup_{x \in X \setminus X^L} A(x)$$ \quad (54)

In the remaining part of this proof we shall show that the vector $\{\chi_a\}$ defined by Equation 54 satisfies the aforestated requirements (i) and (ii).

Clearly, Constraint 21 is immediately satisfied by Constraint 53 and the definition of $\{\chi_a\}$. Next we prove the feasibility of $\{\chi_a\}$ with respect to Constraint 19. Hence, consider a node
\( x \in X \setminus \{x^0 \cup X^L\} \). For it, we have that:

\[
\sum_{a \in A(x)} \chi_a = \sum_{a \in A(x)} \sum_{s \in S \setminus \{s^T\}: a \in A(s)} q(s, a) \quad \text{(from Eq. 54)}
\]

\[
= \sum_{s \in S \setminus \{s^T\}: x(s) = x} \sum_{a \in A(s)} q(s, a) \quad \text{(by term rearrangement)}
\]

\[
= \sum_{s \in S \setminus \{s^T\}: x(s) = x} \sum_{s' \in S \setminus \{s^T\}} \sum_{a \in A(s')} t(s', a, s) \cdot q(s', a) \quad \text{(from Eq. 52)}
\]

\[
= \sum_{a \in A(s)} p(x; a) \sum_{s' \in S \setminus \{s^T\}} q(s', a) \quad \text{(from Eq. 1 and term rearrangement)}
\]

\[
= \sum_{a \in A(s)} p(x; a) \cdot \chi_a \quad \text{(from Eq. 54)}
\]

To prove the satisfaction of Constraint 20 by the vector \( \{\chi_a\} \), first notice that this constraint is trivially satisfied for all non-target nodes \( x \in X^L \). Hence, consider a node \( x \in X^L \) with \( N_x > 0 \). Then, by working as in the proof of the validity of Constraint 19, we can easily establish that

\[
\sum_{a \in A(s)} p(x; a) \cdot \chi_a = \sum_{s \in S \setminus \{s^T\}: x(s) = x} \sum_{a \in A(s)} q(s, a) \quad \text{(55)}
\]

In the STD of the underlying SSP problem, consider the arc set \( C_x(N_x) \), consisting of all the arcs that lead from any state \( s \in S_x(N_x) \equiv \{(x, N^c) : N^c_x = N_x\} \) to the resultant state \( s' = (x^0, N^c - 1_x) \), where \( 1_x \) denotes the unit vector of dimensionality \( |X^L| \) and with the non-zero component corresponding to node \( x \). Clearly, since \( x \) is a target node, \( C_x(N_x) \) is non-empty. Furthermore, since this set aggregates all the possible transitions through which the visitation requirements for \( x \) are reduced from \( N_x \) to \( N_x - 1 \), it defines a cut on the underlying graph defined by \( S \) and \( A(s) \), \( s \in S \). This last observation when combined with the fact that \( \{q(s, a)\} \) can be interpreted as a flow that conveys a unit load from state \( s^0 \) to state \( s^T \), imply that

\[
\sum_{(s, a) \in C_x(N_x)} q(s, a) = 1 \quad \text{(56)}
\]

In the same way, we can define the arc sets \( C_x(N_x - k) \), \( k \in \{1, \ldots, N_x - 1\} \), each consisting of all the arcs that lead from any state \( s \in S_x(N_x - k) \equiv \{(x, N^c) : N^c_x = N_x - k\} \) to the state \( s' = (x^0, N^c - 1_x) \), and establish that

\[
\sum_{(s, a) \in C_x(N_x - k)} q(s, a) = 1, \quad \forall k \in \{1, \ldots, N_x - 1\} \quad \text{(57)}
\]

But then, the satisfaction of Constraint 20 results immediately from the fact that each of the summations appearing in Equations 56 and 57 is subsumed in the double summation that appears in the right-hand-side of Equation 55.
It remains to show that
\[
\sum_{a \in A(x^0)} \chi_a = \sum_{s \in S \setminus \{s^T\} : x(s) \in X^L} \sum_{a \in A(s)} q(s, a)
\]

The validity of this equation is established as follows:

\[
\sum_{a \in A(x^0)} \chi_a = \sum_{s \in S \setminus \{s^T\} : x(s) = x^0} \sum_{a \in A(s)} q(s, a) \quad \text{(as in the proof of Constraint 19)}
\]
\[
= \sum_{s \in S \setminus \{s^T, s^0\} : x(s) = x^0} \sum_{a \in A(s)} q(s, a) + \sum_{a \in A(s^0)} q(s^0, a)
\]
\[
= \sum_{s \in S \setminus \{s^T, s^0\} : x(s) = x^0} \sum_{s' \in S \setminus \{s^T\} a \in A(s')} t(s', a, s) \cdot q(s', a) + \sum_{s' \in S \setminus \{s^T\} a \in A(s')} t(s', a, s^0) \cdot q(s', a) \quad \text{(from Eq. 52)}
\]
\[
= 1 + \sum_{s \in S \setminus \{s^T, s^0\} : x(s) = x^0} \sum_{s' \in S \setminus \{s^T\} a \in A(s')} t(s', a, s) \cdot q(s', a)
\]
\[
= 1 + \sum_{s \in S : x(s) = x^0} \sum_{s' \in S \setminus \{s^T\} a \in A(s')} t(s', a, s) \cdot q(s', a)
\]
\[
- 1 \quad \text{(since} \sum_{s' \in S \setminus \{s^T\} a \in A(s')} t(s', a, (x^0, 0)) \cdot q(s', a) = 1)\]
\[
= \sum_{s \in S \setminus \{s^T\} : x(s) \in X^L} \sum_{a \in A(s)} q(s, a) \quad \text{(from Eq. 1)}
\]

\[\square\]

A.3 Proof of Lemma 1

For all \( n \in \mathbb{Z}^+ \) with \( \sqrt{n} > \frac{1}{\rho} \), we can write

\[
E[\exp\left\{ \frac{1}{\rho} \sum_{i=1}^{n-N} \frac{\Xi_i - 1}{\sqrt{n}} \right\}] = e^{-\frac{\sqrt{n}}{\rho} (E[\exp\left\{ \frac{1}{\rho} \sum_{i=1}^{n-N} \frac{\Xi_i - 1}{\sqrt{n}} \right\}])^{n-N}}
\]
\[
= e^{-\frac{\sqrt{n}}{\rho} \left( \frac{1}{1 - \frac{1}{\rho \sqrt{n}}} \right)^{n-N}}
\]
\[
= e^{-\frac{\sqrt{n}}{\rho} \cdot \left( \frac{\rho \sqrt{n}}{\rho \sqrt{n} - 1} \right)^{n-N}}
\]
\[
= e^{-\frac{\sqrt{n}}{\rho} \cdot e^{n-N \ln\left( \frac{\rho \sqrt{n}}{\rho \sqrt{n} - 1} \right)}} \quad (58)
\]
But

\[
\lim_{n \to \infty} \frac{\sqrt{n}}{\rho} + n \cdot \ln\left(\frac{\rho \sqrt{n}}{\rho \sqrt{n} - 1}\right) = \lim_{n \to \infty} \frac{-\frac{1}{\rho \sqrt{n}} + \ln\left(\frac{\rho \sqrt{n}}{\rho \sqrt{n} - 1}\right)}{1/n} \\
= \lim_{n \to \infty} \frac{\frac{1}{2\rho} \cdot \frac{1}{\sqrt{n}} - \frac{1}{2n} \cdot \frac{1}{\rho \sqrt{n} - 1}}{-n^{-2}} \\
= \lim_{n \to \infty} \frac{1}{2} \cdot \frac{\sqrt{n}}{\rho \cdot (\rho \sqrt{n} - 1)} \\
= \frac{1}{2\rho^2}
\]  

(59)

where the second equality above is obtained through application of L’Hôpital’s rule. The result now follows from Equations 58 and 59. □

Acknowledgement

This work was partially supported by NSF grants DMI-MES-0318657 and CMMI-0619978.

References


