RISK-AVERSE DYNAMIC PROGRAMMING FOR TREE-BASED CONTROLLED MARKOV DECISION PROCESSES

COLE FRANKLIN, DENISSE GARNICA, HUNG T. C. LE, BENJAMIN MAY

Abstract. In this paper, we generalize the work in [Rus10] to rooted trees. Specifically, we generalize a risk-averse optimization problem from Markov decision models on chains to Markov decision models on both finite and infinite trees, as well as the dynamic programming solution to solve them.

Contents

1. Introduction

Decision-making problems under uncertainty often aim to minimize not only expected costs but also the risk associated with unfavorable outcomes. In such settings, the objective is to design policies that remain robust even when outcomes deviate from their expectations. However, solving these risk minimization problems directly is challenging, as the number of possible decision sequences grows exponentially with the time horizon or the size of the decision space. To address this, dynamic programming provides a principled framework that decomposes the global optimization problem into smaller, tractable subproblems, enabling efficient computation of optimal or risk-sensitive policies without exhaustive enumeration.

*These authors contributed equally to this paper.

In [Rus10], Ruszczynski developed dynamic programming equations for minimizing risk in sequential decision processes that are appropriately modeled as Markov decision processes. Our goal in this paper is to generalize Ruszczynski's work to find dynamic programming equations for minimizing risk on decision processes which are not necessarily sequential.

In Section 2 we introduce the notation we use for trees. In section 3 we introduce our Controlled Markov Models on trees. In Section 4 we give our notions of conditional and dynamic risk measures on trees, and also develop a notion of time-consistency of dynamic risk-measures on finite trees which allows us to represent certain dynamic risk measures recursively. In Section 5 we introduce Markov risk measures on finite trees. In Section 6 we provide dynamic programming equations to minimize risk on finite trees. In Section 7 we introduce notation relevant to infinite trees and construct risk-measures on infinite trees, and introduce an infinite tree version of the problem. In Section 8 we find Dynamic Programming equations which solve the infinite tree problem. In Section 9 we provide a policy iteration method for solving the infinite tree problem. In Section 10 we provide a method that converges to the solution of an equation that is relevant to the policy iteration method. In Section 11 we discuss why the problem does not generalize well to general DAGs (Directed Acyclic Graphs).

2. Trees

Throughout, let $\mathcal{T} = (\{v_\alpha\}, \{e_{\alpha\beta}\})$ denote a directed tree with root node v_0, that is in the definition of the order below, $v_0 < v_\alpha$ for all $\alpha \neq 0$. By abuse of notation, we think of $\mathcal{T}$ as the set of vertices.

Definition 2.1 (Order)**.** We write $v_\alpha \leq v_\beta$ if there is a (directed) path from v_α to v_β. This is easily seen to be a partial order on the set of nodes of $\mathcal{T}$.

Definition 2.2 (Maximal rooted subtree)**.** For a node $v_\alpha \in \mathcal{T}$ we write $\mathcal{T}_\alpha$ for the *maximal subtree of* $\mathcal{T}$ *rooted at* v_α, consisting of all $v_\beta \geq v_\alpha$, i.e.

$$\mathcal{T}_\alpha = \{v_\beta : v_\alpha \leq v_\beta\}.$$

Definition 2.3 (Rooted subtrees)**.** For a node $v_\alpha \in \mathcal{T}$, write $\mathrm{trees}(v_\alpha)$ to denote the collection of all subtrees that are rooted at v_α.

Definition 2.4 (Terminal node of subtree)**.** Let $\mathcal{S} \in \mathrm{trees}(v_\alpha)$. We say that v_τ is a *terminal node of* $\mathcal{S}$ if there is no node $v_\beta \in \mathcal{S}$ with $v_\tau < v_\beta$. Let $\mathrm{term}(\mathcal{S})$ denote the set of terminal nodes of the subtree $\mathcal{S}$.

Definition 2.5 (Parents & children (a happy family))**.** For all $\alpha \neq 0$, let $\mathrm{pa}(v_\alpha)$ be the parent of v_α, the unique node v_β such that $e_{\beta\alpha}$ is an edge. Let $\mathrm{ch}(v_\alpha)$ denote the set of immediate children of the node v_α, the nodes v_γ such that $e_{\alpha\gamma}$ is an edge.

Definition 2.6 (Targeted path)**.** Let $v_\beta \leq v_\tau$. Then define $\mathcal{P}(v_\beta, v_\tau)$ as the unique path from v_β to v_τ. Naturally, then we have $\mathrm{pa}_\mathcal{P} \coloneqq \mathrm{pa}|_{\mathcal{P}-v_\beta}$ is a bijection $\mathcal{P}-\{v_\beta\} \to \mathcal{P}-\{v_\tau\}$. Naturally, the child function of $\mathcal{P}$ is $\mathrm{ch}_\mathcal{P} = \mathrm{pa}_\mathcal{P}{}^{-1}$

Definition 2.7 (Height of node)**.** For $v_\alpha \in \mathcal{T}$, define the height of the node $\mathrm{ht}(v_\alpha)$ as the length of $\mathcal{P}(v_0, v_\alpha)$.

Definition 2.8 (Height of tree)**.** Let $\mathrm{ht}(\mathcal{T}) = \sup\{\mathrm{ht}(v_\alpha) | v_\alpha \in \mathcal{T}\}$ be the height of $\mathcal{T}$.

Definition 2.9 (Height-truncated tree)**.** Define $\mathcal{T}_H \coloneqq \{v_\beta : \mathrm{ht}(v_\beta) \leq H\}$. We abuse the subscript notation with the maximal rooted subtree definition, and make the distinction based on which kind of subscript (H or α) is used.

Notation 2.10. We understand thing indexed α to be the same thing indexed v_α.

3. Controlled Markov models

The motivation for our definitions is that we want to be able to take into account how the probability distribution of the state of a given node changes depending on the state and action taken at its parent node. With this we can more accurately take into account the effects of different decisions (decision making policies). To make it precise, for each non terminal node, we will want to map states, or state-action tuples, to joint distributions of the states of the children nodes. However, since we will be assuming conditional independence of children nodes given the state/action of the parent, the set of joint probability distributions of the states of the children of a node v_α will be equivalent to the set of tuples of the individual child distributions.

Definition 3.1 (Controlled Markov model on tree)**.** A controlled Markov model on a tree $\mathcal{T}$ has a *state space* $(\mathscr{X}, \mathscr{B}(\mathscr{X}))$ and a *control space* $(\mathscr{U}, \mathscr{B}(\mathscr{U})$, where $\mathscr{X}, \mathscr{U}$ are Borel spaces equipped with their respective Borel $\sigma-$algebras. Let $\mathscr{P}$ denote the set of probability measures on $(\mathscr{X}, \mathscr{B}(\mathscr{X}))$.

Without a control, a *stochastic (transition) kernel* for a non-terminal node v_α is simply a measurable function $K^{(\alpha)} : \mathscr{X} \to \mathscr{P}^{\mathrm{ch}(v_\alpha)}$; but with a control, this becomes a controlled kernel as follows.

A *control set* is a measurable multifunction $U : \mathscr{X} \rightrightarrows \mathscr{U}$. We think of each $U(x)$ as the set of controls we are allowed to undertake at state x. The graph of the multifunction U is

$$\mathrm{graph}(U) = \{(x, u) \in \mathscr{X} \times \mathscr{U} : u \in U(x)\}.$$

A controlled Markov model then also has a collection of control sets $\{U^{(\alpha)} : v_\alpha \in \mathcal{T} - \mathrm{term}(\mathcal{T})\}$ that dictate which controls are available at each node and each state. Then associated with those control sets are *controlled (Markov) (transition) kernels.* For a non-terminal node v, a controlled kernel is a measurable function $Q^{(\alpha)} : \mathrm{graph}(U^{(\alpha)}) \to \mathscr{P}^{\mathrm{ch}(v_\alpha)}$, i.e., for each (x, u), its value is a $\mathrm{ch}(v_\alpha)-$indexed collection of probability measures on $(\mathscr{X}, \mathscr{B}(\mathscr{X}))$, one for each children of v_α.

Our optimization problem then involves choosing a sequence of actions to minimize the costs associated with it. In particular, a *cost function* for a non-terminal node v_α is a measurable function $c^{(\alpha)} : \mathrm{graph}(U^{(\alpha)}) \to \mathbb{R}$, and for a terminal v_α it is a measurable $c^{(\alpha)} : \mathscr{X} \to \mathbb{R}$. The last component of our controlled Markov model on tree $\mathcal{T}$ is then a collection of cost functions $\{c^{(\alpha)} : v_\alpha \in \mathcal{T}\}$:

$$\mathfrak{M} = (\mathcal{T}, \mathscr{X}, \mathscr{U}, \{U^{(\alpha)}, Q^{(\alpha)} : v_\alpha \notin \mathrm{term}(\mathcal{T})\}, \{c^{(\alpha)}\})$$

We use superscripts instead of subscripts in anticipation for the need of coordinate indexing.

Definition 3.2 (Admissible state histories)**.** We define the space $\mathscr{H}^{(\alpha)}$ of admissible state histories up to (and including) the node v_α as $\mathscr{X}^{\mathcal{P}(v_0, v_\alpha)}$.

Definition 3.3 (Policy, Markov policy)**.** A *policy* Π is a collection of measurable functions $\{\pi^{(\alpha)} : \mathscr{H}^{(\alpha)} \to \mathscr{U} \mid v_\alpha \in \mathcal{T} - \mathrm{term}(\mathcal{T})\}$ such that $\pi^{(\alpha)}(x_{\mathcal{P}(v_0,v_\alpha)}) \in U^{(\alpha)}(x_\alpha)$ for all $v_\alpha, x_{\mathcal{P}(v_0,v_\alpha)}$. Notice that a general policy at node v might refer to information from previous nodes. So a *Markov policy* is one that does not do so, i.e., a collection of measurable functions $\{\pi^{(\alpha)} : \mathscr{X} \to \mathscr{U}\}$ such that $\pi^{(\alpha)}(x) \in U^{(\alpha)}(x)$ for all $x \in \mathscr{X}$.

Remark 3.4. Just to emphasize, 2 things are Markov here. The model is Markov, in that we have transition kernels $\{Q^{(\alpha)}\}$ that only takes into account the current state and the action undertaken. The policy is also Markov, in that deciding which action to take at node v, $\pi^{(\alpha)}$, only takes into account the current state.

Consider the probability space $(\Omega, \mathscr{F}, \mathbb{P}_0)$, where $\Omega = \mathscr{X}^{\mathcal{T}}$ and $\mathscr{F}$ is the product $\sigma-$algebra. Let $\mathbb{P}_0$ be the distribution of the state $x_0 \in \mathscr{X}$. Suppose we are given a policy $\Pi = \{\pi^{(\alpha)}\}$. The Ionescu-Tulcea extension theorem states that there exists a unique probability measure $\mathbb{P}^\Pi$ on $(\Omega, \mathscr{F})$ such that for all measurable $B \subseteq \mathscr{X}$ and $h_\alpha \in \mathscr{H}^{(\alpha)}$,

$$\mathbb{P}^\Pi(x_0 \in B) = \mathbb{P}(B) \tag{3.5}$$

$$\mathbb{P}^\Pi(x_\alpha \in B | h_\alpha) = Q^{(\alpha)}(B | x_\alpha, \pi^\alpha(h_\alpha)) \tag{3.6}$$

For simplicity we will deal with the case where x_0 is fixed, the general case not being much harder.

We provide the following examples to illustrate the flexibility that graphical models, in our case tree models, offer to model decision processes.

Example 3.7. Below is an toy example of an R&D process.

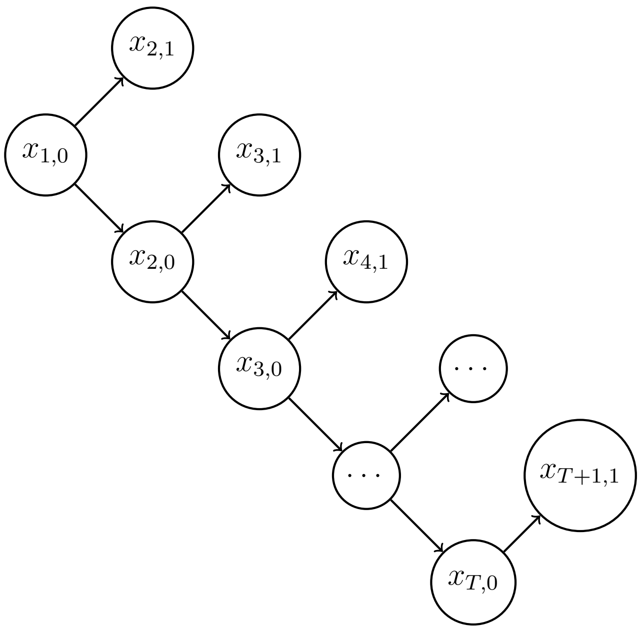

FIGURE 1. A tree decision process

The state space $\mathscr{X} = \mathbb{R}$ denotes the quality of a product, where higher is better; and points $x_{t,0}$ and $x_{t,1}$ can be thought of as at time t, where $x_{t,0}$ denotes the quality of the product in the factory, while $x_{t,1}$ denotes the quality of the shipped product. The action space $\mathscr{U} = \{0, 1\}$ where 0 denotes doing nothing, while 1 denotes investing in R&D. And we start with $x_{1,0} = 0$ deterministic.

The controlled Markov transition kernel can for now be independent of t for all t:

$$Q^{t,0}(x, 0) = Unif[x-2, x], \qquad Q^{t,0}(x, 1) = \mathcal{N}(x+2, 1)$$

$$Q^{t,1}(x, 0) = \mathcal{N}(x, 1), \qquad Q^{t,1}(x, 1) = \mathcal{N}(x+1, 1)$$

where doing nothing is penalized (say, with competitors' products improving), and investing in R&D generally improves the product. The costs associated then can be

$$c^{t,0}(x, 0) = 0, \quad c^{t,0}(x, 1) = 1, \quad c^{t,1}(x) = e^{-x}$$

where it costs nothing to do nothing, 1 to invest in R&D, and it is exponentially expensive to end up with a negative terminal node.

One can also examine the infinite version of this example, where there is no final x_T.

4. Dynamic risk measures on finite trees

Let us first consider the assessment of risk of costs on a finite tree.

Consider the probability space $(\Omega, \mathscr{F}, \mathbb{P}_0)$, where $\Omega = \mathscr{X}^{\mathcal{T}}$ and $\mathscr{F}$ is the product $\sigma-$algebra, are as above, along with a filtration $\mathscr{F}_\alpha$ such that $\mathscr{F}_\alpha \subset \mathscr{F}_\beta$ if $v_\alpha \leq v_\beta$. Let a collection of random variables Z_α adapted to $\mathscr{F}_\alpha$ be given (i.e. let $\sigma(Z_\alpha) \subset \mathscr{F}_\alpha$). Also denote $Z_\alpha \equiv Z_{v_\alpha}$. Assume that $\mathscr{F}_0 = \{\varnothing, \Omega\}$, i.e, Z_0 is deterministic. Throughout, we think of Z_α's as costs.

Fix $p \in [1, \infty]$. For every node v_α, let $\mathscr{Z}_\alpha = \mathcal{L}^p(\Omega, \mathscr{F}_\alpha, \mathbb{P}_0)$. For a subtree $\mathcal{S}$ of $\mathcal{T}$ let $\mathscr{Z}_{\mathcal{S}} = \prod_{v_\beta \in \mathcal{S}} \mathscr{Z}_\beta$. In particular, $\mathscr{Z}_{\mathcal{T}_\alpha} = \prod_{v_\beta \geq v_\alpha} \mathscr{Z}_\beta$.

Definition 4.1 (Conditional risk measure on rooted subtree)**.** A mapping $\rho_{\mathcal{T}_\alpha} : \mathscr{Z}_{\mathcal{T}_\alpha} \to \mathscr{Z}_\alpha$ is called a *conditional risk measure* if it is nondecreasing.

Definition 4.2 (Subset-indexed random variable)**.** If $\mathcal{S} \subseteq \mathcal{S}'$ are subtrees of $\mathcal{T}$ (or more generally subsets) and $Z \in \mathscr{Z}_{\mathcal{S}'}$ then we define $Z_{\mathcal{S}} \in \mathscr{Z}_{\mathcal{S}}$ by $(Z_{\mathcal{S}})_\alpha = Z_\alpha$ for $v_\alpha \in \mathcal{S}$.

Definition 4.3 (Dynamic risk measure)**.** A *dynamic risk measure* is a collection of conditional risk measures $\rho_{\mathcal{T}_\alpha}$ for all $v_\alpha \in \mathcal{T}$. Occasionally, a dynamic risk measure $\{\rho_{\mathcal{T}_\alpha}\}_{v_\alpha \in \mathcal{T}}$ will be written as $\rho_\bullet$ for ease of use.

Definition 4.4 (Time-consistency)**.** A dynamic risk measure $\{\rho_{\mathcal{T}_\alpha}\}_{v_\alpha \in \mathcal{T}}$ is said to be *time-consistent* if for any $v_\alpha \in \mathcal{T}$ and any $v_\beta > v_\alpha$, the following implication holds: If $Z, W \in \mathscr{Z}_{\mathcal{T}_\alpha}$ are such that

(1) $Z_{\mathcal{T}_\alpha - \mathcal{T}_\beta} = W_{\mathcal{T}_\alpha - \mathcal{T}_\beta}$ and,

(2)

$$\rho_{\mathcal{T}_\beta}(Z_{\mathcal{T}_\beta}) \leq \rho_{\mathcal{T}_\beta}(W_{\mathcal{T}_\beta})$$

then in fact

$$\rho_{\mathcal{T}_\alpha}(Z_{\mathcal{T}_\alpha}) \leq \rho_{\mathcal{T}_\alpha}(W_{\mathcal{T}_\alpha})$$

Proposition 4.5. *Then the following condition is equivalent to time-consistency of a dynamic risk measure* $\{\rho_{\mathcal{T}_\alpha}\}$ *: For any* $v_\alpha \in \mathcal{T}$ *and any* $\mathcal{S} \in \mathrm{trees}(v_\alpha)$*, if* $Z, W \in \mathscr{Z}_{\mathcal{T}_\alpha}$ *are such that*

(1) $Z_{\mathcal{S}} = W_{\mathcal{S}}$ *and,*

(2) for all $v_\beta \in \mathrm{term}(\mathcal{S})$

$$\rho_{\mathcal{T}_\beta}(Z_{\mathcal{T}_\beta}) \leq \rho_{\mathcal{T}_\beta}(W_{\mathcal{T}_\beta})$$

then in fact

$$\rho_{\mathcal{T}_\alpha}(Z_{\mathcal{T}_\alpha}) \leq \rho_{\mathcal{T}_\alpha}(W_{\mathcal{T}_\alpha})$$

Proof. That the condition implies time consistency is clear by setting $\mathcal{S} = (\mathcal{T}_\alpha - \mathcal{T}_\beta) \cup \{v_\beta\}$. Thus assume that $\{\rho_{\mathcal{T}_\alpha}\}$ is time-consistent and let $v_\alpha, \mathcal{S}, Z, W$ satisfying the assumption of the condition be given. Let $v_1, \ldots, v_k$ be the terminal nodes of $\mathcal{S}$. Define $Z^0 = Z$ and

$$Z^n_\beta = \begin{cases} Z^{n-1}_\beta & v_\beta \in \mathcal{T}_\alpha - \mathcal{T}_n \\ W_\beta & v_\beta \in \mathcal{T}_n \end{cases}.$$

In particular $Z^k = W$ as $Z_{\mathcal{S}} = W_{\mathcal{S}}$. By assumption,

$$\rho_{\mathcal{T}_n}(Z^{n-1}_{\mathcal{T}_n}) = \rho_{\mathcal{T}_n}(Z_{\mathcal{T}_n}) \leq \rho_{\mathcal{T}_n}(W_{\mathcal{T}_n}) = \rho_{\mathcal{T}_n}(Z^n_{\mathcal{T}_n})$$

so by time-consistency

$$\rho_{\mathcal{T}_\alpha}(Z^{n-1}) \leq \rho_{\mathcal{T}_\alpha}(Z^n).$$

Thus

$$\rho_{\mathcal{T}_\alpha}(Z) = \rho_{\mathcal{T}_\alpha}(Z^0) \leq \rho_{\mathcal{T}_\alpha}(Z^1) \leq \cdots \leq \rho_{\mathcal{T}_\alpha}(Z^k) = \rho_{\mathcal{T}_\alpha}(W).$$

□

Remark 4.6. If $\{\rho_{\mathcal{T}_\alpha}\}_{v_\alpha \in \mathcal{T}}$ is time-consistent, and if for any $v_\alpha \in \mathcal{T}, \mathcal{S} \in \text{trees}(v_\alpha)$ we have that $Z_{\mathcal{S}} = W_{\mathcal{S}}$ and for all $v_\beta \in \text{term}(S)$, $\rho_{\mathcal{T}_\beta}(Z_{\mathcal{T}_\beta}) = \rho_{\mathcal{T}_\beta}(W_{\mathcal{T}_\beta})$; then $\rho_{\mathcal{T}_\alpha}(Z_{\mathcal{T}_\alpha}) = \rho_{\mathcal{T}_\alpha}(W_{\mathcal{T}_\alpha})$.

Definition 4.7 (Family of conditional risk measures)**.** Let $\mathcal{S} \in \text{trees}(v_\alpha)$ be a subtree of $\mathcal{T}$ with root v_α. For $Z \in \mathscr{Z}_{\mathcal{S}}$ set $\hat{Z}_\beta = Z_\beta$ if $v_\beta \in \mathcal{S}$ and $\hat{Z}_\beta = 0$ otherwise, then we define a family of conditional risk measures $\rho_{\mathcal{S}} : \mathscr{Z}_{\mathcal{S}} \to \mathscr{Z}_\alpha$ by

$$\rho_{\mathcal{S}}(Z) := \rho_{\mathcal{T}_\alpha}(\hat{Z})$$

Note that if $S = \mathcal{T}_\alpha$ then $\hat{Z} = Z$ so this is well-defined. It is justified to also call $\rho_{\mathcal{S}}$ conditional risk measures, because they inherit the monotonic property from $\rho_{\mathcal{T}_\alpha}$.

Theorem 4.8. *Suppose that a dynamic risk measure $\{\rho_{\mathcal{T}_\alpha} : v_\alpha \in \mathcal{T}\}$ satisfies for all $v_\alpha \in \mathcal{T}$ and $Z \in \mathscr{Z}_{\mathcal{T}_\alpha}$ the following conditions:*

$$\rho_{\mathcal{T}_\alpha}(Z) = Z_\alpha + \rho_{\mathcal{T}_\alpha}(\hat{Z}) \tag{4.9}$$

$$\rho_{\mathcal{T}_\alpha}(0) = 0 \tag{4.10}$$

here $\hat{Z} \in \mathscr{Z}_{\mathcal{T}_\alpha}$ defined as $\hat{Z}_\beta = Z_\beta$ for $\beta > \alpha$ and $\hat{Z}_\alpha = 0$. Then time-consistency is equivalent to the following:
For all $\mathcal{S} \in \text{trees}(v_\alpha)$ subtree of $\mathcal{T}$ rooted at v_α and $Z \in \mathscr{Z}_{\mathcal{T}_\alpha}$,

$$\rho_{\mathcal{T}_\alpha}(Z) = \rho_{\mathcal{S}}(W) \tag{4.11}$$

where $W \in \mathscr{Z}_{\mathcal{S}}$ defined as $W_\beta = Z_\beta$ for $v_\beta \in \mathcal{S} - \text{term}(\mathcal{S})$ and $W_\beta = \rho_{\mathcal{T}_\beta}(Z_{\mathcal{T}_\beta})$ for $v_\beta \in \text{term}(\mathcal{S})$.

Proof. $\boxed{\Rightarrow}$ Assume that $\{\rho_{\mathcal{T}_\alpha}\}_{v_\alpha \in \mathcal{T}}$ is time-consistent. Take any $v_\alpha, \mathcal{S} \in \text{trees}(v_\alpha)$ and $Z \in \mathscr{Z}_{\mathcal{T}_\alpha}$.
Then let $\hat{W} \in \mathscr{Z}_{\mathcal{T}_\alpha}$ such that $\hat{W}_\beta = W_\beta$ for $v_\beta \in \mathcal{S}$, and $W = 0$ on $\mathcal{T}_\alpha - \mathcal{S}$. Then by Definition 4.7,

$$\rho_{\mathcal{T}_\alpha}(\hat{W}) = \rho_{\mathcal{S}}(W).$$

It remains for us to use time-consistency to show that $\rho_{\mathcal{T}_\alpha}(\hat{W}) = \rho_{\mathcal{T}_\alpha}(Z)$; we already know that $\hat{W}_\beta = W_\beta = Z_\beta$ on $\mathcal{S} - \text{term}(\mathcal{S})$, so we now just need to show that for all $v_\beta \in \text{term}(\mathcal{S})$, $\rho_{\mathcal{T}_\beta}(Z_{\mathcal{T}_\beta}) = \rho_{\mathcal{T}_\beta}(\hat{W}_{\mathcal{T}_\beta})$, so that equality holds at $\mathcal{T}_\alpha$.
Indeed, take any $v_\beta \in \text{term}(\mathcal{S})$. Then

$$\begin{aligned}
\rho_{\mathcal{T}_\beta}(\hat{W}_{\mathcal{T}_\beta}) &= \rho_{\mathcal{T}_\beta}(\hat{W}_\beta, 0, \ldots, 0) \\
&= \hat{W}_\beta + \rho_{\mathcal{T}_\beta}(0) && \text{(by (4.9))} \\
&= W_\beta = \rho_{\mathcal{T}_\beta}(Z_{\mathcal{T}_\beta}) && \text{(by (4.10))}
\end{aligned}$$

This is true for any $v_\alpha, \mathcal{S} \in \text{trees}(v_\alpha), v_\beta \in \text{term}(\mathcal{S})$. By Remark 4.6, $\rho_{\mathcal{T}_\alpha}(\hat{W}) = \rho_{\mathcal{T}_\alpha}(Z)$.
□

$\boxed{\Leftarrow}$ Let $v_\alpha \in \mathcal{T}, \mathcal{S} \in \text{trees}(v_\alpha)$ and $Z, W \in \mathscr{Z}_{\mathcal{T}_\alpha}$ satisfy the assumptions of Definition 4.4.

Define $\hat{Z} \in \mathscr{Z}_{\mathcal{S}}$ as $\hat{Z}_\beta = Z_\beta$ for $v_\beta \in \mathcal{S} - \mathrm{term}(S)$ and $\hat{Z}_\beta = \rho_{\mathcal{T}_\beta}(Z_{\mathcal{T}_\beta})$ for $v_\beta \in \mathrm{term}(S)$. Define $\hat{W} \in \mathscr{Z}_{\mathcal{S}}$ similarly for W. Then

$$\rho_{\mathcal{T}_\alpha}(Z) = \rho_{\mathcal{S}}(\hat{Z}), \quad \rho_{\mathcal{T}_\alpha}(W) = \rho_{\mathcal{S}}(\hat{W})$$

by (4.11). But the assumptions of Definition 4.4 imply that $\hat{Z} \leq \hat{W}$, and $\rho_{\mathcal{S}}$ is monotonic, so

$$\rho_{\mathcal{S}}(\hat{Z}) \leq \rho_{\mathcal{S}}(\hat{W}) \Rightarrow \rho_{\mathcal{T}_\alpha}(Z) \leq \rho_{\mathcal{T}_\alpha}(W)$$

This is true for all $v_\alpha \in \mathcal{T}, S \in \mathrm{trees}(v_\alpha)$, and any $Z, W \in \mathscr{Z}_{\mathcal{T}_\alpha}$ that satisfy the assumptions, so indeed $\{\rho_{\mathcal{T}_\alpha}\}_{v_\alpha \in \mathcal{T}}$ is time-consistent. □

Definition 4.12 (One step conditional risk measure for node)**.** The one step conditional risk measure for node v_α is $\rho_\alpha : \prod_{v_\beta \in \mathrm{ch}(v_\alpha)} \mathscr{Z}_\beta \to \mathscr{Z}_\alpha$, defined as

$$\rho_\alpha(Z) \coloneqq \rho_{\mathcal{S}}(0, Z)$$

where $\mathcal{S} = \{v_\alpha\} \cup \mathrm{ch}(v_\alpha)$.

We can then compute $\rho_{\mathcal{T}_\alpha}$ recursively using one step conditional risk measures as follows:

Theorem 4.13. *Let $J[Z]_\beta = Z_\beta$ for $v_\beta \in \mathrm{term}(\mathcal{T})$. Then recursively define $J[Z]_\alpha = Z_\alpha + \rho_\alpha(J[Z]_{\mathrm{ch}(v_\alpha)})$. If we have the assumptions of Theorem 4.8, we then get that for time-consistent $\rho_\bullet$, for all $v_\alpha \in \mathcal{T}$, $J[Z]_\alpha = \rho_{\mathcal{T}_\alpha}(Z)$*

Proof. The equality is immediate for terminal v_β. Then assuming inductively that $J[Z]_\beta = \rho_{\mathcal{T}_\beta}(Z)$ for all children v_β of v_α, for $\mathcal{S} = \{v_\alpha\} \cup \mathrm{ch}(v_\alpha)$ we have

$$\begin{aligned}
\rho_{\mathcal{T}_\alpha}(Z) &= \rho_{\mathcal{S}}(W) && (W \text{ as defined in Theorem 4.8}) \\
&= \rho_{\mathcal{S}}(Z_\alpha, J[Z]_{\mathrm{ch}(v_\alpha)}) && (\text{Induction}) \\
&= Z_\alpha + \rho_{\mathcal{S}}(0, J[Z]_{\mathrm{ch}(v_\alpha)}) && (\text{Translation, (4.9)}) \\
&= Z_\alpha + \rho_\alpha(J[Z]_{\mathrm{ch}(v_\alpha)}) && (\text{Definition 4.12}) \\
&= J[Z]_\alpha
\end{aligned}$$

□

As in [Rus10], we will henceforth assume that the one step measures ρ_α satisfy the following conditions:

(1) (Convexity). For any $\lambda \in (0, 1)$ and $Z, W \in \prod_{v_\beta \in \mathrm{ch}(v_\alpha)} \mathscr{Z}_\beta$,

$$\rho_\alpha(\lambda Z + (1 - \lambda)W) \leq \lambda \rho_\alpha(Z) + (1 - \lambda)\rho_\alpha(W)$$

(2) (Cash invariance). For any $Z \in \mathscr{Z}_\alpha^{\mathrm{ch}(v_\alpha)}$ and $W \in \prod_{v_\beta \in \mathrm{ch}(v_\alpha)} \mathscr{Z}_\beta$,

$$\rho_\alpha(Z + W) = \sum_{\beta \in \mathrm{ch}(v_\alpha)} Z_\beta + \rho_\alpha(W)$$

(3) (Positive homogeneity). For any $\lambda \geq 0$ and $Z \in \prod_{v_\beta \in \mathrm{ch}(v_\alpha)} \mathscr{Z}_\beta$,

$$\rho_\alpha(\lambda Z) = \lambda \rho_\alpha(Z).$$

We note that it follows from Definition 4.1 that each ρ_α is non-decreasing. In fact, with the above properties, we can demonstrate a stronger monotonicity property, with a condition looser than the element-wise condition of Definition 4.1.

Definition 4.14 (Tree splitting)**.** For fixed $v_\alpha \in \mathcal{T}$, a *splitting* on the tree $\mathcal{T}_\alpha$ is a mapping $\zeta(\cdot,\cdot)$ such that for every $v_\beta \in \mathcal{T}_\alpha$, $\zeta(v_\beta, \cdot) : \mathscr{Z}_\beta \to \mathscr{Z}_\beta^{\mathrm{ch}(v_\beta)}$ such that

$$\sum_{v_\lambda \in \mathrm{ch}(v_\beta)} \zeta_{v_\lambda}(v_\beta, Y) = Y, \tag{4.15}$$

where $\zeta_v(\cdot)$ is the $v-$coordinate of $\zeta(\cdot)$, It is to be understood that $\zeta(v_\beta, Y)$ "splits" $Y \in \mathscr{Z}_\beta$ along the children of v_β in a sum-preserving way.

Theorem 4.16. *Let $\rho_\bullet$ be a time-invariant dynamic risk measure satisfying Cash invariance as well as the assumptions of Theorem 4.8. Assume that there is $v_\alpha \in \mathcal{T}$ and $Z, W \in \mathscr{Z}_{\mathcal{T}_\alpha}$ where there exist splittings $\zeta^{(1)}, \zeta^{(2)}$ on the tree $\mathcal{T}_\alpha$ such that for every $v_\tau \in \mathrm{term}(\mathcal{T}_\alpha)$ the following inequality holds*

$$\begin{aligned} &Z_\tau + \zeta^{(1)}_{v_\tau}(\mathrm{pa}_{\mathcal{P}}(v_\tau), Z_{\mathrm{pa}_{\mathcal{P}}(v_\tau)} + \zeta^{(1)}_{\mathrm{pa}_{\mathcal{P}}(v_\tau)}(\mathrm{pa}^2_{\mathcal{P}}(v_\tau), \ldots, Z_{\mathrm{ch}_{\mathcal{P}}(v_\alpha)} + \zeta^{(1)}_{\mathrm{ch}_{\mathcal{P}}(v_\alpha)}(v_\alpha, Z_\alpha)\ldots)) \\ &\le W_\tau + \zeta^{(2)}_{v_\tau}(\mathrm{pa}_{\mathcal{P}}(v_\tau), W_{\mathrm{pa}_{\mathcal{P}}(v_\tau)} + \zeta^{(2)}_{\mathrm{pa}_{\mathcal{P}}(v_\tau)}(\mathrm{pa}^2_{\mathcal{P}}(v_\tau), \ldots, W_{\mathrm{ch}_{\mathcal{P}}(v_\alpha)} + \zeta^{(2)}_{\mathrm{ch}_{\mathcal{P}}(v_\alpha)}(v_\alpha, W_\alpha)\ldots)). \end{aligned} \tag{4.17}$$

where we write $\mathcal{P} := \mathcal{P}(v_\alpha, v_\tau)$. Then

$$\rho_{\mathcal{T}_\alpha}(Z) \le \rho_{\mathcal{T}_\alpha}(W).$$

Remark 4.18. Notice that in the above inequality, our notation contains extra information for sake of interpretability: $\zeta_v(\mathrm{pa}_{\mathcal{P}}(v), Y)$ indicates which portion of Y is split from $\mathrm{pa}(v)$ to v. But only $\zeta(\mathrm{pa}_{\mathcal{P}}(v), \cdot)$ has a $v-$coordinate, so if we omit the first argument of ζ_v and infer it as $\mathrm{pa}_{\mathcal{P}}(v)$, then (4.17) can be written more compactly as

$$\begin{aligned} &Z_\tau + \zeta^{(1)}_{v_\tau}(Z_{\mathrm{pa}_{\mathcal{P}}(v_\tau)} + \zeta^{(1)}_{\mathrm{pa}_{\mathcal{P}}(v_\tau)}(\ldots, Z_{\mathrm{ch}_{\mathcal{P}}(v_\alpha)} + \zeta^{(1)}_{\mathrm{ch}_{\mathcal{P}}(v_\alpha)}(Z_\alpha)\ldots)) \\ &\qquad \le W_\tau + \zeta^{(2)}_{v_\tau}(W_{\mathrm{pa}_{\mathcal{P}}(v_\tau)} + \zeta^{(2)}_{\mathrm{pa}_{\mathcal{P}}(v_\tau)}(\ldots, W_{\mathrm{ch}_{\mathcal{P}}(v_\alpha)} + \zeta^{(2)}_{\mathrm{ch}_{\mathcal{P}}(v_\alpha)}(W_\alpha)\ldots)). \end{aligned} \tag{4.19}$$

Remark 4.20. Furthermore, when one finds/inspects a splitting, say, $\zeta^{(1)}$ on $\mathcal{T}_\alpha$, the only thing that matters for the splitting at the root $\zeta^{(1)}(v_\alpha, \cdot)$ is the value of $\zeta^{(1)}(v_\alpha, Z_\alpha)$. It does not matter what $\zeta^{(1)}(v_\alpha, Z)$ is for $Z \in \mathscr{Z}_\alpha - Z_\alpha$, and we can set it to be anything, say, $(Z, 0, \cdots)$, to complete the definition of $\zeta^{(1)}(v_\alpha, \cdot)$. Similarly, for each non-initial $v_\beta \in \mathcal{T}_\alpha$, the only thing that matters for $\zeta^{(1)}(v_\beta, \cdot)$ is the value of $\zeta^{(1)}(v_\beta, Z_\beta + \zeta^{(1)}_{v_\beta}(\cdots))$. So a splitting on the tree $\mathcal{T}_\alpha$ is essentially constructed from first splitting the root v_α onwards, then at each $v_\beta \in \mathrm{ch}(v_\alpha)$, split $Z_\beta + \zeta^{(1)}_{v_\beta}(v_\alpha, Z_\alpha)$ and so on.

Proof. We will proceed by induction on the height h of the rooted tree. The proposition holds trivially for the base case of $h = 0$ of $\mathcal{T}_\alpha = (\{v_\alpha\}, \varnothing)$: if $Z_\alpha \le W_\alpha$ then by monotonicity, $\rho_{\mathcal{T}_\alpha}(Z_\alpha) = \rho_\alpha(Z_\alpha) \le \rho_\alpha(W_\alpha) = \rho_{\mathcal{T}_\alpha}(W_\alpha)$.

Assume that the proposition holds for all trees of height $h \le N$. Take v_α with $\mathcal{T}_\alpha$ of height $h = N + 1$. By assumption, there exist splittings $\zeta^{(1)}$ and $\zeta^{(2)}$ on $\mathcal{T}_\alpha$ such that (4.17) holds.

For all $v_\beta \in \mathrm{ch}(v_\alpha)$ let $X^{(\beta)} \in \mathscr{Z}_{\mathcal{T}_\beta}$ be such that $X^{(\beta)}_\beta = Z_\beta + \zeta^{(1)}_{v_\beta}(v_\alpha, Z_\alpha)$ and $X^{(\beta)} = Z$ on $T_\beta - v_\beta$. Let $Y^{(\beta)}$ be defined similarly for W with $\zeta^{(2)}$. Notice that $\zeta^{(1)}|_{\mathcal{T}_\beta}$ with $X^{(\beta)}$, and $\zeta^{(2)}|_{\mathcal{T}_\beta}$ with $Y^{(\beta)}$ still satisfy (4.17) on $\mathcal{T}_\beta$. It follows that $\rho_{\mathcal{T}_\beta}(X^{(\beta)}) \le \rho_{\mathcal{T}_\beta}(Y^{(\beta)})$ by our inductive hypothesis, since $h(\mathcal{T}_\beta) \le h(\mathcal{T}_\alpha) - 1 = N$.

Thus, the monotonicity of ρ_α gives

$$\rho_\alpha((\rho_{\mathcal{T}_\beta}(X^{(\beta)}))_{v_\beta\in \mathrm{ch}(v_\alpha)}) \leq \rho_\alpha((\rho_{\mathcal{T}_\beta}(Y^{(\beta)}))_{v_\beta\in \mathrm{ch}(v_\alpha)})$$

$$\Rightarrow \sum_{v_\beta\in \mathrm{ch}(v_\alpha)} \zeta^{(1)}_{v_\beta}(v_\alpha, Z_\alpha) + \rho_\alpha(J[Z]_{\mathrm{ch}(v_\alpha)}) \leq \sum_{v_\beta\in \mathrm{ch}(v_\alpha)} \zeta^{(2)}_{v_\beta}(v_\alpha, W_\alpha) + \rho_\alpha(J[W]_{\mathrm{ch}(v_\alpha)})$$

(Cash invariance)

$$\Rightarrow Z_\alpha + \rho_\alpha(J[Z]_{\mathrm{ch}(v_\alpha)}) \leq W_\alpha + \rho_\alpha(J[Z]_{\mathrm{ch}(v_\alpha)})$$

(Sum-preserving splitting, (4.15))

$$\Rightarrow \rho_{\mathcal{T}_\alpha}(Z) \leq \rho_{\mathcal{T}_\alpha}(W).$$

□

Remark 4.21. The criterion (4.17) is easily satisfied by $Z \leq W \in \mathscr{Z}_{\mathcal{T}_\alpha}$ by various splittings, say,

$$\zeta^{(1)}(v_\beta, Y) = \zeta^{(2)}(v_\beta, Y) = \left(\frac{Y}{|\mathrm{ch}(v_\beta)|}, \frac{Y}{|\mathrm{ch}(v_\beta)|}, \cdots\right),$$

or more generally

$$\zeta^{(1)}(v_\beta, Y) = \zeta^{(2)}(v_\beta, Y) = \lambda(v_\beta)Y + \kappa(v_\beta),$$

where $\lambda(v_\beta), \kappa(v_\beta) \in \mathbb{R}^{\mathrm{ch}(v_\beta)}$, such that $\mathbf{1}\cdot\lambda(v_\beta) = 1, \mathbf{1}\cdot\kappa(v_\beta) = 0$, which would give, at every terminal node and for every v_β, equal scalar weights for Z_β and W_β, as well as equal scalar contributions. Thus the inequality on the sums in (4.17) is easily satisfied since $Z \leq W$ element-wise.

5. Markov risk measures on finite trees

This section combines the formulations of the last 2 sections. We evaluate a Markov policy in a controlled Markov model as formulated in Section 3 through the risk assessment of Section 4 on the collection of costs associated with the policy. For a given policy $\Pi = \{\pi^{(\alpha)} : v_\alpha \in \mathcal{T} - \mathrm{term}(\mathcal{T})\}$, then for each v_α, our cost variable is $Z^\Pi_\alpha = c^{(\alpha)}(x_\alpha, \pi^{(\alpha)}(x_\alpha))$. Then the risk of the policy is

$$J(\Pi) := J[Z^\Pi]_0 = J[Z^\Pi]_{v_0} = \rho_{\mathcal{T}}(Z^\Pi)$$

where we exte the J notation, and assess the risk from the root node v_0 using a dynamic risk measure $\rho_\bullet$ generated by some one-step conditional risk measures $\{\rho_\alpha\}_{v_\alpha\in\mathcal{T}}$. Refer to Theorem 4.13 for the usage of $J[Z]$. However, similar to [Rus10],

> "We cannot expect to obtain a Markov optimal policy, if our attitude to risk may depend on the whole past of the process."

Recall that at node v_α, $\rho_\alpha : \prod_{v_\beta\in\mathrm{ch}(v_\alpha)} \mathscr{Z}_\beta \to \mathscr{Z}_\alpha$ is $\mathscr{F}_\alpha-$measurable and is therefore allowed to depend on all ancestors of v_α, i.e., we are currently allowed to evaluate risk based on what happened in all ancestors of v_α.

> "In order to overcome these difficulties, we consider a new construction of a one-step conditional measure of risk. Its arguments are measurable functions on the state space $\mathscr{X}$." ([Rus10, page 9])

So on top of the model and policy (cf. Remark 3.4), we are now also imposing the Markov property on the one-step conditional measures of risk.

Fix a probability measure $\mathbb{P}_0$ on the space $(\mathscr{X}, \mathscr{B}(\mathscr{X}))$. Write $\mathscr{B}$ for $\mathscr{B}(\mathscr{X})$. Let $\mathscr{W} = \mathcal{L}^p(\mathscr{X}, \mathscr{B}, \mathbb{P}_0), \mathscr{Y} = \mathcal{L}^q(\mathscr{X}, \mathscr{B}, \mathbb{P}_0)$ with $p, q \in [1, \infty]$ satisfying $1/p + 1/q = 1$. We think of $\mathscr{W}$ as the space of state-based costs. Then define

$$\mathscr{M} = \left\{ m \in \mathscr{Y} : \int_{\mathscr{X}} m(x)\mathbb{P}_0(\mathrm{d}x) = 1, m \geq 0 \right\}.$$

We identify an element $m \in \mathscr{M}$ to be a probability measure on $(\mathscr{X}, \mathscr{B})$ with density m with respect to $\mathbb{P}_0$. (This definition also appeared in [Rus10] with $\mathscr{V}$ instead of $\mathscr{W}$, but we change the notation to avoid confusing a cost with a node of the tree.)

Definition 5.1 (Bilinear form)**.** We define a bilinear pairing on $\mathscr{W} \times \mathscr{Y}$ as

$$\langle w, m\rangle := \int_{\mathscr{X}} w(x)m(x)\mathbb{P}_0(\mathrm{d}x) = \mathbb{E}_m(w)$$

and can be thought of as the expected value of cost w with respect to probability measure m. Then for $n \in \mathbb{N}$, the bilinear pairing on $\mathscr{W}^n \times \mathscr{Y}^n$ is

$$\langle \mathbf{w}, \mathbf{m}\rangle := \sum_{i=1}^{n} \langle \mathbf{w}_i, \mathbf{m}_i\rangle = \sum_{i=1}^{n} \int_{\mathscr{X}} \mathbf{w}_i(x)\mathbf{m}_i(x)\mathbb{P}_0(\mathrm{d}x)$$

Definition 5.2 (Risk transition mapping for node)**.** For non-terminal node $v_\alpha \in \mathcal{T}$, a measurable functional $\sigma^{(\alpha)} : \mathscr{W}^{ch(\alpha)} \times \mathscr{X} \times \mathscr{M}^{\mathrm{ch}(\alpha)} \to \mathbb{R}$ is a risk transition mapping associated with the controlled kernel $Q^{(\alpha)} : \mathrm{graph}(U^{(\alpha)}) \to \mathscr{M}^{\mathrm{ch}(\alpha)}$ for node v_α if

(1) For every $x \in \mathscr{X}$ and every $u \in U^{(\alpha)}(x)$ the functional

$$\mathbf{w} \mapsto \sigma^{(\alpha)}(\mathbf{w}, x, Q^{(\alpha)}(x, u))$$

is a coherent measure of risk on $\mathscr{W}^{\mathrm{ch}(\alpha)}$; i.e., it satisfies monotonicity, convexity, positive homogeneity and cash invariance.

(2) For every $\mathbf{w} \in \mathscr{W}^{\mathrm{ch}(\alpha)}$ and every measurable selection $\pi^{(\alpha)}(\cdot)$ of $U^{(\alpha)}(\cdot)$, the function

$$x \mapsto \sigma^{(\alpha)}(\mathbf{w}, x, Q^{(\alpha)}(x, \pi^{(\alpha)}(x)))$$

is an element of $\mathscr{W}$.

Suppose that for every $x \in \mathscr{X}$ and every $\mathbf{m} = Q^{(\alpha)}(x, u) \in \mathscr{M}^{\mathrm{ch}(\alpha)}$ for some $x, u \in U^{(\alpha)}(x)$, the mapping $\sigma^{(\alpha)}(\cdot, x, \mathbf{m})$ is lower semi-continuous with respect to $\mathbf{w}$. Then $\sigma^{(\alpha)}(\cdot, x, \mathbf{m})$, combined with item 1 of Definition 5.2, is a lower-semicontinuous coherent risk measure on $\mathscr{W}^{\mathrm{ch}(v_\alpha)}$. Therefore by [RS06, Theorem 2.2], there exists a closed convex set $\mathscr{A}^{(\alpha)}(x, \mathbf{m}) \subset \mathscr{M}^{\mathrm{ch}(\alpha)}$ such that we have for all $\mathbf{w} \in \mathscr{W}^{\mathrm{ch}(\alpha)}$

$$\sigma^{(\alpha)}(\mathbf{w}, x, \mathbf{m}) = \sup_{\boldsymbol{\mu} \in \mathscr{A}(x,\mathbf{m})} \langle \mathbf{w}, \boldsymbol{\mu}\rangle = \sup_{\boldsymbol{\mu} \in \mathscr{A}(x,\mathbf{m})} \sum_{v_\beta \in \mathrm{ch}(\alpha)} \langle \mathbf{w}_\beta, \boldsymbol{\mu}_\beta\rangle. \tag{5.3}$$

As in [Rus10], if $\sigma^{(\alpha)}(\cdot, x, \mathbf{m})$ is continuous then $\mathscr{A}^{(\alpha)}(x, \mathbf{m})$ is bounded. And if $p < \infty$, monotonicity and convexity imply that $\sigma^{(\alpha)}(\cdot, x, \mathbf{m})$ is continuous [RS06, Proposition 3.1]. Therefore, if $p < \infty$, then $\mathscr{A}^{(\alpha)}(x, \mathbf{m})$ is a closed and bounded subset of $\mathscr{Y}^{\mathrm{ch}(v_\alpha)} = (\mathscr{W}^{\mathrm{ch}(v_\alpha)})^*$, and is therefore weak* compact by Banach-Alaoglu. The sup in (5.3) can then be replaced by a max.

Definition 5.4 (Tree Markov risk measure for node)**.** For non-terminal node $v_\alpha \in \mathcal{T}$, a one step conditional risk measure $\rho_\alpha : \prod_{v_\beta \in \mathrm{ch}(v_\alpha)} \mathscr{Z}_\beta \to \mathscr{Z}_\alpha$ is a *tree Markov risk measure for node* v_α (for the controlled tree Markov model $\{x_\beta\}$) if there exists a risk transition mapping $\sigma^{(\alpha)}$ such that for all $\mathbf{w} \in \mathscr{W}^{\mathrm{ch}(v_\alpha)}$ and all measurable selection $\pi^{(\alpha)} : \mathscr{X} \to \mathscr{U}$ such that $\pi^{(\alpha)}(x_\alpha) \in U^{(\alpha)}(x_\alpha)$ we have

$$\rho_\alpha(\mathbf{w}(\mathbf{x}_{\mathrm{ch}(v_\alpha)})) = \sigma^{(\alpha)}(\mathbf{w}, x_\alpha, Q^{(\alpha)}(x_\alpha, \pi^{(\alpha)}(x_\alpha)))$$

where $\mathbf{x}_{\mathrm{ch}(v_\alpha)}$ is understood as $(x_\beta)_{v_\beta \in \mathrm{ch}(v_\alpha)}$ and $\mathbf{w}(\mathbf{x}_{\mathrm{ch}(v_\alpha)})$ is understood as $(\mathbf{w}_{v_\beta}(\mathbf{x}_{v_\beta}))_{v_\beta \in \mathrm{ch}(v_\alpha)}$. We also make it a point that the selection $\pi^{(\alpha)}$ affects the left hand side by affecting the probability law for $\mathbf{x}_{\mathrm{ch}(v_\alpha)}$.

From the representation in (5.3), we get that for a tree Markov risk measure ρ_α, there exists a multifunction $\mathscr{A}^{(\alpha)} : \mathscr{X} \times \mathscr{M}^{\mathrm{ch}(v_\alpha)} \rightrightarrows \mathscr{M}^{\mathrm{ch}(v_\alpha)}$ such that

$$\rho_\alpha(\mathbf{w}(\mathbf{x}_{\mathrm{ch}(v_\alpha)})) = \sup_{\boldsymbol{\mu} \in \mathscr{A}^{(\alpha)}(x_\alpha, Q^{(\alpha)}(x_\alpha, u_\alpha))} \langle \mathbf{w}, \boldsymbol{\mu} \rangle$$

where each $\mathscr{A}^{(\alpha)}(x_\alpha, Q^{(\alpha)}(x_\alpha, u_\alpha))$ is a closed, convex set.
For our purposes, we only care when the third argument of $\sigma^{(\alpha)}(\mathbf{w}, x, \mathbf{m})$ is some $Q^{(\alpha)}(x, u)$, so we can further abstract the action u out of our risk transition mapping. Define $\psi^{(\alpha)} : \mathscr{W}^{\mathrm{ch}(v_\alpha)} \times \mathrm{graph}(U^{(\alpha)}) \to \mathbb{R}$ as

$$\begin{aligned} \psi^{(\alpha)}(\mathbf{w}, x_\alpha, u_\alpha) &\coloneqq \sigma^{(\alpha)}(\mathbf{w}, x_\alpha, Q^{(\alpha)}(x_\alpha, u_\alpha)) \\ &= \sup_{\boldsymbol{\mu} \in \mathscr{A}^{(\alpha)}(x_\alpha, Q^{(\alpha)}(x_\alpha, u_\alpha))} \langle \mathbf{w}, \boldsymbol{\mu} \rangle \end{aligned}$$

Theorem 5.5. *Suppose $Q^{(\alpha)}$ is continuous. If $\mathscr{A}^{(\alpha)}$ is lower semincontinuous, then for every $\mathbf{w} \in \mathscr{W}^{\mathrm{ch}(\alpha)}$ the function $(x_\alpha, u_\alpha) \to \psi^{(\alpha)}(\mathbf{w}, x_\alpha, u_\alpha)$ is lower semicontinuous. If $p \in [1, \infty)$ and the multifunction $\mathscr{A}^{(\alpha)}$ is upper semicontinuous, then for every $\mathbf{w} \in \mathscr{W}^{\mathrm{ch}(\alpha)}$ the function $(x_\alpha, u_\alpha) \to \psi^{(\alpha)}(\mathbf{w}, x_\alpha, \cdot)$ is upper semicontinuous.*

Proof. This proposition (and its corollary) is nearly identical to [RS06, Proposition 1]. □

Corollary 5.6. *If $Q^{(\alpha)}(x_\alpha, \cdot)$ is continuous and $\mathscr{A}^{(\alpha)}(x_\alpha, \cdot)$ is lower semicontinuous, then the function $\psi^{(\alpha)}(v_\alpha, x_\alpha, \cdot)$ is lower semicontinuous*

6. Finite tree problem

Fix a finite tree $\mathcal{T}$ and let $(\mathscr{X}, \mathscr{U}, U^{(\alpha)}, Q^{(\alpha)}, c^{(\alpha)})$ be a controlled Markov model on $\mathcal{T}$. Consider the problem

$$\inf_{\Pi} J(\Pi, x_0) \tag{6.1}$$

where, for a policy $\Pi = (\pi^{(\alpha)})_{v_\alpha \in \mathcal{T}_*}$, where $\mathcal{T}_* \coloneqq \mathcal{T} - \mathrm{term}(\mathcal{T})$ is the set of nodes at which we can take actions, the node-cost is $Z^{\Pi}_\alpha = c^{(\alpha)}(x_\alpha, \pi^{(\alpha)}(x_\alpha))$ and $J(\Pi, x_0) = J[Z^{\Pi}]_{v_0}$ in the notation of Theorem 4.13, where each one step tree conditional risk measure ρ_α is a **tree Markov risk measure**, as defined above. Only in this case, can we hope to get a Markov policy, as demonstrated by the following theorem.

Theorem 6.2. *([Rus10, Theorem 2]) Assume that*

1. *For all non-terminal v_α, $Q^{(\alpha)}(\cdot, \cdot)$ is strongly continuous.*
2. *For all non-terminal v_α, ρ_α is Markov and the multifunction $\mathscr{A}^{(\alpha)}(\cdot, \cdot) := \partial\sigma^{(\alpha)}(0, \cdot, \cdot)$ (∂ denotes the subgradient) are lower semicontinuous.*
3. *For all non-terminal v_α and all measurable selection $\pi^{(\alpha)}(\cdot)$ in $U^{(\alpha)}(\cdot)$, the function $x \mapsto c^{(\alpha)}(x, \pi^{(\alpha)}(x))$ is in $\mathscr{W}$. And for terminal v_τ, $c^{(\tau)} \in \mathscr{W}$ are measurable and bounded.*
4. *For all non-terminal v_α, the functions $c^{(\alpha)}(\cdot, \cdot)$ are bounded and lower semicontinuous.*
5. *For all non-terminal v_α, the multifunction $U^{(\alpha)}(\cdot)$ measurable and takes compact value.*

Then

(1) (6.1) *has an optimal solution with optimal value* $w_0(x_0)$*, where* w_0 *is the measurable solution of the following dynamic programming equations. For* $v_\tau \in \operatorname{term}(\mathcal{T}), x_\tau \in \mathscr{X}$*,*

$$w_\tau(x_\tau) = c^{(\tau)}(x_\tau), \tag{6.3}$$

and for non-terminal v_α*,* $x_\alpha \in \mathscr{X}$*,*

$$w_\alpha(x_\alpha) = \min_{u_\alpha \in U^{(\alpha)}(x_\alpha)} \left\{ c^{(\alpha)}(x_\alpha, u_\alpha) + \sigma^{(\alpha)}(w_{\operatorname{ch}(v_\alpha)}, x_\alpha, Q^{(\alpha)}(x_\alpha, u_\alpha)) \right\} \tag{6.4}$$

where

$$\sigma^{(\alpha)}(\mathbf{w}, x, Q^{(\alpha)}(x,u)) = \sup_{\boldsymbol{\mu} \in \mathscr{A}^{(\alpha)}(x, Q^{(\alpha)}(x,u))} \langle \mathbf{w}, \boldsymbol{\mu} \rangle. \tag{6.5}$$

(2) *An optimal Markov policy* $\hat{\Pi} = (\hat{\pi}^{(\alpha)})_{v_\alpha \in \mathcal{T}_*}$ *exists and satisfies the equations*

$$\hat{\pi}^{(\alpha)}(x) \in \operatorname*{arg\,min}_{u \in U^{(\alpha)}(x)} \left\{ c^{(\alpha)}(x,u) + \sigma^{(\alpha)}(w_{\operatorname{ch}(v_\alpha)}, x, Q^{(\alpha)}(x,u)) \right\} \tag{6.6}$$

for all $x \in \mathscr{X}$ *and non-terminal* v_α*.*

(3) *Conversely, a measurable solution to the above four sets of equations defines an optimal Markov policy* $\hat{\Pi}$*.*

Proof. We begin by proving some technical lemmas.

Lemma 6.7. *Let* $w : \mathscr{X}^{\operatorname{ch}(v_\alpha)} \to \mathbb{R}$ *be measurable and bounded. Then for all* v_α*, the functions*

$$\psi^{(\alpha)}(x) = \inf_{u \in U^{(\alpha)}(x)} \left\{ c^{(\alpha)}(x,u) + \sigma^{(\alpha)}(w, x, Q^{(\alpha)}(x,u)) \right\}$$

are measurable and bounded. Additionally, measurable selectors $\pi^{(\alpha)}$ *of* $U^{(\alpha)}$ *exists, such that*

$$\psi^{(\alpha)}(x) = c^{(\alpha)}(x, \pi^{(\alpha)}(x)) + \sigma^{(\alpha)}(w, x, Q^{(\alpha)}(x, \pi^{(\alpha)}(x))).$$

Proof. This follows by the same proof as in [Rus14, Proposition 2]. □

We prove by induction on the height H of the tree $\mathcal{T}$.
If $H = 1$, then $\mathcal{T} = \{v_0\} \cup \operatorname{ch}(v_0)$ and $\mathcal{T}_* = \{v_0\}$. Then clearly, with $w^{(\tau)} := c^{(\tau)}$ for all $v_\tau \in \operatorname{ch}(v_0)$ (as defined in (6.3)), then

$$\begin{aligned}
J(\Pi, x_0) &= c^{(0)}(x_0, \pi^{(0)}(x_0)) + \rho_0((c^{(\tau)}(x_\tau))_{v_\tau \in \operatorname{ch}(v_0)}) \\
&= c^{(0)}(x_0, \pi^{(0)}(x_0)) + \sigma^{(0)}(w_{\operatorname{ch}(v_0)}, x_0, Q^{(0)}(x_0, \pi^{(0)}(x_0))) \\
\Rightarrow \inf J(\Pi, x_0) &= \inf_{\pi^{(0)}(x_0)} c^{(0)}(x_0, \pi^{(0)}(x_0)) + \sigma^{(0)}(w_{\operatorname{ch}(v_0)}, x_0, Q^{(0)}(x_0, \pi^{(0)}(x_0))) \\
&= \inf_{u_0 \in U^{(0)}(x_0)} c^{(0)}(x_0, u_0) + \sigma^{(0)}(w_{\operatorname{ch}(v_0)}, x_0, Q^{(0)}(x_0, u_0)) \\
&= w_0(x_0)
\end{aligned}$$

By Lemma 6.7, the infimum is attained by a measurable Markov policy $\hat{\pi}^{(0)}$ as in (6.6).

Suppose the theorem holds for all tree of height $\leq H$. Let us then consider tree $\mathcal{T}$ of height $H+1$ and controlled Markov model $\mathfrak{M} = (\mathscr{X}, \mathscr{U}, U^{(\alpha)}, Q^{(\alpha)}, c^{(\alpha)})_{v_\alpha \in \mathcal{T}}$ on $\mathcal{T}$. Let $\mathcal{R} = \mathcal{T}_H - \operatorname{term}(\mathcal{T}) = \mathcal{T}_* - (\mathcal{T}_H)_*$, i.e., the nodes that have $(H+1)$-high children. Let us then define a slightly different controlled Markov model on $\mathcal{T}_H$:

$$\overline{\mathfrak{M}} = (\mathscr{X}, \mathscr{U}, U^{(\alpha)}, Q^{(\alpha)}, \overline{c}^{(\alpha)})_{v_\alpha \in \mathcal{T}_H}$$

where we essentially restrict to $\mathcal{T}_H$ and only change the cost functions as $\bar{c}^{(\alpha)}(x_\alpha) := w_\alpha(x_\alpha)$ for all $v_\alpha \in \mathcal{T}_H$. In particular, for $v_\alpha \in \mathcal{T}_H - \mathcal{R}$, the nodes that don't have $(H+1)-$high children, $\bar{c}^{(\alpha)} = w_\alpha = c^{(\alpha)}$, while for $v_\alpha \in \mathcal{R}$,

$$\begin{aligned}\bar{c}^{(\alpha)}(x_\alpha) &= w_\alpha(x_\alpha) \\ &= \inf_{u_\alpha \in U^{(\alpha)}(x_\alpha)} \left[c^{(\alpha)}(x_\alpha, u_\alpha) + \sigma^{(\alpha)}(w_{\mathrm{ch}(v_\alpha)}, x_\alpha, Q^{(\alpha)}(x, u_\alpha))\right]\end{aligned} \tag{6.8}$$

By Lemma 6.7 the inf is attained by a measurable Markov policy $\pi^{(\alpha)}$. We also note that the usage of $w_{\mathrm{ch}(v_\alpha)}$ in (6.8) uses the assumption (4) that the $\mathcal{T}-$terminal costs are in $\mathscr{W}$ (hence $w_{\mathrm{ch}(v_\alpha)} = (c^{(\tau)}(v_\tau))_{v_\tau \in \mathrm{ch}(v_\alpha)}$ allowed as first argument of $\sigma^{(\alpha)}$), and $c^{(\alpha)}(\cdot, u_\alpha) \in \mathscr{W}$ implies $\bar{c}^{(\alpha)}(\cdot) \in \mathscr{W}$ too. It is easily seen that assumption (4) is also used in all later iterations of the minimizing objective, so that $\sigma^{(\alpha)}$ is well-defined.

Due to time-consistency, we then have that

$$\min_{\Pi \text{ on } \mathcal{T}} J_{\mathcal{T}}(\Pi, x_0) = \min_{\Pi \text{ on } (\mathcal{T}_H)_*} J_{\mathcal{T}_H}(\Pi, x_0) \tag{6.9}$$

where we evaluate $J_{\mathcal{T}}$ on the original model $\mathfrak{M}$ and $J_{\mathcal{T}_H}$ on $\overline{\mathfrak{M}}$. We claim that the RHS is exactly $w_0(x_0)$. Indeed this is true. Define $\bar{w}$ analogously for $\overline{\mathfrak{M}}$. Then by induction hypothesis, the RHS has an optimal value $\overline{w}_0(x_0)$ that satisfies the dynamic programming equations as above. But these dynamic programming equations start with costs $\bar{c}^{(\alpha)}(x_\alpha) = w_\alpha(x_\alpha)$ for all $\mathcal{T}_H$ and then progress in the same way as they do on $\mathcal{T}$ back to v_0, so by an easy induction, $\overline{w}_\alpha = w_\alpha$ for all v_α, in particular, $\overline{w}_0 = w_0$, and we are done for (1) and (2). Following the proof, (3) is also clear.

Remark 6.10. Considering only when $\mathscr{X}, \mathscr{U}$ is finite, and when all the one-steps are identical, the method of solving for w_0 using (6.3)- (6.6) is $O(|\mathscr{X}||\mathscr{U}||\mathcal{T}|)$ in units of evaluation of the one-step risk measure in the tree.

□

7. Discounted risk measures on infinite-height trees

In this section, we investigate $\mathcal{T}$ of infinite height.

Definition 7.1. An infinite tree $\mathcal{T}$ is *pseudo-repeating* with respect to a finite-height sub tree $\mathcal{T}'$ containing v_0, if that for all $v_\alpha \in \mathrm{term}(\mathcal{T}') - \mathrm{term}(\mathcal{T})$ we have an isomorphism $\psi_\alpha : \mathcal{T}_\alpha \to \mathcal{T}$ as directed graphs.
Additionally, we consider $(\mathscr{X}, \mathscr{U}, U^{(\alpha)}, Q^{(\alpha)}, c^{(\alpha)})$ repeatable with respect to $\mathcal{T}'$ if

$$(\mathscr{X}, \mathscr{U}, U^{(\alpha)}, Q^{(\alpha)}, c^{(\alpha)}) =^{\psi_\beta} (\mathscr{X}, \mathscr{U}, U^{(\alpha)}|_{\mathcal{T}_\beta}, Q^{(\alpha)}|_{\mathcal{T}_\beta}, c^{(\alpha)}|_{\mathcal{T}_\beta})$$

for all $v_\beta \in \mathrm{term}(\mathcal{T}') - \mathrm{term}(\mathcal{T})$, where by $=^{\psi_\beta}$ we mean that $U^{(\alpha)} = U^{(\psi_\beta(\alpha))}$ and likewise for $Q^{(\alpha)}$ and $c^{(\alpha)}$.

Definition 7.2 (Tree equivalence relation)**.** Additionally, we get an equivalence relation $\sim_{\mathcal{T}'}$ as the equivalence relation generated by identifying elements by the $\psi_\alpha, v_\alpha \in \mathrm{term}(\mathcal{T}') - \mathrm{term}(\mathcal{T})$ from Definition 7.1

Warning: $\sim_{\mathcal{T}'}$ depends on the choice of isomorphism ψ_α, but any two choices are related by an automorphism of $\mathcal{T}$.

Definition 7.3 (Repeating tree)**.** If $\mathcal{T}$ is psuedo-repeating with respect to $\mathcal{T}'$ and furthermore $\mathcal{T}' - (\mathrm{term}(\mathcal{T}') - \mathrm{term}(\mathcal{T}))$ form a system of representatives for $\sim_{\mathcal{T}'}$, then we say that $\mathcal{T}$ is *repeating* with respect to $\mathcal{T}'$.

Remark 7.4. When $|\mathcal{T}'| \geq 2$, $\mathcal{T}$ is repeating with respect to $\mathcal{T}'$ if and only if $\mathcal{T}$ is pseudo-repeating with respect to $\mathcal{T}'$ and every node in $\mathcal{T}$ is either in $\mathcal{T}'$, or is descended from an element of $\text{term}\mathcal{T}' - \text{term}\mathcal{T}$

Proof. Let us first prove the backwards direction. Let $v_\beta \in \mathcal{T} - \{v_0\}$, and let v_α be its most recent ancestor to be in $[v_0]$. Let ψ be an isomorphism between $\mathcal{T}$ and $\mathcal{T}_\alpha$ such that $\psi(v_1) = v_2 \Rightarrow v_2 \in [v_1]$. It follows then that v_β must be in $\psi(\mathcal{T}')$ since v_β cannot be descended from any element of $\psi(\text{term}\mathcal{T}' - \text{term}\mathcal{T})$. Thus $v_\beta \in [v_c]$ for some $v_c \in \mathcal{T}'$. If $v_c \in \text{term}\mathcal{T}' - \text{term}\mathcal{T}$ then $v_\beta \in [v_0]$. We thus get that $\mathcal{T}$ is repeating with respect to $\mathcal{T}'$.

For the other direction, let $v_\beta \in \mathcal{T}$ such that v_β is not a descendent of any node in $\text{term}\mathcal{T}' - \text{term}\mathcal{T}$, and such that v_β is itself not an element of $\text{term}\mathcal{T}' - \text{term}\mathcal{T}$. It follows then that $v_\beta \notin \mathcal{T}_\alpha$ for all $v_\alpha \in \text{term}\mathcal{T}' - \text{term}\mathcal{T}$. Thus $v_\beta \neq \psi_\alpha(v_c)$ for any $v_c \in \mathcal{T}'$, $v_\alpha \in \text{term}\mathcal{T}' - \text{term}\mathcal{T}$, and any isomorphism ψ_α between $\mathcal{T}$ and $\mathcal{T}_\alpha$. Thus, since $v_\beta \in [v_c]$ for some $v_c \in \mathcal{T}'$, we get that $v_\beta = v_c \in \mathcal{T}'$

□

Remark 7.5. More generally, we may consider $\mathcal{T}$ to be *eventually repeating* if there exists a finite-height sub tree $\mathcal{T}_{init}$ with the same root node such that, for all $v_\alpha \in \text{term}\mathcal{T}_{init} - \text{term}\mathcal{T}$ we have $\mathcal{T}_\alpha$ is $\mathcal{T}'_\alpha$−repeating for some finite-height subtree $\mathcal{T}'_\alpha$. The work done on repeating trees can easily be adapted to this scenario by mimicking Section 6.

Definition 7.6 (Self-similar)**.** If $\mathcal{T}$ is $\{v_0\} \cup \text{ch}(v_0)$−repeating, we say that $\mathcal{T}$ is self-similar.

We now consider risk measures for an infinite tree — but only on certain well-behaved ones. We first ask that each node only have finitely many children.

Definition 7.7. For a given tree $\mathcal{T}$, define

$$n(h) := \#\{v \in \mathcal{T} : \text{ht}(v) = h\}$$

as the number of nodes at height h.

We note that since a node has finitely many children, $n(h) < \infty$.

Definition 7.8 (Growth rate bound, growth rate, bounded growth rate)**.** A *growth rate bound* of a tree $\mathcal{T}$ is any $k > 0$ such that $n(h) = O(k^h)$. As $n(h) < \infty$, this is the same as there existing $C > 0$ such that $n(h) \leq Ck^h$ for all h, since $k^h > 0$. Define

$$\text{igb}(\mathcal{T}) := \inf\{k : k \text{ is a growth rate bound of } \mathcal{T}\}$$

to be the *growth rate* of $\mathcal{T}$. If $\text{igb}(\mathcal{T}) < \infty$ we say that $\mathcal{T}$ has a *bounded growth rate.* Warning: $\text{igb}(\mathcal{T})$ is generally not a growth rate bound.

Remark 7.9. When $k \neq 1$, it would be equivalent to ask that $\#\{v \in \mathcal{T} : \text{ht}(v) \leq H\} = O(k^H)$. Let $C > 0$ be such that $n(h) \leq Ck^h$. Then

$$\#\{v \in \mathcal{T} : \text{ht}(v) \leq H\} = \sum_{h=0}^{H} n(h) \leq C \sum_{h=0}^{H} k^h = C\frac{k^{H+1} - 1}{k - 1} = O(k^H).$$

Fix $p \in [1, \infty]$ and an infinite height tree $\mathcal{T}$ with bounded growth rate. Let $\{\mathscr{F}_\alpha : v_\alpha \in \mathcal{T}\}$ be a σ-algebras on Ω such that $\mathscr{F}_\alpha \subset \mathscr{F}_\beta$ if $v_\alpha \leq v_\beta$. For every node $v_\alpha \in \mathcal{T}$, let $\mathscr{Z}_\alpha = \mathcal{L}^p(\Omega, \mathscr{F}_\alpha, \mathbb{P}_0)$, and define the infinite–horizon cost space $\mathscr{Z} = \prod_{v_\alpha \in \mathcal{T}} \mathscr{Z}_\alpha$. Then $Z \in \mathscr{Z}$ is *almost surely bounded* if

$$\sup_{v_\alpha \in \mathcal{T}} \operatorname{ess\,sup} |Z_\alpha| < \infty.$$

To assess risks of Z on $\mathcal{T}$ with $K = \mathrm{igb}(\mathcal{T}) < \infty$, consider a collection of one step conditional risk mappings $\rho_\alpha : \prod_{v_\beta \in \mathrm{ch}(v_\alpha)} \mathscr{Z}_\beta \to \mathscr{Z}_\alpha$, from which we will "build" our infinite-horizon risk measure. To ensure convergence, we need an appropriate discounting factor $\gamma \in \left(0, \frac{1}{K}\right)$. Then for $H \in \mathbb{N}$, define

$$\rho_H^\gamma : \mathscr{Z}_H \to \mathbb{R}, \quad \mathscr{Z}_H := \prod_{v_\beta : \mathrm{ht}(v_\beta) \leq H} \mathscr{Z}_\beta$$

as

$$\rho_H^\gamma(Z_{\mathcal{T}_H}) := J_{\mathcal{T}_H}[\hat{Z}]_{v_0} \tag{7.10}$$

where $\hat{Z}_\alpha = \gamma^{\mathrm{ht}(v_\alpha)} Z_\alpha$ for $v_\alpha \in \mathcal{T}_H$, that is we discount Z by height, and assess $\hat{Z}$ on the tree $\mathcal{T}_H$.

For a finite tree S define

$$\rho_S^\gamma : \mathscr{Z}_S \to \mathbb{R}$$

by

$$\rho_S^\gamma(Z_S) := J_S[\hat{Z}]_{v_0} \tag{7.11}$$

where $\hat{Z}_\alpha = \gamma^{\mathrm{ht}(v_\alpha)} Z_\alpha$ for $v_\alpha \in S$.

Definition 7.12 (Infinite horizon discounted risk measure)**.** The *infinite horizon discounted risk measure* $\varrho^\gamma : \mathscr{Z} \to \mathbb{R}$ is defined as

$$\varrho^\gamma(Z) := \lim_{H \to \infty} \rho_H^\gamma(Z_{\mathcal{T}_H}). \tag{7.13}$$

Theorem 7.14. *Let $K = \mathrm{igb}(\mathcal{T})$. Then $\varrho^\gamma(Z)$ is well-defined for all $\gamma \in \left(0, \frac{1}{K}\right)$ and all almost surely bounded $Z \in \mathscr{Z}$, and satisfies*

(1) *(Convexity) For all $Z, W \in \mathscr{Z}$ and $\lambda \in [0,1]$, $\varrho^\gamma(\lambda Z + (1-\lambda)W) \leq \lambda \varrho^\gamma(Z) + (1-\lambda)\varrho^\gamma(W)$.*

(2) *(Non-decreasing) For all $Z, W \in \mathscr{Z}$, if $Z_\alpha \leq W_\alpha$ for all $v_\alpha \in \mathcal{T}$ then $\varrho^\gamma(Z) \leq \varrho^\gamma(W)$.*

(3) *(Discounted cash invariance) For all $Z \in \mathscr{Z}$, $v_\alpha \in \mathcal{T}$ and all $W_{\mathrm{ch}(v_\alpha)} \in \mathscr{Z}_\alpha^{\mathrm{ch}(v_\alpha)}$, define $X, Y \in \mathscr{Z}$ as*

$$X_\beta := \begin{cases} Z_\beta + W_\beta & v_\beta \in \mathrm{ch}(v_\alpha) \\ Z_\beta & v_\beta \notin \mathrm{ch}(v_\alpha) \end{cases}, \quad Y_\beta := \begin{cases} Z_\alpha + \gamma \sum_{\gamma \in \mathrm{ch}(v_\alpha)} W_\gamma & v_\beta = v_\alpha \\ Z_\beta & v_\beta \neq v_\alpha \end{cases}$$

then

$$\varrho^\gamma(X) = \varrho^\gamma(Y).$$

Call X and Y a cash-discounted pair at v_α if there exist such Z, W.

(4) *(Positive homogeneity) For all $Z \in \mathscr{Z}$, $\lambda \geq 0$, $\varrho^\gamma(\lambda Z) = \lambda \varrho^\gamma(Z)$.*

Definition 7.15. For $H \in \mathbb{N}$, the *height truncation* of a $Z_{\mathcal{T}_{H+1}} \in \mathscr{Z}_{\mathcal{T}_{H+1}}$ at height H is defined as $V(Z) = V(Z)_{\mathcal{T}_H} \in \mathscr{Z}_H$ where

$$V(Z)_\beta := \begin{cases} Z_\beta + \gamma \rho_\beta(Z_{\mathrm{ch}(v_\beta)}) & \mathrm{ht}(v_\beta) = H \text{ and } \mathrm{ch}(v_\beta) \neq \varnothing \\ Z_\beta & \text{otherwise} \end{cases}. \tag{7.16}$$

Lemma 7.17. *For all $H \in \mathbb{N}$ and $Z_{\mathcal{T}_{H+1}} \in \mathscr{Z}_{\mathcal{T}_{H+1}}$,*

$$\rho_{H+1}^\gamma(Z_{\mathcal{T}_{H+1}}) = \rho_H^\gamma(V(Z)_{\mathcal{T}_H}).$$

Proof. (of Lemma 7.17) This is true by definition of ρ_H^γ in (7.10) and Theorem 4.13. $\square$

Lemma 7.18 (Finite height discounted cash invariance)**.** *For height H, for all $Z \in \mathscr{Z}_{\mathcal{T}_H}$, all $v_\alpha \in \mathcal{T}$ with $h(v_\alpha) \le H-1$, and all $W_{\text{ch}(v_\alpha)} \in \mathscr{Z}_\alpha^{\text{ch}(v_\alpha)}$, define $X_{\mathcal{T}_H}, Y_{\mathcal{T}_H} \in \mathcal{T}_H$ as above then*

$$\rho_H^\gamma(X_{\mathcal{T}_H}) = \rho_H^\gamma(Y_{\mathcal{T}_H}).$$

Proof. (of Lemma 7.18) We prove the lemma by induction. For $H = 1$, the lemma (trivially) holds.

Suppose that the lemma holds for all $H \le N$, we want to show the lemma holds for $N+1$ too. Let $V(X)_{\mathcal{T}_N}, V(Y)_{\mathcal{T}_N}$ be the height truncation of X and Y. Claim that

$$\rho_{N+1}^\gamma(X) = \rho_N^\gamma(V(X)) = \rho_N^\gamma(V(Y)) = \rho_{N+1}^\gamma(Y).$$

The first and third equalities hold due to Lemma 7.17. The second equality holds due to the following reasoning.

If $h(v_\alpha) \le N-1$ in the statement of Lemma 7.18 then notice by definition, X and Y have the same values at nodes of height $N+1$, because they only modify Z on $\text{ch}(v_\alpha)$ and v_α respectively. The truncation operation only changes values at the nodes v_τ at height N that have children: $V(X)_\tau = X_\tau + \gamma\rho_\tau(X_{\text{ch}(v_\tau)}), V(Y)_\tau = Y_\tau + \gamma\rho_\tau(Y_{\text{ch}(v_\tau)})$. We said above that X and Y are equal at nodes of height $N+1$, so

$$X_{\text{ch}(v_\tau)} = Y_{\text{ch}(v_\tau)} = Z_{\text{ch}(v_\tau)} \Rightarrow V(X)_\tau = X_\tau + \rho_\tau(Z_{\text{ch}(v_\tau)}), V(Y)_\tau = Y_\tau + \rho_\tau(Z_{\text{ch}(v_\tau)}).$$

So $V(X)_\tau$ and $V(Y)_\tau$ offset X_τ and Y_τ by the same amount, and this is the only modification they make to X and Y. Therefore, since X and Y were a cash-discounted pair at v_α, $V(X)$ and $V(Y)$ are one too, and the second equality holds by induction hypothesis.

Otherwise, if $h(v_\alpha) = N$ then it is no longer true that X and Y have the same values at nodes of height $N+1$, but it is slightly different: $V(X)_\beta = V(Y)_\beta = V(Z)_\beta$ for all $v_\beta \ne v_\alpha$, since X and Y only modify Z on v_α and $\text{ch}(v_\alpha)$. So it remains to be seen if $V(X)_\alpha = V(Y)_\alpha$, and indeed they are equal:

$$\begin{aligned} V(X)_\alpha &= Z_\alpha + \gamma\rho_\alpha((Z+W)_{\text{ch}(v_\alpha)}) \\ &= Z_\alpha + \gamma\mathbf{1}\cdot W_{\text{ch}(v_\alpha)} + \gamma\rho_\alpha(Z_{\text{ch}(v_\alpha)}) = V(Y)_\alpha \end{aligned}$$

and we are done.

Therefore by induction, Lemma 7.18 holds for all H. ☐

Lemma 7.19. *For any $Z \in \mathscr{Z}$, for any height H, define $\hat{Z} \in \mathscr{Z}$ as*

$$\hat{Z}_\alpha := \begin{cases} Z_\alpha + C_\alpha & \text{ht}(v_\alpha) = H \\ Z_\alpha & \textit{otherwise} \end{cases}$$

where $C_\alpha \in \mathbb{R}$ for every $\text{ht}(v_\alpha) = H$. Then

$$\rho_H^\gamma(\hat{Z}_{\mathcal{T}_H}) = \rho_H^\gamma(Z_{\mathcal{T}_H}) + \gamma^H \sum_{\text{ht}(v_\alpha)=H} C_\alpha.$$

Proof. (of Lemma 7.19) The result follows immediately by repeatedly applying Lemma 7.18 and cash invariance. ☐

Proof. (of Theorem 7.14) We first want to show that each ρ_H^γ satisfies the conditions. By Lemma 7.18, each ρ_H^γ satisfies discounted cash invariance. We shall prove the other 3 by induction.

For $H = 1$, these are just one step conditional risk measures, so they satisfy all 3. Suppose that for all $H \le N$, all ρ_H^γ satisfy all 3, and we want to show ρ_{N+1}^γ satisfies them too.

By Lemma 7.17, we have

$$\rho^\gamma_{N+1}(Z_{\mathcal{T}_{N+1}}) = \rho^\gamma_N(V(Z)_{\mathcal{T}_N}). \tag{7.20}$$

So ρ^γ_N is convex, non-decreasing and positive homogeneous by induction hypothesis, and the height truncation $Z_{\mathcal{T}_{N+1}} \mapsto V(Z)_{\mathcal{T}_N}$ is also convex, non-decreasing and positive homogeneous owing to those properties of ρ_β. Therefore their composition ρ^γ_{N+1} is also convex, non-decreasing and positive homogeneous.

Therefore, by induction, each ρ^γ_H satisfies all 4 conditions.

We now want to show that the limit in (7.13) exists.

We have that $\gamma \in \left(0, \frac{1}{K}\right)$ and $K = \mathrm{igb}(\mathcal{T})$, so there exists growth rate bound $K_0 > K$ such that $\delta := \gamma K_0 < 1$. K_0 is a growth rate bound, so there exists B such that $n(h) \le B K_0^h$ for all h.

Fix H, let $C = \sup \operatorname{ess\,sup} |Z_\alpha| < \infty$. We want to estimate $\rho^\gamma_{H+1}(Z_{\mathcal{T}_{H+1}})$.

Define $\hat{Z}$ as

$$\hat{Z}_\tau = \begin{cases} C & \mathrm{ht}(v_\tau) = H+1 \\ Z_\tau & \text{otherwise} \end{cases}$$

then by monotonicity and Lemma 7.17,

$$\rho^\gamma_{H+1}(Z) \le \rho^\gamma_{H+1}(\hat{Z}) = \rho^\gamma_H(V(\hat{Z})).$$

For all v_α of height H, we have the height-truncation of $\hat{Z}$ at height H, $V(\hat{Z})_{\mathcal{T}_H}$, as

$$\begin{aligned} V(\hat{Z})_\alpha &= \hat{Z}_\alpha + \gamma \rho_\alpha(\hat{Z}_{\mathrm{ch}(v_\alpha)}) \\ &= Z_\alpha + \gamma \rho_\alpha((C)_{\mathrm{ch}(v_\alpha)}) \\ &= Z_\alpha + \gamma C |\mathrm{ch}(v_\alpha)| \end{aligned}$$

where we note that this is valid even in the case where $\mathrm{ch}(v_\alpha) = \varnothing$. Then applying Lemma 7.19 on Z and $V(\hat{Z})$ we get

$$\begin{aligned} \rho^\gamma_H(V(\hat{Z})) &= \rho^\gamma_H(Z) + \gamma^H \sum_{\mathrm{ht}(v_\alpha)=H} \gamma C |\mathrm{ch}(v_\alpha)| \\ &= \rho^\gamma_H(Z) + \gamma^{H+1} n(H+1) \\ &\le \rho^\gamma_H(Z) + B_0 \delta^{H+1} \end{aligned}$$

so

$$\rho^\gamma_{H+1}(Z) \le \rho^\gamma_H(Z) + B_0 \delta^{H+1}.$$

By symmetry, we can also define $\hat{Z}$ with $-C$ instead of C, to get that

$$|\rho^\gamma_{H+1}(Z) - \rho^\gamma_H(Z)| \le B_0 \delta^{H+1}$$

so the sequence in (7.13) is Cauchy with $\delta < 1$; hence the limit $\varrho^\gamma(\cdot)$ exists. Since each ρ^γ_H satisfies the 4 conditions, their limit $\varrho^\gamma(\cdot)$ satisfies them too. □

Remark 7.21. Instead of discounting by height, you could use the following alternative discounting scheme. As always, assume we have a rooted tree $\mathcal{T}$ which is repeating with respect to a subtree $\mathcal{T}'$ which induces an equivalence relation $\mathcal{T}'$. Let $\mathrm{term}\mathcal{T}' - \mathrm{term}\mathcal{T} = \{v_1, \dots, v_m\}$. If $v_\alpha \notin \mathcal{T}_{v_i}$ for $i = 1, \dots, m$ set $h'(v_\alpha) = 0$. Otherwise, there is a unique v_j, n such that $v_\alpha \in \mathcal{T}_{v_j}$ but $\psi^n_j(v_\alpha) \notin \mathcal{T}_{v_j}$. In this case we set $h'(v_\alpha) = n$. $h'(v_\alpha)$ represents the number of repeats of $\mathcal{T}'$ you need to go through to get from v_0 to v_α.

If instead of discounting by $\gamma^{\mathrm{ht}(v_\alpha)}$ you discount by $\gamma^{h'(v_\alpha)}$ for an appropriate γ [1], you get

[1] For instance $\gamma \in (0, \frac{1}{m})$ certainly works as $\#\{v_\alpha | h'(v_\alpha) \le H\} = \#(\mathrm{term}\mathcal{T}' \cap \mathrm{term}\mathcal{T}) \sum_{i=0}^H m^i$, which is $O(m'^H)$ for any $m' > m$.

similar results to those theory developed in sections 6-8, where the discounting is more natural in some situations.

8. Discounted infinite-height problem, value iteration

Definition 8.1 (Stationary, cyclical controlled Markov model on tree)**.** A controlled Markov model $(\mathscr{X}, \mathscr{U}, U^{(\alpha)}, Q^{(\alpha)}, c^{(\alpha)})$ on a tree $\mathcal{T}$ is *stationary* if $U^{(\alpha)}, Q^{(\alpha)}, c^{(\alpha)}$ do not depend on v_α. A priori, a controlled Markov model on tree can only be stationary if the tree is self-similar. A controlled Markov model on a $\mathcal{T}'$-repeating tree $\mathcal{T}$ is *cyclical* with respect to $\mathcal{T}'$, or $\mathcal{T}'$-cyclical, if $(U^{(\alpha)}, Q^{(\alpha)}, c^{(\alpha)})$ depends only on v_α's equivalence class under $\sim_{\mathcal{T}'}$. A controlled Markov model on tree can only be $\mathcal{T}'-$cyclical if the tree is $\mathcal{T}'-$repeating, so we assume so when talking about $\mathcal{T}'$-cyclical Markov models. One can think of a stationary Markov model on a self-similar $\mathcal{T}$ as a cyclical model with respect to $\mathcal{T}' = \{v_0\} \cup \mathrm{ch}(v_0)$.

Definition 8.2 (Stationary, cyclical risk measures on tree)**.** Similarly we call Markov risk measures $\rho^{(\alpha)}$ cyclical (resp. stationary) on a tree $\mathcal{T}$ that is cyclical with respect to $\mathcal{T}'$ (resp. self-similar) if the associated Markov risk mappings also depend only on the equivalence class of a node.

We tackle the optimization problem on a **repeating tree with cyclical risk measures**. Assume that we have a controlled Markov model on a tree $\mathcal{T}$ that is $\mathcal{T}'-$cyclical and has $\mathcal{T}'$-cyclical risk measures, where $\mathcal{T}'$ is some *finite subtree.* Define $\mathcal{T}_* := \mathcal{T}' - (\mathrm{term}(\mathcal{T}) - \mathrm{term}(\mathcal{T}'))$, slightly different from Section 6, to be the set of nodes on $\mathcal{T}'$ that are not initial nodes of the "next instances" of $\mathcal{T}'$, i.e., the set of all equivalence classes, where we identify each equivalence class with its representative in $\mathcal{T}'$. So our collection of control sets, controlled kernels, costs and risk measures is:

$$\{U^{(\alpha)}, Q^{(\alpha)}, c^{(\alpha)}\}_{v_\alpha \in \mathcal{T}_*}, \quad \{\rho_\alpha\}_{v_\alpha \in \mathcal{T}' - \mathrm{term}(\mathcal{T}')}$$

It is in this case where we can hope to find an optimal cyclical Markov policy $\Pi = (\pi^{(\alpha)})_{v_\alpha \in \mathcal{T}' - \mathrm{term}(\mathcal{T}')}$ where the policy at each node also only depends on which equivalence class it is in.

We've said above that the discounted, infinite-horizon risk measure is well-defined (the limit exists) for $\gamma \in \left(0, \frac{1}{\mathrm{igb}(\mathcal{T})}\right)$ — let us now illuminate more structure on $\mathrm{igb}(\mathcal{T})$ of a repeating $\mathcal{T}$.

Lemma 8.3. *If there exists K such that $n(h) = \Theta(K^h)$ then $\mathrm{igb}(\mathcal{T}) = K$ and K is a growth bound on $\mathcal{T}$.*

Proof. $n(h) = \Theta(K^h)$ so $n(h) = O(K^h)$, indeed K is a growth rate bound for $\mathcal{T}$.

Suppose for contradiction there exists $\varepsilon > 0$ such that $K - \varepsilon$ is a growth rate bound on $\mathcal{T}$. Since we also know that $K^h = O(n(h))$, we have that $K^h = O((K-\varepsilon)^h)$. That is absurd, and thus $K = \mathrm{igb}(\mathcal{T})$.

□

Proposition 8.4. *Let $\mathcal{T}$ be a $\mathcal{T}'-$repeating tree and $K = \mathrm{igb}\mathcal{T}$. Then*

$$\sum_{v_\alpha \in \mathrm{term}\mathcal{T}' - \mathrm{term}\mathcal{T}} \left(\frac{1}{K}\right)^{\mathrm{ht}(v_\alpha)} = 1.$$

Equivalently,

$$K^H = \sum_{v_\alpha \in \mathrm{term}\mathcal{T}' - \mathrm{term}\mathcal{T}} K^{H - \mathrm{ht}(v_\alpha)}$$

where $H = \mathrm{ht}(\mathcal{T}')$. Conversely, if $K > 0$ satisfies the equations above for a $\mathcal{T}'-$repeating tree, then $K = \mathrm{igb}(\mathcal{T})$.

Proof. Let $H = \mathrm{ht}(\mathcal{T}')$.
For all v_α of height h $> H$, we have that $P(v_0, v_\alpha)$ includes precisely one node in $\mathrm{term}\mathcal{T}' - \mathrm{term}\mathcal{T}$.
Thus we have the recurrence relation

$$n(h) = \sum_{v_\alpha \in \mathrm{term}\mathcal{T}' - \mathrm{term}\mathcal{T}} n(h - \mathrm{ht}(v_\alpha))$$

for all $h > H$.
Notice that

$$\sum_{v_\alpha \in \mathrm{term}\mathcal{T}' - \mathrm{term}\mathcal{T}} z^{\mathrm{ht}(v_\alpha)} = 1$$

has exactly one strictly positive solution, because the expression vanishes at $z = 0$ and increases to ∞ as $z \to \infty$. We write this solution as $\frac{1}{g}$, where necessarily $g \geq 1$ as at $z = 1$ the expression is $|\mathrm{term}\mathcal{T}' - \mathrm{term}\mathcal{T}|$. Thus g is the unique strictly positive real number satisfying

$$g^h = \sum_{v_\alpha \in \mathrm{term}\mathcal{T}' - \mathrm{term}\mathcal{T}} g^{h - \mathrm{ht}(v_\alpha)}$$

for all h
Let $P = \max\{n(i) : i \in \{0, 1, ..., H\}\}$. Let a_h follow the same recurrence relation as $n(h)$ with prescribed starting values of $a_h = g^h$ for $h = 0, 1, ..., H$. It follows that $a_h = g^h$ for all h, and that Ba_h follows that same recurrence relation for all $B \in \mathbb{R}$. Furthermore, we also notice that $Pa_h \geq P \geq n(h)$ for $h = 0, 1, , ..., H$, and thus $Pa_h \geq n(h)$ for all h. Similarly, $\frac{a_h}{g^H} \leq n(h)$ for all h. It directly follows that $n(h) = \Theta(a_h) = \Theta(g^h)$, and by Lemma 8.3 we have $K = g$.
The reverse direction is clear from the uniqueness of the positive root $1/g$. □

Remark 8.5. In the previous proof, Lemma 8.3 also gives us that $\mathrm{igb}(\mathcal{T})$ is a growth bound on $\mathcal{T}$ when $\mathcal{T}$ is repeating with respect to a sub tree.

Corollary 8.6. *The set of growth bounds on $\mathcal{T}'-$repeating $\mathcal{T}$ is*

$$\left\{ k > 0 : k^H \geq \sum_{v_\alpha \in \mathrm{term}\mathcal{T}' - \mathrm{term}\mathcal{T}} k^{H - \mathrm{ht}(v_\alpha)} \right\}.$$

Corollary 8.7. *Let $K = \mathrm{igb}\mathcal{T}$. Then for $\gamma > 0$,*

$$\sum_{v_\alpha \in \mathrm{term}\mathcal{T}' - \mathrm{term}\mathcal{T}} \gamma^{\mathrm{ht}(v_\alpha)} < 1 \Leftrightarrow \gamma < \frac{1}{K}.$$

Definition 8.8. We define the "modulus" that corresponds to γ

$$M_\gamma := \sum_{v_\alpha \in \mathrm{term}\mathcal{T}' - \mathrm{term}\mathcal{T}} \gamma^{\mathrm{ht}(v_\alpha)}$$

Then the corollary above says $\gamma \in \left(0, \frac{1}{\mathrm{igb}(\mathcal{T})}\right) \Leftrightarrow M_\gamma \in (0, 1)$.

We have remarked on the extra properties of K, and indeed the convergence of the discounted infinite-horizon risk measure enjoys extra properties as well, where we can take the limit of a different sequence of risks.

Let $K = \text{igb}(\mathcal{T})$ and $\gamma \in \left(0, \frac{1}{K}\right)$ be the discounting factor. Defining

$$\begin{aligned} J(\Pi, x_0) &= \varrho^\gamma(c^{(0)}(x_0, u_0), \{c^\alpha(x_\alpha, u_\alpha) : v_{\alpha_1} \in \text{ch}(v_0)\}, \dots) \\ &= c^{(0)}(x_0, u_0) + \rho_0(\{\gamma c^{\alpha_1}(x_{\alpha_1}, u_{\alpha_1} + \rho_{\alpha_1}(\{\gamma c^{(\alpha_2)}(x_{\alpha_2}, u_{\alpha_2}) + \rho_{\alpha_2}(\cdots) : v_{\alpha_2} \in \text{ch}(\alpha_1)\}) : \\ &\qquad v_{\alpha_1} \in \text{ch}(v_0)\}), \end{aligned}$$

the optimization problem we will be considering for a general (measurable) policy Π (not necessarily a cyclical one) is

$$\inf_\Pi J(\Pi, x_0). \tag{8.9}$$

We will show that under certain conditions, min is achieved and there indeed exists a cyclical Markov policy that is optimal.

Our presentation closely follows that of [Rus10].

For a cyclical Markov policy $\Pi = (\pi^{(\alpha)})_{v_\alpha \in \mathcal{T}' - \text{term}(\mathcal{T}')}$, define the operator $\mathscr{D}_\Pi : \mathscr{W} \to \mathscr{W}$ as

$$\mathscr{D}_\Pi w_0(x_0) = J_{\mathcal{T}'}(\Pi, x_0)$$

where the recursive cost J is taken (undiscounted) on finite $\mathcal{T}'$, with cost function

$$c_*^{(\alpha)} = \gamma^{\text{ht}(v_\alpha)} w$$

for $v_\alpha \in \mathcal{T}' - \mathcal{T}_* = \text{term}(\mathcal{T}') - \text{term}(\mathcal{T})$, and

$$c_*^{(\alpha)} = \gamma^{\text{ht}(v_\alpha)} c^{(\alpha)}$$

for $v_\alpha \in \mathcal{T}_*$. What $\mathscr{D}_\Pi$ does is to evaluate Π on the finite subtree $\mathcal{T}'$, with height-discounted costs for nodes in $\mathcal{T}_*$, and crucially with height-discounted w_0 as terminal costs for nodes in $\mathcal{T}' - \mathcal{T}_* = \text{term}(\mathcal{T}') - \text{term}(\mathcal{T})$.

Furthermore, define the operator $\mathscr{D} : \mathscr{L}_\infty(\mathscr{X}, \mathscr{B}, \mathbb{P}_0) \to \mathscr{L}_\infty(\mathscr{X}, \mathscr{B}, \mathbb{P}_0)$ as

$$\mathscr{D} w_0(x_0) = \min_\Pi \mathscr{D}_\Pi w_0(x_0) \tag{8.10}$$

When the Markov model restricted to $\mathcal{T}'$ but using c_* satisfies certain assumptions, for instance in the setting of Theorem 8.11, we get that the minimum in (8.10) is attained.

Theorem 8.11. *([Rus10, Theorem 4]) Let $(\mathscr{X}, \mathscr{U}, U^{(\alpha)}, Q^{(\alpha)}, c^{(\alpha)})$ be a controlled Markov model on a tree $\mathcal{T}$ that is $\mathcal{T}'-$cyclical. Assume that we have the following for all $v_\alpha \in \mathcal{T}_*$:*

(1) $Q^{(\alpha)}$ is strongly continuous, i.e. for all bounded measurable functions w on $\mathscr{X}$ and all convergent sequences $(x_k, u_k) \in$ graph $U^{(\alpha)}$ (say converging to (x, u))

$$\int w(y) Q^{(\alpha)}(dy|x_k, u_k) \to^{k\to\infty} \int w(y) Q(^{(\alpha)} dy|x, u).$$

(2) The conditional risk mappings $\rho^{(\alpha)}$ are Markov and for all $x_\alpha \in \mathscr{X}$ the multi-function $\mathscr{A}^{(\alpha)}(\cdot, \cdot) = \partial \sigma^{(\alpha)}(0, \cdot, \cdot)$ is lower semicontinuous, where ∂ denotes the subdifferential

(3) The functions $c^{(\alpha)}$ are bounded, non-negative, and lower semicontinuous.

(4) For all $x \in \mathscr{X}$ the set $U^{(\alpha)}(x)$ is compact and $U^{(\alpha)}$ is measurable.

Then (8.9) has an optimal solution and its optimal value $\hat{w}(x)$, as a function of the initial state $x_0 = x \in \mathscr{X}$, is the unique solution to

$$w = \mathscr{D} w. \tag{8.12}$$

Moreover, an optimal cyclical Markov policy $\hat{\Pi} = (\hat{\pi}^{(\alpha)})_{v_\alpha \in \mathcal{T}_*}$ *exists and satisfies the equation*

$$\hat{\Pi}(x) \in \arg\min_{\Pi} \mathscr{D}_\Pi w. \tag{8.13}$$

Conversely, every bounded solution of (8.12) is the optimal value of (8.9), and every measurable solution of (8.13) defines an optimal cyclical Markov policy.

Remark 8.14. Assumptions imply that for bounded $w \in \mathscr{W}^{ch(v_\alpha)}$, $(x,u) \mapsto \sigma^{(\alpha)}(w, x, Q^{(\alpha)}(x,u))$ is lower semicontinuous on graph$(U^{(\alpha)})$ as in [Rus14, Prop 1].

Remark 8.15. Assumptions also imply that the minimum in (8.10) is attained for $w \in \mathscr{Z}_\infty(\mathscr{X}, \mathscr{B}, \mathbb{P}_0)$ via Theorem 6.2, and thus $\mathscr{D}$ is well defined.

Lemma 8.16. $\mathscr{D}_\Pi, \mathscr{D}$ *are nondecreasing.*

Proof. $\mathscr{D}_\Pi w$ is the risk evaluation on $\mathcal{T}'$ with discounted costs and w as terminal costs, so its monotonicity follows from the monotonicity of risk transition mappings (Definition 5.2). It then follows for $\mathscr{D}$ by taking the min over Π. ☐

Lemma 8.17. *The operators* $\mathscr{D}_\Pi, \mathscr{D}$ *are contraction mappings on* $\mathcal{L}_\infty(\mathscr{X}, \mathscr{B}, \mathbb{P}_0)$ *with modulus* $M_\gamma < 1$.

Proof. Let $H = \mathrm{ht}(\mathcal{T}')$. Recall that $\mathcal{T}'_\alpha$ is the maximal subtree of $\mathcal{T}'$ rooted at v_α. Fix Π, and let $w, \overline{w} \in \mathcal{L}_\infty(\mathscr{X}, \mathscr{B}, \mathbb{P}_0)$. Following our notation in Theorem 4.13, let $J[Z_{\Pi,w}](x_0)$ be the recursive costs of the model on finite $\mathcal{T}'$ controlled by Π with w as terminal costs for the nodes in term$\mathcal{T}'$ − term$\mathcal{T}$, and with initial state x_0. Similarly define $J[Z_{\Pi,\overline{w}}](x_0)$ for $\overline{w}$. Then $\mathscr{D}_\Pi w(x_0) = J[Z_{\Pi,w}](x_0)_{v_0}$, and similar for $\overline{w}$. We want to show by (backward) induction on the height of the node that for all $v_\alpha \in \mathcal{T}'$,

$$\|J[Z_{\Pi,w}]_\alpha - J[Z_{\Pi,\overline{w}}]_\alpha\|_\infty \le \sum_{v_\beta \in \mathrm{term}(\mathcal{T}'_\alpha) - \mathrm{term}\mathcal{T}} \gamma^{\mathrm{ht}(v_\beta)} \|w - \overline{w}\|_\infty, \tag{8.18}$$

where we recall that $\mathcal{T}'_\alpha$ denotes the maximal subtee of $\mathcal{T}'$ that is rooted at v_α.
We first remark that this is true for all $v_\alpha \in \mathrm{term}(\mathcal{T}')$. If $v_\alpha \in \mathrm{term}(\mathcal{T}') \cap \mathrm{term}(\mathcal{T})$ then $J[Z_{\Pi,w}](x_0)_\alpha = J[Z_{\Pi,w}](x_0)_\alpha = c^{(\alpha)}$ so the LHS $= 0$ and the inequality is trivially satisfied. Otherwise, if $v_\alpha \in \mathrm{term}(\mathcal{T}') - \mathrm{term}(\mathcal{T})$, then

$$\|J[Z_{\Pi,w}](x_0)_\alpha - J[Z_{\Pi,w}](x_0)_\alpha\|_\infty = \|\gamma^{\mathrm{ht}(v_\alpha)} w - \gamma^{\mathrm{ht}(v_\alpha)} \overline{w}\| = \gamma^{\mathrm{ht}(v_\alpha)} \|w - \overline{w}\|_\infty$$

and the inequality is satisfied.
We start the backward induction with the base step with all nodes at height H. $H = \mathrm{ht}(\mathcal{T}')$ so they are terminal nodes of $\mathcal{T}'$. By the previous remark, (8.18) holds for all of them.
For the induction step, we first prove a property about the risk transition mappings $\sigma^{(\alpha)}$. We have that for any $\mathbf{d}, \overline{\mathbf{d}} \in \mathcal{L}_\infty^{\mathrm{ch}(v_\alpha)}(\mathscr{X}, \mathscr{B}, \mathbb{P}_0)$ and for any $\boldsymbol{\mu} \in \mathscr{M}^{\mathrm{ch}(v_\alpha)}$,

$$\langle \mathbf{d}, \boldsymbol{\mu} \rangle - \langle \overline{\mathbf{d}}, \boldsymbol{\mu} \rangle = \sum_{v_\beta \in \mathrm{ch}(v_\alpha)} \langle \mathbf{d}_\beta - \overline{\mathbf{d}}_\beta, \boldsymbol{\mu}_\beta \rangle \le \sum_{v_\beta \in \mathrm{ch}(v_\alpha)} \|\mathbf{d}_\beta - \overline{\mathbf{d}}_\beta\|_\infty$$

so by taking the sup over $\boldsymbol{\mu} \in \mathscr{A}^{(\alpha)}(x_\alpha, Q^{(\alpha)}(x_\alpha, \pi^{(\alpha)}(x_\alpha)))$, it follows that

$$\sigma^{(\alpha)}(\mathbf{d}, x_\alpha, Q^{(\alpha)}(x_\alpha, \pi^{(\alpha)}(x_\alpha)) \le \sigma^{(\alpha)}(\overline{\mathbf{d}}, x_\alpha, Q^{(\alpha)}(x_\alpha, \pi^{(\alpha)}(x_\alpha)) + \sum_{v_\beta \in \mathrm{ch}(v_\alpha)} \|\mathbf{d}_\beta - \overline{\mathbf{d}}_\beta\|_\infty$$

By symmetry, it follows that

$$|\sigma^{(\alpha)}(\mathbf{d}, x_\alpha, Q^{(\alpha)}(x_\alpha, \pi^{(\alpha)}(x_\alpha)) - \sigma^{(\alpha)}(\overline{\mathbf{d}}, x_\alpha, Q^{(\alpha)}(x_\alpha, \pi^{(\alpha)}(x_\alpha))| \leq \sum_{v_\beta \in \mathrm{ch}(v_\alpha)} \|\mathbf{d}_\beta - \overline{\mathbf{d}}_\beta\|_\infty.$$

Suppose that the claim (8.18) is true for all nodes of height h (some $h \geq 1$) and above. Take node v_α of height $h-1$. If $v_\alpha \in \mathrm{term}(\mathcal{T}')$ then we know that (8.18) holds as above. Otherwise, it has children at height h, so for all x_0,

$$\begin{aligned}
|J[Z_{\Pi,w}](x_0)_\alpha - J[Z_{\Pi,\overline{w}}](x_0)_\alpha| &= |\sigma^{(\alpha)}(J[Z_{\Pi,w}](x_0)_{\mathrm{ch}(v_\alpha)}, x_\alpha, Q^{(\alpha)}(x_\alpha, \pi^{(\alpha)}(x_\alpha)) \\
&\quad - \sigma^{(\alpha)}(J[Z_{\Pi,\overline{w}}](x_0)_{\mathrm{ch}(v_\alpha)}, x_\alpha, Q^{(\alpha)}(x_\alpha, \pi^{(\alpha)}(x_\alpha))| \\
&\leq \sum_{v_\kappa \in \mathrm{ch}(v_\alpha)} \|J[Z_{\Pi,w}]_\kappa - J[Z_{\Pi,\overline{w}}]_\kappa\|_\infty \\
&\leq \sum_{v_\kappa \in \mathrm{ch}(v_\alpha)} \left(\sum_{v_\beta \in \mathrm{term}(\mathcal{T}'_\kappa) - \mathrm{term}\mathcal{T}} \gamma^{\mathrm{ht}(v_\beta)} \|w - \overline{w}\|_\infty \right) \\
&\qquad \text{(By induction hypothesis)} \\
&= \sum_{v_\beta \in \mathrm{term}(\mathcal{T}'_\alpha) - \mathrm{term}\mathcal{T}} \gamma^{\mathrm{ht}(v_\beta)} \|w - \overline{w}\|_\infty
\end{aligned}$$

so indeed

$$\|J[Z_{\Pi,w}]_\alpha - J[Z_{\Pi,\overline{w}}]_\alpha\|_\infty \leq \sum_{v_\beta \in \mathrm{term}(\mathcal{T}'_\alpha) - \mathrm{term}\mathcal{T}} \gamma^{\mathrm{ht}(v_\beta)} \|w - \overline{w}\|_\infty$$

and we're done; by induction, (8.18) holds for all $v_\alpha \in \mathcal{T}'$. In particular it holds for v_0

$$\begin{aligned}
\|\mathscr{D}_\Pi w - \mathscr{D}_\Pi \overline{w}\|_\infty = \|J[Z_{\Pi,w}]_{v_0} - J[Z_{\Pi,\overline{w}}]_{v_0}\|_\infty &\leq \sum_{v_\beta \in \mathrm{term}(\mathcal{T}') - \mathrm{term}\mathcal{T}} \gamma^{\mathrm{ht}(v_\beta)} \|w - \overline{w}\|_\infty \\
&= M_\gamma \|w - \overline{w}\|_\infty
\end{aligned}$$

as required. Taking the min over Π yields the contraction inequality for $\mathscr{D}$. □

Definition 8.19. For an infinite repeating tree $\mathcal{T}$, with respect to a given equivalence relation $\sim_{\mathcal{T}'}$, for $t \geq 1$ we define $\mathcal{T}^{(\sim_{\mathcal{T}'}, t)}$ to be $\{v_0\} \cup \bigcup_{v_\alpha : h'(v_\alpha) \leq t-1} \mathrm{ch}(v_\alpha)$ where h' is as in Remark 7.21. When $\mathcal{T}'$ is clear from context this will be denoted $\mathcal{T}^{(t)}$.

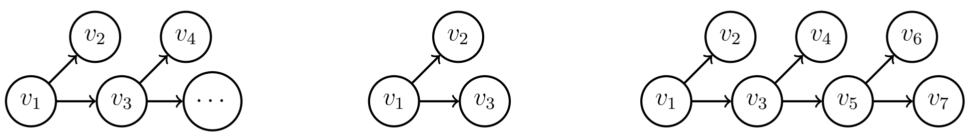

Figure 2. Example $\mathcal{T}, \mathcal{T}', \mathcal{T}^{(3)}$

Lemma 8.20. *Let $Z \geq 0, \gamma > 0$. Then the following are equivalent*

(1) $\varrho^\gamma(Z)$ is well defined.

(2) $\lim_{t\to\infty} \rho^\gamma_{\mathcal{T}^{(t)} - (\mathrm{term}\mathcal{T}^{(t)} - \mathrm{term}\mathcal{T})}(Z_{\mathcal{T}^{(t)} - (\mathrm{term}\mathcal{T}^{(t)} - \mathrm{term}\mathcal{T})})$ is well defined.

Furthermore, when they are both well defined, they are equal.

Proof. Let H be the height of $\mathcal{T}'$. Then it follows that

$$\begin{aligned}&\rho^{\gamma}_{\mathcal{T}^{(\lfloor \frac{h}{H}\rfloor)}-(\mathrm{term}\mathcal{T}^{(\lfloor \frac{h}{H}\rfloor)}-\mathrm{term}\mathcal{T})}(Z_{\mathcal{T}^{(t)}-(\mathrm{term}\mathcal{T}^{(\lfloor \frac{h}{H}\rfloor)}-\mathrm{term}\mathcal{T})})\\ &\leq \rho^{\gamma}_{h}(Z_{\mathcal{T}_h})\\ &\leq \rho^{\gamma}_{\mathcal{T}^{(h)}-(\mathrm{term}\mathcal{T}^{(h)}-\mathrm{term}\mathcal{T})}(Z_{\mathcal{T}^{(h)}-(\mathrm{term}\mathcal{T}^{(h)}-\mathrm{term}\mathcal{T})})\\ &\leq \rho^{\gamma}_{hH}(Z_{\mathcal{T}_{hH}})\end{aligned}$$

and the result follows by the Squeeze Theorem. □

Lemma 8.21. *(1) If $w \in \mathscr{L}_\infty(\mathscr{X},\mathscr{B},\mathbb{P}_0)_+$ and $w \geq \mathscr{D}w$ then $w \geq \hat{w}$ where $\hat{w} = \inf_\Pi J(\Pi, x_0)$*

(2) If $w \in \mathscr{L}_\infty(\mathscr{X},\mathscr{B},\mathbb{P}_0)$ and $w \leq \mathscr{D}w$ then $w \leq \hat{w}$

Proof. We will begin by proving (1). Since $w \geq \mathscr{D}w$, we know there exists Π such that $\mathscr{D}w = \mathscr{D}_\Pi w \Rightarrow w \geq \mathscr{D}_\Pi w$.
We then get that

$$w \geq [\mathscr{D}_\Pi]^t w$$

for $t = 1, 2, \ldots$. $[\mathscr{D}_\Pi]^t w$ represents the cost (with discounting that comes with the operator $\mathscr{D}_\Pi$) of the finite tree problem on the subtree $\mathcal{T}^{(t)}$ with the cyclical Markov policy Π restricted to $\mathcal{T}^{(t)}$, only that we change the terminal costs to be $c^{(\alpha)} = w_0$ for $v_\alpha \in \mathrm{term}\mathcal{T}^{(t)} - \mathrm{term}\mathcal{T}$, i.e., the cost sequence in this Markov process is $Z_\alpha = c(x_\alpha, \pi^{(\alpha)}(x_\alpha))$ for $v_\alpha \in \mathcal{T}^{(t)} - \mathrm{term}\mathcal{T}^{(t)}$, $Z_\alpha = c(x_\alpha)$ for $v_\alpha \in \mathrm{term}\mathcal{T}^{(t)} \cap \mathrm{term}\mathcal{T}$, and $Z_\alpha = w_0$ for $v_\alpha \in \mathrm{term}\mathcal{T}^{(t)} - \mathrm{term}\mathcal{T}$ where we omit the superscript $^{(\alpha)}$ for sake of brevity. Then

$$w \geq [\mathscr{D}_\Pi]^t w = \rho^{\gamma}_{\mathcal{T}^{(t)}}(Z_{\mathcal{T}^{(t)}}) \geq \rho^{\gamma}_{\mathcal{T}^{(t)}-(\mathrm{term}\mathcal{T}^{(t)}-\mathrm{term}\mathcal{T})}(Z_{\mathcal{T}^{(t)}-(\mathrm{term}\mathcal{T}^{(t)}-\mathrm{term}\mathcal{T})})$$

as $w \geq 0$ and ρ is monotonic. Taking the limit as $t \to \infty$ we get by Lemma 8.20 that $w(x_0) \geq J(\Pi, x_0) \geq \hat{w}(x_0)$.

Let us now prove (2). Consider an arbitrary policy (not necessarily Markov) Π with cost sequence Z where $Z_\alpha = c(x_\alpha, u_\alpha)$ on the tree $\mathcal{T}^{(t)} - (\mathrm{term}\mathcal{T}^{(t)} - \mathrm{term}\mathcal{T})$ and $Z_\alpha = w$ on $\mathrm{term}\mathcal{T}^{(t)} - \mathrm{term}\mathcal{T}$. Let Y be given by $Y_\alpha = c(x_\alpha, u_\alpha)$ on the tree $\mathcal{T}^{(t-1)} - (\mathrm{term}\mathcal{T}^{(t-1)} - \mathrm{term}\mathcal{T})$ and $Y_\alpha = w$ on $\mathrm{term}\mathcal{T}^{(t-1)} - \mathrm{term}\mathcal{T}$. Then $\rho^{\gamma}_{\mathcal{T}^{(t)}}(Z) \geq \rho^{\gamma}_{\mathcal{T}^{(t-1)}}(Y)$ since $\mathscr{D}_\Pi w \geq \mathscr{D}w \geq w$. Thus, by induction, we get that

$$w \leq \rho^{\gamma}_{\mathcal{T}^{(t)}}(Z) \tag{8.22}$$

Let $B = \operatorname{ess\,sup}_{x_0 \in \mathscr{X}} |w(x_0)| < \infty$. Then by (7.11), finite height discounted cash invariance and monotonicity, we have that

$$\begin{aligned}\rho^{\gamma}_{\mathcal{T}^{(t)}-(\mathrm{term}\mathcal{T}^{(t)}-\mathrm{term}\mathcal{T})}(Z_{\mathcal{T}^{(t)}-(\mathrm{term}\mathcal{T}^{(t)}-\mathrm{term}\mathcal{T})}) &\geq \rho^{\gamma}_{\mathcal{T}^{(t)}}(Z_{\mathcal{T}^{(t)}}) - B\left(\sum_{v_\alpha \in \mathrm{term}\mathcal{T}^{(t)}-\mathrm{term}\mathcal{T}} \gamma^{\mathrm{ht}(v_\alpha)}\right)\\ &= \rho^{\gamma}_{\mathcal{T}^{(t)}}(Z_{\mathcal{T}^{(t)}}) - B\left(\sum_{v_\alpha \in \mathrm{term}\mathcal{T}'-\mathrm{term}\mathcal{T}} \gamma^{\mathrm{ht}(v_\alpha)}\right)^t\\ &= \rho^{\gamma}_{\mathcal{T}^{(t)}}(Z_{\mathcal{T}^{(t)}}) - BM^t_\gamma\end{aligned}$$

Thus by (8.22), we have

$$w \leq \rho^{\gamma}_{\mathcal{T}^{(t)}-(\mathrm{term}\mathcal{T}^{(t)}-\mathrm{term}\mathcal{T})}(Z_{\mathcal{T}^{(t)}-(\mathrm{term}\mathcal{T}^{(t)}-\mathrm{term}\mathcal{T})}) + BM^t_\gamma \tag{8.23}$$

Thus, by taking $t \to \infty$ for (8.23) and using Lemma 8.20 we get that $w(x_0) \leq J(\Pi, x_0)$ for all policies Π. Taking the inf over all Π yields the desired inequality. □

Lemma 8.24. *([Rus10, Lemma 4]) For a cyclic policy* Π*,* $J(\Pi, x_0)$ *is the unique bounded solution to*

$$w = \mathscr{D}_\Pi w. \tag{8.25}$$

Proof. As $\mathscr{D}_\Pi$ is a contraction mapping on $\mathcal{L}_\infty(\mathscr{X}, \mathscr{B}, \mathbb{P}_0)$, we get that (8.25) has a unique bounded solution w. For all t, we have that $[\mathscr{D}_\Pi]^t w = J_{\mathcal{T}^{(t)}}(\Pi|_{\mathcal{T}^{(t)}}, x_0)$ where the recursive cost J is taken with w as the cost for $v_\alpha \in \text{term}\mathcal{T}^{(t)} - \text{term}\mathcal{T}$. Thus for all t, $w(x_0) = [\mathscr{D}_\Pi]^t w(x_0) = J_{\mathcal{T}^{(t)}}(\Pi|_{\mathcal{T}^{(t)}}, x_0))$, and we will show that this equals the infinite-horizon $J(\Pi, x_0)$ through the following approximation. Using (8.18) with $\mathcal{T}^{(t)}$ in place of $\mathcal{T}'$, we get that

$$\begin{aligned}
&|J_{\mathcal{T}^{(t)}}(\Pi|_{\mathcal{T}^{(t)}}, x_0) - J_{\mathcal{T}^{(t)}-(\text{term}\mathcal{T}^{(t)}-\text{term}\mathcal{T})}(\Pi_{\mathcal{T}^{(t)}-(\text{term}\mathcal{T}^{(t)}-\text{term}\mathcal{T})}, x_0)| \\
\leq &\sum_{v_\alpha \in \text{term}\mathcal{T}^{(t)}-\text{term}\mathcal{T}} \gamma^{\text{ht}(v_\alpha)} ||w - 0||_\infty \\
&\qquad\qquad (\text{the truncated process} = \text{untruncated with cost} = 0 \text{ on truncation}) \\
= &M_\gamma^t \|w\|_\infty \xrightarrow{t\to\infty} 0.
\end{aligned}$$

Thus

$$\lim_{t\to\infty} J_{\mathcal{T}^{(t)}}(\Pi|_{\mathcal{T}^{(t)}}, x_0) = \lim_{t\to\infty} J_{\mathcal{T}^{(t)}-(\text{term}\mathcal{T}'-\text{term}\mathcal{T})}(\Pi_{\mathcal{T}^{(t)}-(\text{term}\mathcal{T}'-\text{term}\mathcal{T})}, x_0) = J(\Pi, x_0),$$

with the last equality using Lemma 8.20. Therefore

$$w(x_0) = \lim_{t\to\infty} J_{\mathcal{T}^{(t)}}(\Pi|_{\mathcal{T}^{(t)}}, x_0) = J(\Pi, x_0)$$

as desired. □

Proof. (of Theorem 8.11) Using that $\mathscr{D}$ is a contraction mapping on $\mathscr{L}_\infty(\mathscr{X}, \mathscr{B}, \mathbb{P}_0)$, we know that there exists $w \in \mathscr{L}_\infty(\mathscr{X}, \mathscr{B}, \mathbb{P}_0)$ such that $w = \mathscr{D}w$. Then, by Lemma 8.21 we have that $\hat{w} = w$.

A priori, this $\hat{w} = w$ might not be achieved by any cyclical Markov policy, but by Theorem 6.2 we get that there exists a cyclical Markov policy $\hat{\Pi}$ that satisfies

$$\hat{w} = \mathscr{D}\hat{w} = \mathscr{D}_{\hat{\Pi}}\hat{w}.$$

By (8.25) it then follows that $\hat{w}$ is the infinite-horizon cost of $\hat{\Pi}$, so $\hat{\Pi}$ is indeed an optimal policy.

The converse direction is clear. □

Like in [Rus10], the dynamic programming equations allow us to calculate/approximate $\hat{w}$ by starting at some $w^1 \in \mathscr{L}_\infty(\mathscr{X}, \mathscr{B}, \mathbb{P}_0)$ and generate the sequence

$$w^{k+1} = \mathscr{D}w^k, \quad k = 1, 2, \ldots.$$

The following theorem is essentially identicial to [Rus10, Theorem 5] in both statement and proof, close to being an exact quote. However, we do not include an equivalent statement to his regarding the special case when $v^1 \geq \sup(\ldots)$, as we believe that part of his statement to be false.

Theorem 8.26. *If the assumptions of Theorem 8.11 are met, then the sequence of functions* w^k *generated by the value iteration method is convergent linearly in* $\mathscr{L}_\infty(\mathscr{X}, \mathscr{B}, \mathbb{P}_0)$ *to the optimal value function* $\hat{w}$*, with the quotient* M_γ*. Moreover, if* $w^1 = 0$*, then the sequence* $\{w^k\}$ *is nondecreasing.*

Proof. Linear convergence of the sequence $\{w^k\}$ follows from the fact that $\mathscr{D}$ is a contraction mapping with modulus $M_\gamma < 1$. If $w^1 = 0$ then $w^2 \geq 0 = w^1$, so the monotonicity of $\mathscr{D}$ (Lemma 8.16) implies that $\{w^k\}$ are nondecreasing. □

Remark 8.27. When $\mathcal{T}$ with root v_0 is an eventually repeating tree (there exists finite-height $\mathcal{T}_{init}$ such that for every $v_\alpha \in \mathcal{T}_{init}$, there exists $\mathcal{T}'_\alpha$ that is repeated in $\mathcal{T}_\alpha$), we can get optimal *eventually cyclical* $\hat{\Pi}$ and $w_0 = J(\hat{\Pi}, x_0)$ by using Theorem 6.2 on the finite $\mathcal{T}_{init}$ where for $v_\alpha \in \operatorname{term}\mathcal{T}_{init} - \operatorname{term}\mathcal{T}$, $w_\alpha(x_\alpha) = \min_\Pi J_{\mathcal{T}_\alpha}(\Pi, x_\alpha)$ and the optimal cyclical Markov policy for $\mathcal{T}_\alpha$ are as solved with Theorem 8.11.

As a consequence of Theorem 8.11 we get

Corollary 8.28. *Let $(\mathscr{X}, \mathscr{U}, U, Q, c)$ be a stationary controlled Markov model on a self-similar tree $\mathcal{T}$, where Q, U, c, ρ satisfy the analogous versions of (1)-(4) in Theorem 8.11. Then (8.9) has an optimal solution and its optimal value $\hat{w}(x)$, as a function of the initial state $x_0 = x \in \mathscr{X}$, satisfies the equation:*

$$w(x) = \min_{u \in U(x)} \left\{ c(x,u) + \gamma\sigma((w)_{v_\alpha \in \operatorname{ch}(v_0)}, x, Q(x,u)) \right\}. \tag{8.29}$$

Moreover, an optimal stationary Markov policy $\hat{\Pi} = \hat{\pi}$ exists and satisfies the equation

$$\hat{\pi}(x) \in \operatorname*{arg\,min}_{u \in U(x)} \left\{ c(x,u) + \gamma\sigma((\hat{w})_{v_\alpha \in \operatorname{ch}(v_0)}, x, Q(x,u)) \right\}. \tag{8.30}$$

Conversely, every bounded solution of (8.29) is the optimal value of (8.9), and every measurable solution of (8.30) defines an optimal stationary Markov policy.

9. Policy iteration

The policy iteration method is as follows: For $k = 0, 1, 2, \ldots$, given Π^k, we find $w^k \in \mathscr{L}_\infty(\mathscr{X}, \mathscr{B}, \mathbb{P}_0)$ that solves

$$w^k = \mathscr{D}_{\Pi^k} w^k, \tag{9.1}$$

and then find the cyclical Markov policy Π^{k+1} that solves

$$\mathscr{D}_{\Pi^{k+1}} w^k = \mathscr{D} w^k. \tag{9.2}$$

Theorem 9.3. *Given the assumptions of Theorem 8.11 we have that the sequence of functions $w^k, k = 0, 1, 2, \ldots$ are nonincreasing and convergent to the unique bounded $\hat{w}$ solving $\hat{w} = \mathscr{D}\hat{w}$.*

Proof. Proceed identically to [Rus10, Theorem 6] using Π in place of π. □

It is worth noting that the proof shows that convergence is in fact linear and is convergent in finitely many steps if there are finitely many possible policies [Rus10].

10. Specialized nonsmooth Newton method

We also adapt a specialized nonsmooth Newton method to solve (9.1). Fix the policy Π^k, so we want to solve

$$w = D_{\Pi^k} w, \tag{10.1}$$

where we suppress $w^k = w$ to use the superscript in the section below to index another iterative procedure. The above equation means solving for $\{w_\alpha\}$:

$$w_\alpha(x_\alpha) = \overline{c^{(\alpha)}}(x_\alpha) + \gamma \sup_{\boldsymbol{\mu} \in \overline{\mathscr{A}^{(\alpha)}}(x_\alpha)} \langle \mathbf{w}_{\operatorname{ch}(v_\alpha)}, \boldsymbol{\mu} \rangle$$

for each $v_\alpha \in \mathcal{T}' - \text{term}\mathcal{T}'$, where

$$\overline{c^{(\alpha)}}(x_\alpha) = c^{(\alpha)}(x_\alpha, \pi^k_\alpha(x_\alpha)), \quad \overline{\mathscr{A}^{(\alpha)}}(x_\alpha) = \mathscr{A}^{(\alpha)}(x_\alpha, Q^{(\alpha)}(x_\alpha, \pi^k_\alpha(x_\alpha))),$$

and $w_\alpha = w_0$ for $v_\alpha \in \text{term}\mathcal{T}' - \text{term}\mathcal{T}$, and $w_\alpha = c^{(\alpha)}$ for $v_\alpha \in \text{term}\mathcal{T}' \cap \text{term}\mathcal{T}$. Then the solution for (10.1) is w_0. Our iterative process to solve this is as follows. We begin with some initial $\mathbf{w}^1_{\mathcal{T}'}$. Then at step l of this method (l is not related to k), for each $v_\alpha \in \mathcal{T}' - \text{term}\mathcal{T}'$ and $x \in \mathscr{X}$ we find

$$\boldsymbol{\mu}^l_\alpha(x) \in \underset{\boldsymbol{\mu} \in \overline{\mathscr{A}^{(\alpha)}}(x)}{\arg\max} \langle \mathbf{w}^l_{\text{ch}(v_\alpha)}, \boldsymbol{\mu} \rangle.$$

With the assumptions of Theorem 8.11, as well as the assumption that p $\in [1, \infty)$, we have that $\overline{\mathscr{A}^{(\alpha)}}(x)$ is weakly compact and thus there is an optimal $\boldsymbol{\mu}^l_\alpha(x)$. Then, the next iterate $\mathbf{w}^{l+1}$ is obtained by solving the system:

$$w^{l+1}_\alpha(x) = \overline{c^{(\alpha)}}(x) + \gamma \langle \mathbf{w}^{l+1}_{\text{ch}(v_\alpha)}, \boldsymbol{\mu}^l_\alpha \rangle$$

for each $x \in \mathscr{X}, v_\alpha \in \mathcal{T}' - \text{term}\mathcal{T}'$ where $w^{l+1}_\alpha = w^{l+1}_0$ for $v_\alpha \in \text{term}\mathcal{T}' - \text{term}\mathcal{T}$ and that $w^{l+1}_\alpha = c^{(\alpha)}$ for $v_\alpha \in \text{term}\mathcal{T}' \cap \text{term}\mathcal{T}$. Due to these conditions on term$\mathcal{T}'$, solving the above system is further reduced to solving for w^{l+1}_0 that satisfies

$$w^{l+1}_0(x_0) = J_{\mathcal{T}'}(\Pi^k, x_0) + \sum_{v_\alpha \in \text{term}\mathcal{T}' - \text{term}\mathcal{T}} \gamma^{\text{ht}(v_\alpha)} \mathbb{E}_{\boldsymbol{\mu}^l_{(0)}(x_0)}[\ldots \mathbb{E}_{\boldsymbol{\mu}^l_{\text{pa}(\text{pa}(v_\alpha))}(x_{\text{pa}(\text{pa}(v_\alpha))})}[\mathbb{E}_{\boldsymbol{\mu}^l_{\text{pa}(v_\alpha)}(x_{\text{pa}(v_\alpha)})}[w^{l+1}_0(x_\alpha)]]\ldots] \quad (10.2)$$

for all $x_0 \in \mathscr{X}$, where $J_{\mathcal{T}'}$ is using $\rho_\alpha(\cdot) = \langle \cdot, \boldsymbol{\mu}^l_\alpha \rangle$ and $c^{(\alpha)}_* = 0$ for $v_\alpha \in \text{term}\mathcal{T}' - \text{term}\mathcal{T}$— since we can obtain all of $\mathbf{w}^{l+1}$ from w^{l+1}_0 as some nested expectation as above. Then continue another iteration with the solved $\mathbf{w}^{l+1}$.

Theorem 10.3. *Assume the conditions of Theorem 8.11 and let* $\mathscr{W} = \mathscr{L}_p(\mathscr{X}, \mathscr{B}, \mathscr{P}_0)$ *and* $p \in [1, \infty)$. *We then have that, for any starting* $\mathbf{w}^1$, *the sequence* $\{w^l_0\}$ *generated by the Netwon method converges to* w^*_0 *where* w^*_0 *solves (10.1), and that the sequence is (not strictly) increasing for* $l = 2, 3, 4, \ldots$

Proof. Let $\mathscr{M}^l : \mathscr{W} \to \mathscr{W}$ be given by

$$[\mathscr{M}^l w](x_0) = J_{\mathcal{T}'}(\Pi^k, x_0) + \sum_{v_\alpha \in \text{term}\mathcal{T}' - \text{term}\mathcal{T}} \gamma^{\text{ht}(v_\alpha)} \mathbb{E}_{\boldsymbol{\mu}^l_{(0)}(x_0)}[\ldots \mathbb{E}_{\boldsymbol{\mu}^l_{\text{pa}(\text{pa}(v_\alpha))}(x_{\text{pa}(\text{pa}(v_\alpha))})}[\mathbb{E}_{\boldsymbol{\mu}^l_{\text{pa}(v_\alpha)}(x_{\text{pa}(v_\alpha)})}[w(x_\alpha)]]\ldots]$$

Notice that for $w, p \in \mathscr{L}_\infty(\mathscr{X}, \mathscr{B}, \mathscr{P}_0)$ we have

$$|[\mathscr{M}^l w](x) - [\mathscr{M}^l p](x)| \leq \sum_{v_\alpha \in \text{term}\mathcal{T}' - \text{term}\mathcal{T}} \gamma^{\text{ht}(v_\alpha)} ||w - p||_\infty$$

since, for all μ, $\mathbb{E}_\mu$ is lipschitz on $\mathscr{L}_\infty(\mathscr{X}, \mathscr{B}, \mathscr{P}_0)$ with constant 1. We then get that $\mathscr{M}^l$ is a contraction mapping on $\mathscr{L}_\infty(\mathscr{X}, \mathscr{B}, \mathscr{P}_0)$.

Thus (10.2) has the unique bounded solution w^{l+1}_0. We additionally have that $\mathscr{M}^l(w) \leq \mathscr{D}_{\Pi^k}(w)$ for all $w \in \mathscr{W}$. We also have that $\mathscr{M}^l$ is (not strictly) increasing due to basic properties of expected value. Thus $[\mathscr{M}^l]^2 w \leq \mathscr{M}^l \mathscr{D}_{\Pi^k} w \leq [\mathscr{D}_{\Pi^k}]^2 w$, and by induction we have that $[\mathscr{M}^l]^t w \leq [\mathscr{D}_{\Pi^k}]^t w$ for all t. Thus $w^{l+1}_0 = \lim_{t \to \infty} [\mathscr{M}^l]^t w \leq \lim_{t \to \infty} [\mathscr{D}_{\Pi^k}]^t w = \overline{w_0}$ where $\overline{w_0}$ solves (10.1). Thus

$$w^l_0 \leq \overline{w_0}$$

for all $l \geq 2$. For the remainder of the proof always assume $l \geq 2$.
Thus

$$w_0^l \leq \mathscr{D}_{\Pi^k} w_0^l = \mathscr{M}^l w_0^l$$

As $\mathscr{M}^l$ is a montone increasing function we get via induction that $w_0^l \leq \mathscr{D}_{\Pi^k} w_0^l \leq [\mathscr{M}^l]^t(w_0^l)$ for all t. Taking the limit as $t \to \infty$ we get $w_0^l \leq \mathscr{D}_{\Pi^k} w_0^l \leq w_0^{l+1}$.
We have thus shown that $\{w_0^l\}$ is a bounded increasing sequence which thus converges to some $w_0^* \in \mathscr{Z}_\infty(\mathscr{X}, \mathscr{B}, \mathscr{P}_0)$. Additionally, it must be that $w_0^* \leq \mathscr{D}_{\Pi^k} w_0^* \leq w_0^*$.
Thus w_0^* solves (10.1).

□

11. Why not DAGs?

One natural question is if the methods used in this paper could be extended to graphical models on a general directed acyclic graph (DAG). To the authors of this paper, however, any straightforward generalization of the methods of this paper seem destined to fail. We give some quick heuristic arguments for this point:

11.1. **Problems with the setup.** Letting $G = (\{v_\alpha\}, \{e_{\alpha\beta}\})$ be a finite DAG, the definitions of Section 3 generalize in a straightforward manner. The most obvious generalizations of the definition of dynamic risk measures, or even one step risk measures, for a very general DAG are problematic. Consider the following DAG, In this case the

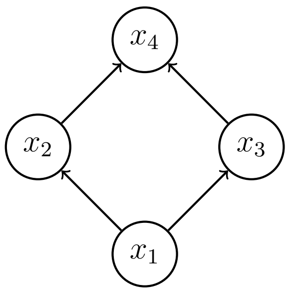

Figure 3. A simple DAG

obvious notion of a dynamic risk measure risk measure, would match the definition in Section 4. The obvious notion of the one-step risk mappings would be maps

$$\rho_{2/3} : \mathscr{Z}_4 \to \mathscr{Z}_{2/3}$$
$$\rho_1 : \mathscr{Z}_2 \times \mathscr{Z}_3 \to \mathscr{Z}_1$$

A natural generalization of (6.1) would then be to find the policy minimizing

$$Z_1 + \rho_1(Z_2 + \rho_2(Z_4 + c), Z_3 + \rho_3(Z_4 + c))$$

where Z_i is the total cost at node i.
Any natural measure of risk for use in our optimization problem should satisfy some form of money-invariance and normalization, as a gaurenteed cost should correspond to a fixed risk. The notion of money invarient for ρ_1 should be the same as in our previous cases, however, the right notion of money invarience for $\rho_{2/3}$ seems elusive. If we insist that each of $\rho_{2/3}$ be money invariant, we run into the problem that for $c \in \mathbb{R}$

$$\rho_1(\rho_2(c), \rho_3(c)) = 2c,$$

we are double counting the cost. This will certainly problematic results for the optimization problem. Thus we are lead to the principle that for $Z \in \mathscr{Z}_4, W \in \mathscr{Z}_2, W' \in \mathscr{Z}_3, c \in \mathscr{Z}_1$ we should certainly have that

$$\rho_1(W + \rho_2(Z + c), W' + \rho_3(Z + c)) = \rho_1(W + \rho_2(Z), W' + \rho_3(Z)) + c.$$

Equivalently, assuming the normal form of money-invarience for ρ_1, we should have

$$\rho_2(Z + c) = \rho_2(Z) + \lambda c, \qquad \rho_2(Z + c) = \rho_2(Z') + \lambda' c$$

and $\lambda + \lambda' = 1$. In order to maintain monotonicity, we must have $\lambda, \lambda' \geq 0$. In particular, one of $0 \leq \lambda, \lambda' < 1$, so we may assume $0 \leq \lambda < 1$. Now assume that $U_2(x) = \{0, 1\}$ where

$$c^{(2)}(x, u) = \begin{cases} 0 & u = 0 \\ \frac{1+\lambda}{2} & \end{cases},$$

$Z^{u=0} - Z^{u=1} = 1$ almost surely. Then we would have

$$c^{(2)}(x, 1) + \rho_2(Z^{u=1}) = \frac{1+\lambda}{2} + \rho_2(Z^{u=0} - 1) = \frac{1+\lambda}{2} - \lambda + \rho_2(Z^{u=0}) > c^{(2)}(x, 0) + \rho_2(Z^{u=0})$$

so $u = 0$ is the best decision at node 2 according to the obvious heuristic using ρ_2, even though $u = 1$ necessarily reduces overall costs by $\frac{1-\lambda}{2}$. So, it seems, any reasonable version of the one-step problem, which uses only information from the history of node 2 and a random variable at node 4, would lead to an sub-optimal solution.

A related idea, is that instead of modifying the requirement on money invariance for $\rho_{2/3}$ we modify the expression we consider at node 1, to something of the form

$$\rho_1(W + \lambda\rho_2(Z + c), W' + \lambda'\rho_3(Z + c))^2$$

but this ignores the philosophy of dynamic programming as we are treating the two parts each argument on a separate footing, negating the advantage we could gain from making it into one-step problems.

11.2. **Is the problem with loops?** The specific problem we encountered in the previous subsection might be narrowed down to the fact that there is an (undirected) loop in the graph. Thus we might try to (first) consider *poly-trees*, DAGs with no undirected loops / whose underlying undirected graph is a tree. At this point we notice, a different problem: when there is not a root, i.e. a node v with $v \leq v'$ for all other nodes v', what should the optimization problem be?

In general, this seems difficult, but maybe it could be some risk-measure applied to appropriate values at each initial node?

As again we see no way forward, we restrict ourselves yet again to try and find a tractable special case. Thus we shall henceforth consider perhaps the simplest type of polytree other than a directed rooted tree, namely the **anti-tree** and look for especially simple problems on them.

Definition 11.1. We say a graph G is an *anti-tree* if the graph G^{op}, obtained from G by reversing the direction of all edges, is a directed rooted tree. In this case we also call the root v_0 of G^{op} the root of G.

We can then hope that we can define some optimization problem that is "concentrated" (in some sense) at the root of the anti-tree, to avoid the above loop-introducing problem. Anti-trees are also very natural to consider in the context of our decision problem, they represent processes where many factors affect the outcome of one particular event which

[2] One can't put λ in front of the W as well as that would violate cash invariance.

is of concern.

Thus let G be an anti-tree with a controlled Markov model. The most natural definition of one-step risk measures on G would be monotone mappings $\rho_\alpha : \mathscr{Z}_\alpha \to \mathscr{Z}_{\text{pa}(\alpha)}$ (which in turn is equivalent to mappings $\rho_{\alpha,\beta} : \mathscr{Z}_\alpha \to \mathscr{Z}_\beta$ for each parent β of a node α). Then the one-step optimization problem at the node α would be something like

$$\min_{u \in U(x_\alpha)} c^\alpha(x_\alpha, u) + \rho_{\text{ch}(\alpha),\alpha}(Z)$$

11.3. **Intractability of the DAG setting.** At this point one could think that the above state problem is natural and even tractable. However, this problem, or in fact any which only takes into account the history of a given node, does not seem to lead to global minima. To see this, let us consider an even simpler version of the problem, suppose that the costs $Z_\alpha = 0$ for $\alpha \neq 0$, and assume that we want to simply minimize $\mathbb{E}_\Pi[Z_0]$ over all allowable choice of policy Π ($\mathbb{E}_\Pi$ denotes the expectation over the probability distribution determined by a policy Π).

Suppose that the recursive optimizations are of the form

$$\min_{\pi^{(\alpha)}} f_\alpha(x_\alpha, x_{\text{pa}(x_\alpha)}, \pi^{(\alpha)}(x_\alpha), \pi^{(\text{pa}(\alpha))}(x_\alpha))$$

for each node α where f_α is perhaps some complicated function. However, only considering the information of the parents of a node alpha can clearly lead to suboptimal solutions, in a pseudo prisoners dilema. For instance you clearly have to consider what policies for v_α play well other polices of the other parents of $\text{ch}(v_\alpha)$. Continuing in this was leads us to the conclusion that you have to consider at least all nodes v_β which are parents to some descendent of v_α.

At this point it is no longer true that the optimization problem at node v_α is independent of the problem at a node v_β if $v_\alpha \notin \mathcal{T}_\beta$ and $v_\beta \notin \mathcal{T}_\alpha$. Therefore, in particular in order to ensure we find the global minimum, we must simultaneously optimizing all break the problem into smaller recursive problems to reduce the computational complexity. Since even this simpler version of the problem seems impossible, we should not expect the more general problem to be solvable in polynomial time (in some appropriate sense), at least with any direct generalization of the methods of this paper.

12. Acknowledgment

This paper was written as part of the Fields Undergraduate Summer Research Program 2025. We would like to thank the Fields Institute for this wonderful opportunity, as well as Professors Sebastian Jaimungal and Piotr Zwiernik for their invaluable mentorship throughout the program.

References

[RS06] Andrzej Ruszczyński and Alexander Shapiro. "Optimization of Convex Risk Functions". In: *Mathematics of Operations Research* 31.3 (2006), pp. 433–452. ISSN: 0364765X, 15265471. URL: http://www.jstor.org/stable/25151738 (visited on 06/20/2024).

[Rus10] Andrzej Ruszczyński. "Risk-averse dynamic programming for Markov decision processes". In: *Mathematical Programming* 125.2 (Oct. 1, 2010), pp. 235–261. DOI: 10.1007/s10107-010-0393-3. URL: https://doi.org/10.1007/s10107-010-0393-3.

[Rus14] Andrzej Ruszczyński. "Erratum to: Risk-averse dynamic programming for Markov decision processes". In: *Mathematical Programming* 145 (May 2014), pp. 601–604. DOI: 10.1007/s10107-010-0393-3.