MARKOV DECISION PROCESSES WITH SET-VALUED TRANSITIONS: A CASE STUDY IN AUTOMATED REPLACEMENT MODELS

Ricardo Shirota Filho*, Fabio Gagliardi Cozman*

*Escola Politécnica, Universidade de São Paulo
Av. Prof. Mello Moraes, 2231
Cidade Universitária, CEP 05508-900 - São Paulo, SP, Brazil

Emails: ricardo.shirota@poli.usp.br, fgcozman@usp.br

Abstract— This paper explores a recently defined generalization of Markov Decision Processes called *Markov Decision Processes with Set-valued Transitions* (MDPST). We apply MDPSTs to the analysis of a well-known model for bus engine replacement based on MDPs. While MDPs are already extensively employed for automation and control, the discussion in this paper suggests that MDPSTs can have even broader applicability, as MDPSTs benefit from the ability to represent lack of probabilistic precision.

Keywords— Markov Decision Process, Automated Decision Making, Bus Engine Replacement.

Resumo— Este artigo explora uma generalização de Processos de Decisão Markovianos chamada *Processos de Decisão Markovianos com Transições de Conjuntos (MDPSTs).* Aplicamos MDPSTs à análise de um modelo bastante conhecido baseado em MDPs para substituição de motores de ônibus. Enquanto MDPs são extensivamente aplicados em controle e automação, a discussão neste artigo sugere que MDPSTs podem ter maior aplicabilidade, pois estes possuem habilidade de representar imprecisão em valores de probabilidade.

Keywords— Processo de Decisão Markoviano, Automação de Tomada de Decisão, Substituição de Motores de Ônibus.

1 Introduction

Markov Decision Processes (MDPs) have a long and respectful history in Operations Research. Adopted as models for solving sequential decision problems, they have been applied to many situations, such as inventory management, production planning, layout design and maintenance and replacement problems. Research in control engineering applies the MDP formulation to stochastic modeling, making it an ideal framework for many problems in automation. Recent work by (Boutilier et al., 1999) has shown how MDPs provide an excellent formulation for planning problems, opening new frontiers for their use as a decision tool. Furthermore, many advances have been achieved in the last few decades in both applied and theoretical fronts on MDPs, leading to applications in areas such as ecology, economics and communication engineering (Puterman, 1994). In the majority of these applications, the main tool to find optimal decisions in MDPs is dynamic programming (Russell and Norvig, 1995).

Early applications of MDPs involve inventory management and maintenance and replacement problems (Puterman, 1994). In this paper we discuss a well-known model formulated by John Rust on bus engine replacement, based on years of data from Harold Zurcher, superintendent of maintenance at the Madison Metropolitan Bus Company (Rust, 1987). We intend to examine some of the limitations of the MDP-based formulation, and to examine the benefits of applying a formulation based on *Markov Decision Processes with Set-valued Transitions (MDPSTs).* Rust's formulation for Zurcher's behavior provides for a nice case study, due to its simple state and transition probability model.

As mentioned, Rust adopts the MDP formulation in order to obtain a model that describes the behavior of Zurcher's decisions on bus engine replacement. The hypothesis is that his behavior can be described as a simple regenerative optimal stopping model, with Zurcher's decisions coinciding with an optimal stopping rule. Due to his background as an economist, Rust uses a bottom-up approach to model the replacement investment along with ten years of monthly data on bus mileage and engine replacements. He seeks an estimate for the unknown primitive parameters which specify Zurcher's expectations of the future value of the state variables, the expected costs of regular bus maintenance, and his perceptions of the customer goodwill costs of unexpected failures. These estimates are then used to test the consistency of Zurcher's behavior with the model.

However, Rust did not have tools and sufficient data to consider all the details inherent to the problem. This forced him to simplify his model and to make additional assumptions in order to obtain a solution. In this paper, we take Rust's original problem as a case study and ask: would MDPSTs be a suitable tool to capture all relevant aspects of Zurcker's behavior? The idea here is to focus on whether MDPSTs would, *in principle*, offer a more adequate framework for knowledge representation in this problem. We specifically question whether we need to assume, as Rust does, that precise transition probabilities must be employed for all actions and states.

In the next section, we provide a brief review of the model formulated in (Rust, 1987), along with an introduction to the MDP/MDPST model. Section 3 discusses the model built by Rust and the possible advantages of adopting an MDPST model. Finally, Section 4 provides a general overview and final comments on this work.

2 Background

This section presents a brief review of the literature on MDPs and the recently proposed MDPSTs (the latter based on work by (Trevizan et al., 2007)), and of the empirical model proposed by (Rust, 1987).

2.1 MDPs

Markov Decision Processes (MDPs) are used in many fields to encode possibly infinite sequences of decisions under uncertainty. For historical review, basic technical development, and substantial reference to related literature, the reader may consult books by (Puterman, 1994) and (Bertsekas, 1995). The following paragraphs present the general MDP formulation, consistent with that adopted in (Rust, 1987).

In general, MDPs are described by:

- a set $\mathcal{T}$ of *stages*; a decision is made at each stage.
- a set $\mathcal{S}$ of *states*.
- a set of *actions* $\mathcal{A}$; the set of actions may be indexed by states.
- a conditional probability distribution P_t that specifies the probability of transition from state s to state r given action a at stage t.
- a *reward* function R_t that indicates how much is gained (or lost, by using a negative value) when action a is selected in state s at stage t.

The state obtained at stage t, in a particular realization of the process, is referred to as s_t; likewise, the action selected at stage t is referred to as a_t.

The history h_t of an MDP at stage t is the sequence of states and actions visited by the process, $[s_1, a_1, \ldots, a_{t-1}, s_t]$. The *Markov assumption* that is adopted for MDPs is that $P(s_t|h_{t-1}, a_t) = P(s_t|s_{t-1}, a_t)$; consequently:

$$P(h_t|s_1) = P(s_t|s_{t-1}, a_{t-1})\, P(s_{t-1}|s_{t-2}, a_{t-2}) \ldots \times P(s_3|s_2, a_2)\, P(s_2|s_1, a_1)\,. \quad (1)$$

A *decision rule* d_t indicates the action that is to be taken in state s at stage t. A *policy* π is a sequence of decision rules, one for each stage. A policy may be *deterministic* or *randomized*; that is, it may prescribe actions with certainty, or rather it may just prescribe a probability distribution over the actions. A policy may also be *history-dependent* or not; that is, it may depend on all states and actions visited in previous stages, or just on the current state. A policy that is not history-dependent is called *Markovian*. A Markovian policy induces a probability distribution over histories through Expression (1).

It can also be assumed that an MDP with infinite horizon (that is, with infinite $\mathcal{T}$) may always stop with some probability. In fact, in many cases it is assumed that the process stops with geometric probability: the process stops at stage t with probability $(1-\gamma)\gamma^{t-1}$ (independently of all other aspects of the process). Then γ is called the *discount* factor of the MDP (Puterman, 1994, p. 125).

The evaluation of a policy π in an MDP can be given by its expected reward:

$$V_\pi(s) = E_{s,\pi}\left[E_T\left[\sum_{t=1}^{T} R(s_t, a_t)\right]\right]; \quad (2)$$

that is, the expectation of the expected reward assuming the process stops at stage T. Now if the process has a geometric probability of stopping at T, with parameter γ, we have (Puterman, 1994, p. 125):

$$V_{\pi,\gamma}(s) = E_{s,\pi}\left[\sum_{t=1}^{\infty} \gamma^{t-1} R(s_t, a_t)\right]. \quad (3)$$

We refer to $V_{\pi,\gamma}(s)$ as the expected total discounted reward. There are other criteria to evaluate policies in MDPs; for example, the expected total reward $E_{s,\pi}[\sum_{t=1}^{\infty} R(s_t, a_t)]$, and the average reward $\lim_{T\to\infty}(1/T)E_{s,\pi}\left[\sum_{t=1}^{T} R(s_t, a_t)\right]$ (Bertsekas, 1995; Puterman, 1994). These criteria may be useful in specific problems but they are usually less realistic than Expression (2) and the associated discounted reward (3). We focus on the latter in this paper.

Additional realism and flexibility can be attached to MDPs by allowing imprecision and indeterminacy in the assessment of transition probabilities. A decision process with states, actions, stages and rewards as described before, but where a set of probability distributions is associated with each transition, has been called a *Markov Decision Process with Imprecise Probabilities* (MDPIP) by White III and Eldeib (White III and El-Deib, 1994). Satia and Lave Jr. use instead the name *MDP with Uncertain Transition Probabilities* (Satia and Lave Jr., 1970), in what may be the first thorough analysis of this model in the literature; Harmanec uses the term *generalized MDP* to refer to MDPIPs (Harmanec, 2002).

MDPIPs can represent incomplete and ambiguous beliefs about transitions between states; conflicting assessments by a group of experts; and situations where one wishes to investigate the effect of perturbations in a "base" model. MDPIPs have also been investigated as representations for abstracted processes, where details about transition probabilities are replaced by an enveloping set of distributions (Givan et al., 2000; Ha and Haddawy, 1996). Similar models are encoded by the *controlled Markov set-chains* by Kurano et al. (Kurano et al., 1998; Hosaka et al., 2002). Slightly less related are the vector-valued MDPs by Wakuta (Wakuta, 1995). Some of these efforts have also adopted *interval-valued rewards*; in this paper it is focused on imprecision/indeterminacy only in transition probabilities.

2.2 MDPSTs

(Trevizan et al., 2007) proposed a new class of MDPIPs pursuing important applications in the field of artificial intelligence planning. As we will suggest in the following section, this class is also useful for applications in classical MDP problems when there are unobservable variables or unknown or ambiguous data, as in the maintenance and replacement problem here considered.

The general formulation is similar to that observed for classical MDPs. Following the notation adopted in Section 2.1, an MDPST is also composed by a set of stages $\mathcal{T}$, a set of states $\mathcal{S}$, a set of actions $\mathcal{A}$ and a reward function R. However, the state transition function $F(s,a)$ maps states $s \in \mathcal{S}$ and actions $a \in \mathcal{A}$ into reachable subsets of $\mathcal{S}$, i.e., into nonempty subsets of $\mathcal{S}$, and a set of mass assignments $m(k|s,a)$ for all $s \in S$, $a \in \mathcal{A}$, and $k \in F(s,a)$. In MDPSTs the optimal policy induces a value function that is the unique solution of (Trevizan et al., 2007):

$$V^*(s) = \max_{a\in\mathcal{A}} \left(R(s,a) + \right.$$
$$\left. +\gamma \sum_{k\in F(s,a)} m(k|s,a) \min_{r\in k} V^*(r) \right). \quad (4)$$

In this case we adopt the Γ-maximin criterion for the evaluation of policies. This means we select the policy that yields the largest value of min $V^*(s)$, where the minimum applies to all transition probabilities. Other well studied criteria are Γ-maximax, E-admissibility and maximality.

Many known algorithms are available for MDPSTs. In fact, every algorithm for MDPIP can be directly applied to MDPSTs (Trevizan et al., 2007). The currently known algorithms include value iteration, policy iteration, modified policy iteration, Harmanec's interval dominance algorithm (Harmanec, 2002) and different variants of mathematical programming.

2.3 The empirical model of Harold Zurcher

As the superintendent of maintenance at the Madison Metropolitan Bus Company, Harold Zurcher had the responsibility of keeping the company's bus fleet in an acceptable operational condition. Zurcher classified the necessary maintenance (to keep the fleet in order) in three categories (1. routine, periodic maintenance, 2. replacement or repair of individual components at time of failure and 3. major engine overhaul and/or replacement), however Rust focuses only on the third component of investment.

In this approach, Zurcher must decide whether it is cost-efficient to simply replace or repair a failed bus engine component, or engage a full engine replacement. On a bus with a relatively low mileage, it is reasonable to proceed with a minimal maintenance procedure, substituting just the damaged component. However, in a bus with relatively high mileage it is expected that other components will shortly fail, making sense to replace the entire engine.

John Rust initially formulates Harold Zurcher's behavior as a simple regenerative optimal stopping model of bus engine replacement. Using a bottom-up approach he derives a joint stochastic process $\{a_t, s_t\}$ to explain the bus data, where $a_t = 1$ if a replacement occurs at time t and $a_t = 0$ otherwise, and s_t denotes observed state variables associated with the replacement investment decision. Bus data provided by Zurcher consists of "[...]monthly observations on the mileage (odometer reading) on each bus, plus a maintenance diary which records the date, mileage and list of components repaired or replaced each time a bus visits the company shop" (Rust, 1987, p. 999). Hence the decisions are made monthly, and state variables s_t represent the accumulated mileage since replacement.

Rust proposes a decomposition of the cost function as follows:

$$c(s,\theta_1) = m(s,\theta_{11}) + \mu(s,\theta_{12})b(s,\theta_{13}), \quad (5)$$

where $m(s,\theta_{11})$ is the conditional expectation of regular maintenance and operating expenses (maintenance, fuel and insurance costs), $\mu(s,\theta_{12})$ is the conditional probability of unexpected engine failures, $b(s,\theta_{13})$ is the conditional expectation of towing costs, repair costs and the perceived dollar cost of lost customer goodwill in the event of an unexpected engine failure, and $\theta_1 = (\theta_{11}, \theta_{12}, \theta_{13})$ is the unobservable variables associated to the cost function. However, the absence of specific data on maintenance and operating cost or on the occurrence of unexpected breakdowns made it impossible to estimate m, μ and b separately, forcing Rust to use a generic estimate of the sum, c.

The cost function is used to determine the utility function:

$$u(s_t, a_t, \theta_1) = \begin{cases} -c(s_t, \theta_1) & if \quad a_t = 0, \\ -[\overline{P} - \underline{P} + c(0, \theta_1)] & if \quad a_t = 1. \end{cases} \tag{6}$$

This function models the property that if the decision to keep the current engine is made ($a_t = 0$), the system incurs only the expected operating cost $c(s_t, \theta_1)$. If, on the other hand, it is decided that the engine should be replaced, then the old bus engine is cannibalized for scrap value $\underline{P}$, a new or rebuild engine is installed at cost $\overline{P}$ and incurs the operating cost $c(0, \theta_1)$. The utility function is then used to determine the value function V_θ which is the unique solution to Bellman's equation:

$$V_{opt}(s) = \max_{a \in \mathcal{A}} \left(R(s, a) + \gamma \sum_r p(r|s, a) V_{opt}(r) \right). \tag{7}$$

Rust proves that under certain conditions there is an optimal stationary Markovian replacement policy (Rust, 1985):

$$a_t = \begin{cases} 1 & if \quad s_t > \rho(\theta_1, \theta_2), \\ 0 & if \quad s_t \leq \rho(\theta_1, \theta_2), \end{cases} \tag{8}$$

in which the constant ρ represents a threshold value of mileage (optimal stopping barrier) such that whenever current mileage on the bus s_t exceeds ρ, it is optimal to incur the replacement costs $RC = (\overline{P} - \underline{P})$ and replace the old bus engine with a new one (Rust, 1987).

3 A modified model for bus engine replacement

Rust provides a simple model for bus engine replacement. This model was shown to be statistically consistent with the replacement data provided by Zurcher, however this result should be interpreted with caution. In this section, we present and discuss some aspects of the model for which assumptions have been made, but not enough support provided.

We begin our discussion by the assumption that the transition probabilities have a functional form. Rust attempts to prove this by running adherence tests of the selected distributions on the available data (the replacement data) in order to check their consistency. This procedure has two outcomes, it can accept the adherence of the distribution to the data, or it can reject it. Of the 8 particular parametric functional forms tested, none were rejected, leading us to conclude that any one of them could equally represent the data. Nevertheless, Rust insists on selecting one specific functional form. He then uses "more intuitive criteria in order to select a 'best fit' model from the array of alternative functional forms" (Rust, 1987, p. 1020). This was achieved by obtaining a 'compromise' between choosing the functional form with highest likelihood value and choosing a 'parsimonious' one, yet consistent with his priors and other non-quantitative information about the replacement model. What is not explained is how this 'compromise' is composed, and what is considered to be a 'parsimonious' functional form.

While the exact criterion used to select the transition probability distribution is not explained, we can try to understand why Rust adopted this procedure. MDPIPs are known since early 1970s, hence available when Zurcher's bus engine replacement model was formulated. But it was only recently that MDPIPs began to receive its deserved attention. If used as the base model for the formulation of this model, it would have not been necessary to assume a functional form for the transition probabilities, given that MDPIPs allow indeterminacy in their assessment. This way, Rust was left with the classical MDP formulation and forced to use precise probability assignments. Since the data flatly refuted a precise value for the transition probabilities, he was forced to obtain a distribution consistent with the data in order to build a model. This does not, however, justify the use of one specific functional form in preference of the remaining tested forms, given that they were all considered adherent to the data by the likelihood ratio test. This makes MDPIP/MDPST the ideal choice for modeling this situation, given that the model builder does not have to make random suppositions about the functional form of the transition probabilities.

Another assumption made by Rust is the proposition that the system states can be modeled solely based on the accumulated mileage. It is perhaps too simplistic to assume that the engine condition is influenced only by the odometer reading. Many other variables can be shown to affect the condition of the engine, like maintenance history, fuel and air quality and even subjective quantities like bus driver attitude (driving aggressiveness). Rust was not unaware of this problem, but did not formulate a more realistic model since he could not obtain a solution for it.

We observe two main difficulties in adding these additional parameters to model the system states. The first one, which was also mentioned by Rust, is the unavailability of the necessary information about these extra parameters. As noted in (Rust, 1987), the data provided by Zurcher consisted only on monthly observations on the mileage on the bus odometer, plus a maintenance diary containing records on date, mileage and list of components repaired or replaced each time the bus visited the company shop. No other information was available. Hence, Rust was unable to add these extra parameters, since he could not model them. Nevertheless, they continue to be important in the description of the engine's condition, and should be included (possibly as unobservable parameters) in order to obtain a more realistic model.

The second problem for increasing the details of the state space model is the fact that each new state variable increases the size of the space. This increases proportionally the time needed to obtain an optimal solution (i.e. policy iteration is $O(n^3)$, which means that solution is obtained in time proportional to the third power of the state space size). Advances in the last few decades have made it possible for researchers to solve problems much more complex than those computed by Rust, allowing us to incorporate these additional details. Not only has computer technology improved, but new algorithms have also been proposed, leading to faster solutions. It can also be mentioned the formulation of strategies for state space reduction, like the factored representation briefly commented in (Boutilier et al., 1999).

When considering the above mentioned observations, the classical MDP framework is no longer suitable. Instead, the MDPST[1] framework is a better option. In order to overcome the limitations imposed by the MDP framework, the technical aspects of the possible MDPST formulation is presented in the following paragraphs.

In a general manner the difference in between the MDP and the MDPST model relies on the state transition function and the probability assignments. This will provide the necessary tools to incorporate improvements to solve both problems appointed in the beginning of this section.

Recalling Section 2.2, which introduced the concept of MDPST, the state transition function $F(s, a)$ maps states s and actions a into nonempty subsets of $\mathcal{S}$, and a set of mass assignments $m(k|s, a)$ for all s, $a \in \mathcal{A}$ and $k \in F(s, a)$. This formulation allows us to elaborate a model in which the questioned assumptions are no longer necessary:

- **Transition probability:** In MDPSTs, the transition probabilities are given by a set of mass assignments. This allows one to model uncertainty in the transition probabilities. It is achieved by enumerating constraints on mass assignments, and the probabilities can be any value which respects these constraints. In this specific example, each mass assignment could be given by an upper and a lower probability value, representing the maximum and the minimum expected probability of transition from one certain state to another.

- **State transition function:** In MDPs, the state transition function maps states and actions into states. In a different manner, in MDPSTs the state transition function maps states and actions into *nonempty subsets of* $\mathcal{S}$. This means that a transition from a state, given a feasible action, can be mapped into a *set of states*, instead of only one specific state (although the set can be singleton, in which case it becomes identical to the classical MDP). This allows one to add unobservable or unknown state variables, with transitions being mapped into the set of all states with values applicable to these variables.

[1]Since MDPSTs are a class of MDPIPs, this can also be done using the latter. However, MDPIPs are considerably more complicated and, in this case, do not provide additional benefits (when compared to MDPSTs), the reason why we have chosen to adopt the MDPST formulation.

Many classical algorithms for solving MDPs are able to compute solutions for MDPSTs, as seen in Section 2.2. (Trevizan et al., 2007) shows that any algorithm used to compute optimal policies for MDPIPs can be applied to MDPST. This includes well known policy and value iteration, along as variations of mathematical programs (multilinear and integer programs).

4 Conclusion

The analysis provided by Rust in his paper about Harold Zurcher's bus engine replacement model is interesting but still incomplete. Despite statistically consistent with the replacement data available on the bus fleet, Rust's formulation may not represent the actual engine failure process as one would like.

In order to implement an update and an improvement of this model, we have suggested an MDPST formulation (Trevizan et al., 2007). This not only allows one to include unobservable or unknown parameters (as imprecision on the transition of state variables), but also the imprecise probabilities associated with ambiguous or incomplete data. Trevizan et al. initially formulated the model for use in AI planning problems, nonetheless it can be applied to classical MDP problems, when these problems involve unobservabilities or incompleteness of data.

However, even this MDPST formulation of Zurcher's bus engine replacement model is not complete. Other aspects can also be improved, which were not considered in our proposed model. Rust, for example, comments on his doubts about the optimal stationary policy adopted in his paper. The data shows that engine replacement was conducted from a minimum of 82,400 to a maximum of 387,300 miles, a variance too large to be consistent with the adopted optimal threshold replacement policy. On a preliminary research on this issue, it was found that in some cases of MDPs (as in constrained MDP, where additional constraints are defined on the value function), it is common not to obtain a pure policy as an optimal solution (Puterman, 1994). In these cases a *mixed strategy* is the usual outcome. If this constrained model could be proved to correctly represent this replacement model, a mixed strategy can explain such a large variance. However, it should also be examined whether Zurcher actually performs optimally (thus if the threshould replacement policy is really not optimal).

The cost function adopted by Rust was a rough estimate. Their functional form was also used to represent the possible cost values; again no proofs were presented to show that this is true. Work by (Wakuta, 1995), (Kurano et al., 1998) and (Hosaka et al., 2002) provide different classes of MDPs which are focused on imprecision on the reward function, and could possibly be applied to this situation.

Finally, it must be noticed that there are different approaches to maintenance problems, i.e. work by Jayakumar and Asgarpoor (Jayakumar and Asgarpoor, 2004; Jayakumar and Asgarpoor, 2006). In this work, states represent levels of deterioration and actions include different levels of maintenance. Costs and rewards are given to maintenance level selected and condition of equipment respectfully, and the goal is to obtain the most cost-efficient maintenance level to be carried out at each deterioration level.

Acknowledgments

This work has been financially supported by FAPESP grant 2004/09568-0; the first author is supported by FAPESP grant 05/58090-9; the second author is partially supported by CNPq grant 3000183/98-4.

References

Bertsekas, D. P. (1995). *Dynamic Programming and Optimal Control (Vol. 1, 2)*, Athena Scientific, Belmont, Massachusetts.

Boutilier, C., Hanks, S. and Dean, T. (1999). Decision-theoretic planning: Structural assumptions and computational leverage, *Journal of Artificial Intelligence Research* **11**: 1–94.

Givan, R., Leach, S. M. and Dean, T. (2000). Bounded-parameter Markov decision processes, *Artificial Intelligence* **122**(1-2): 71–109.

Ha, V. and Haddawy, P. (1996). Theoretical foundations for abstraction-based probabilistic planning, *in* E. Horvitz and F. Jensen (eds), *Conference on Uncertainty in Artificial Intelligence*, Morgan Kaufmann, San Francisco, California, United States, pp. 291–298.

Harmanec, D. (2002). Generalizing Markov decision processes to imprecise probabilities, *Journal of Statistical Planning and Inference* **105**: 199–213.

Hosaka, M., Nakagami, J. and Kurano, M. (2002). Controlled Markov set-chains with set-valued rewards – the average case, *International Transactions in Operations Researach* **9**: 113–123.

Jayakumar, A. and Asgarpoor, S. (2004). Maintenance optimization of equipment by linear programming, *8th International Conference on Probabilistic Methods Applied to Power Systems*, Ames, Iowa, pp. 145–149.

Jayakumar, A. and Asgarpoor, S. (2006). Maintenance optimization of equipment by linear programming, *Probability in the Engineering and Informational Sciences* **20**: 183–193.

Kurano, M., Song, J., Hosaka, M. and Huang, Y. (1998). Controlled Markov set-chains with discounting, *Journal Applied Probability* **35**: 293–302.

Puterman, M. L. (1994). *Markov Decision Processes*, Wiley series in probability and mathematical statistics, John Wiley and Sons, New York.

Russell, S. J. and Norvig, P. (1995). *Artificial Intelligence: a Modern Approach*, Prentice Hall, New Jersey.

Rust, J. (1985). Stationary equilibrium in a market for durable assets, *Econometrica* **53**(4): 783–805.

Rust, J. (1987). Optimal replacement of gmc bus engines: an empirical model of Harold Zurcher, *Econometrica* **55**(5): 999–1033.

Satia, J. K. and Lave Jr., R. E. (1970). Markovian decision processes with uncertain transition probabilities, *Operations Research* **21**: 728–740.

Trevizan, F. W., Cozman, F. G. and de Barros, L. N. (2007). Planning under risk and Knightian uncertainty, *20th IJCAI*, pp. 2023–2028.

Wakuta, K. (1995). Vector-valued Markov decision processes and the system of linear inequalities, *Stochastic Processes and their Applications* **56**: 159–169.

White III, C. C. and El-Deib, H. K. (1994). Markov decision processes with imprecise transition probabilities, *Operations Research* **42**(4): 739–749.