QUANTITATIVE FINANCE
RESEARCH CENTRE

UNIVERSITY OF
TECHNOLOGY SYDNEY

THINK.CHANGE.DO

QUANTITATIVE FINANCE RESEARCH CENTRE

Research Paper 358 March 2015

Stochastic Switching for Partially Observable Dynamics and Optimal Asset Allocation

Juri Hinz

ISSN 1441-8010 www.qfrc.uts.edu.au

Stochastic switching for partially observable dynamics and optimal asset allocation

Juri Hinz
juri.hinz@uts.edu.au

March 10, 2015

Abstract

In industrial applications, optimal control problems frequently appear in the context of decisions-making under incomplete information. In such framework, decisions must be adapted dynamically to account for possible regime changes of the underlying dynamics. Using stochastic filtering theory, Markovian evolution can be modeled in terms of latent variables, which naturally leads to high dimensional state space, making practical solutions to these control problems notoriously challenging. In our approach, we utilize a specific structure of this problem class to present a solution in terms of simple, reliable, and fast algorithms.

1 Introduction

The main problem in the dynamical decision-making process is to determine how to update the information and to apply a control action in order to reach an optimal result over a given time period. These questions are often stated within decision theory for partially observable Markov processes, which deals with discrete-time *optimal stochastic control* under incomplete information. Although theoretical fundamentals of these problems are well-understood, (see [1], [2], [6], and [12]) numerical solutions in practical applications remain persistently challenging and require using heuristics (see [11], [3]). The main trust of this contribution is to present a working approach to a particular problem class. We present a class of optimal stochastic control problems under incomplete information, which fall within the scope of the so-called *convex switching systems* (see [9]) and possess efficient algorithmic solutions.

Important applications can be modeled as stochastic control problems whose state dynamics is linear. It turns out that by exploiting this issue, efficient numerical schemes, including solution assessment, can be obtained. Let us introduce a specific optimal stochastic switching problem whose state evolution consists of one discrete and one continuous component. We thus suppose that the state space $E = P \times \mathbb{R}^d$ is the product of a finite space P and the Euclidean space $\mathbb{R}^d$. Furthermore, assume that the discrete component $p \in P$ is deterministically driven by a finite set A of actions in terms of a function $\alpha : P \times A \to A$, $(p, a) \to \alpha(p, a)$, where $\alpha(p, a) \in P$ is the new value of the discrete component of the state if its previous discrete component value was p and the action $a \in A$ was taken. Furthermore, we assume that the continuous state component evolves as an uncontrolled Markov process $(Z_t)_{t=0}^T$ on $\mathbb{R}^d$ whose evolution is driven by random linear transformations $Z_{t+1} = W_{t+1} Z_t$ with random $d \times d$ disturbance matrices $(W_t)_{t=1}^T$ which are independent and integrable. Now we turn to the specification of control costs. Assume that taking an action $a \in A$ causes an immediate reward $r_t(p, z, a)$ which depends on the state $(p, z) \in E$ and on the action $a \in A$ through given reward functions $r_t : E \times A \to \mathbb{R}$ which may be time $t = 0, \ldots, T-1$ dependent. When the system arrives at the last time step $t = T$ in the state $(p, z) \in E$, the agent collects the *scrap value* $r_T(p, z)$, described by a pre-specified scrap function $r_T : E \to \mathbb{R}$. At each time $t = 0, \ldots, T-1$ the *decision rule* π_t is given by a mapping $\pi_t : E \to A$, prescribing at time t in the state $(p, z) \in E$ the action $\pi_t(p, z) \in A$. A sequence $\pi = (\pi_t)_{t=0}^{T-1}$ of decision rules is called *policy*. When controlling the system by policy $\pi = (\pi_t)_{t=0}^{T-1}$, the positions $(p_t^\pi)_{t=0}^T$ and the actions $(a_t^\pi)_{t=0}^{T-1}$ evolve randomly

$$a_t^\pi = \pi_t(p_t^\pi, Z_t), \qquad p_{t+1}^\pi = \alpha(p_t^\pi, a_t^\pi), \qquad Z_{t+1} = W_{t+1} Z_t, \qquad t = 0, \ldots, T-1.$$

Having started at initial values $p_0^\pi = p_0$ and $Z_0 = z_0$, the goal of the controller is to maximize the expectation of the reward

$$v_0^\pi(p_0, z_0) = \mathbb{E}\left(\sum_{t=0}^{T-1} r_t(p_t^\pi, Z_t, a_t^\pi) + r_T(p_T^\pi, Z_T) \right)$$

accumulated within the entire time depending on the choice of the policy $\pi = (\pi_t)_{t=0}^{T-1}$. Following [9], we call such Markov decision problem a *convex switching systems*, if

$$\begin{array}{c} r_t(p, \cdot, a), r_T(p, \cdot) \text{ for all } t = 0, \ldots, T-1,\ a \in A,\ p \in P \\ \text{are convex and globally Lipschitz continuous.} \end{array} \tag{1}$$

Example: The simplest example of a convex switching system is given by an American Put in discrete time. Here, the discounted asset price $(Z_t)_{t=0}^T$ at time steps $0, \dots, T$ is modeled by a sampled geometric Brownian motion following

$$Z_{t+1} = W_{t+1} Z_t, \quad t = 0, \dots, T-1, Z_0 \in \mathbb{R}_+ \tag{2}$$

where $(W_t)_{t=1}^T$ are independent identically distributed random variables following log-normal distribution. The price of an option with strike price K for the interest rate $\lambda \geq 0$ and with maturity date T is given by the solution to the optimal stopping problem

$$\sup\{\mathbb{E}(\max(e^{-\lambda\tau}(K - Z_\tau), 0)) : \ \tau \text{ is } \{0, 1, \dots, T\}\text{-valued stopping time}\}.$$

This switching system is defined by two positions and two actions $P = \{1, 2\}$, $A = \{1, 2\}$. Here, the positions "stopped" and "goes" are represented by $p = 1$, $p = 2$ respectively and the actions "stop" and "go" denoted by $a = 1$ and $a = 2$. With this interpretation, the position change is given by

$$(\alpha(p, a))_{p,a=1}^2 = \begin{bmatrix} \alpha(1,1) & \alpha(2,1) \\ \alpha(1,2) & \alpha(2,2) \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}.$$

The reward at time $t = 0, \dots, T-1$ and the scrap value are defined by

$$r_t(p, z, a) = e^{-\lambda t}(K - z)^+(p - \alpha(p, a)), \tag{3}$$

$$r_T(p, z) = e^{-\lambda T}(K - z)^+(p - \alpha(p, 1)), \tag{4}$$

for all $p \in P$, $a \in A$, $z \in \mathbb{R}$.

Note that the state process of a convex switching system follows a *controlled Markov process* on $E = P \times \mathbb{R}^d$ governed by the family

$$\mathcal{K}_t^a v(p, z) = \mathbb{E}(v(\alpha(p, a), W_{t+1} z)), \qquad t = 0, \dots, T-1, \ \ a \in A, \tag{5}$$

of transition kernels acting on functions $v : E = P \times \mathbb{R}^d \to \mathbb{R}$ where the above expectation exists. Using these kernels, the policy value is obtained by the *policy valuation* algorithm

$$v_T^\pi = r_T, \qquad v_t^\pi(p, z) = r_t(p, z, \pi_t(p, z)) + \mathcal{K}_t^{\pi_t(p,z)} v_{t+1}^\pi(p, z).$$

To obtain a policy $\pi^* = (\pi_t^*)_{t=0}^{T-1}$, which maximizes the total expected reward, one introduces for each $t = 0, \dots, T-1$ the so-called *Bellman operator*

$$\mathcal{T}_t v(p, z) = \max_{a \in A} \left(r_t(p, z, a) + \mathcal{K}_t^a v(p, z) \right), \quad (p, z) \in E \tag{6}$$

acting on each function $v : E \to \mathbb{R}$ where the above expectation exists. Next consider the *Bellman recursion*, also referred to as the *backward induction*:

$$v_T = r_T, \quad v_t = \mathcal{T}_t v_{t+1}, \qquad \text{for } t = T-1, \dots, 0. \tag{7}$$

For convex switching systems, there exists a recursive solution $(v_t^*)_{t=0}^T$ to the Bellman recursion

$$\begin{aligned} v_T(p,z) &= r_T(p,z) && (8)\\ v_t(p,z) &= \max_{a\in A}(r_t(p,z) + \mathbb{E}(v_{t+1}(\alpha(p,a), W_{t+1}z))) && (9) \end{aligned}$$

for $t = T-1,\dots,0$, $p \in P$, and $z \in \mathbb{R}^d$. The functions $(v_t^*)_{t=0}^T$ resulting form backward induction are called *value functions*, they determine an optimal policy $\pi^* = (\pi_t^*)_{t=0}^{T-1}$ via

$$\pi_t^*(p,z) = \text{argmax}_{a\in A}\left(r_t(p,z,a) + \mathbb{E}(v_{t+1}^*(\alpha(p,a), W_{t+1}z))\right) \qquad (10)$$

for $p \in P$, $z \in \mathbb{R}^d$, for all $t = 0,\dots,T-1$.

2 Solution techniques

The first step in obtaining a numerical solution to the backward induction (7) is an appropriate discretization of the Bellman operator (6). For this reason, we consider a modified Bellman operator

$$\mathcal{T}_t^n v(p,z) = \max_{a\in A}(r_t(p,.,a) + \sum_{k=1}^{n} \nu_{t+1}^n(k) v(\alpha(p,a), W_{t+1}(k)z))$$

with integration replaced by its numerical counterpart, defined in terms of appropriate distribution sampling $(W_t(k))_{k=1}^n$ to each disturbance W_t, $t = 1,\dots,T$, with corresponding integration weights $(\nu_t^n(k))_{k=1}^n$. In the resulting modified backward induction

$$v_T^{(n)} = r_T, \qquad v_t^{(n)} = \mathcal{T}_t^n v_{t+1}^{(n)}, \qquad t = T-1,\dots 0 \qquad (11)$$

the functions $(v_t^{(n)})_{t=0}^T$ need to be described by algorithmically tractable objects. Note that since all reward $(r_t(p,\cdot,a))_{t=0}^{T-1}$ and scrap $r_T(p,\cdot)$ functions are convex in the continuous variable, the modified value functions (11) are also convex. For these functions, we suggest an approximation in terms of piecewise linear and convex functions as follows: Introduce the so-called *sub-gradient envelope* $\mathcal{S}_G f$ of a convex function $f : \mathbb{R}^d \to \mathbb{R}$ on a grid $G \subset \mathbb{R}^d$ as a maximum

$$\mathcal{S}_G f = \vee_{g\in G}(\triangledown_g f)$$

of sub-gradients $\triangledown_g f$ of f on all grid points $g \in G$. Using sub-gradient envelope operation, we define the double-modified Bellman operator as

$$\mathcal{T}_t^{m,n} v(p,.) = \mathcal{S}_{G^m} \max_{a\in A}\left(r_t(p,.,a) + \sum_{k=1}^{n} \nu_{t+1}^n(k) v(\alpha(p,a), W_{t+1}(k).)\right),$$

where the operator $\mathcal{S}_{G^m}$ stands for the sub-gradient envelope on the grid $G^m = \{g^1, \dots, g^m\}$, as defined above. The corresponding double-modified backward induction

$$\begin{aligned} v_T^{(m,n)}(p, \cdot) &= \mathcal{S}_{G^m} r_T, (p, \cdot), \quad p \in P & (12) \\ v_t^{(m,n)} &= \mathcal{T}_t^{m,n} v_{t+1}^{(m,n)}, \qquad t = T-1, \dots 0. & (13) \end{aligned}$$

yields the double-modified value functions $(v_t^{(n,m)})_{t=0}^T$. This scheme enjoys excellent asymptotic properties. Under slight additional assumptions on distribution sampling and grid tightening, [9] shows that the double-modified value functions converge to the true value functions almost surely uniformly on compact sets.

In this work we focus on the algorithmic properties of the scheme (12), (13). Since the double-modified value functions $(v_t^{(m,n)})_{t=0}^T$ are piece-wise linear and convex, they can be expressed in a compact form, using matrix representations. Note that a piecewise convex function f can be described by a matrix by representing all linear functionals participating in the matrix representation as matrix rows. To denote this relation, let us agree on the following notation: Given a function f and a matrix F such that $f(z) = \max(Fz)$ holds for all $z \in \mathbb{R}^d$, then we write $f \sim F$. Let us emphasize that the sub-gradient envelope operation $\mathcal{S}_G$ is reflected on the side of matrix representatives by a specific row-rearrangement operator Υ_G, in the following sense

$$f \sim F \quad \Rightarrow \quad \mathcal{S}_G f \sim \Upsilon_G[F].$$

Thereby, the row-rearrangement operator Υ_G of the grid $G = \{g^1, \dots, g^m\} \subset \mathbb{R}^d$ acts on each matrix L with d columns as

$$(\Upsilon_G F)_{i,\cdot} = F_{\mathrm{argmax}(Fg^i),\cdot} \qquad \text{for all } i = 1, \dots, m.$$

For piecewise convex functions, the result of maximization, summation, and composition with linear mappings, followed by sub-gradient envelope can be obtained using their matrix representatives. More precisely, for convex piecewise linear functions $(f_i)_{i=1}^n$ given in terms of their matrix representatives $(F_i)_{i=1}^n$, meaning that $f_i \sim F_i$ for $i = 1, \dots, n$, we obtain

$$\begin{aligned} \mathcal{S}_G(\sum_{i=1}^n f_i) &\sim \sum_{i=1}^n \Upsilon_G[F_i], & (14) \\ \mathcal{S}_G(\vee_{i=1}^n f_i) &\sim \Upsilon_G[\sqcup_{i=1}^n F_i], & (15) \\ \mathcal{S}_G(f_i(W\cdot)) &\sim \Upsilon_G[F_i W], \qquad i = 1, \dots, n, & (16) \end{aligned}$$

where the operator $\sqcup$ stands for binding matrices by rows.

In the context of convex switching systems, the double-modified backward induction involves only maximizations, summations and compositions with linear mappings applied to piecewise linear convex functions, thus it can be rewritten in terms of matrix operations, giving the following algorithm:

Pre-calculations: For a convex switching system and $G^m = \{g^1, \ldots, g^m\}$, implement the row-rearrangement operator $\Upsilon = \Upsilon_{G^m}$ and the row maximization operator $\sqcup_{a \in A}$. For $t = 1, \ldots, T$, determine a distribution sampling $(W_t(k))_{k=1}^n$ of each disturbance W_t with the corresponding weights $(\nu_t(k))_{k=1}^n$. Given reward $(r_t)_{t=0}^{T-1}$ and scrap r_T functions, determine the matrix representatives of their sub-gradient envelopes

$$\mathcal{S}_{G^m} r_t(p, \cdot, a) \sim R_t(p, a), \qquad \mathcal{S}_{G^m} r_T(p, \cdot) \sim R_T(p)$$

for $t = 0, \ldots, T-1$, $p \in P$ and $a \in A$. Denoting the matrix representatives of each (approximate) value function by

$$v_t^{(m,n)}(p, \cdot) \sim V_t(p) \quad \text{for } t = 0, \ldots, T,\ p \in P.$$

These matrix representatives are obtained via:

Initialization: Start with the matrices

$$V_T(p) = R_T(p), \qquad \text{for all } p \in P. \tag{17}$$

Recursion: For $t = T-1, \ldots, 0$ calculate for $p \in P$

$$V_t(p) = \sqcup_{a \in A} \big(R_t(p, a) + \sum_{k=1}^{n} \nu_{t+1}(k) \Upsilon [V_{t+1}(\alpha(p, a)) \cdot W_{t+1}(k)] \big) \tag{18}$$

Example: Let us illustrate this algorithm using the optimal stopping problem from American Put from above. To be able representing the functions (3) and (4), we embed the state space $\mathbb{R}_+$ into $\mathbb{R}^2$ amending the price component z by one $(z, 1)$. This procedure yields matrices $R_t(p, a) = e^{-\lambda t}(p - \alpha(p, a))\Pi$, $R_T(p, a) = e^{-\lambda T}(p - \alpha(p, 1))\Pi$ for all $p \in P, a \in A$ and $t = 0, \ldots, T-1$ where

$$\Pi = \begin{bmatrix} -1 & K \\ 0 & 0 \end{bmatrix} \qquad \text{with strike price } K \in \mathbb{R}_+.$$

To describe the dynamics in $\mathbb{R}^2$ sucht hat the embedded first component follows a geometric Browinan motion and the amended second component is fixed, we introduce the disturbance matrices

$$W_t = \begin{bmatrix} \varepsilon_t & 0 \\ 0 & 1 \end{bmatrix} \quad \text{with i.i.d random variables } (\varepsilon_t)_{t=1}^T. \tag{19}$$

Thereby, for $t = 1, \ldots, T$, the random variables $\ln(\varepsilon_t)$ follow normal distribution with mean $\lambda - \frac{\sigma^2}{2}$ and variance $\sigma^2 > 0$. Now we turn to the distribution sampling. Using an ordinary Monte-Carlo sampling of the disturbance matrices, define for each $t = 1, \ldots, T$, the sequence $(\varepsilon_t(k))_{k=1}^n$ which consists of independent realization copies of ε_t. The matrices $W_t(k)$ are defined as in (19), replacing the variable ε_t by the realization $\varepsilon_t(k)$ for $k = 1, \ldots n$. For optimal stopping problem, the recursion (18) boils down to a comparison of the so-called continuation value to the current payoff, hence one needs to determine the matrix $V_t(2)$ using for $t = T-1, \ldots, 0$ the recursion

$$V_T(2) := e^{-\lambda T}\Pi, \qquad V_t(2) = \Upsilon[e^{-\lambda t}\Pi] \sqcup (\frac{1}{n}\sum_{k=1}^{n} \Upsilon[V_{t+1}(2)W_{t+1}(k)]).$$

3 Switching under incomplete information.

The subsequent work is devoted to an application of the above technique to solve optimal switching problems under partial observation. The so-called partially observable Markov decision processes, whose applications enjoys unprecedented popularity [10], [7] have a long history in decision-making. Starting with [13], [14] the optimal control problems have been addressed in [8], [4], and [1], among others.

For sake of concreteness, let us formulate our approach under the assumption that the underlying stochastic driver follows the so-called partially observable Markov processes (POMPs), which is usually addressed under the framework of hidden-Markov modeling (HMM).

The idea POMPs is to realize a time series $(y_t)_{t=0}^T$ in such a way that it behaves as it was driven by a background device which may operate in different regimes. Thereby, one supposes that the operating regime is not directly observed and evolves like a Markov chain $(x_t)_{t=0}^T$ on a finite space which is identified with the set $\{e_1, \ldots, e_d\}$ of unit vectors in $\mathbb{R}^d$. In some situations, the hidden process $(x_t)_{t=0}^T$ can be given a physical meaning, but for many cases it just describes the evolution of latent variables. The basic advantage thereby is that it is possible to trace the evolution of the hidden states indirectly, based on the observation of $(y_t)_{t=0}^T$, using efficient recursive schemes for calculation of the so-called *hidden state estimate*

$$\widehat{x}_t = \mathbb{E}(x_t \,|\, y_j, j \leq t) \qquad t = 0, \ldots T.$$

Thereby, at each time $t = 0, \ldots, T-1$, the probability vector $\widehat{x}_t$ describes the distribution of x_t conditioned on the past observation of $(y_j)_{j=0}^t$. More importantly, such approach *reproduces a Markovian dynamics* in the following sense: Although $(y_t)_{t=0}^T$ is not Markovian in general, it turns out that the observations $(y_t)_{t=0}^T$ equipped with latent variables $(\widehat{x}_t)_{t=0}^T$ form a two-component process such that

$$\text{the evolution} \quad (\widehat{x}_t, y_t)_{t=0}^T \quad \text{is Markovian.} \tag{20}$$

From this perspective, modeling a time series by POMP yields a technique to address control problems in certain non-Markovian situations. Namely, having assumed that the stochastic driver $(y_t)_{t=0}^T$ of our control problem can be described as observations $(y_t)_{t=0}^T$ of a POMP, the multi-variate Markovian dynamics $(\widehat{x}_t, y_t)_{t=0}^T$ can be constructed in order to treat the original problem in the standard settings of optimal stochastic control for Markovian processes.

In what follows, we show how certain POMP control problems can be solved within the framework of convex switching systems. To some extent, this is a surprising result, since the dynamics under partial observation involves a regular a Bayesian information update, which introduces a non-linearity by re-normalization. That is, although we consider control problems which do not meet assumptions required for convex switching systems, a specific state space extension transforms them into convex switching framework.

Let us introduce the ingredients required therefore. Assume that an unobservable global regime evolves like a Markov chain $(x_t)_{t=0}^T$ on the set $\mathcal{X} = \{e_1, \ldots, e_d\}$ of unit vectors in $\mathbb{R}^d$, while the information available to the controller is gained from the observation of the process $(y_t)_{t=0}^T$ which takes values in a measure space $\mathcal{Y}$. As in the standard setting of POMP, it is assumed that the transition to x_{t+1} and the generation of output y_{t+1} occur independently, given current state x_t. More precisely, the joint evolution $((x_t, y_t))_{t=0}^T$ follows a Markov process whose transition kernels Q_t for $t = 0, \ldots, T-1$ are acting on functions $\phi : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ as

$$\int \phi(x', y') Q_t(\mathrm{d}(x', y') \,|\, (x, y)) = \sum_{x' \in \mathcal{X}} \int_{\mathcal{Y}} \phi(x', y') \Gamma_{x,x'} \mu_x(\mathrm{d}y'). \tag{21}$$

Thereby, the stochastic matrix $\Gamma = (\Gamma_{x,x'})_{x,x' \in \mathcal{X}}$ describes the transition from x_t to x_{t+1} whereas μ_x denotes the distribution of the observation y_{t+1} conditioned on $x_t = x \in \mathcal{X}$. Assuming that for each $x \in \mathcal{X}$ the distribution μ_x is absolutely continuous with respect to a reference measure μ on $\mathcal{Y}$, we introduce the densities

$$\nu_x(y) = \frac{\mathrm{d}\mu_x}{\mathrm{d}\mu}(y), \quad y \in \mathcal{Y}, \; x \in \mathcal{X},$$

to write the distributions as

$$\mu_x(\mathrm{d}y) = \nu_x(y) \mu(\mathrm{d}y) \qquad x \in \mathcal{X}.$$

Using the reference measure μ, the transition kernel (21) of $(x_t, y_t)_{t=0}^T$ is written as

$$\int \phi(x', y') Q_t(\mathrm{d}(x', y') \,|\, (x, y)) = \sum_{x' \in \mathcal{X}} \int_{\mathcal{Y}} \phi(x', y') \Gamma_{x,x'} \nu_x(y) \mu(\mathrm{d}y'),$$

for all $t = 0, \ldots, T-1$, $x \in \mathcal{X}$ and $y \in \mathcal{Y}$. As indicated above, it turns out that $(\widehat{x}_t, y_t)_{t=0}^T$ follows a Markov process on the state space $\widehat{\mathcal{X}} \times \mathcal{Y}$, driven by transition kernels $\widehat{Q}_t$ which act for $t = 0, \ldots, T-1$ on functions $\phi : \widehat{\mathcal{X}} \times \mathcal{Y} \to \mathbb{R}$ as

$$\int_{\widehat{\mathcal{X}} \times \mathcal{Y}} \phi(\widehat{x}', y') \widehat{Q}_t(\mathrm{d}(\widehat{x}', y') \,|\, (\widehat{x}, y)) = \int_{\mathcal{Y}} \phi\left(\frac{\Gamma^\top \mathcal{V}(y')\widehat{x}}{\|\mathcal{V}(y')\widehat{x}\|}, y' \right) \|\mathcal{V}(y')\widehat{x}\| \mu(\mathrm{d}y'). \quad (22)$$

In this formula, $\mathcal{V}(y)$ stands for the diagonal matrix whose diagonal elements are given by $(\nu_x(y))_{x \in \mathcal{X}}$ for $y \in \mathcal{Y}$, and the norm is defined as $\|z\| = \sum_{i=1}^n |z_i|$, each $z \in \mathbb{R}^d$.

As in the case of convex switching systems, we assume that the discrete state component $p \in P$ is deterministically controlled by actions $a \in A$ using a given function $\alpha : P \times A \to A$ where the sets P and A finite. Now, let us turn to the definition of our control costs. Naturally, the reward earned at time t is dependent on the observation y_t. However, it is more convenient to model the expectation of the next-step reward, conditioned on the situation at time t. That is, given $t = 0, \ldots, T-1$, we aim to model the conditioned next-step reward expectation as a function of the state distribution $\widehat{x}$, of the position p, and action a chosen at the time t. Note that with this modeling, the observation y_t recorded at time t indirectly influences the next-step reward expectation through conditioned distribution $\widehat{x}_t$. Let us agree on the following definition

Definition 1. *With notations as above, a partially observable switching problem is a stochastic control problem whose controlled Markov evolution on the state space $P \times \mathcal{X} \times \mathcal{Y}$ is governed by transition kernels*

$$\mathcal{K}_t^a \phi(p, \widehat{x}, y) = \int_{\mathcal{Y}} \phi\left(\alpha(p, a), \frac{\Gamma^\top \mathcal{V}(y')\widehat{x}}{\|\mathcal{V}(y')\widehat{x}\|}, y' \right) \|\mathcal{V}(y')\widehat{x}\| \mu(\mathrm{d}y') \quad a \in A, \quad (23)$$

for $t = 0, \ldots, T-1$, acting on functions ϕ on $P \times \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ where the above integral exists. Furthermore, the reward and scarp values are given as functions, which do not depend on $y \in \mathcal{Y}$:

$$r_t : P \times \mathcal{X} \times A \to \mathbb{R}, \qquad r_T : P \times \mathcal{X} \to \mathbb{R}, \quad t = 0, \ldots, T-1. \quad (24)$$

Note that in (23) the resulting function $\mathcal{K}_t^a\phi$ was not depending on the last component $y \in Y$. In the case that the argument function ϕ also does not depend on the last component $y \in Y$, we agree to write

$$\mathcal{K}_t^a\phi(p,\widehat{x}) = \int_{\mathcal{Y}} \phi\left(\alpha(p,a), \frac{\Gamma^\top \mathcal{V}(y')\widehat{x}}{\|\mathcal{V}(y')\widehat{x}\|}\right) \|\mathcal{V}(y')\widehat{x}\|\mu(\mathrm{d}y'). \tag{25}$$

Note that the value functions of a partially observable Markov switching problem as defined above do not depend on the observation component. This is verified inductively. Using (24) and following for $t = T-1,\dots,0$ the backward induction with kernel action (25), we obtain:

$$v_T = r_T, \quad v_t(p,\widehat{x}) = \max_{a\in A}\left(r_t(p,\widehat{x},a) + \mathcal{K}_t^a v_{t+1}(p,\widehat{x})\right), \qquad \widehat{x} \in \mathcal{X}, p \in P. \tag{26}$$

Now let us introduce a function extension technique. Given a function $f : P \times \widehat{\mathcal{X}} \to \mathbb{R}$ introduce its positive-homogeneous extension $\tilde{f} : P \times \mathbb{R}_+^d \to \mathbb{R}$ by

$$\tilde{f}(p,x) := \|x\| f(p, \frac{x}{\|x\|}) \qquad x \in \mathbb{R}_+^d, \quad p \in P.$$

Note that with this definition, the values of continuous component $\widehat{x} \in \widehat{\mathcal{X}}$ are extended to the entire cone $\mathbb{R}_+^d \supset \widehat{\mathcal{X}}$ and the extension $\tilde{f}$ is indeed positive-homogeneous $\tilde{f}(p,\lambda x) = \lambda \tilde{f}(p,x)$ for all $x \in \mathbb{R}_+^d$, $\lambda \in \mathbb{R}_+$, for each $p \in P$. Similarly, for a function $f : P \times \widehat{\mathcal{X}} \times A \to \mathbb{R}$ we introduce the positive-homogeneous extension $f : P \times \mathbb{R}_+^d \times A \to \mathbb{R}$ by $\tilde{f}(p,x,a) := \|x\| f(p, \frac{x}{\|x\|}, a)$ for all $x \in \mathbb{R}_+^d$, $p \in P$ and $a \in A$.

Lemma 1. *Given a partially observable switching problem with notations as above, consider a function ϕ on $P \times \widehat{\mathcal{X}}$ with positive-homogeneous extension $\tilde{\phi}$, then it holds that*

$$\mathcal{K}_t^a\phi(p,\widehat{x}) = \mathbb{E}(\tilde{\phi}(\alpha(p,a), W_{t+1}\widehat{x})) \quad \widehat{x} \in \widehat{\mathcal{X}},\ a \in A,\ t = 0,\dots,T-1, \tag{27}$$

where the W_{t+1} is given as a matrix-valued function

$$W_{t+1} = \Gamma^\top \mathcal{V}(Y_{t+1}), \qquad a \in A,\ t = 0,\dots,T-1 \tag{28}$$

of a random variable Y_{t+1} whose distribution equals to the reference measure μ.

Proof. Using (25) we verify the assertions (27) and (28) for each $p \in P$,

$a \in A$, $t = 0, \dots, T-1$ as

$$
\begin{aligned}
(\mathcal{K}^a_t \phi)(p, \widehat{x}) &= \int_Y \phi\left(\alpha(p,a), \frac{\Gamma^\top \mathcal{V}(y')\widehat{x}}{\|\Gamma^\top \mathcal{V}(y')\widehat{x}\|}\right) \|\mathcal{V}(y')\widehat{x}\| \mu(\mathrm{d}y') \\
&= \int_Y \tilde{\phi}\left(\alpha(p,a), \Gamma^\top \mathcal{V}(y')\widehat{x}\right) \|\Gamma^\top \mathcal{V}(y')\widehat{x}\|^{-1} \|\mathcal{V}(y')\widehat{x}\| \mu(\mathrm{d}y') \\
&= \int_Y \tilde{\phi}\left(\alpha(p,a), \Gamma^\top \mathcal{V}(y')\widehat{x}\right) \mu(\mathrm{d}y') = \mathbb{E}(\tilde{\phi}(\alpha(p,a), W_{t+1}\widehat{x})).
\end{aligned}
$$

□

Let us define the extended transition kernels by

$$
\tilde{\mathcal{K}}^a_t \tilde{\phi}(p, x) = \mathbb{E}(\tilde{\phi}(\alpha(p,a), W_{t+1}x)) \qquad x \in \mathbb{R}^d, \;\; a \in A, \;\; 0 = 1, \dots T-1, \tag{29}
$$

where the disturbances $(W_t)_{t=1}^T$ are given by (28) in terms of identically distributed random variables $(Y_t)_{t=1}^T$, each following reference distribution μ. The following result shows that the original backward induction (26) can be solved using extended transition kernels (29) instead of the original kernels (23).

Proposition 1. *Given a partially observable Markov switching problem, consider its value functions $(v_t)_{t=0}^T$ returned by backward induction (26) with rewards, scrap values and transition kernels given by (24) and (25). Furthermore , consider functions $(\tilde{v}_t)_{t=0}^T$ on $P \times \mathbb{R}^d_+$ obtained recursively by*

$$
\begin{aligned}
\tilde{v}_T &= \tilde{r}_T, & (30) \\
\tilde{v}_t(p, x) &= \max_{a \in A} \left(\tilde{r}_t(p, x, a) + \tilde{\mathcal{K}}^a_t \tilde{v}_{t+1}(\alpha(p,a), x)\right), & (31)
\end{aligned}
$$

for $x \in \mathbb{R}^d_+$, $t = T-1, \dots, 0$, with extended reward $(\tilde{r}_t)_{t=0}^{T-1}$, scrap $\tilde{r}_T$ functions and transition kernels $\tilde{\mathcal{K}}^a_t$ as defined above. Then it holds that

$$
\tilde{v}_t \text{ is the positive-homogeneous extension of } v_t \text{ for all } t = 0, \dots, T. \tag{32}
$$

Proof. Let us proceed inductively, starting at $t = T$, where our induction assumption (32) holds by the initialization in (26) and (30). Having supposed that $\tilde{v}_{t+1}$ is the positive-homogeneous extension of v_{t+1}, we use Lemma 1 to conclude that $\tilde{\mathcal{K}}^a_t \tilde{v}_{t+1}$ is the positive-homogeneous extension of $\mathcal{K}^a_t v_{t+1}$ for each $a \in A$. Applying summations of and maximizations in (31), we verify that $\tilde{v}_t$ is the positive-homogeneous extension of v_t, as required. □

Since for each $t = 0, \ldots, T-1$ the transition kernel $\tilde{\mathcal{K}}_t^a$ from (29) acts in terms of disturbances, the backward induction (30), (31)

$$\begin{aligned} \tilde{v}_T(p, x) &= \tilde{r}_T(p, x) & (33) \\ \tilde{v}_t(p, x) &= \max_{a \in A}(\tilde{r}_t(p, x) + \mathbb{E}(\tilde{v}_{t+1}(\alpha(p, a), W_{t+1}x))) & (34) \end{aligned}$$

for $t = T-1, \ldots, 0$, $p \in P$, $x \in \mathbb{R}^d$, as required in (8), (9) for convex switching systems. To ensure additional convexity conditions (1) required for the convex switching framework, further assumptions on extensions $\tilde{r}_t$ and $\tilde{r}_T$ for $t = 0, \ldots, T-1$ must be imposed explicitly. For simplicity, let us agree that

$$\begin{array}{c} \text{positive-homogeneous extenstions } (\tilde{r}_t(p, \cdot, a))_{t=0}^{T-1} \text{ and } \tilde{r}_T(p, \cdot) \text{ for} \\ p \in P \ a \in A \text{ of (24) are convex and globally Lipschitz continuous.} \end{array} \quad (35)$$

It turns out that this assertion is fulfilled if for $t = 0, \ldots, T-1$ the original functions (24) are convex and globally Lipschitz in the continuous component.

4 Example: An adaptive investment strategy.

We now illustrate our technique using a simplified problem of dynamic fund allocation. In our approach we consider optimization of an investment strategy for a single risky asset under the assumption that the increments of the sampled asset price process follows a hidden Markov dynamics (see [5]). To obtain such as price evolution, we introduce a random time sampling of the *continuous* price process $(S(t))_{t \geq 0}$, which is inspired by the so-called *Point&Figure Chart* technique.

Suppose trading shall occur only at times where a notable price change may require a position re-balancing. Thereby, the price evolution is sampled as follows: Having fixed a price change step $\Delta > 0$ and starting the observations at the initial time $\tau_0 = 0$, one writes into a *Point&Figure Chart* one of the symbols **x** or **o** at the first time τ_1 where the asset price leaves the interval $[S(\tau_0) - \Delta, S(\tau_0) + \Delta]$. If the price increases to the upper bound $S(\tau_0) + \Delta$ one writes the symbol **x**, otherwise the symbol **o** is written. Repeating the same procedure with the next interval $[S(\tau_1) - \Delta, S(\tau_1) + \Delta]$ and proceeding further, a sequence of stopping times $(\tau_k)_{k \in \mathbb{N}}$ is determined, with the symbols **x** or **o** at each time, which are arranged in a diagram as shown in Figure 1. Assume that the trading occurs only at $(\tau_k)_{k \in \mathbb{N}}$, and that at each time τ_k the trading decision is based only on the observation of the sampled price history $S(\tau_0), \ldots, S(\tau_k)$. Since the price process $(S(t))_{t \geq 0}$ is continuous, the

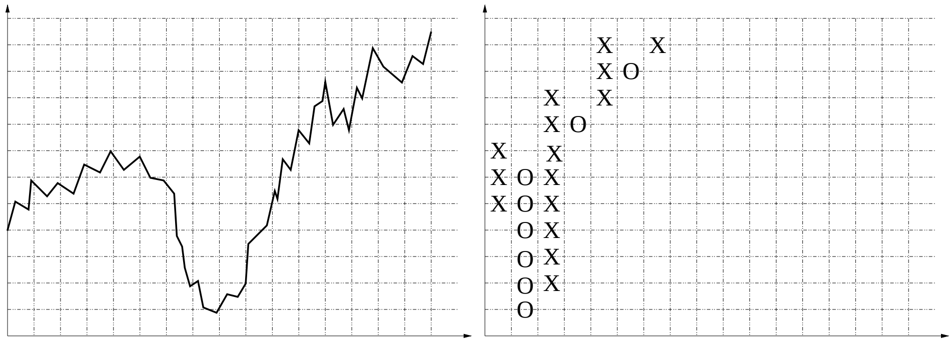

Figure 1: Asset price and its P&F diagram.

stochastic driver of our model is given by the binary increment process

$$y_t = S(\tau_t) - S(\tau_{t-1}), \qquad t = 1, \dots, T,$$

which takes values in the set $\mathcal{Y} = \{-\Delta, \Delta\}$. This process is modeled as the observable part $(y_t)_{t=1}^T$ of a hidden Markov dynamics $(x_t, y_t)_{t=0}^T$. For the sake of concreteness, we suppose that the hidden regimes $\mathcal{X}$ can be identified with some background market situations. As a simple illustration, we consider a two-state $\mathcal{X} = \{e_1, e_2\}$ regime switching with transition matrix

$$\Gamma = \begin{bmatrix} p_1 & (1-p_1) \\ (1-p_2) & p_2 \end{bmatrix}$$

and assume that if the market is in the state $x_t = e_1$ then the next price increment y_{t+1} takes values in $\mathcal{Y} = \{-\Delta, \Delta\}$ with probabilities q_1 and $(1-q_1)$ respectively. Similarly, conditioned on the current state $x_t = e_2$ we have the probabilities $(1-q_2)$ and q_2 for the observation y_{t+1} of the next price move. Choosing the reference measure μ as the uniform distribution on $\mathcal{Y}$ by $\mu(\{\Delta\}) = \mu(\{-\Delta\}) = 1/2$, we obtain the following diagonal density matrices

$$\mathcal{V}(-\Delta) = 2 \begin{bmatrix} q_1 & 0 \\ & 1-q_1 \end{bmatrix}, \qquad \mathcal{V}(\Delta) = 2 \begin{bmatrix} 1-q_2 & 0 \\ 0 & q_2 \end{bmatrix},$$

which gives merely two disturbance matrix realizations $\Gamma^\top \mathcal{V}(-\Delta)$, $\Gamma^\top \mathcal{V}(\Delta)$. According to (28), we define the disturbance matrices by $(W_t = \Gamma^\top \mathcal{V}(Y_t))_{t=1}^T$, using independent identically distributed random variables $(Y_t)_{t=1}^T$, whose distribution is the reference measure μ.

Now, we introduce the position control α for our dynamic asset allocation problem. Consider a situation where the asset position can either be short, neutral, or long, labeled by the numbers $p = 1, 2, 3$ respectively. At each time $t = 0, \dots, T$, the controller must make a decision whether the next position shall be long short, or neutral. Given the set $P = \{1, 2, 3\}$ of all possible positions, we introduce the action set as $A = \{1, 2, 3\}$ where a stands for the targeted position after re-allocation, in which case the position control function α is determined by the following matrix:

$$\begin{bmatrix} \alpha(1,1) & \alpha(1,2) & \alpha(1,3) \\ \alpha(2,1) & \alpha(2,2) & \alpha(2,3) \\ \alpha(3,1) & \alpha(3,2) & \alpha(3,3) \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \\ 1 & 2 & 3 \end{bmatrix}.$$

Finally, let us turn to the definition of the reward and the scrap functions. In this example, we model the payoff in terms of affine linear reward function

$$r_t(p, \widehat{x}, a) = r(p, \widehat{x}, a) = (p-1)\rho^\top \widehat{x} - c|p - \alpha(p,a)|, \quad t = 0, \dots T-1 \quad (36)$$

for all $x \in \widehat{X}$, and $a \in A$. Here $c(p - \alpha(p,a))$ represents the proportional transaction costs determined by a parameter $c > 0$ and the term $(p-1)\rho^\top \widehat{x}$ stands for the expected revenue from holding position p from time t to $t+1$, if the distribution of the market state is described by the probability vector $\widehat{x} \in \widehat{\mathcal{X}}$. Thereby, the vector $\rho \in \mathbb{R}^2$ given by

$$\rho = \Delta[1 - 2q_1, 2q_2 - 1]^\top.$$

such that $(p-2)\rho^\top \widehat{x}$ describes for the return, expected from the realization of next price movement y_{t+1}, conditioned on the information $\widehat{x}_t$ available at time t, for a given portfolio position $p \in P$. Assuming that at the end $t = T$, all asset positions must be closed, we define the scrap value as

$$r_T(p, \widehat{x}) = r(p, \widehat{x}, 2) \quad \text{for } t = 0, \dots, T-1, \ p \in P, \ \widehat{x} \in \widehat{\mathcal{X}}. \quad (37)$$

With these definitions, the assumption (24) is satisfied. Furthermore, we easily meet the assumption (35) in view of the following consideration: Note that since all entries of the probability vector $\widehat{x} \in \widehat{\mathcal{X}}$ sum up to one $\vec{1}^\top \widehat{x} = 1$, the constant transaction cost term in (36) can be re-written for $t = 0, \dots, T-1$, $p \in P$, $\widehat{x} \in \widehat{\mathcal{X}}$ as

$$r_t(p, \widehat{x}, a) = R(p,a)\widehat{x}, \quad r_T(p, \widehat{x}) = R(p,1)\widehat{x} \quad (38)$$

with

$$R(p,a) = ((p-2)\rho - c|p - \alpha(p,a)|\vec{1})^\top$$

for $t = 0, \dots, T-1$, $p \in P$, $a \in A$. Because of this linearity, we observe that the positive-homogeneous extensions are obtained by the same formula

$$\tilde{r}_t(p, x, a) = R(p,a)x, \quad \tilde{r}_T(p, x) = R(p,1)x$$

for all $x \in \mathbb{R}^2$, $t = 0, \ldots, T-1$, $p \in P$, $a \in A$, which satisfy (35). Since now our problem fulfills all assumptions required for the convex switching algorithm, we propose an approximate solution via (17) and (18).
Initialization: Having defined the row-rearrangement operator Υ for a grid $G \subset \mathcal{X}$, initialize the matrices representing the value functions form the scrap matrices given in (38)

$$V_T(p) = R(p, 1), \qquad \text{for all } p \in P. \tag{39}$$

Recursion: For $t = T-1, \ldots, 0$ use reward matrices from (38) to calculate for $p \in P$

$$V^E_{t+1}(p) = \frac{1}{2}\big(\Upsilon[V_{t+1}(p) \cdot \Gamma^\top \mathcal{V}(-\Delta)] + \Upsilon[V_{t+1}(p) \cdot \Gamma^\top \mathcal{V}(\Delta)]\big), \tag{40}$$

$$V_t(p) = \sqcup_{a \in A}\big(R(p, a) + V^E_{t+1}(\alpha(p, a))\big), \tag{41}$$

Note that because there are only two disturbance realizations, we perform an exact integration as in (9) using two weights of size 1/2 in the reference measure μ.

Let us consider a numerical illustration. A hidden Markov model with parameters $p_1 = p_2 = 0.8$ and $q_1 = q_2 = 0.9$ generates increments of size $\Delta = 1$ of the asset price whose typical evolution is depicted in the Figure 2. Having supposed transaction costs $c = 0.05$ and introducing an equally

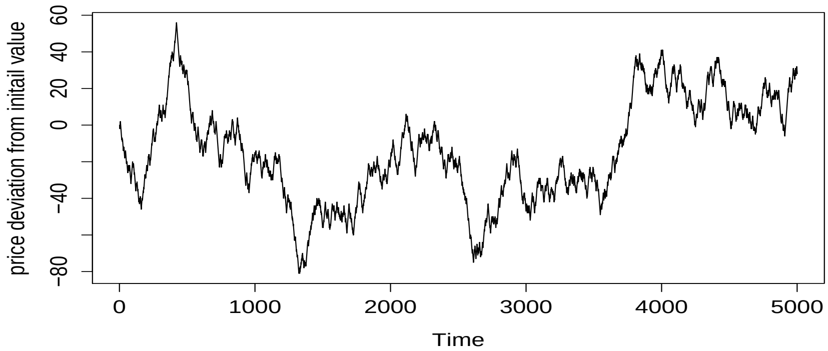

Figure 2: Asset price evolution, adjusted to start at the origin.

spaced grid G of size 101 as

$$G = \{\frac{k}{100}e_1 + (1 - \frac{k}{100})e_2 \,:\, k = 0, \ldots, 100\} \subset \widehat{\mathcal{X}}$$

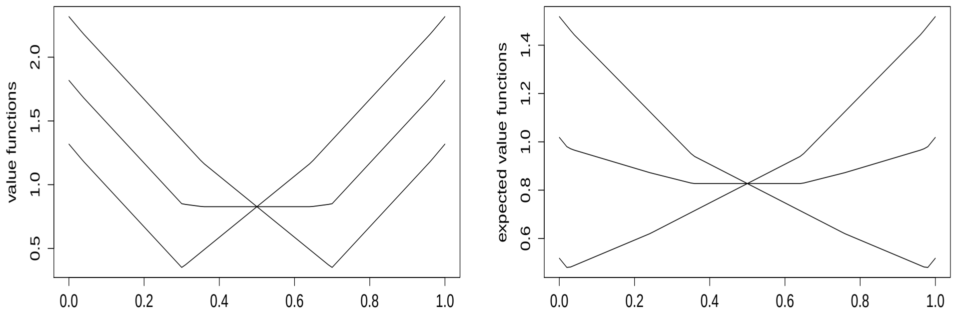

Figure 3: Functions represented by $(V_1(p))_{p=1}^3$ and $(V_2^E(p))_{p=1}^3$.

we determine the matrix representatives $(V_t(p))_{p=1}^3$ of the value functions and their expectations for $t = 10, \ldots, 1$ from the recursions (39) – (41). For the value functions depicted in the Figure 3, an candidate of the optimal decision rule π_1^* is determined by (10) as

$$\pi_1^*(\widehat{x}) = \text{argmax}_{a \in A} \left(\max R(p,a)\widehat{x} + \max V_2^E(\alpha(p,a))\widehat{x} \right), \quad \widehat{x} \in \widehat{\mathcal{X}}.$$

Finally, the Figure 4 shows a joint evolution of the asset price (adjusted to start at the origin, blue line), the portfolio positions obtained by subsequent application of the decision rule π_1^* (gray oscillating line) and the corresponding wealth (green increasing curve). To depict all three plots in the same

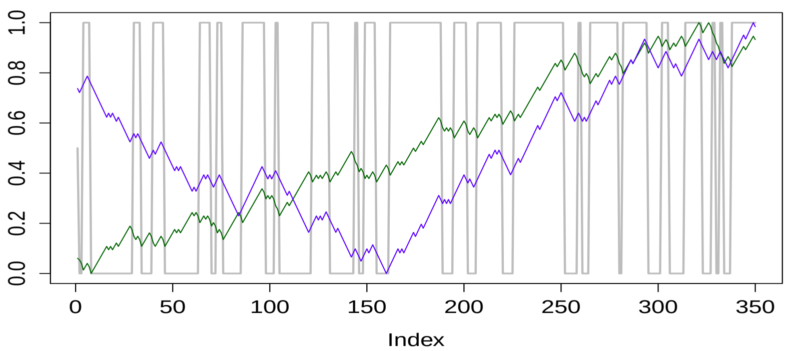

Figure 4: Asset price, portfolio positions and wealth

graph, we have scaled each curve to the interval $[0, 1]$.

5 Conclusion

In this work, we present a novel approach to solving specific switching problems under partial information and show how to apply these results to optimal dynamic asset allocation.

References

[1] N. Bäuerle and U. Rieder. *Markov Decision Processes with Applications to Finance*. Springer, Heidelberg, 2011.

[2] D. P. Bertsekas. *Dynamic Programming and Optimal Control*. Athena Scientific, 2005.

[3] D. P. Bertsekas and J. N. Tsitsiklis. *Neuro-Dynamic Programming*. Athena Scientific, 1996.

[4] P. Kasyanov E. Feinberg and M. Zgurovsky. Partially observable total-cost markov decision process with weakly continuous transition probabilities. *arXiv:1401.2168*, 2014.

[5] R. Elliott and J. Hinz. Portfolio optimization, hidden Markov models, and technical analysis of p&f-charts. 5(4):1–15, 2002.

[6] E. A. Feinberg and A. Shwartz. *Handbook of Markov Decision Processes*. Kluwer Academic, 2002.

[7] M. Hauskrecht. Value-function approximations for partially observable Markov decision processes. *Journal of Artificial Intelligence Research*, 13:33–94, 2000.

[8] O. Hernández-Lerma. *Adaptive Markov Control Processes*. Springer, 1989.

[9] J. Hinz. Optimal stochastic switching under convexity assumptions. *SIAM Journal on Control and Optimization*, 52(1):164–188, 2014.

[10] George E. Monahan. A survey of partially observable Markov decision processes: Theory, models, and algorithms. *Management Science*, 28:1–16, 1982.

[11] W. B. Powell. *Approximate dynamic programming: Solving the curses of dimensionality.* Wiley, 2007.

[12] M.L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming.* Wiley, New York, 1994.

[13] D. Rhenius. Incomplete information in markovian decision models. *Ann. Statist.*, 2:1327 – 1334, 1974.

[14] A. Yushkevich. Reduction of a controlled markov model with incomplete data to a problem with complete information in the case of borel state and control spaces. *Theory Prob. Appl.*, 21:153–158, 1976.