"Markov control with rare state observation": Sensitivity analysis with respect to optimal treatment strategies against HIV-1

Stefanie Winkelmann * Christof Schütte † Max von Kleist ‡

December 17, 2012

Abstract

We present the theory of "Markov decision processes (MDP) with rare state observation" and apply it to optimal treatment scheduling and diagnostic testing to mitigate HIV-1 drug resistance development in resource-poor countries. The developed theory assumes that the state of the process is hidden and can only be determined by making an examination. Each examination produces costs which enter into the considered cost functional so that the resulting optimization problem includes finding optimal examination times. This is a realistic ansatz: In many real world applications, like HIV-1 treatment scheduling, the information about the disease evolution involves substantial costs, such that examination and control are intimately connected.

However, a perfect compliance with the optimal strategy can rarely be achieved. This may be particularly true for HIV-1 resistance testing in resource-constrained countries. In the present work, we therefore analyze the sensitivity of the costs with respect to deviations from the optimal *examination times* both analytically and for the considered application. We discover continuity in the cost-functional with respect to the examination times. For the HIV-application, moreover, sensitivity towards small deviations from the optimal examination rule depends on the disease state.

Furthermore, we compare the optimal rare-control strategy to (i) constant control strategies (one action for the remaining time) and to (ii) the permanent control of the original, fully observed MDP. This comparison is done in terms of expected costs and in terms of life-prolongation. The proposed rare-control strategy offers a clear benefit over a constant control, stressing the usefulness of medical testing and informed decision making.

*Dep. of Mathematics and Computer Science, Arnimallee 6, D-14195 Berlin, Germany (stefanie.winkelmann@fu-berlin.de).

†Dep. of Mathematics and Computer Science, Arnimallee 6, D-14195 Berlin, Germany, (christof.schuette@fu-berlin.de).

‡Dep. of Mathematics and Computer Science, Arnimallee 6, D-14195 Berlin, Germany, (vkleist@zedat.fu-berlin.de).

This indicates that lower-priced medical tests could improve HIV treatment in resource-constrained settings and warrants further investigation.
words:278/250

keywords: information costs, hidden state, bellman equation, optimal therapeutic strategies, diagnostic frequency, resource-poor

AMS subject classifications: 49N30, 60J27, 60J28, 93B07, 90C40, 93E20

1 Introduction

The theory of Markov decision processes (MDP) is a well established tool to analyze situations in which the dynamics of a stochastic process may be influenced by a decision maker. A basic component of a Markov control model is the observability of the process: in standard Markov control theory the process is assumed to be observable at all times, while in the theory of partially observable Markov decision processes the information about the process is incomplete. In both cases, the degree of information is predefined and cannot be influenced by the decision maker. However, in many real world applications (like medical therapies, asset management...) it is possible to decide whether to deduce information or not - and this information is in general not gratis. Instead, the problem is to find the right balance between optimal interaction and reduction of *information costs*.

In this article, we present a model for Markov decision processes with *information costs*. The process is assumed to be continuous in time, while the state space is discrete. We define a suitable cost criterion including the costs of the process and the costs for information and denote the corresponding Bellman equation with reference to [1]. In this model, a control strategy has to declare for each state x not only an action a, but also a *lag time* τ until the next state observation.
Given the optimal strategy, we analyze how small deviations from this strategy affect the cost criterion. The research question is motivated by the fact that a perfect compliance with the optimal strategy may not be accomplished in many real world applications. A meaningful example is the treatment of HIV-1 in Africa, which will serve as an application of the presented control model. Due to limited infrastructure it may not be possible to follow a recommended diagnostic surveillance scheme accurately. In this case, knowledge of potential invariability with respect to patient health damage is required.
Many researchers have studied optimal treatment strategies against HIV, e.g. [2–4]. However, the question of sensitivity with respect to deviations from the optimal examination rule has not been addressed, and rare-state examination/medical testing has not been a part of the previous control approaches. We believe, however, that it is an important pre-requisite for the implementation of such strategies.
After exposing the fundamental components of the HIV-model presented in [1], we specify the optimal therapeutic strategy and analyze the sensitivity of the

optimal costs with respect to changes in the examination *lag times* τ. Furthermore, we assess the benefit of information/interaction by comparing the optimal treatment- and examination strategy of our framework with two opposed modifications of the problem: (i) the case of *constant control* which consists of maintaining one action for the remaining time without any further state examination, and (ii) the case of *permanent control*, which assumes full observability and continuous interaction as in original MDP.
In order to analyze the differences between these approaches, we decompose the costs into components of state-, action- and information costs. As a second criterion for the quality of a therapeutic strategy we consider the probability of death after fixed time intervals for the different approaches.

2 Theory of Markov control with rare state observation

In the following, we describe the Markov control model derived in [1] and complement the theory by a sensitivity analysis with respect to deviations from the optimal strategy.

2.1 The control model

We consider a continuous time Markov control process $(X_t)_{t\geq 0}$ on a discrete state space $\mathcal{S}$. There is a finite set $\mathcal{A}$ of actions that are available in order to influence the process. Given action $a\in\mathcal{A}$, the dynamics of the process are defined by the generator L_a where $L_a(x,y)\geq 0$ is the transition rate for a transition from $x\in\mathcal{S}$ to $y\in\mathcal{S}$, $y\neq x$, while $L_a(x,x)$ satisfies $L_a(x,x)=-\sum_{y\neq x} L_a(x,y)$. The cost function $c:\mathcal{S}\times\mathcal{A}\to[0,\infty)$ denotes the costs produced by the process per unit of time depending on the actual state and the chosen action. In the application, the states describe the health status of a patient, while the actions refer to different medical treatments. The cost function measures the costs of the treatments as well as the health damage to the patient.

In opposition to original Markov control theory, we assume that the process cannot be continuously observed and influenced. Instead, each examination of the process produces costs $k_{\text{info}}>0$ which enter the cost functional, such that a fundamental part of the optimization problem is to determine optimal *examination times*. Controlling the process then proceeds according to the following structure: Starting with some known state $X_0\in\mathcal{S}$ at some examination time $t_0\geq 0$ one chooses an action $a\in\mathcal{A}$ as well as an *examination lag time* $\tau(X_0)>0$ defining the next examination time $t_1=t_0+\tau(X_0)>t_0$. During the time interval $(t_0,t_1]$ the (random) behavior of the process (X_t) is fully described by the infinitesimal generator L_a and produces costs according to the cost function $c(\cdot,a)$. We do not observe this behavior but only determine the state X_{t_1} of the process at time t_1. For this information

expenses k_{info} accrue. Knowing the new state X_{t_1} at time t_1, we choose again an action and a lag time and the procedure restarts. During a time interval $[t_j, t_{j+1})$ the action is fixed, i.e., it can only be changed after examination.

In this context, a *strategy* is a function

$$u : \mathcal{S} \to \mathcal{A} \times (0,\infty], \quad x \mapsto u(x) = \big(a(x), \tau(x)\big) \tag{2.1}$$

giving for each state $x \in \mathcal{S}$ both an action $a \in \mathcal{A}$ and an examination lag time $\tau > 0$. The lag time τ is allowed to be infinite which is appropriate in situations where state examinations/interaction cannot change anything (e.g. because the actual state is absorbing) and which in the same time will guaranty the existence of an optimal strategy.
The set of all strategies is denoted by $\mathcal{U}$.

2.2 Cost criterion and Bellman equation

As an optimality criterion we choose expected discounted costs over an infinite-horizon. Given a strategy $u \in \mathcal{U}$, an initial state $x \in \mathcal{S}$ and a discount factor $\lambda > 0$, these are defined by

$$J(x,u) = \mathbb{E}_x^u \left(\sum_{j=0}^{\infty} e^{-\lambda t_j} \left(C\big(X_{t_j}, a(X_{t_j}), \tau(X_{t_j})\big) + e^{-\lambda \tau(X_{t_j})} k_{\text{info}} \right) \right) \tag{2.2}$$

where $\mathbb{E}_x^u$ stands for the expectation value with respect to the measure determined by x and u, while

$$C(x,a,\tau) := \mathbb{E}_x^a \left(\int_0^{\tau} e^{-\lambda s} c(X_s, a) ds \right) \tag{2.3}$$

are the expected discounted costs for the time interval $(0,\tau]$ when starting in state $x \in \mathcal{S}$ and choosing action $a \in \mathcal{A}$. Moreover, it holds $t_{j+1} = t_j + \tau(X_{t_j})$.

The corresponding value function

$$V(x) := \min_{u \in \mathcal{U}} J(x,u) \tag{2.4}$$

is characterized by the Bellman equation

$$V(x) = \min_{a \in \mathcal{A}, \tau \in [0,\infty]} \left(C(x,a,\tau) + e^{-\lambda \tau} \left(k_{\text{info}} + T_{a,\tau} V(x) \right) \right) \tag{2.5}$$

where $T_{a,\tau} : \mathbb{R}^{|\mathcal{S}|} \to \mathbb{R}^{|\mathcal{S}|}$, $T_{a,\tau} v := e^{L_a \tau} \cdot v$ is the transition matrix on $\mathcal{S}$ for some fixed lag time $\tau > 0$ and action a, see [1].[1]

[1] For $\tau = \infty$ the right handside of 2.5 is given by $C(x,a,\infty) := \mathbb{E}_x^a \big(\int_0^{\infty} e^{-\lambda s} c(X_s, a) ds \big)$.

2.3 Sensitivity analysis w.r.t. deviation from optimal examination rule

In the following we will be interested in computing the cost functional $J(x,u)$ for strategies that slightly differ from the optimal strategy u^*. In order to give a compact formula for J we introduce a few notations:
Given a strategy $u(x)=(a(x),\tau(x))$, define a discount vector $e_\tau \in \mathbb{R}^{\mathcal{S}}$ by $e_\tau(x) := e^{-\lambda\tau(x)}$ and a diagonal discount matrix $D_\tau \in \mathbb{R}^{\mathcal{S},\mathcal{S}}$ by $D_\tau(x,x) := e_\tau(x)$ and $D_\tau(x,y) := 0$ for $x \neq y$. Let $P \in \mathbb{R}^{\mathcal{S},\mathcal{S}}$ with $P(x,y) := \left(e^{L_{a(x)}\cdot\tau(x)}\right)(x,y)$ be the transition rule under u for the observed process $X_{t_0}, X_{t_1}, ...$, and define $C_u \in \mathbb{R}^{\mathcal{S}}$ by $C_u(x) := C(x,a(x),\tau(x))$.

Lemma 2.1. *The cost functional $J = J(\cdot,u)$ (as a vector in $\mathbb{R}^{\mathcal{S}}$) for a strategy $u \in \mathcal{U}$ is given by*

$$J = (Id - D_\tau P)^{-1}(C_u + k_{\text{info}} e_\tau). \tag{2.6}$$

Proof. In analogy to the Bellmann equation (see eq. (2.5)), J fulfills the recursion

$$J(x) = C(x,a(x),\tau(x)) + e^{-\lambda\tau(x)}\left(k_{\text{info}} + T_{a(x),\tau(x)}J(x)\right)$$

which can be written in the form

$$J = C_u + k_{\text{info}} e_\tau + D_\tau P J.$$

This is equivalent to

$$(Id - D_\tau P)J = C_u + k_{\text{info}} e_\tau.$$

It remains to take the inverse of $Id - D_\tau P$, which exists by the following argumentation. If the matrix $Id - D_\tau P$ was not invertible, the equation

$$(Id - D_\tau P)v = 0 \tag{2.7}$$

would have a solution $v \in \mathbb{R}^{|\mathcal{S}|} \neq 0$. As D_τ is a diagonal matrix with diagonal entries $0 < e^{-\lambda\tau(x)} < 1$, it's inverse D_τ^{-1} exists and is again diagonal with $D_\tau^{-1}(x,x) = e^{\lambda\tau(x)} > 1$. We rewrite (2.7) as $Pv = D_\tau^{-1}v$ and take the maximum norm on both sides. As P is a transition matrix, the entries of Pv are convex combinations of the entries of v, such that it holds $||Pv||_\infty \leq ||v||_\infty$. On the other hand, it holds $||D_\tau^{-1}v||_\infty > ||v||_\infty$, as each entry of v is multiplied by a constant > 1. Together we get

$$||v||_\infty \geq ||Pv||_\infty = ||D_\tau^{-1}v||_\infty > ||v||_\infty,$$

a contradiction to $v \neq 0$. □

Theorem 2.2 (Continuity of J with respect to τ)**.** *The cost functional J as a function of the strategy $u(x) = (a(x),\tau(x))$ is continuous with respect to the parameter $\tau(x)$ for all $x \in \mathcal{S}$.*

Remark 2.3. *The continuity of J with respect to some $\tau(x)$, $x \in \mathcal{S}$ fixed, refers to all components $J(y,u)$, $y \in \mathcal{S}$, of J. In other words, "small" modifications in $\tau(x)$ lead to "small" modifications in $J(y,u)$ for <u>all</u> states $y \in \mathcal{S}$, and not only for $y = x$.*

Proof. The continuity of J with respect to τ follows from the continuity of the expressions $e^{-\lambda\tau}$, $e^{L\tau}$ and $C(x,a,\tau)$ which are the τ-dependent components in

$$J = (Id - D_\tau P)^{-1}(C_u + k_{\text{info}} e_\tau),$$

compare equation 2.6. □

3 Application

Given the theoretical ansatz of section 2, we will now formulate a model for HIV-dynamics and analyze the valuefunction with respect to deviations from the optimal strategy.

3.1 HIV Model

Action Space. In line with [1] we choose the set of treatments $\mathcal{A} = \{a_\emptyset, a_1, a_2\}$, where $a_\emptyset$ denotes the absence of medical intervention, while a_1 and a_2 denote the application of two distinct treatment lines. This choice is motivated by the fact that in the sequel we will focus on HIV treatment in resource-constrained settings in which only two treatment lines are available (a_1 & a_2).

State Space. In brief, the HIV-model contains four lumped states for each virus type: The respective virus type can either be absent, or present in low-, medium- or high copy numbers, denoted by 0, ℓ, m and h respectively. The ℓ-states are reflecting states, which are justified by the inability to eradicate HIV [5, 6] and the h-states are reflecting states, because there is a maximum carrying capacity of the system.

According to their treatment susceptibility, our model further distinguishes 4 viral strains M ("mutants"): a strain WT (wild type) that is susceptible to all treatment lines, a strain R1 which is susceptible to treatment 2 (a_2), but unaffected by (resistant to) treatment 1 (a_1), a strain R2 that is susceptible to a_1, but unaffected by a_2 and a highly resistant strain HR which is resistant to all treatments ($a_1 \& a_2$).

We consider all permutations of viral strains $M \in \{\text{WT}, \text{R1}, \text{R2}, \text{HR}\}$ and respective copy numbers $n_C(M) \in \{0, \ell, m, h\}$ and patient death ✠, resulting in state space dimension $|\mathcal{S}| = 4^4 + 1 = 257$, with $\mathcal{S} = \{0, \ell, m, h\}^4 \cup ✠$.

In order to describe a state $x \in \mathcal{S}$ we will use the compact vector notation

$$x = \big[n_C(\text{WT}), n_C(\text{R1}), n_C(\text{R2}), n_C(\text{HR})\big].$$

For example, the state $x = \big[0, h, 0, \ell\big]$ describes the absence of wild type strains, a $h$igh number of R1-mutants, the absence of R2-mutants and a ℓow number of highly resistant mutants. The proposed Markov model of HIV-dynamics [1] is particularly suited to describe the long term dynamics of drug resistance development after treatment application.

Generator Entries. The distinct treatments $a \in \mathcal{A}$ are related to distinct generators L_a of our HIV-model. The basic transitions between copy number states $n_C(M)$ are shown in Fig. 1 and exemplified for the wild type strain WT below.

$$\big[\ell,*,*,*\big] \underset{\delta_m}{\overset{k_\ell(a,\mathrm{WT})}{\rightleftarrows}} \big[m,*,*,*\big], \quad \big[m,*,*,*\big] \underset{\delta_h}{\overset{k^{\mathrm{WT}}_{m,a}}{\rightleftarrows}} \big[h,*,*,*\big], \tag{3.8}$$

$$\big[h,*,*,*\big] \xrightarrow{d_h} \maltese, \qquad \big[m,*,*,*\big] \xrightarrow{d_m} \maltese, \qquad \big[\ell,*,*,*\big] \xrightarrow{d_\ell} \maltese, \tag{3.9}$$

where $*$ indicates an arbitrary number of the respective virus strain (R1,R2 and HR in the example above). The parameters $k_{\ell,a}$ and $k_{m,a}$ denote the reaction propensities of going from copy number ℓ to copy number m and from copy number m to copy number h respectively (viral growth), which are decreased depending on the treatment $a \in \{a_\emptyset, a_1, a_2\}$. The parameters δ_m and δ_h are independent of the treatment and denote the reaction propensities for going from copy number m to copy number ℓ and from copy number h to copy number m respectively (virus elimination). The parameters $d_h > d_m > d_\ell$ denote the propensity for the death of the patient. These parameters are unaffected by the treatments, as well [1]. We assume that high viral burden (states h and m respectively) increases the risk of death, whereas d_ℓ equals the propensity for "natural death". The propensity for death was computed according to $d = 1/(\text{residual life expectancy})$, and is exemplified in [1].

The considered transitions (mutations) between viral strains M are depicted in Fig 1. Specifically, mutation generates a ℓow number of viral particles from either a $m$edium or $h$igh number of viruses belonging to a distinct strain. Exemplified for the wild type strain WT those are:

$$\begin{aligned}
\big[h,0,*,*\big] &\xrightarrow{\mu^{\mathrm{R1}}_h(a,\mathrm{WT})} \big[h,\ell,*,*\big], \quad \big[m,0,*,*\big] \xrightarrow{\mu_{m,\mathrm{R1}}(1-\eta(a,\mathrm{WT}))} \big[m,\ell,*,*\big] \\
\big[h,*,0,*\big] &\xrightarrow{\mu^{\mathrm{WT,R2}}_{h,a}} \big[h,*,\ell,*\big], \quad \big[m,*,0,*\big] \xrightarrow{\mu^{\mathrm{WT}\to\mathrm{R2}}_{m,a}} \big[m,*,\ell,*\big] \\
\big[0,h,*,*\big] &\xrightarrow{\mu^{h}_{\mathrm{R1}\to\mathrm{WT}}(a)} \big[\ell,h,*,*\big], \quad \big[0,m,*,*\big] \xrightarrow{\mu^{\mathrm{R1,WT}}_{m,a}} \big[\ell,m,*,*\big] \\
\big[0,*,h,*\big] &\xrightarrow{\mu^{\mathrm{R2,WT}}_{h,a}} \big[\ell,*,h,*\big], \quad \big[0,*,m,*\big] \xrightarrow{\mu^{\mathrm{R2,WT}}_{m,a}} \big[\ell,*,m,*\big].
\end{aligned}$$

where the first two lines indicate mutation arising from the wild type strain and the remaining two lines indicate mutations yielding the wild type strain. The parameters $\mu_{h,\mathrm{R1},a}$ and $\mu_{h,\mathrm{R2},a}$ denote the propensity for the emergence- and disappearance of a mutation that confers drug resistance to treatment 1 or 2 (a_1,a_2), respectively, emanating from copy number state h. Analogously $\mu_{m,\mathrm{R1},a}$ and $\mu_{m,\mathrm{R2},a}$ denote the propensity for the emergence- and disappearance of a mutation emanating from copy number states m. Note, that we consider only the following mutations: $\mathrm{WT} \leftrightarrow \mathrm{R1}$, $\mathrm{WT} \leftrightarrow \mathrm{R2}$, $\mathrm{R1} \leftrightarrow \mathrm{HR}$ and $\mathrm{R2} \leftrightarrow \mathrm{HR}$, which is motivated by the fact, that a direct transition from $\mathrm{WT} \leftrightarrow \mathrm{HR}$ is very unlikely, because the genetic distance between the two viral strains is too large to be overcome at once.

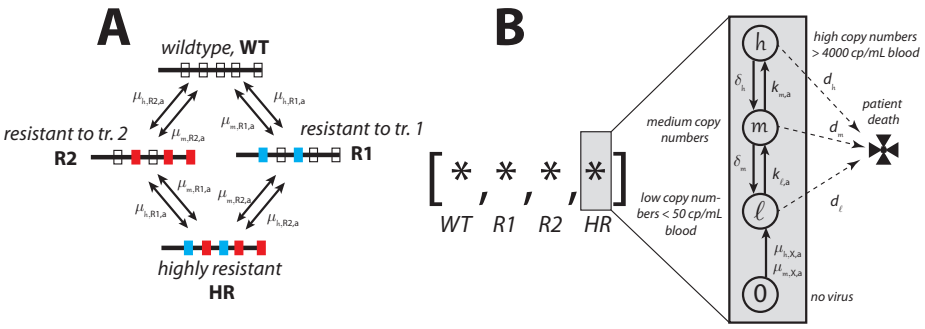

Figure 1: **Simplified HIV Model** A: Transitions in between viral strains M in terms of mutations in the virus' genome. The horizontal line with the small boxes shall schematically represent the viral genome and the codons (boxes) that are relevant for drug resistance development. A blank box shall indicate the absence of any mutations, whereas the blue- and red coloring indicates a pattern of mutations that confers resistance to treatment line one (blue coloring, virus type **R1**) and -two (red coloring, virus type **R2**). B: Transitions between copy number states n_C.

The effect of treatment a is considered in the following way:

$$k_{\ell,a}^{M} = \Big(1-\eta(a,M)\Big)k_{\ell,\emptyset}, \quad k_{m,a}^{M} = \Big(1-\eta(a,M)\Big)k_{m,\emptyset}, \tag{3.10}$$

$$\mu_{h,a}^{M,\tilde{M}} = \Big(1-\eta(a,M)\Big)\mu_{h,\emptyset}^{\tilde{M}}, \quad \mu_{m,a}^{M,\tilde{M}} = \Big(1-\eta(a,M)\Big)\mu_{m,\emptyset}^{\tilde{M}}. \tag{3.11}$$

The parameter $\eta(a,M)$ is a constant that denotes the efficacy of treatment $a \in \{a_\emptyset, a_1, a_2\}$ on viral strain $M \in \{\mathrm{WT}, \mathrm{R1}, \mathrm{R2}, \mathrm{HR}\}$; i.e if strain M is susceptible to treatment, then $0 < \eta \leq 1$ and if the viral strain M is insusceptible to treatment then $\eta = 0$. The index $\emptyset$ denotes the absence of medical intervention ($a_\emptyset$), hence $\eta = 0$. The parameters $k_{\ell,\emptyset}$, $k_{m,\emptyset}$, $\mu_{h,\emptyset}$ and $\mu_{m,\emptyset}$ denote the growth rates and mutation rate in the absence of medical intervention (see Table A1).

The model building process as well as the process of parameter estimation for the model are described in [1] in more detail. Final model parameters are shown in the Table A1 (appendix).

Costs. We assume the cost function c to be of the form

$$c(x,a) = c_{\mathcal{S}}(x) + c_{\mathcal{A}}(a)$$

where $c_{\mathcal{S}}$ measures the costs related to the health damage of the individual while $c_{\mathcal{A}}$ describes the treatment costs. The aim is thus to find a cost-effective strategy that ensures good health of a patient. We parameterized our model in terms of values that are representative for South Africa. The costs $c_{\mathcal{S}}(x)$ of being in the respective states $x \in \mathcal{S}$ were computed based on the average productivity loss times the average daily monetary contribution of one individual (assessed in terms of daily per capita GDP), where death is interpreted in terms of a

complete loss in productivity, see [1]. The examination costs k_{info} accrue each time a medical test is made. The direct costs for treatment and examination are displayed in appendix Table A1 together with the indirect costs for health damage.
Given this structure of the cost function, also the value function V (optimal total discounted costs) can be split up into parts of state costs $V_{\mathcal{S}}$ (health damage of a patient), action costs $V_{\mathcal{A}}$ (costs for medical treatment) and information costs V_{info} (costs for medical tests), see [1].

3.2 Optimal strategy in a resource-poor setting

We applied the theory described in Section 2 to the model presented in Section 3.1 using the cost-parameters in Table A1 (appendix), which are representative for South Africa. We computed the cost-optimal treatment- and diagnostic strategy using a modified version of the standard policy-iteration algorithm [7]. In our application, we set $\tau_{\min} = 1$ days and $\tau_{\max} = 5000$ days in order to numerically solve the optimization problem. The computed optimal strategy is shown in Table 1. In brief, there are two states, in which diagnostic testing is indicated: state $\big[\{m\},0,0,0\big]$ and $\big[\{h\},0,0,0\big]$. A switch to treatment line a_2 is indicated, if viral strain R1 is present, i.e. after resistance development to treatment line a_1. I.e., our computation indicates that it would be cost-optimal to implement a sparse diagnostic surveillance depending on the health status of the patient. Currently, this is not standard-of-care in South Africa.

	state			
	$\big[\{m\},0,0,0\big]$	$\big[\{h\},0,0,0\big]$	$\big[*,\{\ell,m,h\},0,0\big]$	otherwise
action	a_1	a_1	a_2	a_1
τ	11	45	$\geq \tau_{\max}$	$\geq \tau_{\max}$

Table 1: **Optimal strategy.** Calculated optimal strategy for the resource-poor settings (South Africa) giving the treatment, the *examination lag time* τ (in days) and the valuefunction (in US$) depending on the state of the patient. For clarity reasons, states are merged according to their related treatment choice.

However, an exact implementation of the proposed diagnostic surveillance may not be possible, because patients may not be able to perfectly comply with the indicated scheme (shown in Table 1). Implementation of a diagnostic surveillance scheme may be further complicated in resource-constrained settings due to infra-structural deficiencies. We will in the next section assess the sensitivity of the valuefunction with respect to deviations from the optimal examination rule and in the section thereafter, we will compare the "optimal strategy with rare state observations" with the two opposed cases in which the action is either constant over all times (which corresponds to infinite lag times) or continuously adapted to the process as in original Markov control theory without examination costs (which stands for infinitesimally small lag times).

3.3 Sensitivity analysis w.r.t. deviation from optimal examination rule

We have shown in Theorem 2.2 that the cost functional J is continuous with respect to the lag time parameters $\tau(x)$ for all $x \in \mathcal{S}$. In our application, we have only two states ($[m,0,0,0]$ and $[h,0,0,0]$) for which diagnostic testing is indicated. We therefore computed the impact of τ-variations around the optimum in Fig. 2A & B for the indicated states by using eq. (2.6). For state $[h,0,0,0]$, the total costs sharply rise if τ is decreased or -increased (solid blue line in Fig. 2A) in relation to its optimum value τ^* (solid dot). The increase of $J(x)$ upon increasing values of τ are paralleled by an increase in $J_{\mathcal{S}}(x)$ (the "pure" state costs; dashed red line). When decreasing τ, the opposite is true, namely $J_{\mathcal{S}}(x)$ decreases, but the overall costs $J(x)$ increase. Note that the slope of $J_{\mathcal{S}}(x)$ (dashed red line) corresponds to the cost-increase attributed to patient health damage.

Although we observe a very sensitive response towards changes in τ for state $[h,0,0,0]$, for the other state, $[m,0,0,0]$, we get much less sensitivity towards deviations from τ^* (see Fig. 2B, solid blue line and solid black dot). In particular, upon increases in τ, total- (solid blue line) and state costs (dashed red line) are only marginally increased.

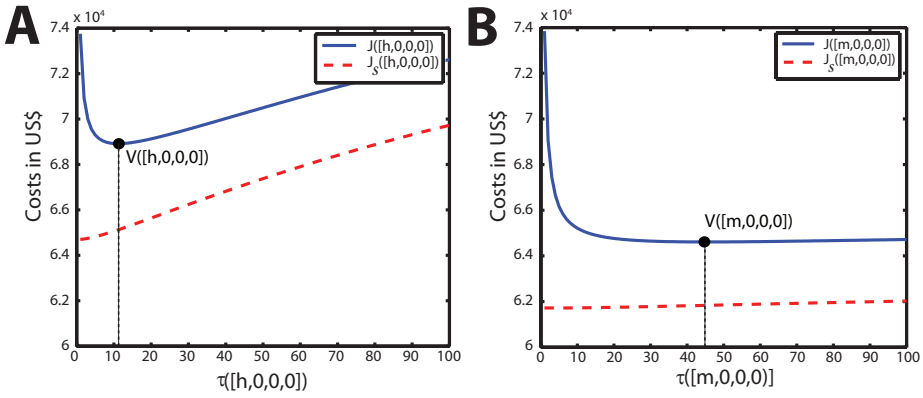

Figure 2: **Sensitivity with respect to τ.** Cost functional $J(x,u)$ and $J_{\mathcal{S}}(x,u)$ for $x=[h,0,0,0]$ (left graphic) and $x=[m,0,0,0]$ (right graphic) with u varying in $\tau([h,0,0,0])$ or $\tau([m,0,0,0])$ (while being optimal in all other parameters).

A summary in terms of a two-dimensional contour plot, which takes variation in both $\tau\Big([h,0,0,0]\Big)$ and $\tau\Big([m,0,0,0]\Big)$ simultaneously into account, is shown in Fig. 3 and confirms the observations made from Fig. 2A&B, indicating that if patients have a high viral load (state $[h,0,0,0]$), they should strictly comply with the optimal strategy. If we focus on the potential health damage to the patient (dashed red lines in Fig. 2), we can conclude that there is little margin if diagnostic testing is behind schedule for patients that are in state

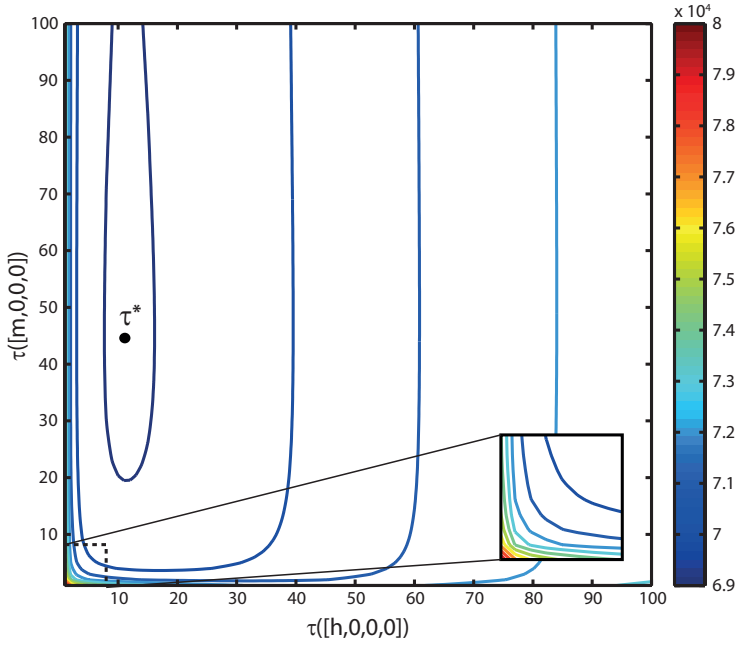

Figure 3: **Sensitivity with respect to τ.** Cost functional $J(x,u)$ for $x=[h,0,0,0]$ and u varying in $\tau([h,0,0,0])$ (x-axis) and $\tau([m,0,0,0])$ (y-axis), while being optimal in all other parameters.

$[h,0,0,0]$ (high virus load). For state $[m,0,0,0]$ belated diagnosis will have little consequences for the health of the patient. In a setting with constrained resources this means that patients with high virus loads should be prioritized for subsequent diagnosis over patients with less virus.

4 Comparison with constant control and original Markov control theory

In this section we will compare the optimal strategy given in section 3.2 to two extreme cases: Namely, (i) the process under constant control, i.e. the action is fixed for all times, and (ii) the case where we can permanently observe the process and adapt the action, i.e. as in the original Markov control theory. The resulting costs and the probability of death (which treatment is intended to prevent) will be computed.

(i) The process under constant control refers to the condition in which an action is initially chosen and maintained for the remaining time. It can be associated to infinitesimal large information costs ($k_{\text{info}}=\infty$) which make testing unaffordable ($\tau(x)=\infty$ for all $x\in\mathcal{S}$). In the presented model this situation refers

to the choice of either $a_\emptyset$, a_1, or a_2 for all times. The corresponding costs are given by

$$J_a(x) := \mathbb{E}_x^a \left(\int_0^\infty e^{-\lambda s} c(X_s, a)\, ds \right) \tag{4.12}$$

with $a \in \{a_\emptyset, a_1, a_2\}$. Especially, the choice of $a = a_\emptyset$ stands for "natural" disease process without medical intervention.

(ii) We assume that the process is all the time freely observable ($k_{\text{info}} = 0$) and that actions can immediately be adapted. In our model, this immediately results in vanishing lag times, i.e. $\tau(x) = 0$ for all $x \in \mathcal{S}$, such that the discrete structure of the cost functional (eq. (2.2)) gets lost and the given Bellman equation (2.5) is not anymore suited to characterize the optimal strategy. Instead, this situation corresponds to an original (continuous time) Markov control process. Here, a deterministic stationary strategy is given by a function $f : \mathcal{S} \to \mathcal{A}$, declaring for each state which action to chose. The corresponding costs are given by

$$\hat{J}(x, f) := \mathbb{E}_x^f \left(\int_0^\infty e^{-\lambda s} c(X_s, f(X_s))\, ds \right) \tag{4.13}$$

fulfilling

$$\lambda \hat{J}(x, f) = c(x, f(x)) + L_{f(x)} \hat{J}(x, f), \tag{4.14}$$

see [8]. The optimal strategy f^* and the value function $\hat{V}(x) = \hat{J}(x, f^*) = \min_f \hat{J}(x, f)$ are characterized by the Bellman equation

$$\lambda \hat{V}(x) = \min_{a \in \mathcal{A}} \left(c(x, a) + L_a \hat{V}(x) \right), \tag{4.15}$$

with $f^*(x) = \operatorname{argmin}_{a \in \mathcal{A}} \left(c(x, a) + L_a \hat{V}(x) \right)$, see [8].

The cost functional $\hat{J}$ correlates with the cost functional J defined in 2.2 in the following way: Given a strategy $f : \mathcal{S} \to \mathcal{A}$ of the original Markov control process, consider the strategy $u \in \mathcal{U}$ with $u(x) = (f(x), \tau_0)$ for a $\tau_0 > 0$ independent of x. Setting $k_{\text{info}} = 0$ in 2.6 and taking the limit $\tau_0 \to 0$ gives

$$\begin{aligned} J &= (Id - D_{\tau_0} P)^{-1} C_u \\ &= \frac{\tau_0}{\tau_0} (Id - D_{\tau_0} P)^{-1} C_u \\ &= \left(\frac{1}{\tau_0} (Id - D_{\tau_0} P) \right)^{-1} \frac{1}{\tau_0} C_u \\ &\overset{\tau_0 \to 0}{\longrightarrow} (\lambda Id - L_f)^{-1} c_f \\ &= \hat{J} \end{aligned}$$

with $L_f(x, y) = L_{f(x)}(x, y)$ and $c_f(x) = c(x, f(x))$, where the limit is determined by considering the series expansion of the matrix exponential $P(x, y) = e^{L_{f(x)} \tau_0}(x, y)$ and the last equality follows by (4.14).

	$k_{\text{info}}=\infty$ constant control			$k_{\text{info}}=500$ MDP wt. rare obs.	$k_{\text{info}}=0$ original MDP
	$a_\emptyset$	a_1	a_2		
total costs ($V_{\mathcal{S}}+V_{\mathcal{A}}+V_{\text{info}}$)	107 350	76 790	70 030	69 149	61 420
netto costs ($V_{\mathcal{S}}+V_{\mathcal{A}}$)	107 350	76 790	70 030	66 855	61 420
state costs ($V_{\mathcal{S}}$)	107 350	76 024	66 940	65 116	59 589
$\mathbb{P}(X_{3y}=✠\|X_0=[h,0,0,0])$	0.44	0.22	0.15	0.15	0.13
$\mathbb{P}(X_{5y}=✠\|X_0=[h,0,0,0])$	0.63	0.34	0.24	0.23	0.21
$\mathbb{P}(X_{15y}=✠\|X_0=[h,0,0,0])$	0.95	0.69	0.62	0.60	0.54

Table 2: The netto costs are given by $J_a([h,0,0,0])$ in the case of constant control, by $V_{\mathcal{S}}([h,0,0,0])+V_{\mathcal{A}}([h,0,0,0])$ in the case of MDP with rare state observation and by $\hat{V}([h,0,0,0])$ in the case of original MDP.
The probability of death $\mathbb{P}(X_t=✠|X_0=[h,0,0,0])$ after 3, 5 or 15 years when starting in state $[h,0,0,0]$ was computed by analytically solving the Kolmogorov equations in the case of constant control and in the original MDP setting, where we used the generator under optimal control $L^*(x,y)=L_{a^*(x)}(x,y)$. In the MDP with rare state observation setting, we approximated $\mathbb{P}(X_t=✠|X_0=[h,0,0,0])$ using a well-established Monte-Carlo-Method [9].

Obviously, it holds

$$J_a(x)\geq V(x)\geq \hat{V}(x)\quad \forall a\in\mathcal{A}, \tag{4.16}$$

where V is the valuefunction defined in (2.4): The first inequality follows from the fact that the strategy of constant control is contained in the set of strategies $\mathcal{U}$ over which we minimize in section 2.2, while the second inequality is due to the fact that a continuous adaption of the optimal action choice combined with cost free state information ($k_{\text{info}}=0$) can only lead to an improvement of the total costs. The same is true if we consider - instead of the total optimal costs V - the netto costs $V_{\text{netto}}=V_{\mathcal{S}}+V_{\mathcal{A}}$ which are the total costs without information costs. As both J_a and $\hat{V}$ do not contain any information costs (in setting (i) there are no tests at all and in setting (ii) information is for free), considering the netto costs V_{netto} instead of V is better suited to make a comparison.
Table 2 shows the (netto) costs and the probability of death for setting (i) and (ii) as well as for $k_{\text{info}}=500$ (compare appendix Table A1).

We can make the following observations. In accord with (4.16), the costs of the optimal MDP scheme with rare state observation go below those of any constant control, while they exceed the costs of the original MDP scheme with permanent optimal control. There is a huge difference between the costs resulting of an absence of medical treatment ($a_\emptyset$ at all times) and those arising under constant control with a_1 or a_2. While the values of the netto costs all significantly differ from each other, the value of the total optimal costs $V([h,0,0,0])=69149$ arising from optimal control with rare state observation is very close to the costs under constant control with a_2. This means that in terms of the total costs, the optimal strategy with rare state observation is

only slightly better than a "blind"/constant control which could challenge the utility of medical testing. In fact, it is the reduction of action- and state costs which justifies the medical tests.

In terms of survival benefit, the optimal MDP scheme with rare state observation is better than the absence of medical intervention ($a_\emptyset$ at all times) and better than a constant treatment with only therapy line a_1, however, it is only slightly better than a constant treatment with a_2 for the time horizon analyzed (3, 5 and 15 years), which is however much more expensive in terms of treatment costs (netto costs − state costs). Also, for larger time horizons, the differences between constant treatment with only a_2 and the optimal MDP scheme with rare state observation are expected to further increase.
Again, the biggest difference in the probability of death can be found when comparing the absence of medical intervention with all other considered strategies. This emphasizes that the fundamental step is to start a medical treatment, while the details of the treatment strategy are secondary.

Best in terms of survival and costs is, as expected, the permanent control of the original Markov Control problem. However, for many applications (as the one considered here) the assumption of continuous and cost-free observation and interaction is unrealistic. It is a matter of fact that information itself has a worth, and it makes sense to take account of this worth. This is exactly what is done by our model, where the value of the information costs on the optimization problem is determined by the value of k_{info}.

5 Conclusion

We presented a model for continuous time Markov decision processes that can be observed and influenced at variable discrete time points. Each observation produces costs which enter the cost functional such that the optimization problem consists of finding for each state an action and a lag time determining the date for the next examination. Given an adequate cost criterion we discovered a continuity with respect to the lag times of all states which means that "small" deviations from the optimal strategy do not lead to a huge increase (jump) in the costs.

We exemplified this continuity for HIV-therapies in Africa where a high prevalence of HIV-infections coheres with a restricted infrastructure, which complicates an exact adherence to testing dates. We found that sensitivity of the expected costs towards "small" deviations from the optimal lag times τ^* depend on the considered state: For the more critical state $[h,0,0,0]$ (indicating a high copy number of wild type virus) the costs sharply increase. However, for the less critical state $[m,0,0,0]$ (indicating a medium copy number of wild type virus) the response towards changes in the lag time is not that problematic. This means that a patient with high virus load should strictly comply with the

next examination date, while a patient with medium virus load is more flexible.

In order to overview the range of possible costs we gave an upper bound by considering constant strategies (no further adjustment of the action) and a lower bound by calculating the costs for the original Markov control problem (continuous interaction). We distinguished state-, action- and information costs and found out that as for the netto costs (state costs + action costs) the optimal strategy of the new model is clearly better than any constant strategy and worse than the optimal continuous strategy (original MDP). The differences in the probability of death are not that significant. It is mainly the absence of medical intervention that dramatically decreases the chances of survival.

Although the permanent control without information costs naturally delivers the best results, a more realistic ansatz is to take into account the costs of state testing - which is done by the proposed model.

As demonstrated in [1], the optimal strategy and the valuefunction strongly depend on the cost parameter k_{info}. An interesting future problem is to find out about monotony and continuity of the time parameters and the valuefunction with respect to this parameter - both within the framework of the new model (at $k_{\text{info}} > 0$) and with respect to the original Markov model (at $k_{\text{info}} = 0$).

References

[1] S. Winkelmann, C. Schütte, and Max von Kleist. Markov control processes with rare state observation: Theory and application to treatment scheduling in HIV-1. *Comm. Math. Sci.*, submitted, 2012.

[2] R. Luo, M. J. Piovoso, J. Martinez-Picado, and R. Zurakowski. Optimal antiviral switching to minimize resistance risk in HIV therapy. *PLoS One*, 6(11):e27047, 2011.

[3] S. M. Shechter, M. D. Bailey, and A. J. Schaefer. A modeling framework for replacing medical therapies. *IIE Transactions*, 40:861–869, 2008.

[4] E. A. Hernandez-Vargas, R. H. Middleton, and P. Colaneri. Optimal and MPC switching strategies for mitigating viral mutation and escape. In *Preprints of the 18th IFAC World Congress Milano (Italy) August 28 - September 2*, 2011.

[5] D. Finzi, J. Blankson, J. D. Siliciano, J. B. Margolick, K. Chadwick, T. Pierson, K. Smith, J. Lisziewicz, F. Lori, C. Flexner, T. C. Quinn, R. E. Chaisson, E. Rosenberg, B. Walker, S. Gange, J. Gallant, and R. F. Siliciano. Latent infection of CD4+ T cells provides a mechanism for lifelong persistence of HIV-1, even in patients on effective combination therapy. *Nat Med*, 5(5):512–517, May 1999.

[6] O. Lambotte, M.-L. Chaix, B. Gubler, N. Nasreddine, C. Wallon, C. Goujard, C. Rouzioux, Y. Taoufik, and J.-F. Delfraissy. The lymphocyte HIV reservoir in patients on long-term HAART is a memory of virus evolution. *AIDS*, 18(8):1147–1158, May 2004.

[7] R. A. Howard. *Dynamic programming and Markov processes*. MIT Press, 1960.

[8] X. Guo and O. Hernandez-Lerma. Continuous-time markov decision processes: Theory and applications. In *Stochastic Modelling and Applied Probability*. Springer, Heidelberg, 2009.

[9] D.T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J Comp Phys*, 22:403–434, 1976.

param.	[unit]	value	param.	[unit]	value	param.	[unit]	value
δ_h	[1/d]	$6.13\cdot10^{-2}$	$\mu_{h,\mathrm{R1},\emptyset}$	[1/d]	1.24	$\eta(a_1,\{\mathrm{WT},\mathrm{R2}\})$		0.979
δ_m	[1/d]	$5.1\cdot10^{-2}$	$\mu_{m,\mathrm{R1},\emptyset}$	[1/d]	$4.34\cdot10^{-2}$	$\eta(a_1,\{\mathrm{R1},\mathrm{HR}\})$		0
$k_{\ell,\emptyset}$	[1/d]	0.13	$\mu_{h,\mathrm{R2},\emptyset}$	[1/d]	$2.41\cdot10^{-4}$	$\eta(a_2,\{\mathrm{WT},\mathrm{R1}\})$		0.966
$k_{m,\emptyset}$	[1/d]	0.13	$\mu_{m,\mathrm{R2},\emptyset}$	[1/d]	$2.33\cdot10^{-2}$	$\eta(a_2,\{\mathrm{R2},\mathrm{HR}\})$		0
d_ℓ	[1/d]	$9.4\cdot10^{-5}$	$c_{\mathcal{A}}(a_\emptyset)$	[US$/d]	0	$c_{\mathcal{S}}(\mathcal{L})$	[US$/d]	0
d_m	[1/d]	$2.7\cdot10^{-4}$	$c_{\mathcal{A}}(a_1)$	[US$/d]	0.3	$c_{\mathcal{S}}(\mathcal{M})$	[US$/d]	2.2
d_h	[1/d]	$5.5\cdot10^{-4}$	$c_{\mathcal{A}}(a_2)$	[US$/d]	1.08	$c_{\mathcal{S}}(\mathcal{H})$	[US$/d]	8.8
k_{info}	[US$]	500	$\lambda^{\ddagger}$	[1/d]	$1.75\cdot10^{-4}$	$c_{\mathcal{S}}(✠)$	[US$/d]	22.1

Table A1: **General Model parameters.** $\mathcal{L}$ denotes the set of states for which condition $n_C(M)\leq\ell$ for all possible virus mutants M holds, i.e. $[\leq\ell,\leq\ell,\leq\ell,\leq\ell]$. Set $\mathcal{H}$ is defined as all states where at least one $n_C(M)>m$ and set $\mathcal{M}$ denotes the remaining states (except death). ‡ Assuming an annual inflation of 6.2% for South Africa.