Optimal Control of Partially Observable Piecewise Deterministic Markov Processes

Nicole Bäuerle
based on a joint work with D. Lange

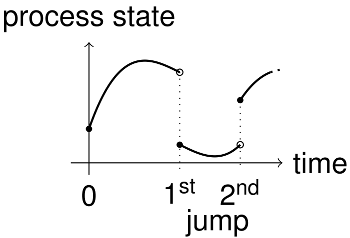
Karlsruhe Institute of Technology

Wien, April 2018

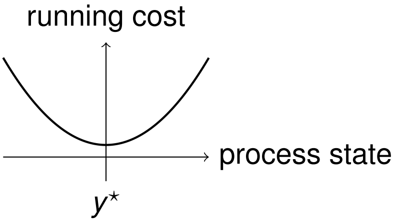
Outline

- Motivation
- PDMP with Partial Observation
- Controlled PDMP with Partial Observation
- Reduction to a Discrete-Time Problem
- The Filter
- Existence Results
- Application

Motivation



A typical path of a PDMP



Ingredients of a PDMP

In order to define a PDMP we need the following data

- The *drift* $\Phi : \mathbb{R}^d \times \mathbb{R}_+ \to \mathbb{R}^d$ is continuous and satisfies for all $y \in \mathbb{R}^d$ and $s, t > 0$: $\Phi(y, t+s) = \Phi(\Phi(y,s), t)$.

Ingredients of a PDMP

In order to define a PDMP we need the following data

- The *drift* $\Phi : \mathbb{R}^d \times \mathbb{R}_+ \to \mathbb{R}^d$ is continuous and satisfies for all $y \in \mathbb{R}^d$ and $s, t > 0$: $\Phi(y, t+s) = \Phi(\Phi(y, s), t)$.
- The jump times $0 := T_0 < T_1 < \dots$ are $\mathbb{R}_+$-valued random variables such that $S_n := T_n - T_{n-1}, n \in \mathbb{N}$, $S_0 := 0$. They are generated by an *intensity* $\lambda : \mathbb{R}^d \to (0, \infty)$.

Ingredients of a PDMP

In order to define a PDMP we need the following data

- The *drift* $\Phi : \mathbb{R}^d \times \mathbb{R}_+ \to \mathbb{R}^d$ is continuous and satisfies for all $y \in \mathbb{R}^d$ and $s, t > 0$: $\Phi(y, t + s) = \Phi(\Phi(y, s), t)$.
- The jump times $0 := T_0 < T_1 < \dots$ are $\mathbb{R}_+$-valued random variables such that $S_n := T_n - T_{n-1}, n \in \mathbb{N}$, $S_0 := 0$. They are generated by an *intensity* $\lambda : \mathbb{R}^d \to (0, \infty)$.
- A *transition kernel* Q from $\mathbb{R}^d$ to $\mathbb{R}^d$ gives the probability $Q(B|y)$ that the process jumps into B given the state y.

Ingredients of a PDMP

In order to define a PDMP we need the following data

- The *drift* $\Phi : \mathbb{R}^d \times \mathbb{R}_+ \to \mathbb{R}^d$ is continuous and satisfies for all $y \in \mathbb{R}^d$ and $s, t > 0$: $\Phi(y, t + s) = \Phi(\Phi(y, s), t)$.
- The jump times $0 := T_0 < T_1 < \dots$ are $\mathbb{R}_+$-valued random variables such that $S_n := T_n - T_{n-1}, n \in \mathbb{N}$, $S_0 := 0$. They are generated by an *intensity* $\lambda : \mathbb{R}^d \to (0, \infty)$.
- A *transition kernel* Q from $\mathbb{R}^d$ to $\mathbb{R}^d$ gives the probability $Q(B|y)$ that the process jumps into B given the state y.

The PDMP is then defined by

$$Y_t := \Phi(Y_{T_n}, t - T_n), \quad \text{for } T_n \le t < T_{n+1}, n \in \mathbb{N}_0.$$

A PDMP with Partial Observation

Let $(\epsilon_n)_{n\in\mathbb{N}}$ be iid $\mathbb{R}^d$-valued random variables. We assume that the controller is only able to observe $X_n := Y_{T_n} + \epsilon_n$.
Define $\Lambda(y,t) := \int_0^t \lambda\big(\Phi(y,s)\big)ds$. Then

$$\mathbb{P}_{x,y}(S_n \le t, Y_{T_n} \in C, X_n \in D \mid S_0, Y_{T_0}, X_0, \dots, S_{n-1}, Y_{T_{n-1}}, X_{n-1})$$
$$= \int_0^t \exp\big(-\Lambda(Y_{T_{n-1}}, s)\big)\lambda\big(\Phi(Y_{T_{n-1}}, s)\big) \int_C Q_\epsilon(D-y')Q\big(dy'|\Phi(Y_{T_{n-1}}, s)\big)ds$$

Policies for POPDMP

- Consider marked point process $(T_n, \hat{Y}_n := Y_{T_n}, X_n)$.

Policies for POPDMP

- Consider marked point process $(T_n, \hat{Y}_n := Y_{T_n}, X_n)$.
- Relaxed controls on action space A:
  $\mathcal{R} := \{r : [0, \infty) \to \mathcal{P}(A) \mid r \text{ measurable}\}$.

Policies for POPDMP

- Consider marked point process $(T_n, \hat{Y}_n := Y_{T_n}, X_n)$.
- Relaxed controls on action space A:
  $\mathcal{R} := \{r : [0, \infty) \to \mathcal{P}(A) \mid r \text{ measurable}\}$.
- Observable histories: $\mathcal{H}_0 := \mathbb{R}^d$ and for $n \in \mathbb{N}$

$$\mathcal{H}_n := \mathcal{H}_{n-1} \times \mathcal{R} \times \mathbb{R}_+ \times \mathbb{R}^d.$$

Element: $h_n = (x_0, r_0, s_1, x_1, \dots, r_{n-1}, s_n, x_n) \in \mathcal{H}_n$

Policies for POPDMP

- Consider marked point process $(T_n, \hat{Y}_n := Y_{T_n}, X_n)$.
- Relaxed controls on action space A:
  $\mathcal{R} := \{r : [0, \infty) \to \mathcal{P}(A) \mid r \text{ measurable}\}$.
- Observable histories: $\mathcal{H}_0 := \mathbb{R}^d$ and for $n \in \mathbb{N}$

$$\mathcal{H}_n := \mathcal{H}_{n-1} \times \mathcal{R} \times \mathbb{R}_+ \times \mathbb{R}^d.$$

  Element: $h_n = (x_0, r_0, s_1, x_1, \dots, r_{n-1}, s_n, x_n) \in \mathcal{H}_n$
- A *discrete time history dependent relaxed control policy* is $\pi^D := (\pi_0^D, \pi_1^D, \dots)$ where $\pi_n^D : \mathcal{H}_n \to \mathcal{R}$ is measurable.

Policies for POPDMP

- Consider marked point process $(T_n, \hat{Y}_n := Y_{T_n}, X_n)$.
- Relaxed controls on action space A:
  $\mathcal{R} := \{r : [0,\infty) \to \mathcal{P}(A) \mid r \text{ measurable}\}$.
- Observable histories: $\mathcal{H}_0 := \mathbb{R}^d$ and for $n \in \mathbb{N}$

$$\mathcal{H}_n := \mathcal{H}_{n-1} \times \mathcal{R} \times \mathbb{R}_+ \times \mathbb{R}^d.$$

  Element: $h_n = (x_0, r_0, s_1, x_1, \dots, r_{n-1}, s_n, x_n) \in \mathcal{H}_n$
- A *discrete time history dependent relaxed control policy* is $\pi^D := (\pi_0^D, \pi_1^D, \dots)$ where $\pi_n^D : \mathcal{H}_n \to \mathcal{R}$ is measurable.
- Continuous control:

$$\pi_t := \sum_{n=0}^{\infty} 1_{\{T_n \le t < T_{n+1}\}}(t) \cdot \pi_n^D(H_n)(t - T_n).$$

Controlled Data

We assume that the policy influences the drift, the intensity and the transition kernel.

Controlled Data

We assume that the policy influences the drift, the intensity and the transition kernel.

- Drift $\Phi : \mathcal{R} \times \mathbb{R}^d \times \mathbb{R}_+ \to \mathbb{R}^d$, notation $\Phi^r(y, t)$.

Controlled Data

We assume that the policy influences the drift, the intensity and the transition kernel.

- Drift $\Phi : \mathcal{R} \times \mathbb{R}^d \times \mathbb{R}_+ \to \mathbb{R}^d$, notation $\Phi^r(y, t)$.
- Jump rate $\lambda^A : \mathbb{R}^d \times A \to (0, \infty)$.

Controlled Data

We assume that the policy influences the drift, the intensity and the transition kernel.

- Drift $\Phi : \mathcal{R} \times \mathbb{R}^d \times \mathbb{R}_+ \to \mathbb{R}^d$, notation $\Phi^r(y, t)$.
- Jump rate $\lambda^A : \mathbb{R}^d \times A \to (0, \infty)$.
- Jump kernel Q^A from $\mathbb{R}^d \times A$ to $\mathbb{R}^d$.

Controlled Process

A *Controlled Partially Observable Piecewise Deterministic Markov Process* with *local characteristics* $(\Phi^r, \lambda^A, Q^A, Q_\epsilon)$ satisfies

$$\mathbb{P}^{\pi^D}_{x,y}(S_n \leq t, \hat{Y}_n \in C, X_n \in D | S_0, \ldots, S_{n-1}, \hat{Y}_{n-1}, X_{n-1}, \pi^D_{n-1})$$
$$= \int_0^t e^{-\Lambda^{\pi^D_{n-1}}(\hat{Y}_{n-1},s)} \int_A \lambda^A(\Phi^{\pi^D_{n-1}}(\hat{Y}_{n-1}, s), a)$$
$$\int_C Q_\epsilon(D - y')Q^A(dy'|\Phi^{\pi^D_{n-1}}(\hat{Y}_{n-1}, s), a)\pi^D_{n-1}(H_{n-1}, s)(da)ds$$

where $\Lambda^r(y, t) := \int_0^t \int_A \lambda^A(\Phi^r(y, s), a)r_s(da)\, ds$. ▸ Toy Example.

The Optimization Problem

For a fixed policy π^D we define the cost as

$$J(x,\pi^D) := \int \mathbb{E}_{x,y}^{\pi^D}\left[\int_0^\infty e^{-\beta t}\int_A c(Y_t,a)\,\pi_t(da)\,dt\right] Q_0(dy|x).$$

The Optimization Problem

For a fixed policy π^D we define the cost as

$$J(x,\pi^D) := \int \mathbb{E}_{x,y}^{\pi^D}\left[\int_0^\infty e^{-\beta t}\int_A c(Y_t,a)\,\pi_t(da)\,dt\right] Q_0(dy|x).$$

The *value function* is defined as

$$J(x) := \inf_{\pi^D} J(x,\pi^D) \qquad \text{for all } \; x \in \mathbb{R}^d.$$

The Optimization Problem

For a fixed policy π^D we define the cost as

$$J(x,\pi^D) := \int \mathbb{E}_{x,y}^{\pi^D}\left[\int_0^\infty e^{-\beta t}\int_A c(Y_t,a)\,\pi_t(da)\,dt\right] Q_0(dy|x).$$

The *value function* is defined as

$$J(x) := \inf_{\pi^D} J(x,\pi^D) \qquad \text{for all } \; x \in \mathbb{R}^d.$$

A policy $\pi^\star$ is optimal if

$$J(x) = J(x,\pi^\star) \qquad \text{for all } \; x \in \mathbb{R}^d.$$

A discrete-time POMDP

We define a POMDP as follows:

A discrete-time POMDP

We define a POMDP as follows:

- State space $\mathbb{R}_+ \times \mathbb{R}^{2d} \ni (s, y, x)$.

A discrete-time POMDP

We define a POMDP as follows:

- State space $\mathbb{R}_+ \times \mathbb{R}^{2d} \ni (s, y, x)$.
- Action space $\mathcal{R} \ni r$.

A discrete-time POMDP

We define a POMDP as follows:

- State space $\mathbb{R}_+ \times \mathbb{R}^{2d} \ni (s, y, x)$.
- Action space $\mathcal{R} \ni r$.
- Transition law with $\Gamma^r(y,t) := \beta t + \int_0^t \int_A \lambda^A(\Phi^r(y,u), a) r_u(da) du$

$$\tilde{Q}\big([0,t] \times C \times D | y, r\big) =$$

$$\int_0^t e^{-\Gamma^r(y,u)} \int_A \lambda^A\big(\Phi^r(y,u), a\big) \int_C Q(D - y') Q^A\big(dy' | \Phi^r(y,u), a\big) r_u(da) du.$$

A discrete-time POMDP

We define a POMDP as follows:

- State space $\mathbb{R}_+ \times \mathbb{R}^{2d} \ni (s, y, x)$.
- Action space $\mathcal{R} \ni r$.
- Transition law with $\Gamma^r(y,t) := \beta t + \int_0^t \int_A \lambda^A(\Phi^r(y,u),a) r_u(da)du$

$$\tilde{Q}\big([0,t] \times C \times D | y, r\big) = \int_0^t e^{-\Gamma^r(y,u)} \int_A \lambda^A\big(\Phi^r(y,u),a\big) \int_C Q(D-y')Q^A\big(dy' | \Phi^r(y,u),a\big) r_u(da)du.$$

- The one-stage cost

$$g(y,r) = \int_0^\infty e^{-\Gamma^r(y,t)} \int_A c(\Phi^r(y,t),a)\, r_t(da)\, dt.$$

The Problem for the discrete-time POMDP

For a fixed policy π^D we define the cost as

$$\tilde{J}(x, \pi^D) := \widetilde{\mathbb{E}}_x^{\pi^D} \left[\sum_{k=0}^{\infty} g(\hat{Y}_k, \pi_k^D(H_k)) \right].$$

The Problem for the discrete-time POMDP

For a fixed policy π^D we define the cost as

$$\tilde{J}(x, \pi^D) := \widetilde{\mathbb{E}}_x^{\pi^D} \left[\sum_{k=0}^{\infty} g(\hat{Y}_k, \pi_k^D(H_k)) \right].$$

The *value function* is defined as

$$\tilde{J}(x) := \inf_{\pi^D} \tilde{J}(x, \pi^D) \quad \forall\, x \in \mathbb{R}^d.$$

The Problem for the discrete-time POMDP

For a fixed policy π^D we define the cost as

$$\tilde{J}(x, \pi^D) := \widetilde{\mathbb{E}}_x^{\pi^D} \left[\sum_{k=0}^{\infty} g(\hat{Y}_k, \pi_k^D(H_k)) \right].$$

The *value function* is defined as

$$\tilde{J}(x) := \inf_{\pi^D} \tilde{J}(x, \pi^D) \quad \forall\, x \in \mathbb{R}^d.$$

A policy $\pi^\star$ is optimal if

$$\tilde{J}(x) = \tilde{J}(x, \pi^\star) \quad \forall\, x \in \mathbb{R}^d.$$

Equivalence of the two Problems

Theorem

Let $x \in \mathbb{R}^d$ be an initial observation and π^D a policy. Then, it holds

$$J(x, \pi^D) = \tilde{J}(x, \pi^D).$$

Assumptions

(C1) The action space A is a compact metric space.

Assumptions

(C1) The action space A is a compact metric space.

(C2) $\lambda^A : \mathbb{R}^d \times A \to (0, \infty)$ is continuous and $0 < \underline{\lambda} < \lambda^A < \bar{\lambda}$.

Assumptions

(C1) The action space A is a compact metric space.

(C2) $\lambda^A : \mathbb{R}^d \times A \to (0, \infty)$ is continuous and $0 < \underline{\lambda} < \lambda^A < \bar{\lambda}$.

(C3) Q^A is weakly continuous, i.e. $(x, a) \mapsto \int v(z) Q^A(dz|x, a)$ is continuous and bounded for all $v : \mathbb{R}^d \to \mathbb{R}$ continuous and bounded.

Assumptions

(C1) The action space A is a compact metric space.

(C2) $\lambda^A : \mathbb{R}^d \times A \to (0, \infty)$ is continuous and $0 < \underline{\lambda} < \lambda^A < \bar{\lambda}$.

(C3) Q^A is weakly continuous, i.e. $(x, a) \mapsto \int v(z) Q^A(dz|x, a)$ is continuous and bounded for all $v : \mathbb{R}^d \to \mathbb{R}$ continuous and bounded.

(B1) There exists $E^0 = \{y^1, \dots, y^d\} \subset \mathbb{R}^d$ s.t. for all $y \in \mathbb{R}^d$ and $a \in A$ we have $Q^A(E^0|y, a) = 1$ and $Q_0(E_0) = 1$.

Assumptions

(C1) The action space A is a compact metric space.

(C2) $\lambda^A : \mathbb{R}^d \times A \to (0, \infty)$ is continuous and $0 < \underline{\lambda} < \lambda^A < \bar{\lambda}$.

(C3) Q^A is weakly continuous, i.e. $(x, a) \mapsto \int v(z) Q^A(dz|x, a)$ is continuous and bounded for all $v : \mathbb{R}^d \to \mathbb{R}$ continuous and bounded.

(B1) There exists $E^0 = \{y^1, \dots, y^d\} \subset \mathbb{R}^d$ s.t. for all $y \in \mathbb{R}^d$ and $a \in A$ we have $Q^A(E^0|y, a) = 1$ and $Q_0(E_0) = 1$.

(B2) Q_ϵ has a bounded density f_ϵ with respect to some σ-finite measure ν.

The Filter

Assumption (B2) implies that the transition law has the density

$$
\begin{aligned}
& \tilde{q}(s, y', x | y, r) \\
= \quad & e^{-\Gamma^r(y,s)} f(x - y') \int_A \lambda^A\big(\Phi^r(y,s), a\big) Q^A\big(y' | \Phi^r(y,s), a\big) r_s(da).
\end{aligned}
$$

The Filter

Assumption (B2) implies that the transition law has the density

$$\begin{aligned} & \tilde{q}(s, y', x | y, r) \\ = \; & e^{-\Gamma^r(y,s)} f(x - y') \int_A \lambda^A\big(\Phi^r(y,s), a\big) Q^A\big(y' | \Phi^r(y,s), a\big) r_s(da). \end{aligned}$$

Define for $y' \in E^0$ the updating operator

$$\Psi(\rho, r, s, x)(y') := \frac{\sum_{y \in E^0} \tilde{q}(s, y', x | y, r)\rho(y)}{\sum_{\hat{y} \in E^0} \sum_{y \in E^0} \tilde{q}(s, \hat{y}, x | y, r)\rho(y)}.$$

and for $h_n = (x_0, r_0, s_1, x_1, \ldots, r_{n-1}, s_n, x_n)$, $\mu_0(x_0) := Q_0(\cdot | x_0)$,

$$\mu_n(\cdot | h_n) = \mu_n(\cdot | h_{n-1}, r_{n-1}, s_n, x_n) \;\; := \;\; \Psi\big(\mu_{n-1}(\cdot | h_{n-1}), r_{n-1}, s_n, x_n\big).$$

The Filter

Theorem

It holds that

$$\begin{aligned} & \tilde{\mathbb{P}}_x^{\pi^D}\left(\hat{Y}_n \in C \mid X_0, R_0, S_1, X_1, \ldots, R_{n-1}, S_n, X_n\right) \\ = \; & \mu_n(C|X_0, R_0, S_1, X_1, \ldots, R_{n-1}, S_n, X_n) \end{aligned}$$

Filtered MDP

We define a POMDP as follows:

Filtered MDP

We define a POMDP as follows:

- State space $\mathcal{P}(E^0) \ni \rho$.

Filtered MDP

We define a POMDP as follows:

- State space $\mathcal{P}(E^0) \ni \rho$.
- Action space $\mathcal{R} \ni r$.

Filtered MDP

We define a POMDP as follows:

- State space $\mathcal{P}(E^0) \ni \rho$.
- Action space $\mathcal{R} \ni r$.
- Transition law

$$\hat{Q}(B|\rho, r) = \int\int \sum_{y \in E^0} 1_B\big(\Psi(\rho, r, s, x)\big)\tilde{q}^{SX}(s, x|y, r)\nu(dx)ds\rho(y),$$

where $\tilde{q}^{SX}(s, x|y, r) := \sum_{y' \in E^0} \tilde{q}(s, y', x|y, r)$.

Filtered MDP

We define a POMDP as follows:

- State space $\mathcal{P}(E^0) \ni \rho$.
- Action space $\mathcal{R} \ni r$.
- Transition law

$$\hat{Q}(B|\rho, r) = \int\int \sum_{y \in E^0} 1_B\big(\Psi(\rho, r, s, x)\big)\tilde{q}^{SX}(s, x|y, r)\nu(dx)ds\rho(y),$$

where $\tilde{q}^{SX}(s, x|y, r) := \sum_{y' \in E^0} \tilde{q}(s, y', x|y, r)$.

- The one-stage cost

$$\hat{g}(\rho, r) := \sum_{y \in E^0} g(y, r)\rho(y).$$

Filtered MDP

Policy $\pi = (f_0, f_1, \dots)$ with $f_n : \mathcal{P}(E^0) \to \mathcal{R}$ can be seen as a special policy $\pi^D \in \Pi^D$ by setting

$$\pi_n^D(h_n) := f_n(\mu_n(\cdot|h_n)).$$

Filtered MDP

Policy $\pi = (f_0, f_1, \dots)$ with $f_n : \mathcal{P}(E^0) \to \mathcal{R}$ can be seen as a special policy $\pi^D \in \Pi^D$ by setting

$$\pi_n^D(h_n) := f_n(\mu_n(\cdot|h_n)).$$

We denote the *cost of policy* π by

$$V(\rho, \pi) := \hat{\mathbb{E}}_\rho^\pi \left[\sum_{n=0}^{\infty} \hat{g}\big(\mu_n, f_n(\mu_n)\big) \right].$$

Filtered MDP

Policy $\pi = (f_0, f_1, \ldots)$ with $f_n : \mathcal{P}(E^0) \to \mathcal{R}$ can be seen as a special policy $\pi^D \in \Pi^D$ by setting

$$\pi_n^D(h_n) := f_n(\mu_n(\cdot|h_n)).$$

We denote the *cost of policy* π by

$$V(\rho, \pi) := \hat{\mathbb{E}}_\rho^\pi \left[\sum_{n=0}^{\infty} \hat{g}\big(\mu_n, f_n(\mu_n)\big) \right].$$

The *value function* is defined as

$$V(\rho) := \inf_{\pi \in \Pi^\infty} V(\rho, \pi) \qquad \text{for all } \rho \in \mathcal{P}(E^0).$$

Equivalence of the two Problems

Theorem

Let $x \in \mathbb{R}^d$ be an initial observation, π and π^D given as on previous slide. Then, it holds

$$V(Q_0(\cdot|x), \pi) = J(x, \pi^D).$$

Further Assumptions

(C4) $r \mapsto \Phi^r(y, t)$ is continuous for all $y \in E^0, t \geq 0$.

Further Assumptions

(C4) $r \mapsto \Phi^r(y, t)$ is continuous for all $y \in E^0, t \geq 0$.

(C5) $c : \mathbb{R}^d \times A \to \mathbb{R}_+$ is lower semi-continuous with respect to the product topology.

Regularization of the Filter

Let $h_\sigma : \mathbb{R} \to \mathbb{R}, \sigma > 0$ be a regularization kernel , i.e.

(i) $h_\sigma(t) \geq 0$ for all $t \in \mathbb{R}$,

(ii) $\int_{\mathbb{R}} h_\sigma(t)dt = 1$,

(iii) $\lim_{\sigma \downarrow 0} \int_{-a}^{a} h_\sigma(t)dt = 1$ for all $a > 0$.

Regularization of the Filter

Let $h_\sigma : \mathbb{R} \to \mathbb{R}, \sigma > 0$ be a regularization kernel , i.e.

(i) $h_\sigma(t) \geq 0$ for all $t \in \mathbb{R}$,

(ii) $\int_{\mathbb{R}} h_\sigma(t)dt = 1$,

(iii) $\lim_{\sigma \downarrow 0} \int_{-a}^{a} h_\sigma(t)dt = 1$ for all $a > 0$.

We use a regularized filter of the form

$$\hat{\Psi}(\rho, r, s, x)(y') := \frac{\int_{\mathbb{R}} \sum_{y \in E^0} \tilde{q}(u, y', x|y, r)\rho(y)h_\sigma(s-u)du}{\sum_{\hat{y} \in E^0} \int_{\mathbb{R}} \sum_{y \in E^0} \tilde{q}(u, \hat{y}, x|y, r)\rho(y)h_\sigma(s-u)du}.$$

Note that we have $\lim_{\sigma \downarrow 0} \hat{\Psi} = \Psi$

Solution of the Problem

Theorem

Under all assumptions we have for the regularized version of the problem

a) $V^\sigma(\rho) = \inf_{r\in\mathcal{R}} \left\{ \hat{g}(\rho, r) + \int_{\mathcal{P}(E^0)} V^\sigma(\rho')\hat{Q}(d\rho'|\rho, r) \right\}.$

b) *There exists an f^* which attains the minimum in a) and $(f^*, f^*, \ldots)$ is optimal for the filtered problem. The optimal policy for the original problem is $(\pi_0^D, \pi_1^D, \ldots)$ with*

$$\begin{aligned} \pi_0^P(x)(t) &= f^\star\big(Q_0(\cdot|x)\big)(t), \quad x \in \mathbb{R}^d \\ \pi_n^P(h_n)(t) &= f^\star\big(\mu_n(\cdot|h_n)\big)(t), \quad h_n \in H_n. \end{aligned}$$

Application: Toy Example

(i) The state space is $\mathbb{R}$, $E^0 := \{-2, 0, 2\}$, the action space is $A := [-1; 1]$.

(ii) The controlled drift is given by

$$\frac{d}{dt}\Phi^r(y,t) = \int_A a r_t(da), \qquad \Phi(y,0) = y.$$

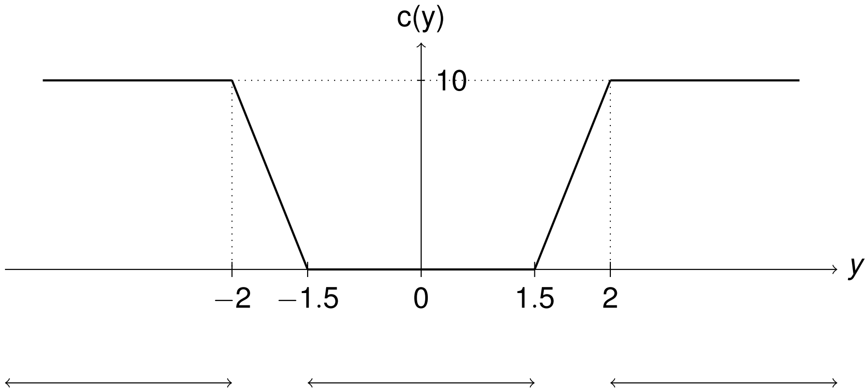
(iii) We set $\lambda^A \equiv 1$ and $\beta := 1$.

(iv) The jump transition kernel Q^A and the cost function are given in the picture on the next slide.

(vii) Signal density: $f_\epsilon(-1) = f_\epsilon(0) = f_\epsilon(1) = \frac{1}{3}$.

back

Application: Transition Kernel and Cost Function



$Q(y;\cdot) = \delta_{-2}(\cdot)$ $\quad$ $Q(y;\cdot) = \delta_0(\cdot)$ $\quad$ $Q(y;\cdot) = \delta_2(\cdot)$

Solution of the Problem: Optimal Policy

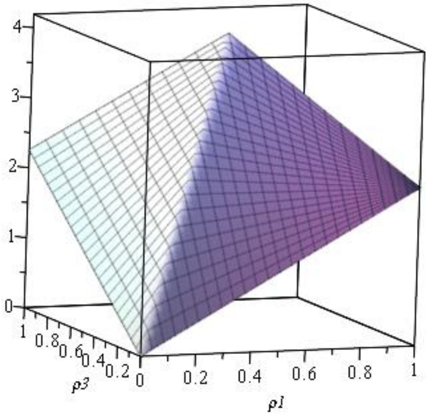
Theorem

In this POPDMP all assumptions which we previously made, are satisfied and there exists an optimal policy.

In this example we have also computed the value function and the optimal policy numerically by value iteration. For the optimal policy we obtained

$$\pi_n(h_n, t) := \begin{cases} 1_{\{t \le \frac{1}{2}\}} & \text{if } \mu_n^1 \ge \mu_n^3, \\ -1_{\{t \le \frac{1}{2}\}} & \text{if } \mu_n^1 \le \mu_n^3. \end{cases}$$

Solution of the Problem: Value Function



References

- N.B. and Lange, D.: Optimal control of partially observable piecewise deterministic Markov processes. *SIAM J. Control Optim.*, **56**, 1441-1462, 2018.
- N.B. and Rieder, U.: Markov Decision Processes with Applications to Finance. *Springer*, 2011.
- Brandejsky, A. and de Saporta, B. and Dufour, F.: Optimal stopping for partially observed piecewise-deterministic Markov processes. *Stochastic Process. Appl.*, **123**, 3201-3238, 2013.
- Davis, M.H.A.: Markov Models and Optimization. *Chapman and Hall*, 1993.
- Yushkevich, A. A.: On reducing a jump controllable Markov model to a model with discrete time. *Theory Probab. Appl.*, **25**, 58-69, 1980.

Thank you very much
for your attention!