Mean-field Markov Decision Process with common noise and open-loop controls

Huyên PHAM

Université Paris Cité, LPSM

Based on joint work with
Médéric MOTTE, Université Paris Cité, LPSM

9th Colloquium on BSDE and mean field systems
Annecy, june 27-july 1, 2022

Mean-field approach to large population stochastic control

• Large number of N interacting dynamic agents/entities

• Agents are cooperative, and act for collective welfare following a center a decision/social planner. Other interpretation:

- **Influencer** controls the state of many individuals in social networks

▸ When $N \to \infty$: optimal control of McKean-Vlasov equation (mean-field control problem). Many papers mostly on continuous-time, and finite horizon

▸ Here, we focus on

- **Discrete time** and possibly discrete space (graphs)
- **Infinite horizon**
- Common noise
- When $N \to \infty$: Conditional McKean-Vlasov Markov Decision Process (CMKV-MDP)

→ Mathematical framework of reinforcement learning (RL) with many interacting cooperative agents (R. Carmona, M. Laurière and Z. Tan 19, Gu, Guo, Wei and Xu 19)

Motivation from targeted advertising application

- A company **C**: Internet retailer, candidate to election
- A social network **SN**
  - N connected users of **SN**: state = customer/voter or not of **C**
  - Users data: cookies (track record of visited web pages)

- **Targeted advertising**:
  - An influencer (Criteo, etc) **I** working for C
  - **I** displays personalized online banner ads to users according to their cookies and public data (no direct access to individual states) → Open-loop control and common noise
  - Objective: optimize the targeted ad strategy, e.g., in order to attract the largest possible clients/voters given ads costs.

Outline

Framework and notations

- A universal probability space $(\Omega, \mathcal{F}, \mathbb{P})$
- **State and action** spaces: $\mathcal{X}$ and A (compact Polish)
  - $\mathcal{P}(\mathcal{X})$, resp. $\mathcal{P}(A)$, resp. $\mathcal{P}(\mathcal{X} \times A)$: set of probability measures on $\mathcal{X}$, resp. A, resp. $\mathcal{X} \times A$, with Wasserstein distance
- Discrete time **transition dynamics**
  - Idiosyncratic noises: $(\varepsilon_t^i)_{t \in \mathbb{N}^*}$, for agent $i \in \mathbb{N}^*$, i.i.d. valued in E
  - Common noise: $(\varepsilon_t^0)_{t \in \mathbb{N}^*}$ for all agents, i.i.d. valued in E^0
  - F: meas. function from $\mathcal{X} \times A \times \mathcal{P}(\mathcal{X} \times A) \times E \times E^0$ into $\mathcal{X}$
- **Reward** on infinite horizon:
  - discount factor $\beta \in [0, 1)$
  - f: meas. bounded function from $\mathcal{X} \times A \times \mathcal{P}(\mathcal{X} \times A)$ into $\mathbb{R}$

Assumptions on transition dynamics and reward

(HF$_{lip}$) There exists K_F s.t. for all $a \in A$, $e^0 \in E^0$, $x, x' \in \mathcal{X}$, $\nu, \nu' \in \mathcal{P}(\mathcal{X} \times A)$,

$$\mathbb{E}\Big[d_{\mathcal{X}}\big(F(x, a, \nu, \varepsilon_1^1, e^0), F(x', a, \nu', \varepsilon_1^1, e^0)\big)\Big] \quad \leq \quad K_F\big(d_{\mathcal{X}}(x, x') + \mathcal{W}(\nu, \nu')\big).$$

(Hf$_{lip}$) There exists K_f s.t. for all $a \in A$, $x, x' \in \mathcal{X}$, $\nu, \nu' \in \mathcal{P}(\mathcal{X} \times A)$,

$$\Big|f(x, a, \nu) - f(x', a, \nu')\Big| \quad \leq \quad K_f\big(d_{\mathcal{X}}(x, x') + \mathcal{W}(\nu, \nu')\big).$$

Remark: Lipschitz assumption in **(HF$_{lip}$)** is made on expectation, not pathwisely.

Information and decentralized open-loop controls in the finite population model

- A subalgebra $\mathcal{G}$ of $\mathcal{F}$ rich enough (used for randomization)
- ξ^i initial state in $\mathcal{X}$ of agent $i = 1, \ldots, N$; independent of $\mathcal{G}$
- **Decentralized open-loop control**: a sequence $\boldsymbol{\alpha} = (\alpha^1, \ldots, \alpha^N)$ of processes valued in A^N, and adapted w.r.t.

$$\mathcal{F}_t^N \quad = \quad \sigma\{\xi^i, (\varepsilon_s^i)_{s \le t}, (\varepsilon_s^0)_{s \le t}, i = 1, \ldots, N\} \vee \mathcal{G}, \qquad t \in \mathbb{N}.$$

Remark: No symmetry assumption in control α^i, $i = 1, \ldots, N$.

Mean-field control problem in the N-population model

- **Mean-field controlled dynamics**: State process $X^{i,N,\boldsymbol{\alpha}}$ of agent i governed by

$$\begin{cases} X_0^{i,N,\boldsymbol{\alpha}} &= \xi^i \\ X_{t+1}^{i,N,\boldsymbol{\alpha}} &= F\big(X_t^{i,N,\boldsymbol{\alpha}}, \alpha_t^i, \frac{1}{N}\sum_{j=1}^N \delta_{(X_t^{j,N,\boldsymbol{\alpha}}, \alpha_t^j)}, \varepsilon_{t+1}^i, \varepsilon_{t+1}^0\big). \end{cases} \tag{1}$$

- **Gain functional** for each agent $i = 1, \ldots, N$:

$$V^{i,N,\boldsymbol{\alpha}} = \mathbb{E}\Big[\sum_{t=0}^{\infty} \beta^t f\big(X_t^{i,N,\boldsymbol{\alpha}}, \alpha_t^i, \frac{1}{N}\sum_{j=1}^N \delta_{(X_t^{j,N,\boldsymbol{\alpha}}, \alpha_t^j)}\big)\Big]$$

- **Optimal gain** for the center of decision (social planner/influencer):

$$V^N(\xi^1, \ldots, \xi^N) = \sup_{\boldsymbol{\alpha}} \frac{1}{N}\sum_{i=1}^N V^{i,N,\boldsymbol{\alpha}}. \tag{2}$$

Remark: Problem (1)-(2) is a standard MDP with state space $\mathcal{X}^N$, action space A^N, for which it is known that optimization over (randomized/relaxed) open-loop control is equivalent to optimization over (randomized/relaxed) feedback control. But hardly tractable when N is large!

Mean-field control problem in the ∞-population model

Formally, we expect (by "propagation of chaos") a problem formulation with:

- **Controlled McKean-Vlasov equation**: state X^α of representative agent governed by

$$\begin{cases} X_0^\alpha &= \xi \\ X_{t+1}^\alpha &= F\big(X_t^\alpha, \alpha_t, \mathbb{P}^0_{(X_t^\alpha,\alpha_t)}, \varepsilon_{t+1}, \varepsilon^0_{t+1}\big). \end{cases} \tag{3}$$

where the control process α is valued in A, and adapted w.r.t.

$$\mathcal{F}_t = \sigma\{\xi, (\varepsilon_s)_{s\le t}, (\varepsilon^0_s)_{s\le t}\} \vee \mathcal{G}, \quad t \in \mathbb{N}.$$

- Gain functional of representative agent

$$V^\alpha = \mathbb{E}\Big[\sum_{t=0}^{\infty} \beta^t f\big(X_t^\alpha, \alpha_t, \mathbb{P}^0_{(X_t^\alpha,\alpha_t)}\big)\Big]$$

- Optimal gain for the center of decision (social planner/influencer):

$$V(\xi) = \sup_\alpha V^\alpha. \tag{4}$$

Remark: Problem (3)-(4) is *a priori* a nonstandard MDP due to $\mathbb{P}^0_{(X_t^\alpha,\alpha_t)}$, and called CMKV-MDP.

Questions addressed in this talk

- **(I)** Can we obtain a tractable resolution of CMKV-MDP?
  - Dynamic programming, Bellman equation?
  - Examples of explicit resolution
- **(II)** Convergence of N-agent MDP to CMKV-MDP?

(1) V^N towards V, and at which rate?

(2) How to get an approximate optimal control for the N-agent problem from an optimal control for the CMKV-MDP? At which accuracy?

Questions addressed in this talk

- **(I)** Can we obtain a tractable resolution of CMKV-MDP?
  - Dynamic programming, Bellman equation?
  - Examples of explicit resolution
- **(II)** Convergence of N-agent MDP to CMKV-MDP?

(1) V^N towards V, and at which rate?

(2) How to get an approximate optimal control for the N-agent problem from an optimal control for the CMKV-MDP? At which accuracy?

▸ Randomization of controls plays a crucial role! This is a noticeable difference with continuous time framework.

Related literature:

(I) In discrete time:

- Bauerle (21) closed-loop policies, Carmona, Laurière, Tan (22)

(II) mostly for continuous time MKV diffusion and for (1)

- Lacker (18), Djete (20): tightness arguments (no rate of convergence)
- Cecchin (21): Finite state: rate of CV $N^{-1/2}$
- Germain, P., Warin (21): BSDE methods under existence of a smooth solution. Rate of convergence N^{-1}
- Gangbo, Mayorga, Swiech (20), Cardaliaguet, Daudin, Jackson, Souganidis (22): viscosity solutions method. Rate of CV: $N^{-\gamma}$ for some $\gamma \in (0, 1]$.

Outline

Reformulation on $\mathcal{P}(\mathcal{X})$ with relaxed/randomized control

- $X = X^{\alpha} \rightsquigarrow$ CMKV dynamics with open-loop control α:

$$X_{t+1} \quad = \quad F(X_t, \alpha_t, \mathbb{P}^0_{(X_t,\alpha_t)}, \varepsilon_{t+1}, \varepsilon^0_{t+1}).$$

- Set $\mu_t = \mathbb{P}^0_{X_t}$. Then (with the pushforward measure notation $\star$):

$$\mu_{t+1} \quad = \quad F(\cdot, \cdot, \mathbb{P}^0_{(X_t,\alpha_t)}, \cdot, \varepsilon^0_{t+1}) \star \left(\mathbb{P}^0_{(X_t,\alpha_t)} \otimes \lambda_{\varepsilon}\right)$$

Reformulation on $\mathcal{P}(\mathcal{X})$ with relaxed/randomized control

- $X = X^\alpha \rightsquigarrow$ CMKV dynamics with open-loop control α:

$$X_{t+1} \quad = \quad F(X_t, \alpha_t, \mathbb{P}^0_{(X_t,\alpha_t)}, \varepsilon_{t+1}, \varepsilon^0_{t+1}).$$

- Set $\mu_t = \mathbb{P}^0_{X_t}$. Then (with the pushforward measure notation $\star$):

$$\mu_{t+1} \quad = \quad F(\cdot,\cdot,\mathbb{P}^0_{(X_t,\alpha_t)},\cdot,\varepsilon^0_{t+1}) \star \left(\mathbb{P}^0_{(X_t,\alpha_t)} \otimes \lambda_\varepsilon\right)$$

Bayes formula: $\mathbb{P}^0_{(X_t,\alpha_t)} = \mu_t \cdot \hat{\alpha}_t$, where $\hat{\alpha}_t$ is the probability kernel on $\mathcal{X} \times A$:

$$\hat{\alpha}_t : x \in \mathcal{X} \quad \longmapsto \quad \mathcal{L}^0\big(\alpha_t | X_t = x\big) \in \mathcal{P}(A)$$

→ **Controlled stochastic Fokker-Planck equation** on $\mathcal{P}(\mathcal{X})$:

$$\mu_{t+1} \quad = \quad \hat{F}(\mu_t, \hat{\alpha}_t, \varepsilon^0_{t+1}), \quad t \in \mathbb{N},$$

with relaxed ($\mathcal{P}(A)$-valued) feedback control $\hat{\alpha}$ valued in $\hat{A} \equiv L^0(\mathcal{X}; \mathcal{P}(A))$, and $\hat{F}(\mu, \hat{a}, e^0) = F(\cdot,\cdot,\mu\cdot\hat{a},\cdot,e^0) \star \big((\mu\cdot\hat{a}) \otimes \lambda_\varepsilon\big)$.

- Similarly and with law of conditional expectations, we have

$$V^\alpha \quad = \quad \mathbb{E}\Big[\sum_{t=0}^{\infty} \beta^t \hat{f}(\mu_t, \hat{\alpha}_t)\Big]$$

for some function $\hat{f} : \mathcal{P}(\mathcal{X}) \times \hat{A} \to \mathbb{R}$ explicitly derived from f.

Bellman operator of the lifted MDP

- Bellman operator $\mathcal{T}$ of the lifted MDP: for $W \in L_m^\infty(\mathcal{P}(\mathcal{X}))$,

$$\mathcal{T}[W](\mu) \quad := \quad \sup_{\hat{a}\in\hat{A}} \mathcal{T}^{\hat{a}}[W](\mu) := \sup_{\hat{a}\in\hat{A}} \left\{ \hat{f}(\mu,\hat{a}) + \beta\mathbb{E}\left[W\left(\hat{F}(\mu,\hat{a},\varepsilon_1^0)\right)\right]\right\}.$$

▸ $\mathcal{T}$ is well-defined and contractive on $L_m^\infty(\mathcal{P}(\mathcal{X}))$ → unique fixed point, denoted $V^\star$: $V^* = \mathcal{T}[V^*]$.

Characterization by Bellman equation on $\mathcal{P}(\mathcal{X})$

Theorem

(i) **Law invariance**. For any ξ, $\tilde{\xi}$ s.t. $\mathbb{P}_\xi = \mathbb{P}_{\tilde{\xi}}$, we have $V(\xi) = V(\tilde{\xi})$. We then define $V(\mu) := V(\xi)$, for $\mu = \mathbb{P}_\xi \in \mathcal{P}(\mathcal{X})$.

(ii) **Dynamic Programming (DP)**. $V = V^\star$, hence satisfies the Bellman fixed point equation:

$$V(\mu) \quad = \quad \mathcal{T}[V](\mu) \; = \; \sup_{\hat{a} \in \hat{A}} \mathcal{T}^{\hat{a}}[V](\mu), \quad \mu \in \mathcal{P}(\mathcal{X}).$$

(iii) For all $\epsilon > 0$, there exists an ϵ-**optimal randomized feedback control** for $V(\xi)$ in the form:

$$\alpha_t^\varepsilon \quad := \quad \mathsf{a}_\epsilon(\mathbb{P}^0_{X_t}, X_t, U_t), \quad t \in \mathbb{N}.$$

where $(U_t)_t$ sequence of i.i.d. $\mathcal{G}$-measurable $\sim \mathcal{U}([0,1])$, for some measurable function $\mathsf{a}_\epsilon(\mu, x, u)$ on $\mathcal{P}(\mathcal{X}) \times \mathcal{X} \times [0,1]$ constructed from the argmax in $\mathcal{T}^{\hat{a}}$.

Outline

Convergence of the N-agent MDP

• Usually based on propagation of chaos on state process $X^{i,N}$ towards X pathwisely or in law (symmetry arguments is crucial), and then deduce convergence of V^N towards V

• Here, we do not have in general propagation of chaos on $X^{i,N,\boldsymbol{\alpha}}$ controlled by $\boldsymbol{\alpha} = (\alpha^1, \ldots, \alpha^N)$, which is not assumed to be symmetric.

Convergence of the N-agent MDP

• Usually based on propagation of chaos on state process $X^{i,N}$ towards X pathwisely or in law (symmetry arguments is crucial), and then deduce convergence of V^N towards V

• Here, we do not have in general propagation of chaos on $X^{i,N,\alpha}$ controlled by $\boldsymbol{\alpha} = (\alpha^1, \ldots, \alpha^N)$, which is not assumed to be symmetric.

• Instead, we prove a propagation of chaos on the Bellman operator of the N-agent MDP defined by

$$\mathcal{T}_N^{\boldsymbol{a}}[W](\boldsymbol{x}) \quad := \quad \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{a}) + \beta \mathbb{E}\Big[W\big(\boldsymbol{F}(\boldsymbol{x}, \boldsymbol{a}, (\varepsilon_1^i)_{i \in [\![1,N]\!]}, \varepsilon_1^0)\big)\Big]$$

where for $\boldsymbol{x} = (x^i)_{i \in [\![1,N]\!]} \in \mathcal{X}^N$, $\boldsymbol{a} = (a^i)_{i \in [\![1,N]\!]} \in A^N$,

$$\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{a}) \quad := \quad \frac{1}{N} \sum_{i=1}^N f\Big(x^i, a^i, \frac{1}{N} \sum_{j=1}^N \delta_{(x^j, a^j)}\Big)$$

$$\boldsymbol{F}(\boldsymbol{x}, \boldsymbol{a}, (e^i)_{i \in [\![1,N]\!]}, e^0) \quad := \quad \Big(F(x^i, a^i, \frac{1}{N} \sum_{j=1}^N \delta_{(x^j, a^j)}, e^i, e^0)\Big)_{i \in [\![1,N]\!]} \in \mathcal{X}^N.$$

Propagation of chaos of the Bellman operator

For V value function on $\mathcal{P}(\mathcal{X})$ of the CMKV-MDP, we set $\check{V}$ defined on $\mathcal{X}^N$ by

$$\check{V}(\boldsymbol{x}) \quad := \quad V(\mu_N[\boldsymbol{x}]), \qquad \text{for } \boldsymbol{x} = (x^i)_{i\in[\![1,N]\!]} \in \mathcal{X}^N,$$

where $\mu_N[\boldsymbol{x}] = \frac{1}{N}\sum_{i=1}^N \delta_{x^i}$.

Proposition

There exists some positive constant C s.t. for all $\boldsymbol{x} = \mathcal{X}^N$, $N \in \mathbb{N}^*$,

$$\Big|\sup_{\boldsymbol{a}\in A^N} \mathcal{T}_N^{\boldsymbol{a}}[\check{V}](\boldsymbol{x}) \; - \; \sup_{\hat{a}\in\hat{A}} \mathcal{T}^{\hat{a}}[V](\mu_N[\boldsymbol{x}])\Big| \quad \leq \quad C M_N{}^{\gamma},$$

where $\gamma = \min\Big[1, \frac{|\ln\beta|}{(\ln 2K_F)_+}\Big]$, *and*

$$M_N \quad := \quad \sup_{\nu\in\mathcal{P}(\mathcal{X}\times A)} \mathbb{E}\big[\mathcal{W}(\nu_N,\nu)\big], \qquad (\nu_N \text{ empirical measure of } \nu).$$

Remark. From Fournier-Guillin 15: $M_N \underset{N\to\infty}{\to} 0$, and for $\mathcal{X}\times A \subset \mathbb{R}^d$

- $M_N = \mathcal{O}(N^{-\frac{1}{2}})$ for $d = 1$
- $M_N = \mathcal{O}(N^{-\frac{1}{2}}\log(1+N))$ for $d = 2$
- $M_N = \mathcal{O}(N^{-\frac{1}{d}})$ for $d \geq 3$

Convergence rate of N-agent MDP

Theorem

1. Value function. There exists some positive constant C s.t. for all $\boldsymbol{x} \in \mathcal{X}^N$,

$$\left|V^N(\boldsymbol{x}) - V(\mu_N[\boldsymbol{x}])\right| \leq CM_N^\gamma.$$

2. Optimal control. Let $\mathrm{a}_\varepsilon : \mathcal{P}(\mathcal{X}) \times \mathcal{X} \times [0,1] \to A$, be an ε-optimal randomized feedback policy for CMKV-MDP. Then, it defines an $(\varepsilon + \mathcal{O}(M_N^\gamma))$-optimal randomized feedback control $\boldsymbol{\alpha}^\varepsilon = (\alpha^{\varepsilon,i})_{i\in[\![1,N]\!]}$ for the N-agent MDP with

$$\alpha_t^{\varepsilon,i} = \mathrm{a}_\varepsilon(\mu_N[\boldsymbol{X}_t], X_t^i, U_t^i), \quad t \in \mathbb{N},\ i = 1,\dots,N,$$

where $\boldsymbol{X} = (X^i)_{i\in[\![1,N]\!]}$ is the state of the N-agent controlled by $\boldsymbol{\alpha}^\varepsilon$, and U_t^i, $i = 1,\dots,N$, $t \in \mathbb{N}$, are i.i.d. $\sim \mathcal{U}([0,1])$.

Interpretation: Policies to agents in the population are applied randomly by the central planner in a non-symmetric assignment.

Outline

Back to targeted advertising example: spaces and noise

- State space $\mathcal{X} = \{0, 1\}$:
  - $x = 1$ (resp. 0): customer (not customer) of company **C**
- Action space $A = \{0, 1\}$:
  - $a = 1$ (resp. 0): **I** displays (or not) an ad
- For each **SN** user i:
  - ε_t^i: uniform r.v. representing e.g. time spent at day t on a forum about product sold by **C**
- For simplicity here, no common noise

Targeted advertising example: dynamics and reward

- State transition function:

$$F(x, a, \mu, e) = \begin{cases} \mathbf{1}_{e>\mu(\{0\})-2\eta a} & \text{if} \quad x = 0 \\ \mathbf{1}_{e<\mu(\{1\})+2\eta a} & \text{if} \quad x = 1. \end{cases}$$

  - Large e: eager to change of operator
  - $\mu(\{0\})$: proportion of **SN** users that are not customers of **C**
  - $\eta > 0$: efficiency of ad for incentive to become or remain a customer of **C**

- Reward function: for $x \in \mathcal{X} = \{0, 1\}, a \in A = \{0, 1\}$,

$$f(x, a) = x - ca,$$

  - $c > 0$: ad cost

Lifted to deterministic control problem on $[0,1]$

- State variable: $p_t \equiv$ proportion of **SN** users that are customers of **C**

- (Relaxed) control variable: $q_t \equiv$ probability of displaying an ad to **SN** users (sending an ad to q_t proportion of the **SN** users)

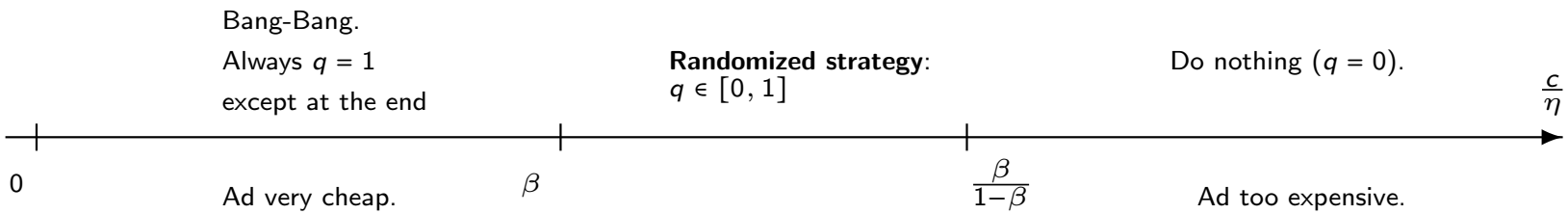
$\rightarrow$

$$p_{t+1} = p_t + q_t \min(\eta, 1 - p_t), \quad t \in \mathbb{N}.$$

- Value function on $[0,1]$:

$$V(p) = \sup_q \sum_{t=0}^{\infty} \beta^t (p_t - c q_t), \quad p \in [0,1].$$

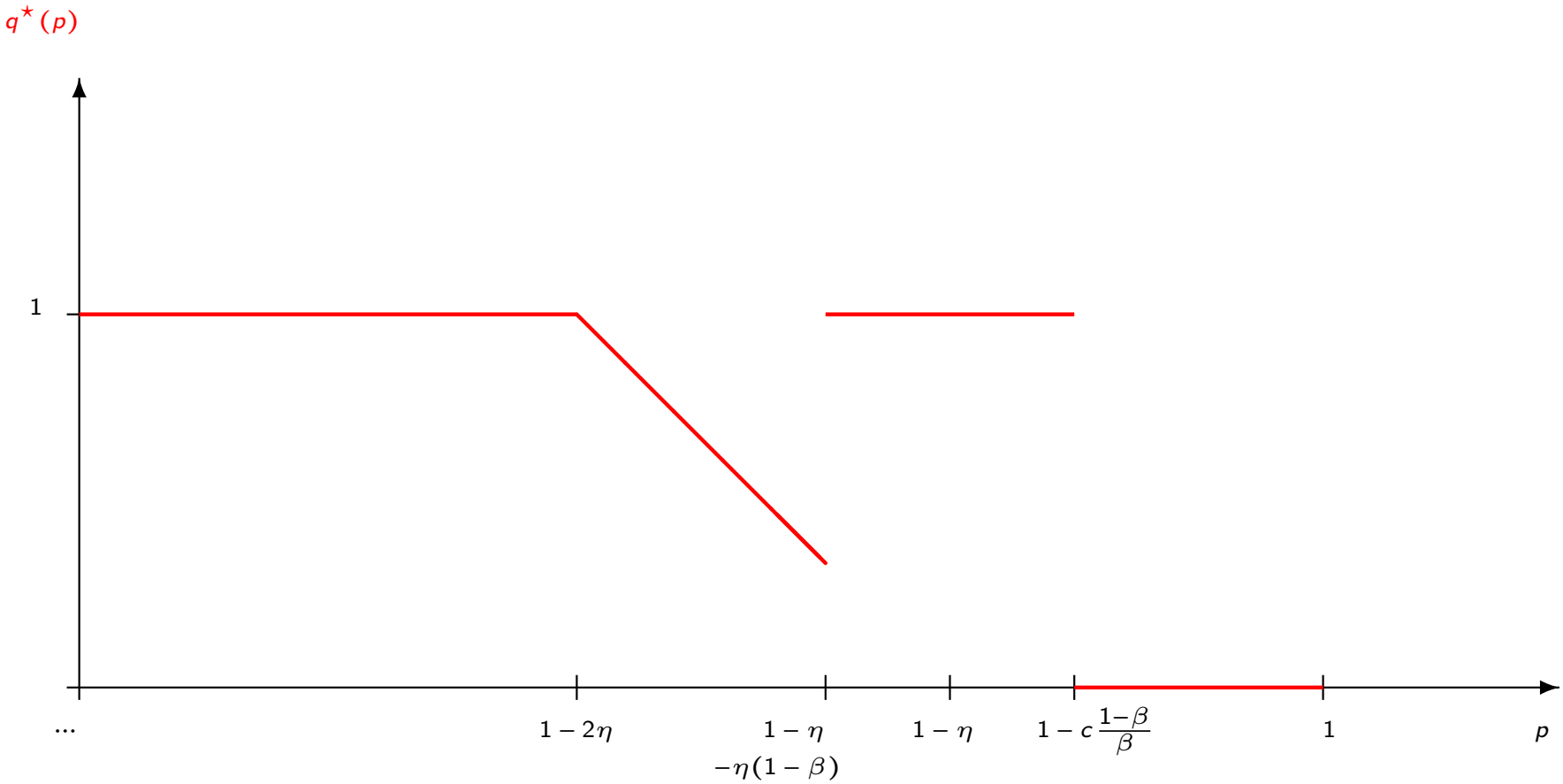
Disjunction depending on the ratio cost/efficiency of ad



Three cases according to the position of $\frac{c}{\eta}$ relative to β and $\frac{\beta}{1-\beta}$.

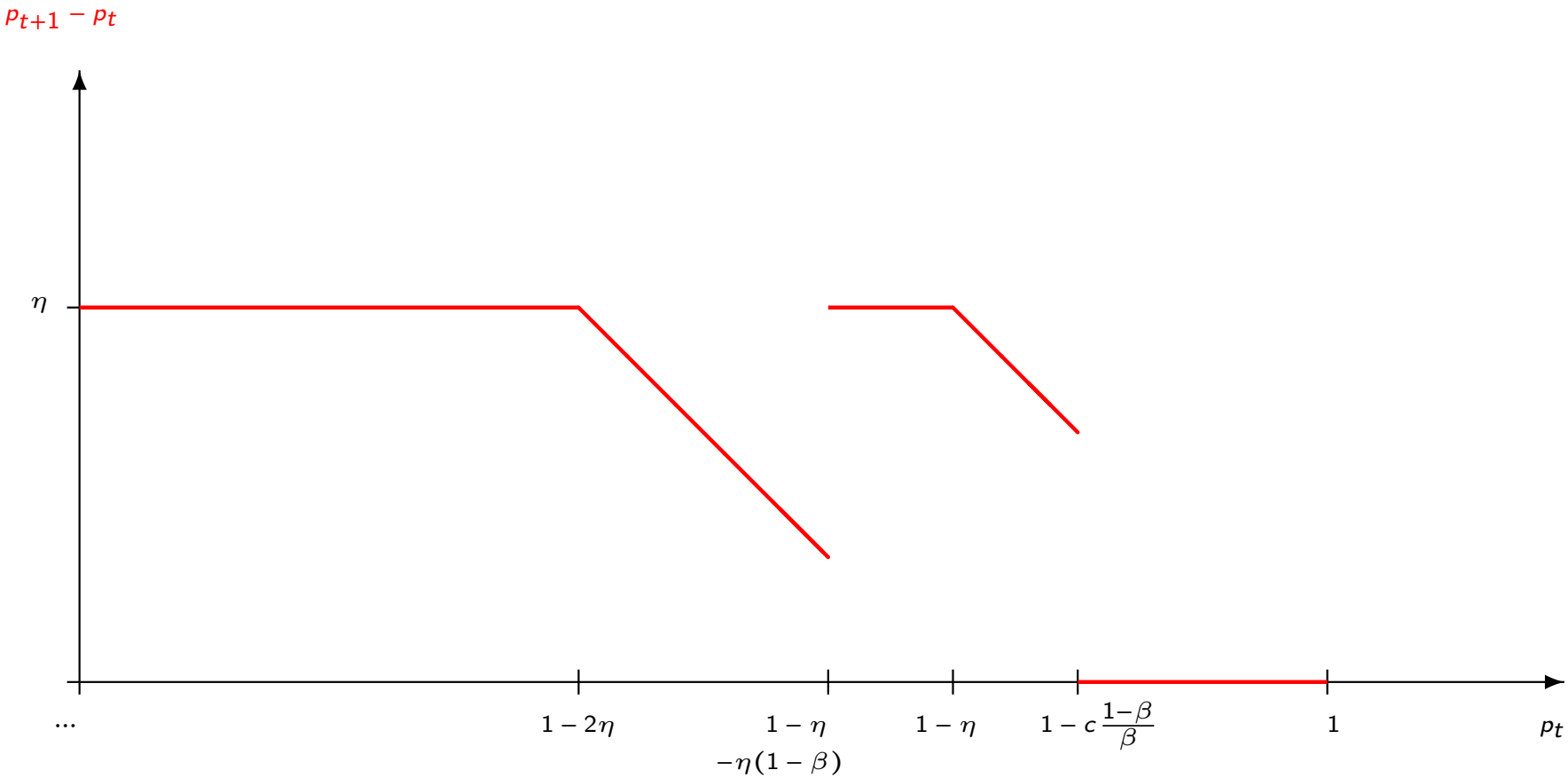
$\rightarrow$ We next focus on the case: $\beta < \frac{c}{\eta} < \frac{\beta}{1-\beta}$.

Optimal control



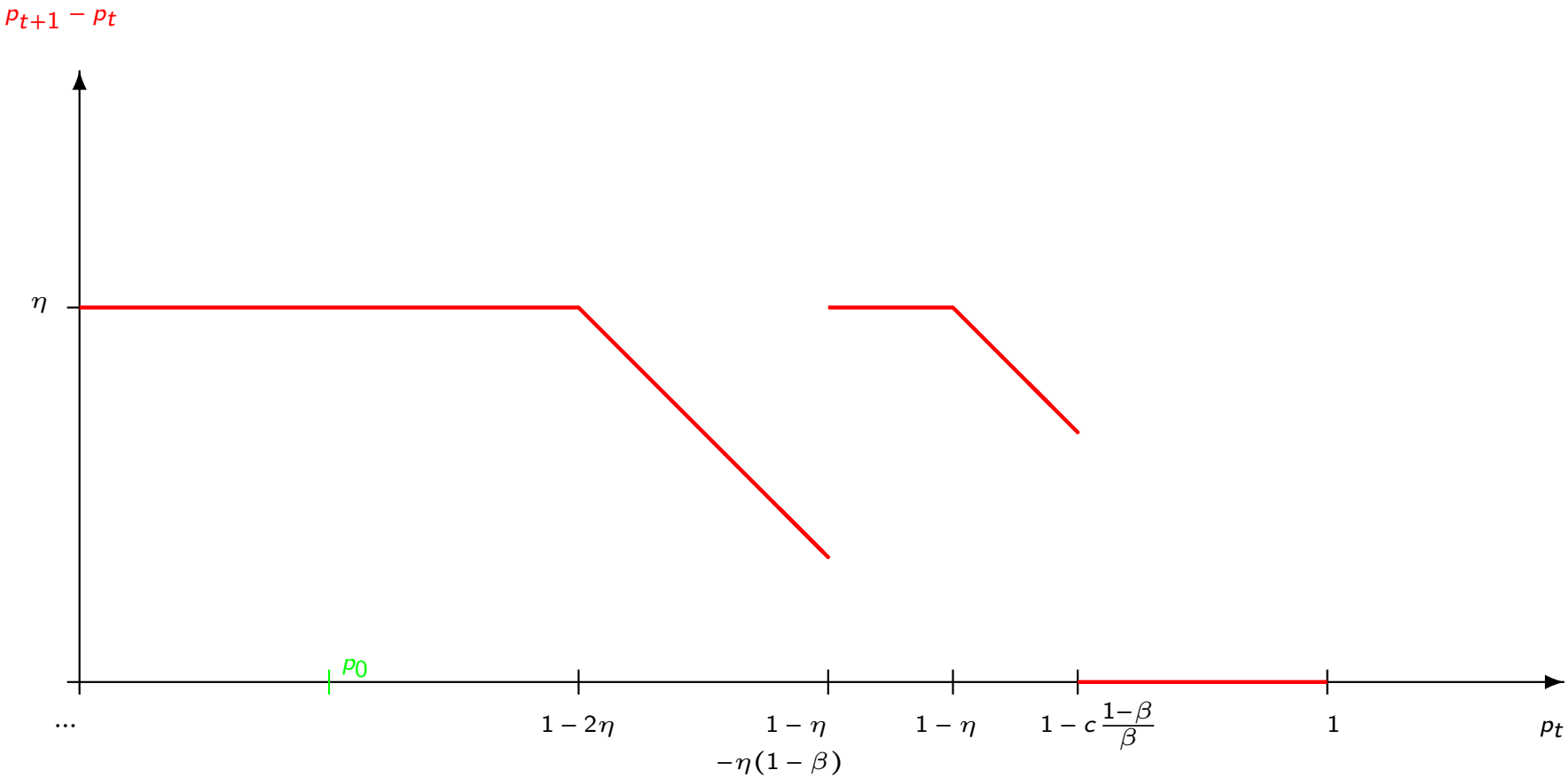
Case $\beta < \frac{c}{\eta} < \frac{\beta}{1-\beta}$. Optimal policy (in red).

Optimal increment of p_t



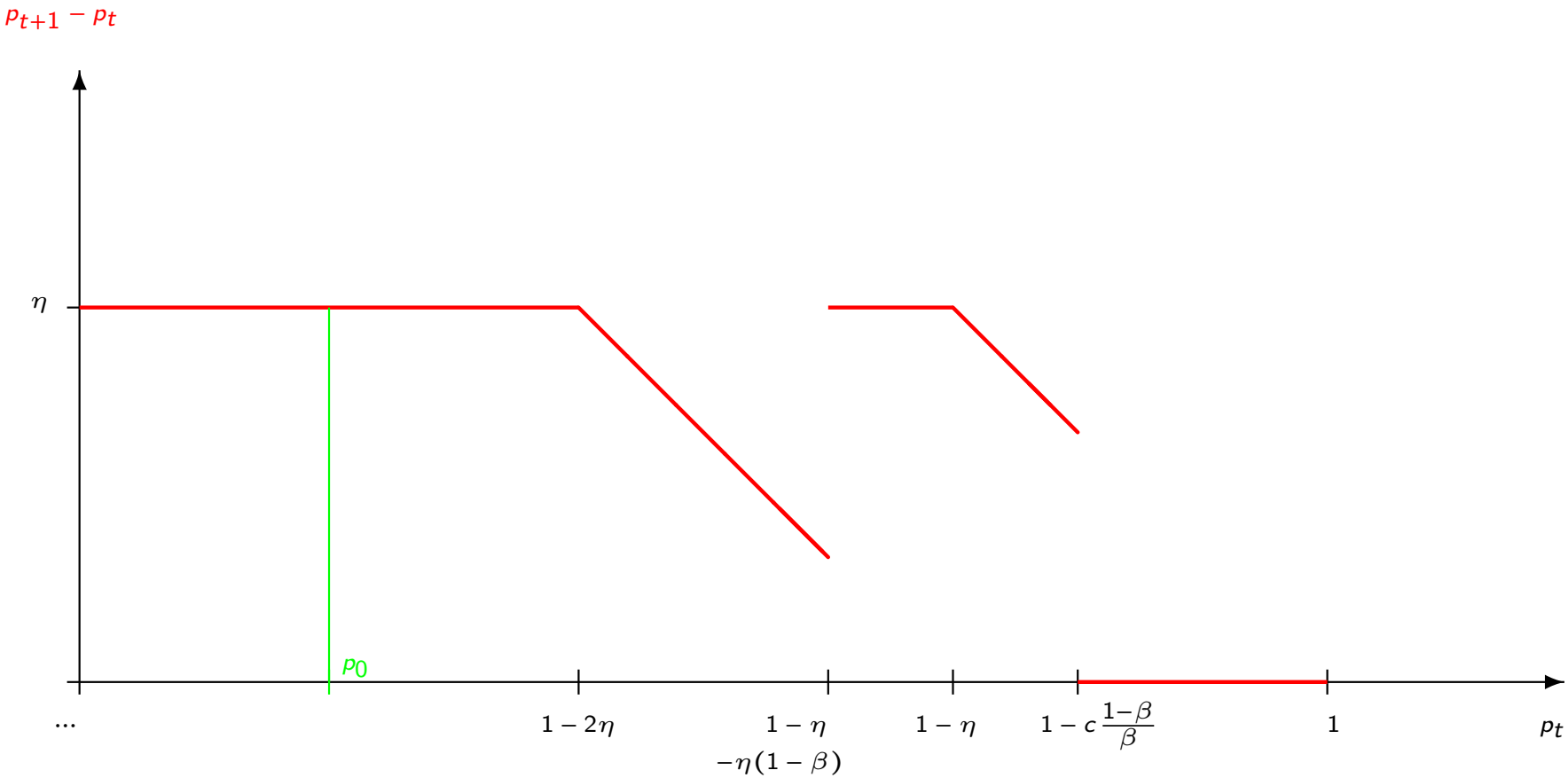
Case $\beta < \frac{c}{\eta} < \frac{\beta}{1-\beta}$. Variational optimal state (in red).

Optimal increment of p_t



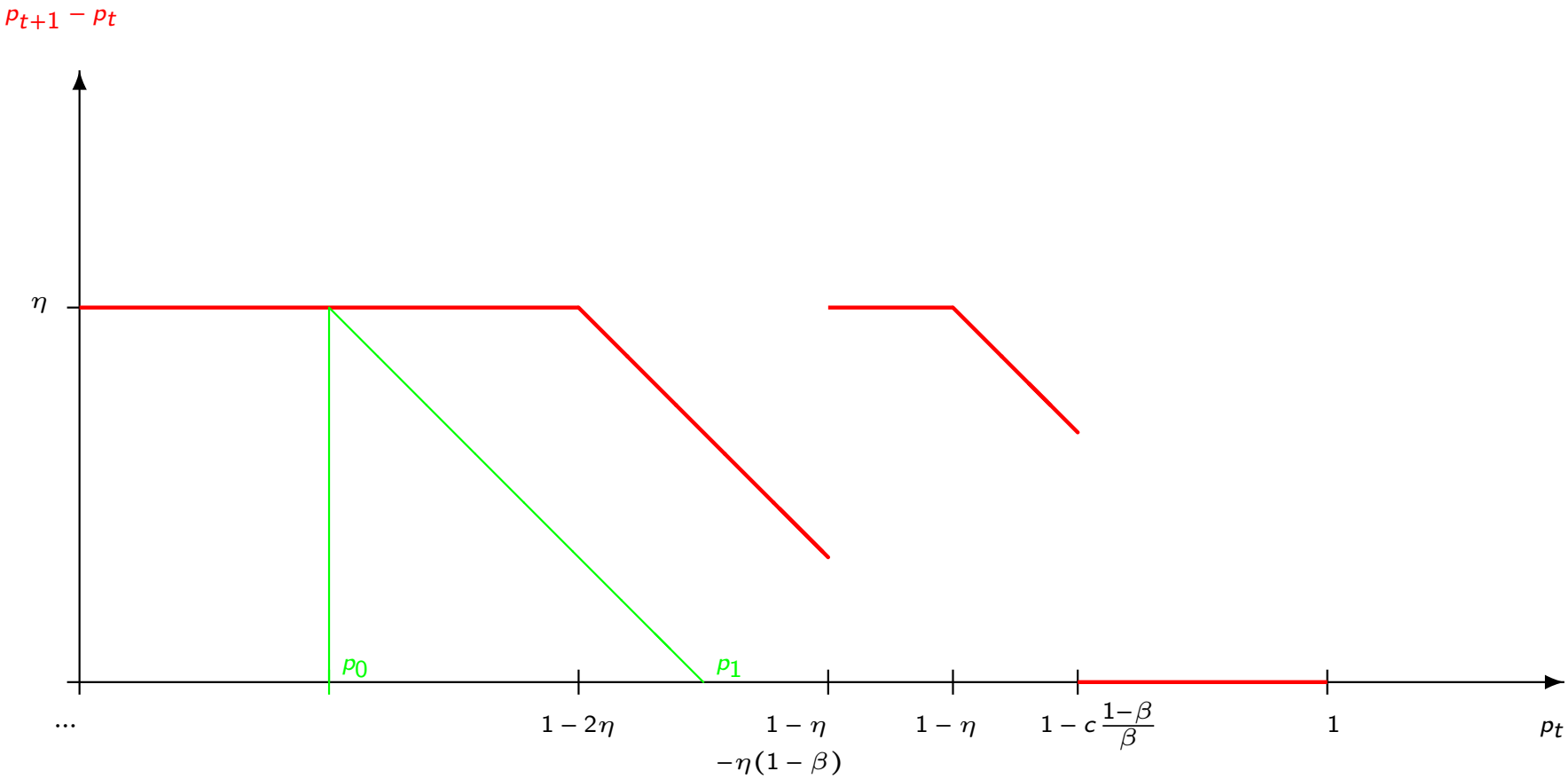
Case $\beta < \frac{c}{\eta} < \frac{\beta}{1-\beta}$. Variational optimal state (in red).

Optimal increment of p_t



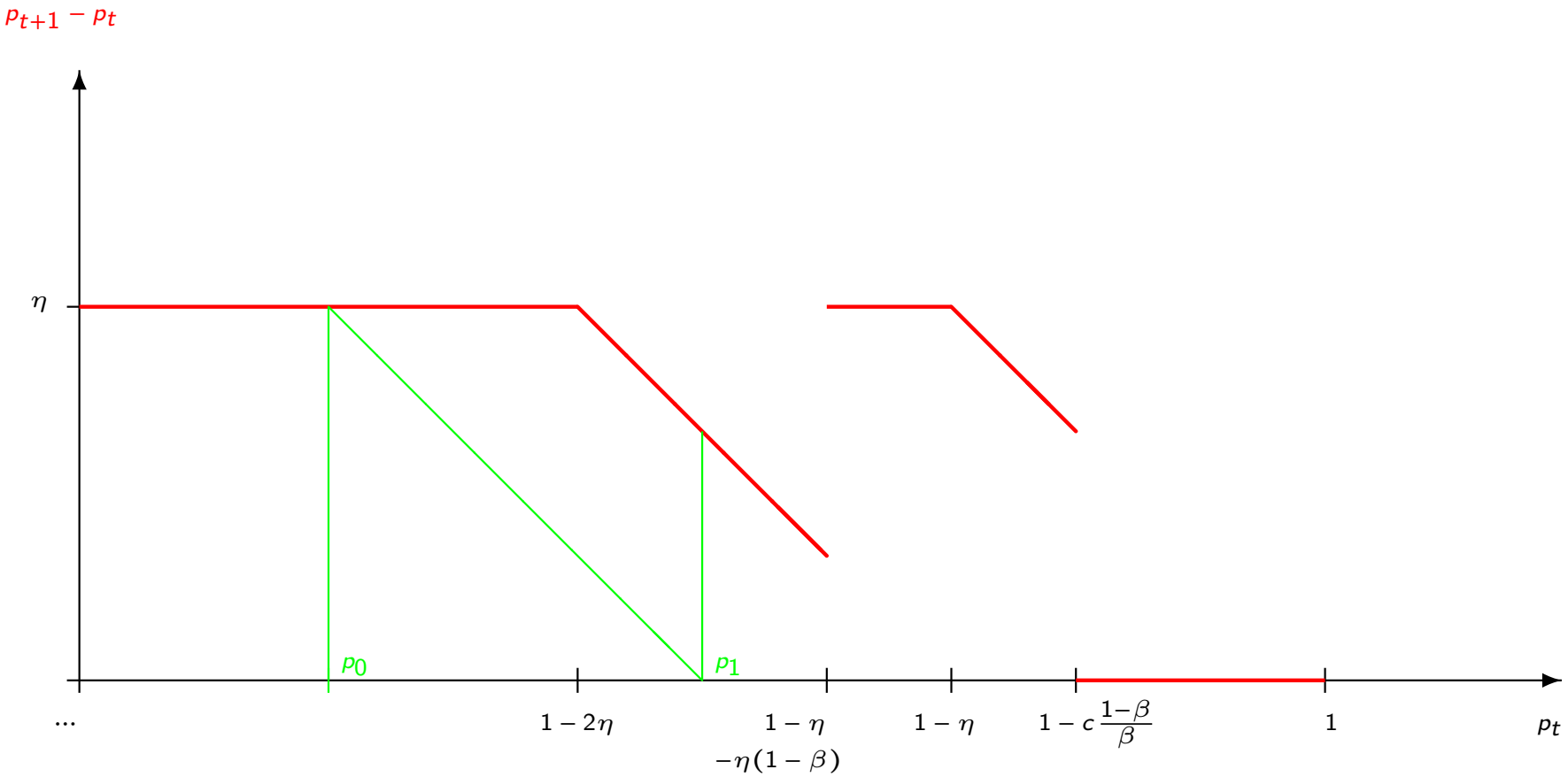
Case $\beta < \frac{c}{\eta} < \frac{\beta}{1-\beta}$. Variational optimal state (in red).

Optimal increment of p_t



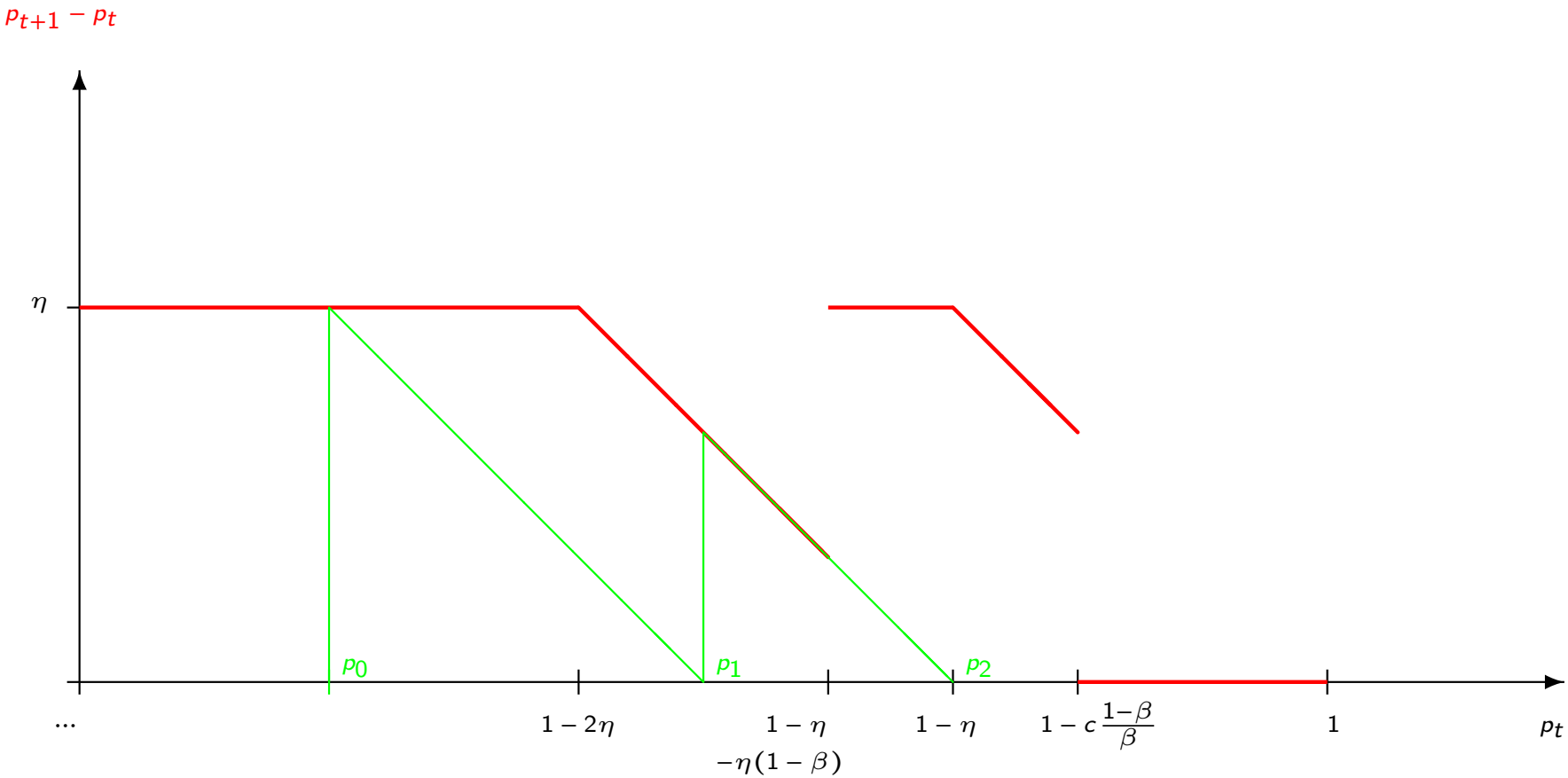
Case $\beta < \frac{c}{\eta} < \frac{\beta}{1-\beta}$. Variational optimal state (in red).

Optimal increment of p_t



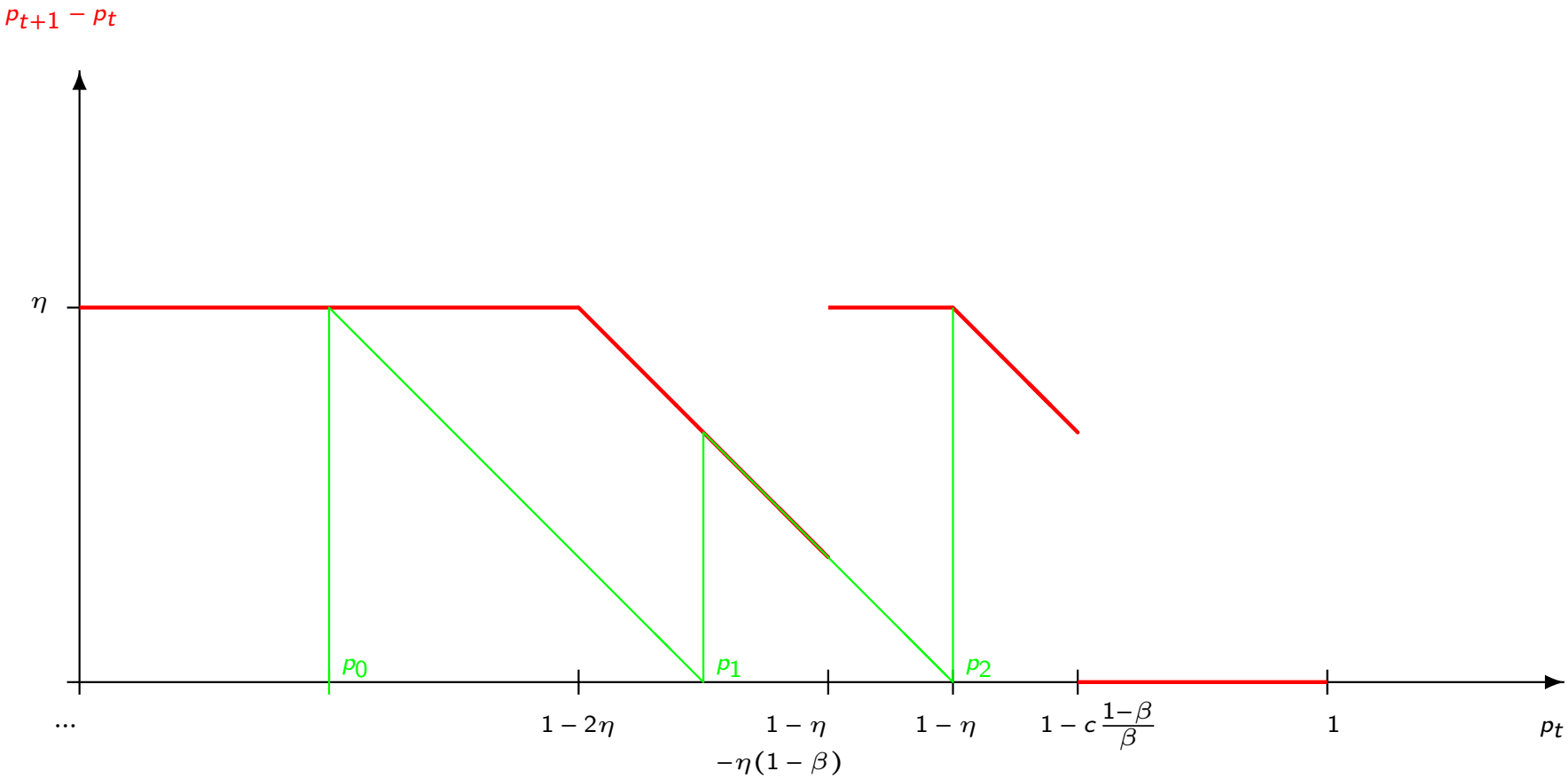
Case $\beta < \frac{c}{\eta} < \frac{\beta}{1-\beta}$. Variational optimal state (in red).

Optimal increment of p_t



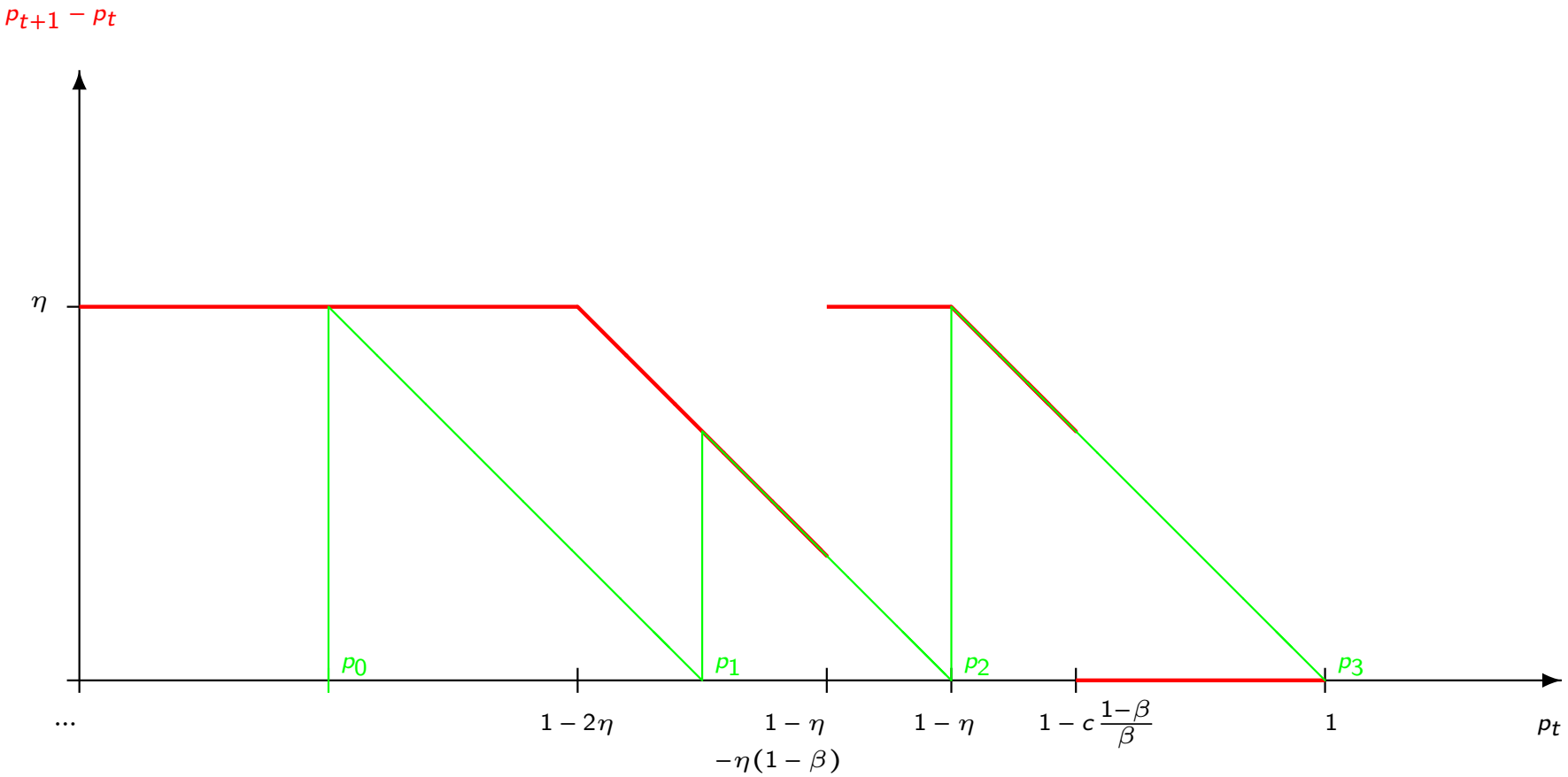
Case $\beta < \frac{c}{\eta} < \frac{\beta}{1-\beta}$. Variational optimal state (in red).

Optimal increment of p_t



Case $\beta < \frac{c}{\eta} < \frac{\beta}{1-\beta}$. Variational optimal state (in red).

Optimal increment of p_t



Case $\beta < \frac{c}{\eta} < \frac{\beta}{1-\beta}$. Variational optimal state (in red).

Summary of main results

- CMKV-MDP lifted to optimization problem on space of laws with relaxed controls:
  - Dynamic Programming Bellman fixed point equation characterizing the value function
  - ε-optimal randomized feedback policy a_ε
- Examples of explicit resolution of the lifted MDP
- N-agent MDP $\underset{N\to\infty}{\longrightarrow}$ CMKV-MDP with explicit rate of convergence.
- (Approximate) optimal randomized feedback control of CMKV-MDP $\to$ Quantitative approximation of optimal control for the N-agent MDP

Some remarks

- Open loop vs feedback controls vs randomized feedback controls
  - In standard MDP, it is well-known that: sup. over open-loop/feedback/randomized feedback controls give same value
  - Here with mean-field dependence, we have:

$$\begin{aligned} \text{Sup. over open-loop control} &= \text{Sup. over randomized feedback control} \\ &> \text{Sup. over feedback control.} \end{aligned}$$

  - This differs from continuous time McKean-Vlasov control where randomization does not yield greater gain

References

• M. Motte, H.P.: Mean-field Markov decision process with common noise and open-loop controls, to appear in *Annals of Applied Probability*.

• M. Motte: Mathematical models for large populations, behavorial economics, and targeted advertising, PhD thesis (2021).

Thank you for your attention