Learning methods for mean-field models: application to power consumption control

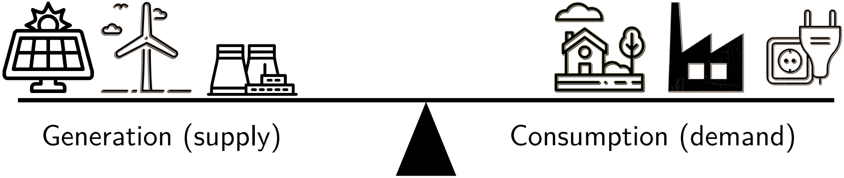
Bianca Marin Moreno[1,2]

Margaux Brégère[1], Pierre Gaillard[2] and Nadia Oudjane[1]

[1]EDF R&D, Palaiseau, France

[2]INRIA Grenoble, THOTH, Grenoble, France

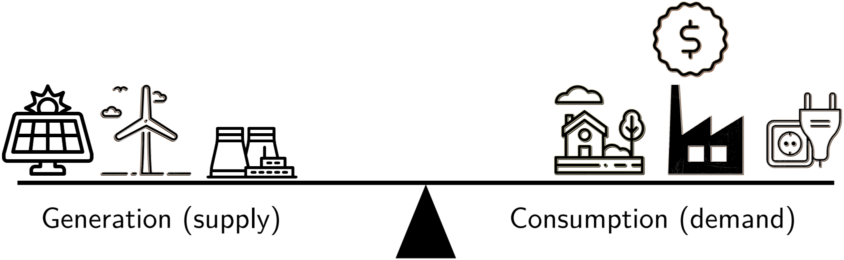
Balancing the power grid



- Difficulties on the supply side:
  - Integration of renewable energy → intermittent nature
  - Energy storage devices and energy importation → costly alternatives

Demand-Side Management

- Solution: adjust energy consumption to better match the energy supply



Demand-Side Management

- TCLs: Thermostatically Controlled Loads
  - Electrical heating or cooling elements controlled by a thermostat: water-heaters, ar conditioners, refrigerators, etc
  - Flexible loads
- Smart meters
  - Allow communication between load and supplier in near real time

Control of a population of water-heaters

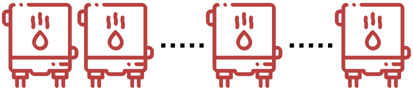
- **Goal:** Control the average consumption of a population of water-heaters (Busic and Meyn, 2016; Bendotti et al., 2021)

Control of a population of water-heaters

- **Goal:** Control the average consumption of a population of water-heaters (Busic and Meyn, 2016; Bendotti et al., 2021)

Individual consumption (time step n) | Average consumption (time step n)

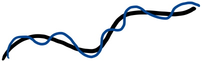
$1 \quad 2 \quad i \quad M$



$ON_n^1 \quad ON_n^2 \quad ON_n^i \quad ON_n^M \implies \frac{1}{M}\sum_{i=1}^{M} ON_n^i$

- in order to track a **reference profile** (γ_n) by sending a control signal (π_n)

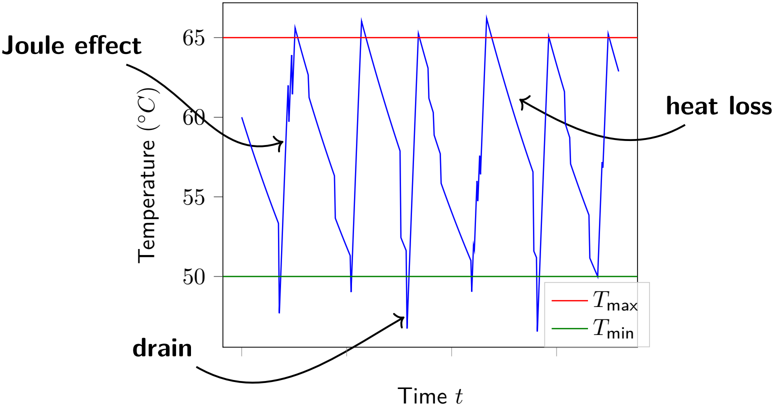
$$\pi_n \implies \begin{cases} \text{device } 1 \to ON_n^1 \\ \vdots \\ \text{device } i \to ON_n^i \\ \vdots \\ \text{device } M \to ON_n^M \end{cases} \implies \underbrace{\frac{1}{M}\sum_{i=1}^{M} ON_n^i}_{\text{average cons.}} \approx \underbrace{\gamma_n}_{\text{target}}$$



Setting and Model

Water-heater uncontrolled dynamics

- $[T_{\text{min}}, T_{\text{max}}]$ = temperature deadband



Water-heater controlled dynamics

- **Goal:** Control the average consumption of a population of water-heaters
- Idea: probability of turning ON/OFF before leaving $[T_{\mathsf{min}},T_{\mathsf{max}}]$
- Formulation as a Markov Decision Process:

Water-heater controlled dynamics

- **Goal:** Control the average consumption of a population of water-heaters
- Idea: probability of turning ON/OFF before leaving $[T_{\mathsf{min}}, T_{\mathsf{max}}]$
- Formulation as a Markov Decision Process:
  - state space $\mathcal{X}$: ON/OFF and temperature

Water-heater controlled dynamics

- **Goal:** Control the average consumption of a population of water-heaters
- Idea: probability of turning ON/OFF before leaving $[T_{\text{min}}, T_{\text{max}}]$
- Formulation as a Markov Decision Process:
  - state space $\mathcal{X}$: ON/OFF and temperature
  - action space $\mathcal{A}$: turn ON/OFF

Water-heater controlled dynamics

- **Goal:** Control the average consumption of a population of water-heaters
- Idea: probability of turning ON/OFF before leaving $[T_{\text{min}}, T_{\text{max}}]$
- Formulation as a Markov Decision Process:
  - state space $\mathcal{X}$: ON/OFF and temperature
  - action space $\mathcal{A}$: turn ON/OFF
  - policy $(\pi_n)_{n \leq N}$ = control signal to learn
    - $\pi_n(a_n|x_n)$ = probability of choosing action a_n (turn ON/OFF) given current state x_n (ON/OFF and temperature)

Water-heater controlled dynamics

- **Goal:** Control the average consumption of a population of water-heaters
- Idea: probability of turning ON/OFF before leaving $[T_{\text{min}}, T_{\text{max}}]$
- Formulation as a Markov Decision Process:
  - state space $\mathcal{X}$: ON/OFF and temperature
  - action space $\mathcal{A}$: turn ON/OFF
  - policy $(\pi_n)_{n \leq N}$ = control signal to learn
    - $\pi_n(a_n|x_n)$ = probability of choosing action a_n (turn ON/OFF) given current state x_n (ON/OFF and temperature)
  - probability kernel $x_{n+1} \sim p_n(\cdot|x_n, a_n)$ (drains)

Optimisation problem and mean field approach

- M water-heaters
- **Goal:** find a control signal (π_n) to approach a **reference profile** (γ_n)

$$\min_{\pi \in (\Delta_{\mathcal{A}})^{\mathcal{X} \times N}} \mathbb{E}\left[\sum_{n=1}^{N}\left(\frac{1}{M}\sum_{i=1}^{M} ON_n^i(\pi) - \gamma_n\right)^2\right]$$

Optimisation problem and mean field approach

- M water-heaters
- **Goal:** find a control signal (π_n) to approach a **reference profile** (γ_n)

$$\min_{\pi \in (\Delta_{\mathcal{A}})^{\mathcal{X} \times N}} \mathbb{E}\left[\sum_{n=1}^{N}\left(\frac{1}{M}\sum_{i=1}^{M} ON_n^i(\pi) - \gamma_n\right)^2\right]$$

- mean field limit $M \to \infty$:
  - $\mu_n^\pi(x,a) = \mathbb{P}(x_n = x, a_n = a | \pi, (p_n)_n)$= state-action distribution induced by π

$$\frac{1}{M}\sum_{i=1}^{M} ON_n^i(\pi) \longrightarrow \underbrace{\mathbb{E}_{\mu_n^\pi}[\{ON_n(\pi)\}]}_{\text{average cons.}}$$

- Control problem $(\mathcal{C})$

$$\min_{\pi \in (\Delta_{\mathcal{A}})^{\mathcal{X} \times N}} F(\mu^\pi) := \sum_{n=1}^{N} (\mathbb{E}_{\mu_n^\pi}[\{ON_n(\pi)\}] - \gamma_n)^2$$

Optimisation vs. Learning

Main problem:

$$\min_{\pi \in (\Delta_{\mathcal{A}})^{\mathcal{X} \times N}} F(\mu^\pi),$$

where $\mu_n^\pi(x, a) := \mathbb{P}(x_n = x, a_n = a | \pi, (p_n)_n)$

Optimisation

- $p = (p_n)_n$ is known
- **Today's talk:** A novel approach to solve the main problem with known p

Optimisation vs. Learning

Main problem:

$$\min_{\pi \in (\Delta_{\mathcal{A}})^{\mathcal{X} \times N}} F(\mu^\pi),$$

where $\mu_n^\pi(x, a) := \mathbb{P}(x_n = x, a_n = a | \pi, (p_n)_n)$

Optimisation

- $p = (p_n)_n$ is known
- **Today's talk:** A novel approach to solve the main problem with known p

Learning

- Reality: $(p_n)_n$ is unknown
  - User's water consumption behavior is unknown
- Challenge: Learning the model while optimizing
- Work in progress

Optimisation vs. Learning

Main problem:

$$\min_{\pi \in (\Delta_{\mathcal{A}})^{\mathcal{X} \times N}} F(\mu^\pi),$$

where $\mu_n^\pi(x,a) := \mathbb{P}(x_n = x, a_n = a | \pi, (p_n)_n)$

Optimisation

- $p = (p_n)_n$ is known
- **Today's talk:** A novel approach to solve the main problem with known p

Learning

- Reality: $(p_n)_n$ is unknown
  - User's water consumption behavior is unknown
- Challenge: Learning the model while optimizing
- Work in progress

Convex/Concave Utility Reinforcement Learning (CURL)

$$\min_{\pi \in (\Delta_{\mathcal{A}})^{\mathcal{X} \times N}} F(\mu^\pi)$$

- This applies to many others machine learning problems:
  - **Reinforcement learning (Sutton and Barto, 2018):** $F(\mu^\pi) := -\langle \mu^\pi, r \rangle$, for a reward function r
  - **Imitation learning (Ghasemipour et al., 2020):** $F(\mu^\pi) := -D_f(\mu^\pi, \mu^*)$, where D_f is a Bregman divergence induced by a function f
  - **Potential games in mean field games (Geist et al., 2022):** when the reward of the game is $-\nabla F(\mu^\pi)$
- Few algorithms in the literature for CURL: Hazan et al. (2019) (Frank-Wolfe), Geist et al. (2022) (Online Mirror Descent/ Fictitious Play)
- We present a new approach for CURL

Algorithmic approaches

Problem reformulation

$$\min_{\pi \in (\Delta_{\mathcal{A}})^{\mathcal{X} \times N}} F(\mu^{\pi}) := \sum_{n=1}^{N} \left(\mathbb{E}_{\mu_n^{\pi}}[\{ON_n(\pi)\}] - \gamma_n\right)^2$$

⚠ gradient on π? convexity?

Problem reformulation

$$\min_{\pi \in (\Delta_{\mathcal{A}})^{\mathcal{X} \times N}} F(\mu^\pi) := \sum_{n=1}^{N} (\mathbb{E}_{\mu_n^\pi}[\{ON_n(\pi)\}] - \gamma_n)^2$$

⚠ gradient on π? convexity?

$$\Longrightarrow \min_{\mu \in ?} F(\mu)$$

gradient on μ! convexity!

Problem reformulation

$$\min_{\pi \in (\Delta_{\mathcal{A}})^{\mathcal{X}\times N}} F(\mu^\pi) := \sum_{n=1}^{N} \left(\mathbb{E}_{\mu_n^\pi}[\{ON_n(\pi)\}] - \gamma_n\right)^2$$

⚠ gradient on π? convexity?

$$\Longrightarrow \min_{\mu \in ?} F(\mu)$$

gradient on μ! convexity!

$$\mathcal{M}_{\mu_0} := \left\{ (\mu_n)_n \Big| \sum_{a'} \mu_n(x', a') = \sum_{x,a} p_n(x'|x,a)\mu_{n-1}(x,a) \right\}$$

$\mu \in \mathcal{M}_{\mu_0} \longrightarrow \pi \in (\Delta_{\mathcal{A}})^{\mathcal{X}\times N}$ such that $\boldsymbol{\mu}^{\boldsymbol{\pi}} = \boldsymbol{\mu}$

Iterative scheme

- Consider the following iterative scheme at iteration k

$$\mu^{k+1} \in \arg\min_{\mu^\pi \in \mathcal{M}_{\mu_0}} \left\{ \langle \nabla F(\mu^k), \mu^\pi \rangle + \frac{1}{\tau_k} \Gamma(\mu^\pi, \mu^k) \right\} \tag{1}$$

- where Γ is a non-standard regularization

$$\Gamma(\mu^\pi, \mu^{\pi'}) := \sum_{n=1}^{N} \mathbb{E}_{(x,a)\sim\mu_n^\pi(\cdot)} \left[\log \left(\frac{\pi_n(a|x)}{\pi_n'(a|x)} \right) \right]$$

Iterative scheme

- Consider the following iterative scheme at iteration k

$$\mu^{k+1} \in \underset{\mu^\pi \in \mathcal{M}_{\mu_0}}{\arg\min} \left\{ \langle \nabla F(\mu^k), \mu^\pi \rangle + \frac{1}{\tau_k} \Gamma(\mu^\pi, \mu^k) \right\} \tag{1}$$

- where Γ is a non-standard regularization

$$\Gamma(\mu^\pi, \mu^{\pi'}) := \sum_{n=1}^{N} \mathbb{E}_{(x,a)\sim\mu_n^\pi(\cdot)} \left[\log \left(\frac{\pi_n(a|x)}{\pi'_n(a|x)} \right) \right]$$

First result:

- Dynamic Programming yielding in a simple closed-form solution for (1): $\mu^{k+1} := \mu^{\pi^{k+1}}$ such that

$$\pi_n^{k+1}(a|x) := \frac{\pi_n^k(a|x) \exp\left(\tau_k \tilde{Q}_n^k(x,a)\right)}{\sum_{a' \in \mathcal{A}} \pi_n^k(a'|x) \exp\left(\tau_k \tilde{Q}_n^k(x,a')\right)}$$

MD-MFC Algorithm

Algorithm MD-MFC

for $k = 0, ..., K-1$ **do**
  $\mu^k = \mu^{\pi^k}$
  Compute $\tilde{Q}_N^k(x,a)$ for all $(x,a) \in \mathcal{X} \times \mathcal{A}$
  for $n = N, ..., 1$ **do**
    $\forall (x,a) \in \mathcal{X} \times \mathcal{A}:$
    $\pi_n^{k+1}(a|x) = \frac{\pi_n^k(a|x)\exp\left(\tau_k \tilde{Q}_n^k(x,a)\right)}{\sum_{a'} \pi_n^k(a'|x)\exp\left(\tau_k \tilde{Q}_n^k(x,a')\right)}$
    Compute $\tilde{Q}_{n-1}^k(x,a)$
  end for
end for
return μ^{π^K}, π^K

Convergence analysis

<u>**Second result**</u>:

Theorem (MD-MFC convergence)

Let π^* *a minimizer and* K *the number of iteration, thus*

$$\min_{0 \leq s \leq K} F(\mu^{\pi^s}) - F(\mu^{\pi^*}) \leq O\big(\frac{1}{\sqrt{K}}\big)$$

Convergence analysis

Second result:

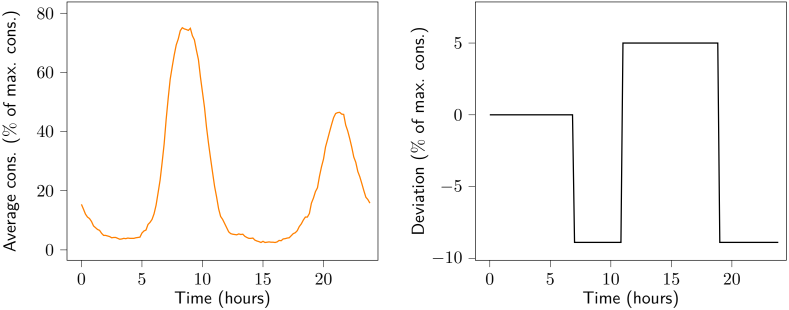
> Theorem (MD-MFC convergence)
>
> *Let π^* a minimizer and K the number of iteration, thus*
>
> $$\min_{0 \le s \le K} F(\mu^{\pi^s}) - F(\mu^{\pi^*}) \le O\left(\frac{1}{\sqrt{K}}\right)$$

Proof idea:

- Γ is a Bregman divergence and is 1-strongly convex with respect to the $\sup_{1 \le n \le N} \| \cdot \|_1$ norm
- $\Rightarrow$ MD-MFC converge as a Mirror Descent algorithm

Experiments

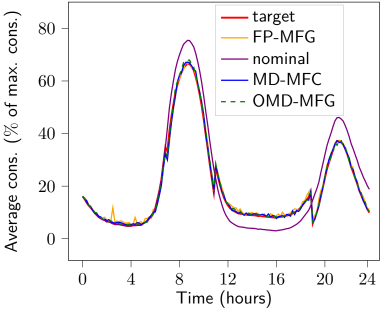
Target = uncontrolled dynamics + deviation



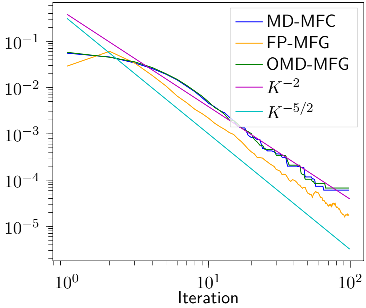
(a) Average consumption. (b) Eight hours step deviation signal.

- Nb of water-heaters = 10^4
- Time horizon = one day
- Time step = 10 minutes
- Heaters are homogeneous and randomly initialised
- Drains adapted from SMACH data

Results



(a) Consumption simulation

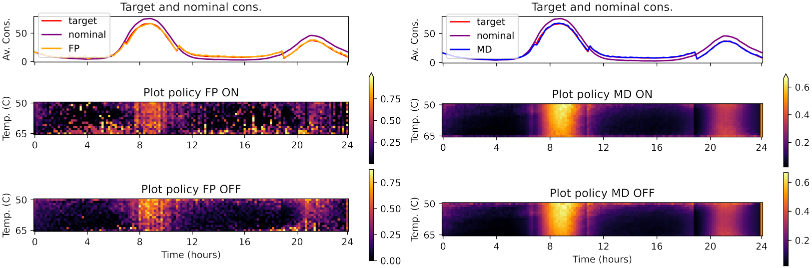


(b) Objective function

- FP-MFG (Perrin et al., 2020), OMD-MFG (Pérolat et al., 2021)

Optimal policy from FP-MFG and MD-MFC

- Different policies may lead to the same consumption
- Regularization in MD provides more interesting solutions from an operational point of view



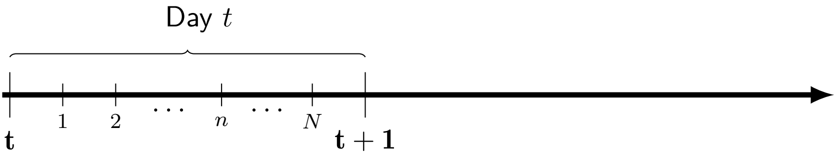
(a) Policy FP-MFG (b) Policy MD-MFC

Figure: Optimal policy FP-MFG and MD-MFC

Work in progress...

Learning: Online protocol idea for unknown dynamics

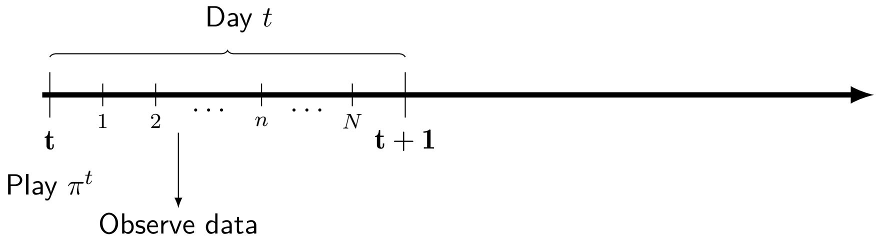
Now we want to calculate a policy every day t over an horizon T, but we need to learn the model dynamics



Play π^t

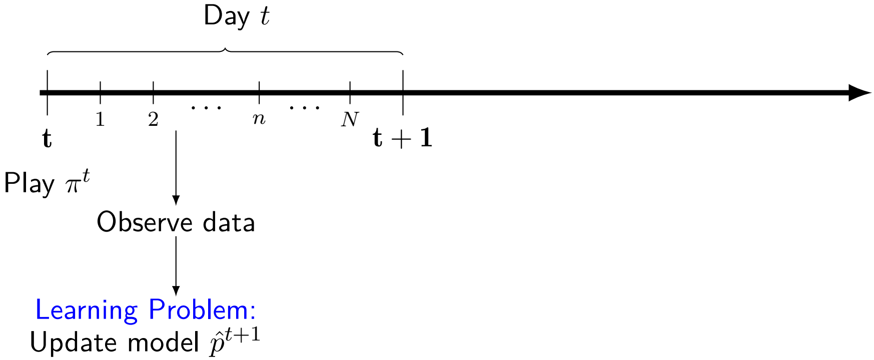
Learning: Online protocol idea for unknown dynamics

Now we want to calculate a policy every day t over an horizon T, but we need to learn the model dynamics



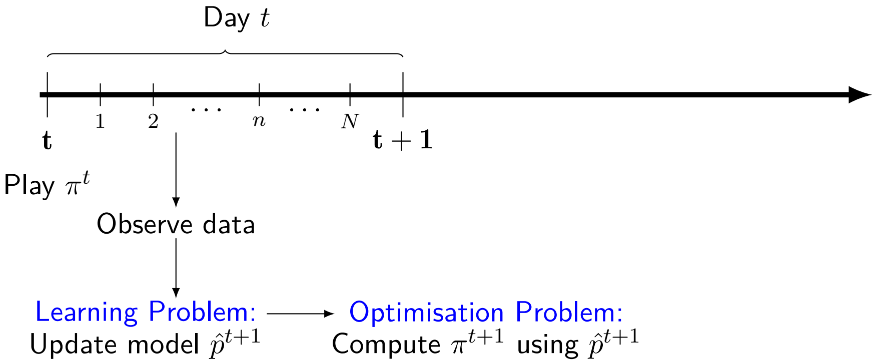
Learning: Online protocol idea for unknown dynamics

Now we want to calculate a policy every day t over an horizon T, but we need to learn the model dynamics



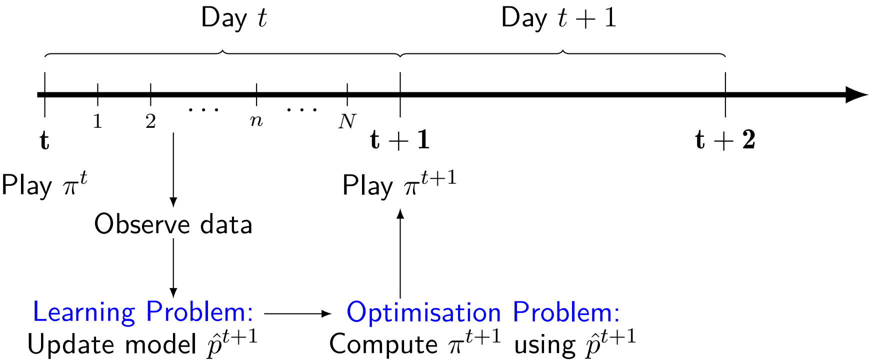
Learning: Online protocol idea for unknown dynamics

Now we want to calculate a policy every day t over an horizon T, but we need to learn the model dynamics



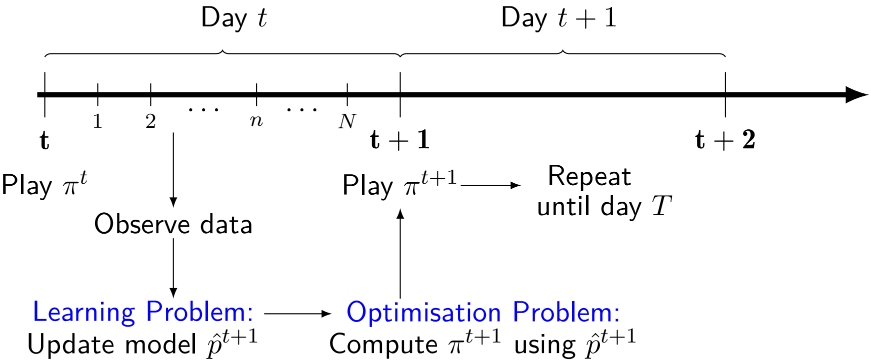
Learning: Online protocol idea for unknown dynamics

Now we want to calculate a policy every day t over an horizon T, but we need to learn the model dynamics



Learning: Online protocol idea for unknown dynamics

Now we want to calculate a policy every day t over an horizon T, but we need to learn the model dynamics



Conclusion

- Goal: Control the average energy consumption of a water-heater's population to better match a target signal
  - Innovative modelling of water-heaters as MDPs
  - New algorithm with theoretical results
  - Experimental results
    - Validating the efficacy of MD-MFC
    - Showing that MD-MFC is relevant to the industrial problem
  - Extension of the algorithm to a more realistic case (unknown dynamics and adversarial objective function)

Conclusion

- Goal: Control the average energy consumption of a water-heater's population to better match a target signal
  - Innovative modelling of water-heaters as MDPs
  - New algorithm with theoretical results
  - Experimental results
    - Validating the efficacy of MD-MFC
    - Showing that MD-MFC is relevant to the industrial problem
  - Extension of the algorithm to a more realistic case (unknown dynamics and adversarial objective function)

Thank you for your attention! Questions?

References

Bendotti, P., Oudjane, N., and Wan, C. (2021). Distributed control of ï¬ exible loads by mean-ï¬ eld inversion.

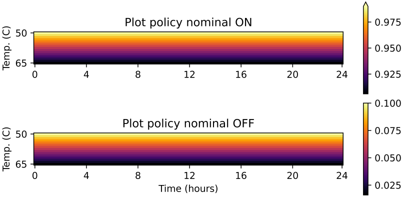
Busic, A. and Meyn, S. (2016). Distributed randomized control for demand dispatch. pages 6964–6971.

Geist, M., Pérolat, J., Laurière, M., Elie, R., Perrin, S., Bachem, O., Munos, R., and Pietquin, O. (2022). Concave utility reinforcement learning: The mean-field game viewpoint. In *Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems*, AAMAS '22, page 489â 497, Richland, SC. International Foundation for Autonomous Agents and Multiagent Systems.

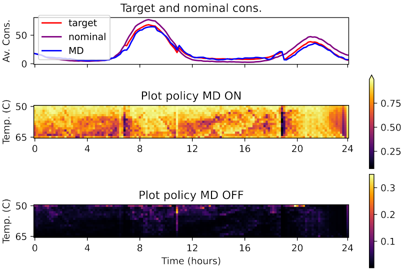
Ghasemipour, S. K. S., Zemel, R., and Gu, S. (2020). A divergence minimization perspective on imitation learning methods. In Kaelbling, L. P., Kragic, D., and Sugiura, K., editors, *Proceedings of the Conference on Robot Learning*, volume 100 of *Proceedings of Machine Learning Research*, pages 1259–1277. PMLR.

Hazan, E., Kakade, S., Singh, K., and Van Soest, A. (2019).

Optimal policy with nominal initialization



(a) Nominal policy deviation



(b) Policy MD-MFC

Figure: Optimal policies for Fictitious Play and Mirror Descent.

Greedy-MD

Greedy Mirror Descent:

- Initialize algorithm with π^1
- For each episode $t \in \{1, ..., T\}$:
  - Play π^t and observe data $(x_1^t, a_1^t, \ldots, x_N^t, a_N^t)$

Greedy-MD

Greedy Mirror Descent:

- Initialize algorithm with π^1
- For each episode $t \in \{1, ..., T\}$:
  - Play π^t and observe data $(x_1^t, a_1^t, \ldots, x_N^t, a_N^t)$
  - Use data to update a probability kernel estimation p^t such that $\|p^t(\cdot|x,a) - p(\cdot|x,a)\|_1 \leq O\left(\frac{1}{\sqrt{t}}\right)$

Greedy-MD

Greedy Mirror Descent:

- Initialize algorithm with π^1
- For each episode $t \in \{1, ..., T\}$:
  - Play π^t and observe data $(x_1^t, a_1^t, \ldots, x_N^t, a_N^t)$
  - Use data to update a probability kernel estimation p^t such that $\|p^t(\cdot|x,a) - p(\cdot|x,a)\|_1 \leq O\left(\frac{1}{\sqrt{t}}\right)$
  - Observe objective function F^t

Greedy-MD

Greedy Mirror Descent:

- Initialize algorithm with π^1
- For each episode $t \in \{1, ..., T\}$:
  - Play π^t and observe data $(x_1^t, a_1^t, \ldots, x_N^t, a_N^t)$
  - Use data to update a probability kernel estimation p^t such that $\|p^t(\cdot|x,a) - p(\cdot|x,a)\|_1 \leq O\left(\frac{1}{\sqrt{t}}\right)$
  - Observe objective function F^t
  - Compute π^{t+1} solving **one iteration of MD-MFC** with F^t, π^t, and p^t

Greedy-MD

Greedy Mirror Descent:

- Initialize algorithm with π^1
- For each episode $t \in \{1, ..., T\}$:
  - Play π^t and observe data $(x_1^t, a_1^t, \ldots, x_N^t, a_N^t)$
  - Use data to update a probability kernel estimation p^t such that $\|p^t(\cdot|x,a) - p(\cdot|x,a)\|_1 \leq O\left(\frac{1}{\sqrt{t}}\right)$
  - Observe objective function F^t
  - Compute π^{t+1} solving **one iteration of MD-MFC** with F^t, π^t, and p^t
- **Greedy Mirror Descent achieves sub-linear regret!**
  - $O(\sqrt{T \log(T)})$