Optimal Experiment Design in Markov Chains

ETH Zürich, Broad Institute

Mojmír Mutný

Andreas Krause

Tadeusz Janik

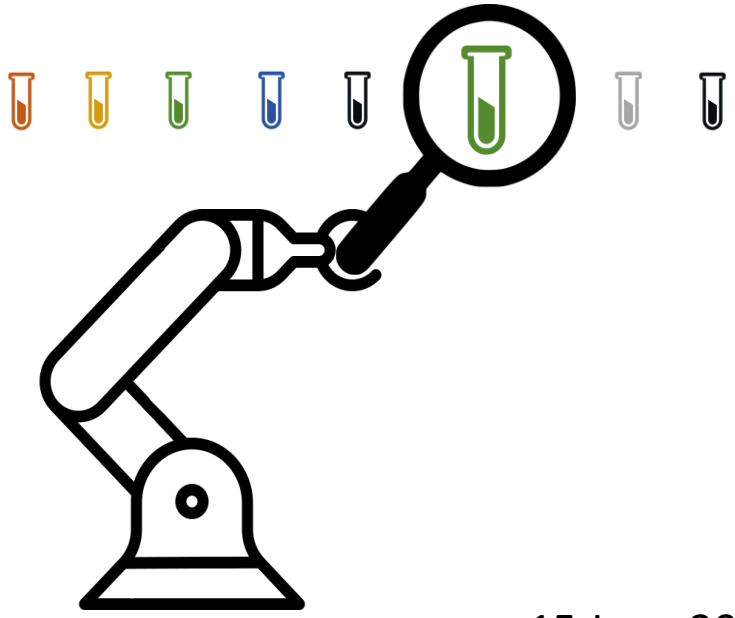
Manish Prajapat

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Jose Folch

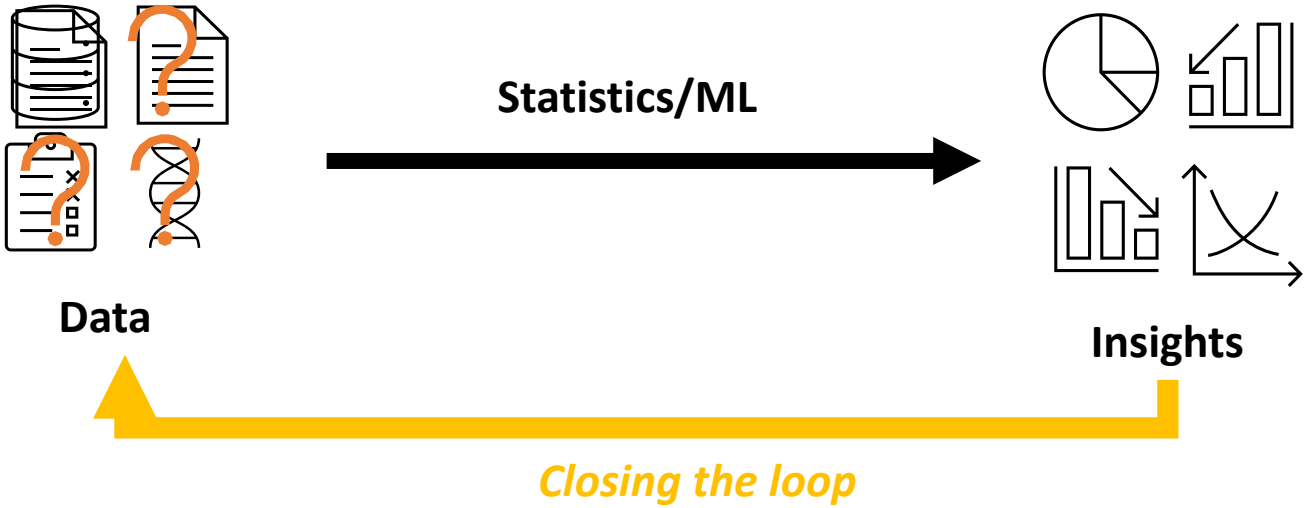
Ruth Misener

Imperial College London



15 June 2026

Statistical Experiment Design?



What if?

- What if we can decide what data we acquire? → **Experiment Design**
- What if the downstream decisions influence what data we gather? → **Adaptive Experiment Design**

What data? Depends…

- How do we model the problem? → Mathematical Model
- What do we want to do with it? → Task
- What data we can get? → Search Space
- What information we already have? → What is uncertain?

In this talk: Experiments are temporally structured decision rules – policies.

Experiment Design and Estimators

General Setup

- **Unknown:** Unknown parameter θ_* (RKHS element)
- **Measurement:** X is a query to $y \sim p(\theta_*^\top \Phi(x))$
- **Evaluation functional** $\mathrm{f(x)} = \Phi(x)^\top \theta_\star$
- **Estimator:** Estimator $\hat{\theta}(\boldsymbol{X}, \boldsymbol{y})$
  - $\boldsymbol{X}$ stacked measurements, $|\mathbf{X}| < \mathrm{T}$
  - and their outcomes $\mathbf{y}$
- **Utility:** $\mathbf{U}(\hat{\theta}, \theta_\star) : \mathcal{H}_k \times \mathcal{H}_k \to \mathbb{S}^{p \times p}$

  Random quantity due to y
- **Expected utility:**

$$\mathbf{E}_{t,\theta_\star}(\mathbf{X}) = \mathbb{E}_{p_{\theta_\star}(y_1|\mathbf{X}_1)\ldots p_{\theta_\star}(y_t|\mathbf{X}_{1:t}, y_{1:t-1})}[\mathbf{U}(\hat{\theta}, \theta_\star)|\mathcal{F}_{t-1}]$$

- **Scalarization** $s(E_{\theta_\star}), \mathrm{where} s : \mathcal{H}_k \to \mathbb{R}$

Experiment Design problem

$$\mathbf{X}^\star = \arg\max_{\mathbf{x}, |\mathbf{X}| \leq T} s(\mathbf{E}_{\theta_\star}(\mathbf{X})).$$

An Example (Regression with Poisson responses)

- **Unknown:** Unknown parameter θ_*
- **Measurement:** x is a query to $y = \theta_*^\top \Phi(x) + \epsilon_H$
- **Estimator:** heteroscedastic least-squares

$$\hat{\theta} \in \arg\min_{\theta \in \mathcal{H}_k} \mathrm{Loss}(\mathbf{X}, y)$$

- **Utility:** covariance of residuals

$$U(\hat{\theta}, \theta_\star) = (\hat{\theta} - \theta_\star)(\hat{\theta} - \theta_\star)^\top$$

- **Expected utility:**

$$\mathbf{E}_{\theta_\star} = \sum_{x \in \mathbf{X}} \frac{\Phi(x)\Phi(x)^\top}{\theta_\star^\top \Phi(x)}$$

- **Scalarization:** A-design

$$s(\mathbf{A}) = \frac{1}{\mathrm{Tr}(\mathbf{A}^{-1})}$$

1. NP-hard Combinatorial problem

2. Depends on the unknown $\theta_\star$

Experiment Design Objectives: An Assortment

1. NP-hard Combinatorial problem

We resort to relaxation techniques!

$$\mathbf{X} = \{x_{t_1}, x_{t_2}, \dots x_{t_T}\}$$

Represent as counts

$$\mathbf{X} = \{(x_1, \#_{x_1}), (x_2, \#_{x_2}), \dots, (x_{|\mathcal{X}|}, \#_{x_{|\mathcal{X}|}})\}$$

Represent as fractional counts

$$\mathbf{X} = \{(x_1, \frac{\#_{x_1}}{T}), (x_2, \frac{\#_{x_2}}{T}), \dots, (x_{|\mathcal{X}|}, \frac{\#_{x_{|\mathcal{X}|}}}{T})\}$$

Relax

$$\mathbf{X}(\eta) = \{(x_1, \eta(x_1)), (x_2, \eta(x_2)), \dots, (x_{|\mathcal{X}|}, \eta(x_{|\mathcal{X}|}))\}$$

$$\eta \in \Delta_{\mathcal{X}}$$

$\eta_\star$

$$\eta^\star = \arg\max_{\eta \in \Delta_{\mathcal{X}}} \; s\Big(\underbrace{E_{\theta_\star}\Big(T \sum_{\mathcal{X}} \eta(x) r\left(\Phi(x)\right)\Big)}_{F(\eta) \quad F_{\theta_\star}(\eta)}\Big)$$

Continuous domain, solve and round to fractional

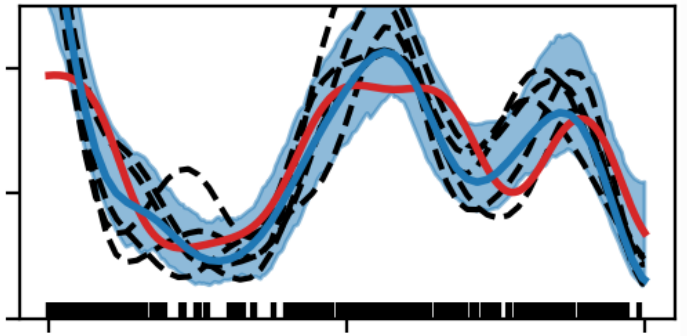
2. Depends on the unknown $\theta_\star$

- Build uncertainty set
- Be robust or average

$$\mathbf{X}^\star = \arg\max_{\mathbf{X}, |\mathbf{X}| \le T} \min_{\theta \in \Theta} s(\mathbf{E}_{\theta_\star}(\mathbf{X}))$$

$$\mathbf{X}^\star = \arg\max_{\mathbf{X}, |\mathbf{X}| \le T} \int_{\theta \in \Theta} s(\mathbf{E}_{\theta_\star}(\mathbf{X}))$$

- Θ contains all plausible models



Markov chains and Experiment Design

Markov chain process

- **Introducing:** state x and action a
- **Known** transition kernel $P(x'|x,a)$
- Observe Unknown function

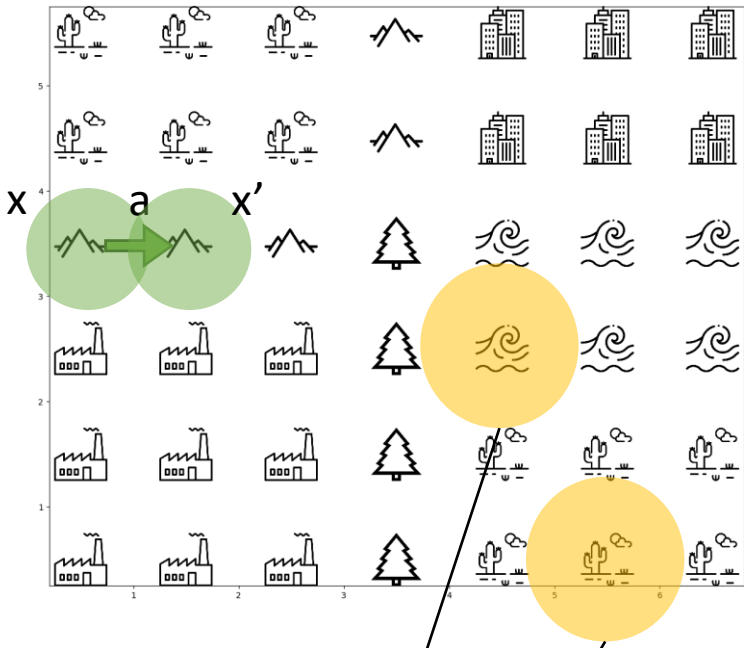
$$y = \Phi(x,a)^\top \theta_\star + \epsilon$$

Interaction

- *Can only execute policy to interact with environment that respects the transitions*
- Classically we could pick any (x,a)!

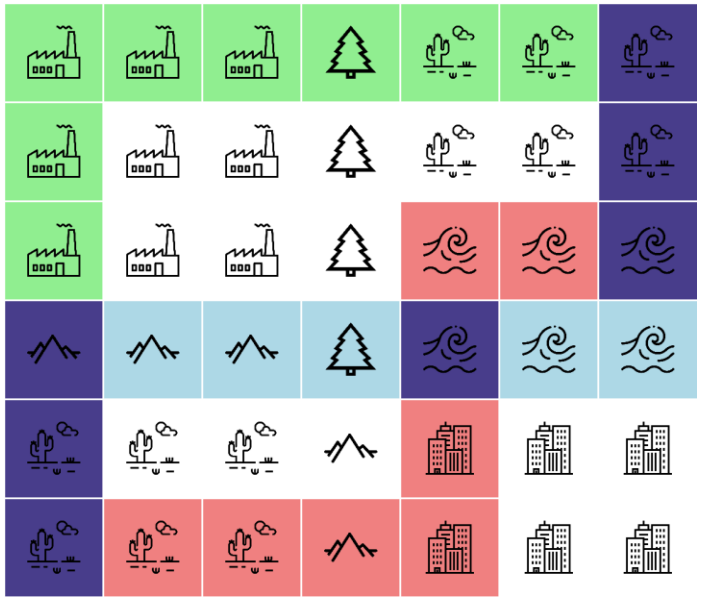
Naïve solution:

- Optimize over trajectories $\eta \in \Delta_{|\mathcal{P}|}$
  - Exponential blowup in variables ✖
  - Stochastic dynamics does not allow for execution of the trajectory ✖



$\Phi(x_1) = (7, 20, \text{“water”})$
$\Phi(x_1) = (\text{pH, salinity, area type})$

$\Phi(x_2) = (8, 14, \text{“arid”})$

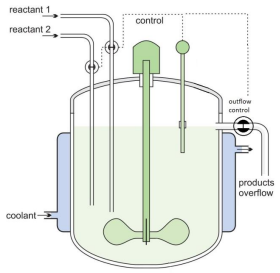


Generating trajectories

Goal: *Design a tractable sequence of policies such that the visited states and executed actions converge to the optimal visitation of the states and actions.*

Modern Experiment Design: Applications

Example 1: Reactor Optimization



Mutny, Janik, Krause (AISTATS 2023)
Folch, …, **Mutny** (NeurIPS 2024)

model

Parametric Differential Equations

randomness

measurement errors

task

$$\arg\max_x f(x)$$

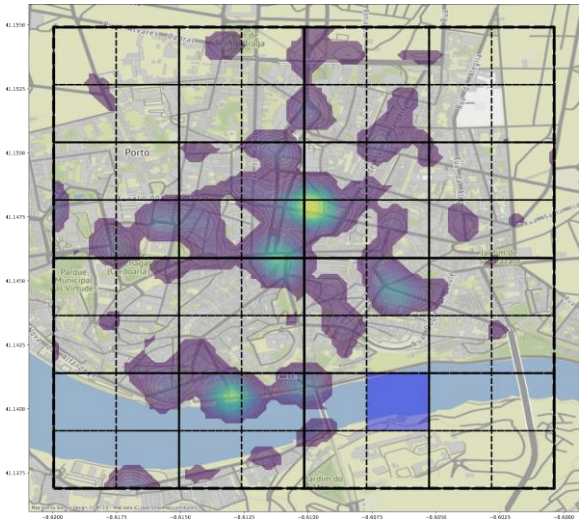
search space

Space of non-Markov policies, parametric controllers

Operating constraint, where next measurement depends on prior state

- Equilibrium constraints
- Smooth ramps
- Safety constraints

Example 2: Spatial Sensing



Mutny & Krause (ICML 2021)
Mutny & Krause (AISTATS 2022)

task

Levels sets identification

randomness

$$n(A) \sim \text{Poisson}\left(\int_A \lambda(x)dx\right)$$

model

Non-parametric
Poisson point process

action space

Hierarchical Borel sets

$$\int_A \lambda(x)dx$$

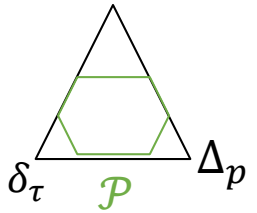
Sensing done with:

- Vehicles – need to respect roads
- Drones – spatial constraints
- Satellites – need to work with positioning

MDP structure of the experiments

How to deal with combinatorial space of all trajectories $\mathcal{P}$?

- η is over probability simplex of trajectories (exponentially many vertices -- very large)
- **Observation 1:** Our objectives depends only on pairs (x, a)



Space of trajectory probabilities

$$\eta^\star = \arg\max_{\eta \in \Delta_{\mathcal{X}}} s\left(E_{\theta_\star}\right.$$

As opposed to $\eta \in \Delta_{|\mathcal{P}|}$

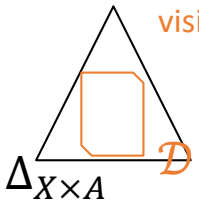
$$d^\star = \arg\max_{d \in \Delta_{\mathcal{X}\times\mathcal{A}}} s\left(E_{\theta_\star}\left(T \sum_{\mathcal{X}\times\mathcal{A}} d(x,a) r\left(\Phi(x,a)\right)\right)\right)$$

Proportion of states visiting x and executing a

- **Observation 2:** Not all states and actions proportions are possible:

Space of state-action visitation distributions

State-action visitation polytope [Putterman (1994)]



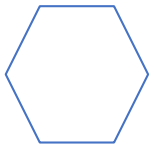
$$\mathcal{D}_h := \{d_h \mid d(x,a) \geq 0,\ \sum_{a,x} d_h(a,x) = 1,\ \sum_a d_h(x',a) = \sum_{x,a} d_{h-1}(x,a) p(x'|x,a)\}.$$

normalization — flow constraint

convex polytope — very well studied in RL — efficient solvers for linear problems

$\pi(a|x)$

Space of policies



Π

Forward propagation

$$d_h(x,a) = \left(\prod_{i=1}^{h} P_{\pi_h} d_0(x)\right) \pi_h(a|x).$$

Marginalization:

$$\pi_h(a|x) = \frac{d_h(a,x)}{\sum_a d_h(a,x)}$$

Lemma: Any d, can be realized by a Markov policy.

MDPs and Relaxations

State-action visitation distribution (Putterman (1994))

$$d^\star = \underset{d \in \mathcal{D} = \cup_{h=1}^{H} \mathcal{D}_h}{\arg\max} \; s\left(E_{\theta_\star}\left(T \sum_{x,a \in \mathcal{X} \times \mathcal{A}} d(x,a) r\left(\Phi(x,a)\right) \right) \right)$$

Optimal visitation of states

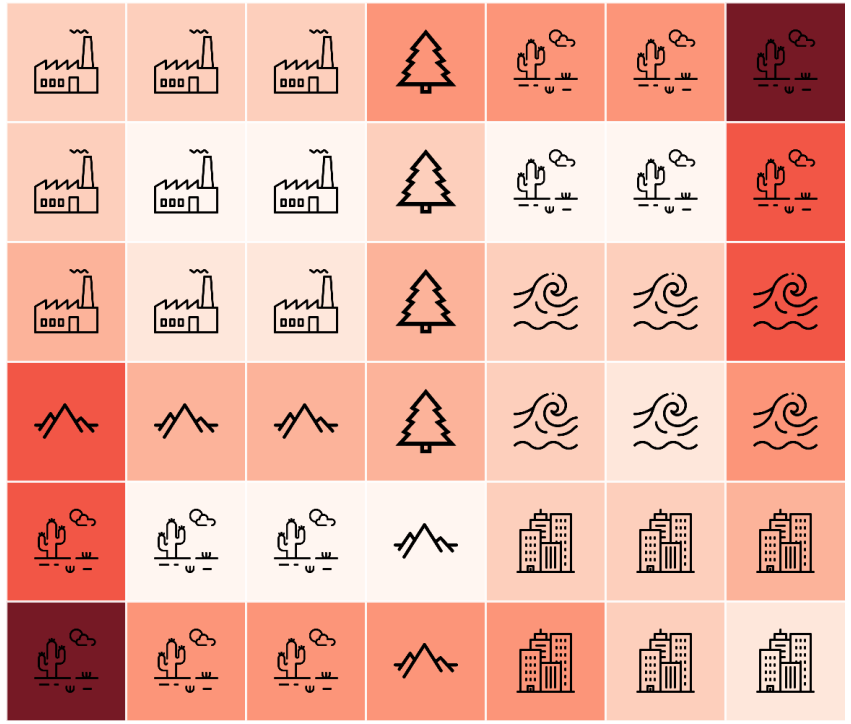
Same type of relaxation := F(d)

Good news: Its convex for many examples

Challenge: Constraint description might be complicated

Solution: Convex RL (Hazan et. al, 2019; Zahavy e.t al., 2021)

- **Franke Wolfe:** Decompose into a **sequence** of classical **RL problems**
- Solving each RL problem is *easy*: Policy iteration, policy gradient etc..



D-design

Mixture policy: $\{\alpha_i, \pi_i\}_{i=1}^n \quad \sum_{i=1}^n \alpha_i = 1, \alpha \geq 0$

$$d_{\pi_{\text{mix},n}} = \sum_{i=1}^n \alpha_i d_{\pi_i}$$

Linearization at current mixture

Reward from RL

Linear Maximization Oracle

$$d_{\pi_{n+1}} = \underset{d \in \mathcal{D}}{\arg\max} \sum_{x,a} \nabla F(d_{\pi_{\text{mix}}})(x,a) d(x,a) = \underset{d_\pi \in \mathcal{D}}{\arg\max} \sum_{x,a} r(x,a) d(x,a)$$

Adaptivity into the picture

So far: No interaction with the system, **Now:** Interaction and unkowns!

In each step we optimize only correction:

- In each round we calculate and execute **a new Markov policy π_t**

$$G_t(d) = \min_{\theta \in \Theta_t} F_\theta \left(\frac{1}{t+1} d + \frac{t}{t+1} \hat{d}_t \right) \qquad d_{\pi_t} = \arg\max_{d \in \mathcal{D}} G_t(d)$$

Robust design

Empirical visitation of past interactions (counts)

Historical sidenote: This idea appears in Pukelsheim's textbook on Bayesian designs

Note:

- Each policy is optimized to be **correction** to the empirical visitation
- The set of policies $\{\pi_1, \pi_2, \ldots, \pi_T\}$ is **history dependent**
- The history dependence is via **the objective**, namely empirical state visitation
- The **uncertainty** is time-dependent
- The overall policy over episodes/rounds is **non-Markovian!** → *but not optimal non-Markovian*

Theory and Convergence

- **Theorem:** "Best possible explanation of a teacher to a student"
- **Theory (Assumption):**
  - Concave $F_\theta(\eta)$ in η
  - Smooth $F_\theta(\eta)$ in θ
  - "Compatibility of the objective with uncertainty reduction"

best empirical

$$F_{\theta_\star}(d_\star) - F_{\theta_\star}(d_\tau)$$

Comparing to the unknown $\theta_\star$

Then as the number of experiments τ,

$$F_{\theta_\star}(d_\star) - F_{\theta_\star}(d_\tau) \leq \frac{1}{\tau}\left(F_{\theta_\star}(\eta^\star) - F_{\theta_\star}(\eta_0) + \sum_{t=1}^{\tau}\epsilon_t^S + \sum_{t=1}^{\tau}\epsilon_t^A + \sum_{t=1}^{\tau}\frac{L}{1+\tau}\right)$$

Initial error (regularization)

Rounding Error $O(\sqrt{\tau})$

Approximation error

$$\epsilon_A = |l_{v_t}^\top(\theta_\star - \bar{\theta}_t)|$$

Confidence set width at time t

Conclusion: As $\Theta_t \to 0$, this error decreases at $O(1/\sqrt{\tau})$ and **optimal constant** in many cases.

Examples:

- **Rounding error with s**ampling from optimized measure → martingale $O(\sqrt{\tau})$
- **Approximation error with** D-design and Gaussian likelihood is zero

Experiment Design with MDPs: Difficulty & Relaxation

Manish Prajapat

The whole process summarized

- In each round we calculate and execute **a Markov policy π_t**
- The set of policies $\{\pi_1, \pi_2, \dots, \pi_T\}$ is **history dependent**
- The history dependence is via **the objective**, namely empirical state visitation

$$G_t(d) = \min_{\theta \in \Theta_t} F_\theta \left(\frac{1}{t+1} d + \frac{t}{t+1} \hat{d}_t \right)$$

Empirical visitation of past interactions (induces history dependence)

- The overall policy over episodes/rounds is **non-Markovian.**

Competitive with respect to the unrelaxed objective?

$$\mathbb{E}_{\eta_i \sim \pi_i^\star, i \in [\tau]} \left[F_{\theta_\star} \left(\frac{1}{\tau} \sum_{i=1}^{\tau} \hat{\eta} i \right) \right] - F_{\theta_\star}(\hat{\eta}_\tau) \overset{\text{Jensen}}{\leq} F_{\theta_\star}(\eta^\star) - F_{\theta_\star}(\hat{\eta}_\tau) \leq \mathcal{O}(\tau^{-1/2})$$

Optimal policy of the unrelaxed objective

Empirical state-visitation (i.e. integral lattice point)

Difficulty of the unrelaxed problem (in terms of combinatorial optimization)

- Results Known $\theta_\star$, Additional structure: **Submodularity (D-design)**
- Optimal policy is non-Markovian and deterministic (interesting)
- Optimal problem is **NP-hard to** mult. **approximate** with constant factor

Def: Experiment Design problem (on MDP)

$$\arg\max_{\pi_1, \dots \pi_T} \mathbb{E}_{\pi_i} [s(\mathbf{E}_{\theta_\star})(\mathbf{X})]$$

$\boldsymbol{X}$ contains concatenated trajectories

Continuous Domains and Model Predictive Control (MPC)

So far, the decision space was discrete

- Continuous decision space → no change in terms of methodology
- Problem is still convex in the probability distribution space
  - infinite dimensional convex optimization problem

$$d^\star = \underset{d \in \mathcal{D} \subset \mathcal{P}(\mathcal{X}, \mathcal{A})}{\arg\max} \; s\left(E_{\theta_\star}\left(T \int_{\mathcal{X}} d(x,a) r\left(\Phi(x,a)\right) dx da\right)\right)$$

  - Need to resort to some form of representation.
    - Mixture of Gaussians, Normalizing flows, Neural Networks
    - **Model Predictive Control: Parametrize in terms of actions**
- **Suppose Linear Dynamics $x_{h+1} = Ax_h + Ba_h$**
- The policy is parametrized by the set of actions a_h

Past visitation of states

$$\underset{a_0,\ldots,a_H}{\arg\max} \sum_{h=0}^{H} \nabla F_{t,0}\left(d_{\pi_{\text{mix},t}}\right)(x_h, a_h)$$

$$\text{s.t. } ||a_h|| \leq a_{\max} x_h \in [-0.5, 0.5]$$

$$x_{h+1} = Ax_h + Ba_h$$

Non-linear MPC

- Non-convex objective
- Adaptivity means: **receding horizon replanning**

Folch et. al. (2024) **Transition Constrained Bayesian Optimization via Markov Decision Processes**

Experiment Design with MDPs: Examples I

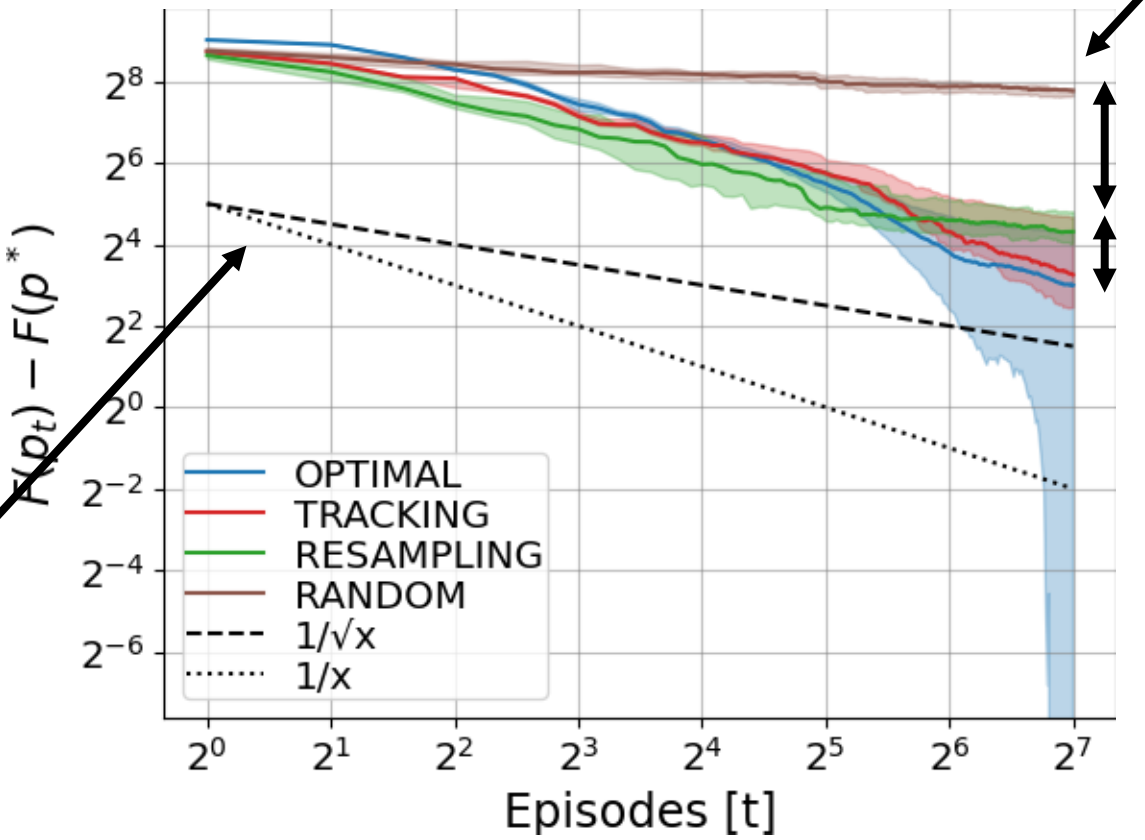
Spatial Sensing

- **Goal:** Estimate the spatial distribution of species
- Drone flying over and local constraints
- Features: elevation and slope

"The y-axis is value of the constant:"

$$\frac{F(p_t)}{t}$$

Random flatlines;
It can't compete with optimal policy



Gap due to ED

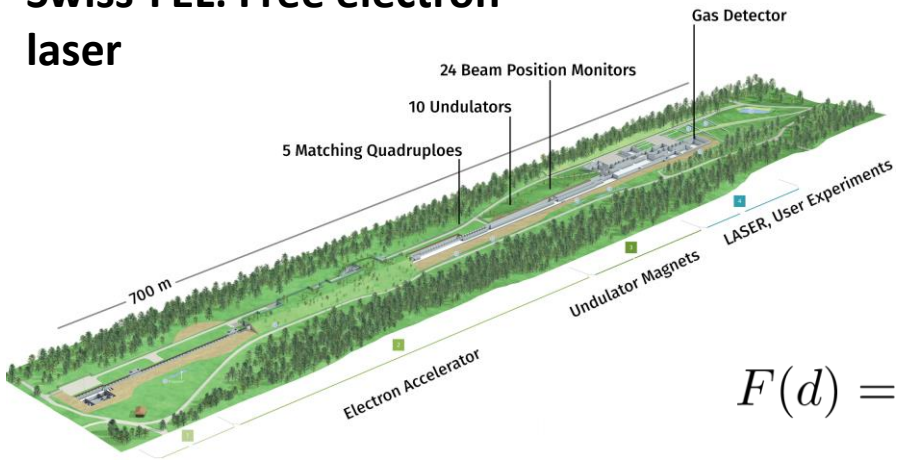
Adaptivity Gap

Theory well-represented

Experiment Design with MDPs: Examples III


Johannes Kirschner
Nicole Hiller
PAUL SCHERRER INSTITUT
PSI

Swiss-FEL: Free electron laser

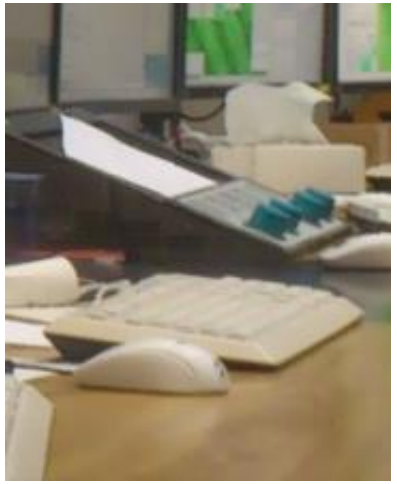


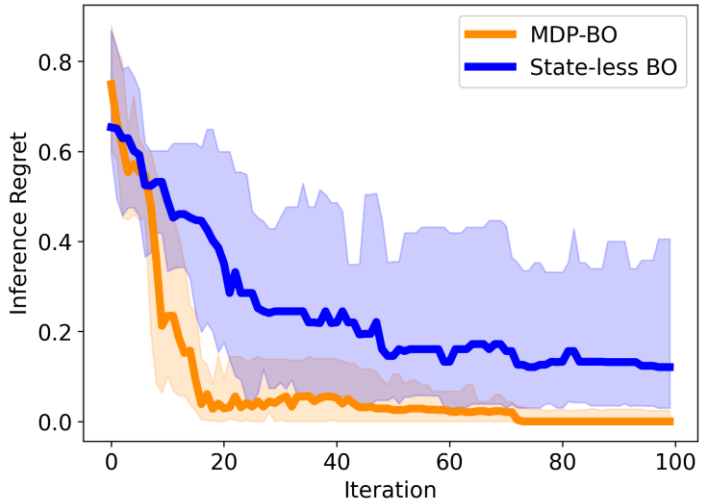
- **Goal:** Identify best-parameter
- By changing parameters, system is not equilibrium.
- The larger the parameter change the more noise

$$F(d) = \max_{z,z' \in \mathcal{Z}} ||\Phi(z) - \Phi(z')||^2_{\mathbf{V}(d)^{-1}}$$

$$\mathbf{V}(\eta) = \left(\sum_{x,a \in \mathcal{X} \times \mathcal{A}} \frac{d(x,a)\Phi(x,a)\Phi(x,a)^\top}{\sigma^2(x,a)} + \lambda \mathbf{I} \right)$$

The noise is **proportional to the distance between x->a**





Experiment Design with MDPs: Examples IV

Optimizing preference models with MDPs

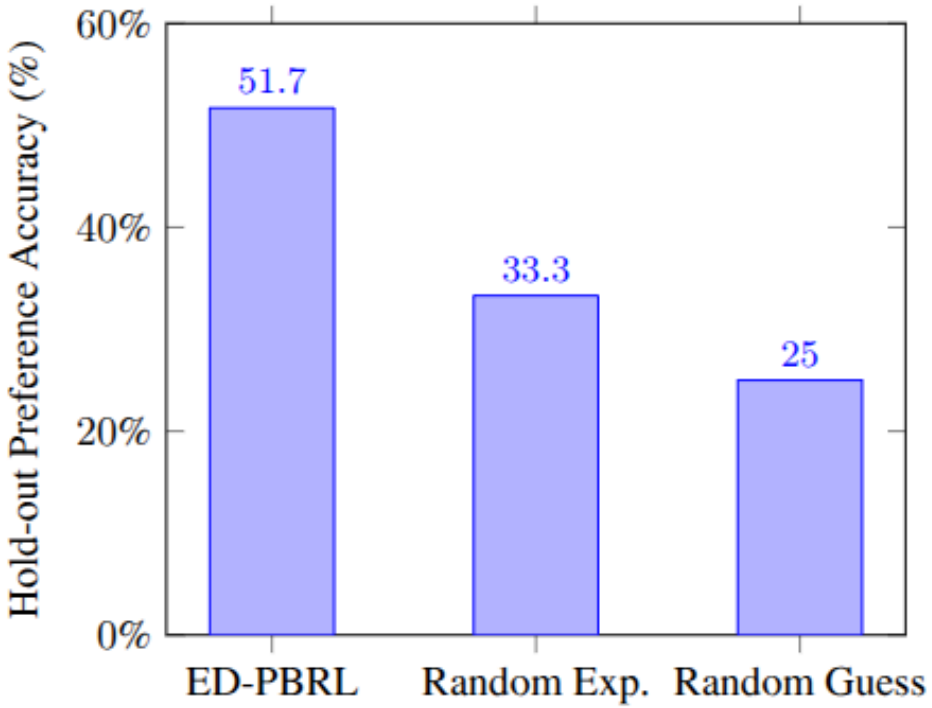
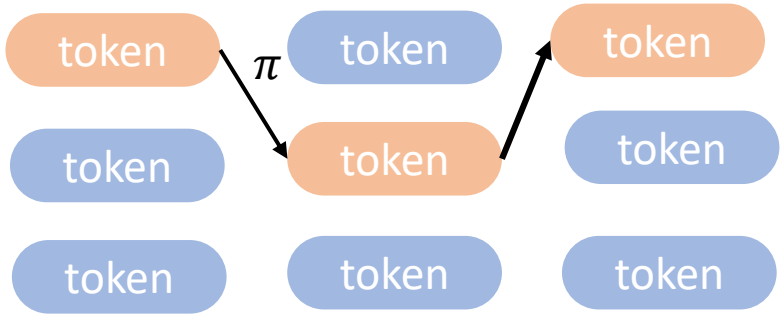


Figure: Human Experiment
Our algorithm – ED-PBRL
- Budget: 100 queries
- Accuracy of model trained on 100 queries
- 20 participants

Q: Can we learn a model?
F(user, prompt) → **#hastag1 #hastag2**

Challenges:
- Per-user interaction
  → **data efficiency**
- Feedback is only relative
  → **preference feedback**
- Structured feedback
  → **LLMs sampling is a Markov chain!**

What prompts we show to the user?



$$\pi_1, \pi_2 = \operatorname{argmin} F(d_{\pi_1}, d_{\pi_2})$$

Thank you for your attention

Andreas Krause

Tadeusz Janik

Manish Prajapat

Antanas Murelis

Johannes Kirschner

ETH

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

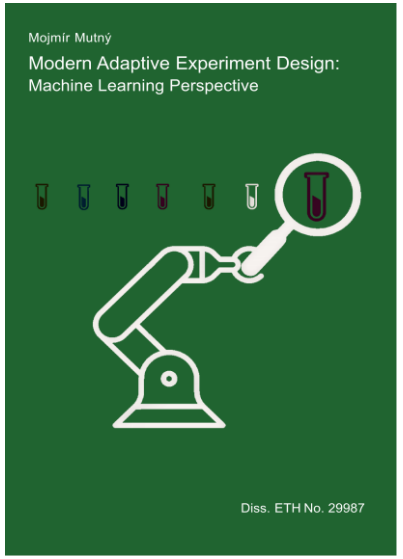
Jose Folch

Ruth Misener

Imperial College London

Nicole Hiller

PAUL SCHERRER INSTITUT PSI



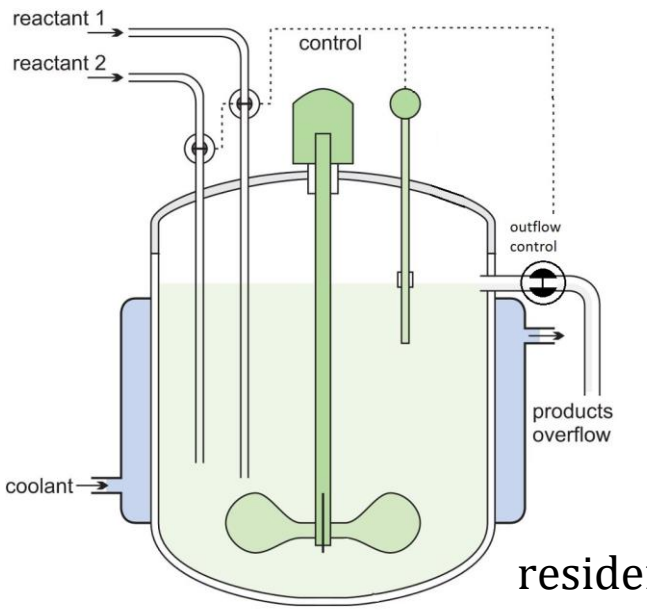
Basic framework in the thesis, more recent works on my website.

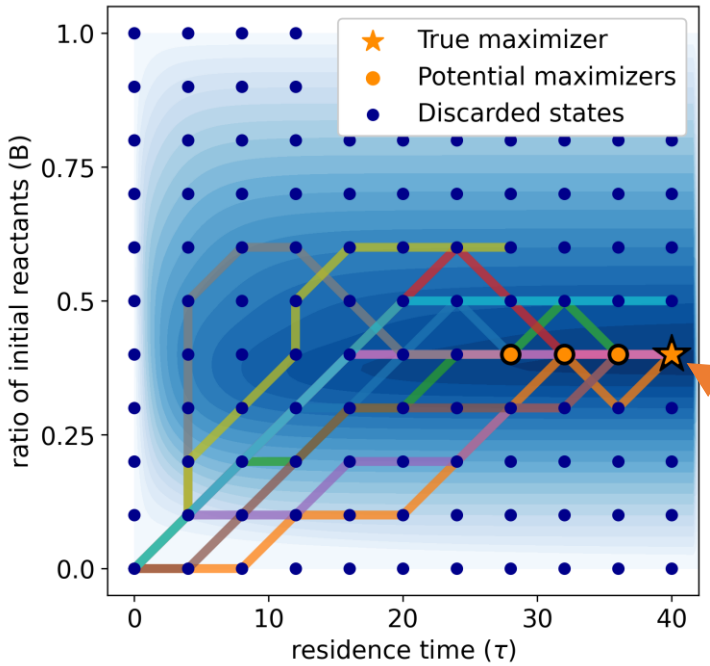
Experiment Design with MDPs: Examples II

Optimizing batch chemical reactors

- Complicated reaction mechanism (non-linear)
  - Not first-order reaction
- **Goal:** Identify best-reaction parameters (maximization identification)
- Opening a valve & changing residence time is continuous
  - → Transient information
  - → Noisy information
  - → Optimal path in parameter space

$$\text{ratio} = \frac{\text{flow reactant 1}}{\text{flow reactant 2}}$$





Optimum