Towards safe control parameter tuning in distributed multi-agent systems

64th IEEE Conference on Decision and Control (CDC 2025)

Abdullah Tokmak[1], T. B. Schön[2], D. Baumann[1]

[1]Aalto University, Espoo, Finland
[2]Uppsala University, Uppsala, Sweden

December 11, 2025

Aalto University
School of Electrical
Engineering

UPPSALA
UNIVERSITET

Motivational example

keysight.com

- Many real-world problems are of **distributed** multi-agent nature
- No central coordinating node
- Optimize agents' parametrized control policies to reach **cooperative goal**

Goal

Optimize control parameters in distributed multi-agent systems (MAS) while ensuring **safety** and **sample efficiency**.

Introduction

Goal

Optimize control parameters in distributed MAS while ensuring **safety** and **sample efficiency**.

- Control policy parameters $a^{(j)} \in \mathcal{A}$ of each agent $j \in \{1, \ldots, \mathrm{N}\}$
- Global black-box reward function f: $\mathcal{A}^N \to \mathbb{R}$, global control parameter $\mathbf{a} \in \mathcal{A}^N$
- Regularity: Function f member of RKHS H_k of kernel k

Sample efficiency and safety guarantees

SafeOpt:[1] **Safe Bayesian optimization** (BO) algorithm with **Gaussian process** (GP) regression that works well in single-agent systems.

[1] Y. Sui et al., "Safe exploration for optimization with Gaussian processes," ICML 2015.

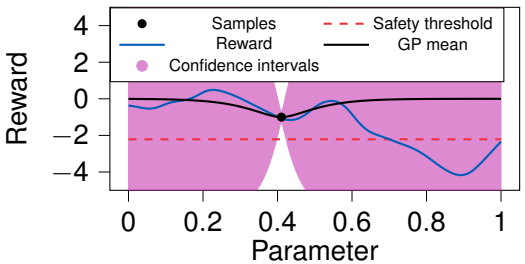
SafeOpt: Safe BO with GP regression

- Episodic setting: Global parameter $\mathbf{a}_t \in \mathcal{A}^N \Rightarrow y_t = f(\mathbf{a}_t) + \text{noise}$ at each iteration
- GPs with kernel k: Uncertainty quantification with confidence intervals

Safe policy optimization problem

$$\max_{\mathbf{a} \in \mathcal{A}^N} f(\mathbf{a}) \quad \text{subject to } f(\mathbf{a}_t) \geq h, \quad \forall t \geq 1$$

1: **for** $t = 1, 2, \ldots$ **do**
2: Build GP with samples
3: $\mathbf{a}_{t+1} \leftarrow$ SafeOpt
4: $y_{t+1} \leftarrow f(\mathbf{a}_{t+1}) + \text{noise}$
5: **return** Best parameter $\mathbf{a}^\star$

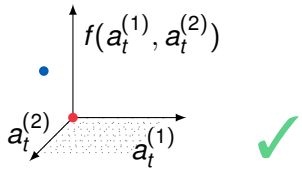


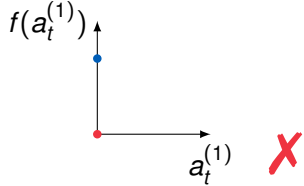
SafeOpt for distributed multi-agent systems

Safe policy optimization problem

$$\max_{\mathbf{a}\in\mathcal{A}^N} f(\mathbf{a}) \quad \text{subject to } f(\mathbf{a}_t) \geq h, \quad \forall t \geq 1$$

- **Samples** $(\mathbf{a}_t, y_t)$ through experiments are central to inference problem
- Agent j observes **local parameter** $a_t^{(j)}$, which is **projection** of $\mathbf{a}_t$
- Local information $(a_t^{(j)}, y_t)$: Observed local function $f^{(j)} : \mathcal{A} \to \mathbb{R}$ **not well-defined**
- Example: $f(0,0) = 0$ and $f(0,1) = 1 \Rightarrow f^{(1)}(0) = 0$ and $f^{(1)}(0) = 1$



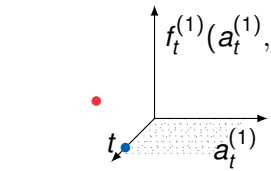


- **Full communication** is **infeasible**: Scalability, communication bandwidth, privacy

Exploiting the iteration variable t

- Construct **well-behaved mapping** using $(a_t^{(j)}, y_t)$
- Explicitly model the **iteration variable** t
- Time-varying local function $f_t^{(j)} : (\mathcal{A}, \mathbb{N}) \to \mathbb{R}$
  $f(0,0) = 0, f(0,1) = 1 \Rightarrow f_t^{(1)}(0,1) = 0, f_t^{(1)}(0,2) = 1$
- Construct GP to model local time-varying $f_t^{(j)}$ using **spatio-temporal kernel**[2]

$$k_t^{(j)}((\mathbf{a}, t), (\mathbf{a}', t')) = \underbrace{k_{\mathrm{S}}^{(j)}(a, a')}_{\text{Observables}} \cdot \underbrace{k_{\mathrm{T}}^{(j)}(t, t')}_{\text{Unobservables}}$$

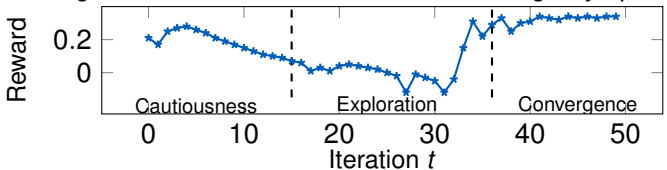


- We interpret **time** t as a **latent variable** with a concrete physical interpretation

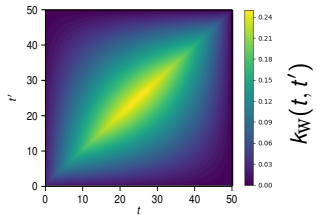
[2] Bogunovic et al., "Time-varying Gaussian process bandit optimization," AISTATS 2016.

Spatio-temporal kernel

- **Spatial part:** Exploit smoothness of global reward function $f \in H_k$
- **Temporal part:** How does **rest of MAS behave** if Agent j remains constant?
  - Example: 3-agent MAS without communication from Agent j's perspective



- Beginning/end: **smoother** sample paths (RBF kernel)
- Midpoint: **rougher** sample paths (Matérn-12 kernel)



$$k_{\mathrm{T}}(t,t') = k_{\mathrm{RBF}}(t,t') + k_{\mathrm{W}}(t,t')k_{\mathrm{Ma12}}(t,t')$$

$$k_{\mathrm{W}}(t,t') = \frac{1}{t_{\mathrm{end}}} \min(t,t') \cdot \min(t_{\mathrm{end}} - t, t_{\mathrm{end}} - t)$$

Re-interpreting the optimization problem

Safe (global) policy optimization problem

$$\max_{\mathbf{a}\in\mathcal{A}^N} f(\mathbf{a}) \quad \text{subject to } f(\mathbf{a}_t) \geq h, \quad \forall t \geq 1$$

$$\Downarrow$$

Time-varying local proxy of safe policy optimization problem

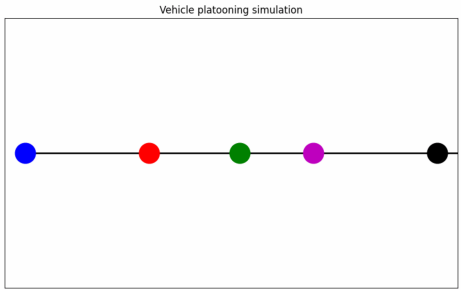
$$\max_{a\in\mathcal{A}} f_t^{(j)}(a,t) \quad \text{subject to } f_t^{(j)}(a_t^{(j)},t) \geq h, \quad \forall t \geq 1, \forall j \in [1,\ldots,N]$$

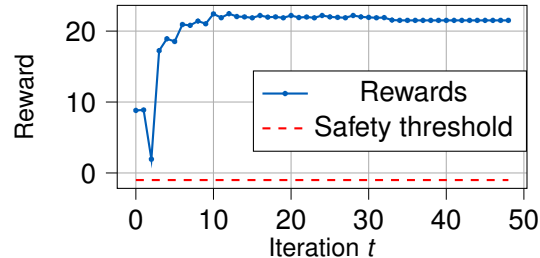
- Safe BO is a **sequential decision-making** problem using GP posterior at iteration t
- **Single agent:** Regression in spatial domain $\mathcal{A}^N \Rightarrow \mathbf{a}_{t+1}$
- **MAS**: Regression in spatial domain $\mathcal{A}$, one-step extrapolation[3] in $t \Rightarrow a_{t+1}^{(j)}$

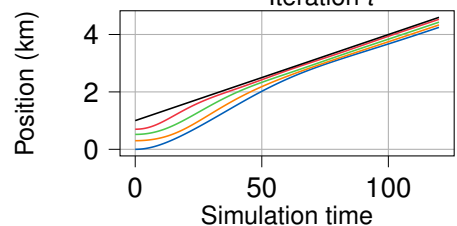
[3] S. Roberts et al., "Gaussian processes for time-series modelling," Philos. Trans. R. Soc. A, 2013.

Vehicle platooning simulation

- Tune synchronization PI controller using safe BO in 5-agent heterogenous MAS
- Bi-directional nearest neighbor communication







~~Towards~~ Safe parameter tuning in distributed MAS

Safe optimization problem

$$\max_{\mathbf{a}\in\mathcal{A}^N} f(\mathbf{a}) \quad \text{subject to } f(\mathbf{a}_t) \geq h$$

Proxy of safe optimization problem

$$\max_{a\in\mathcal{A}} f_t^{(j)}(a,t) \quad \text{subject to } f_t^{(j)}(a_t^{(j)},t) \geq h$$

- Error using proxy instead of original optimization problem **not quantified** here
- **Heuristically**, we derived $(a_t^{(j)}, t) = \Pi_t^{(j)}(\mathbf{a}_t)$ and spatio-temporal kernel $k_t^{(j)}(a_t^{(j)}, t)$
- Find suitable $\Pi_t^{(j)}(\mathbf{a}_t)$ and $k_t^{(j)}$ to minimize kernel discrepancy $\bar{\epsilon}$ such that

$$\sup_{\mathbf{a},\mathbf{a}'\in\mathcal{A}^N} |k(\mathbf{a},\mathbf{a}') - k_t^{(j)}(\Pi_t^{(j)}(\mathbf{a}), \Pi_t^{(j)}(\mathbf{a}'))| \leq \bar{\epsilon}.$$

- With this, we can build **confidence intervals** between global reward function f and local time-varying GP mean[4] $\Rightarrow$SafeOpt-like **safety guarantees**

[4] C. Fiedler et al., "Practical and rigorous uncertainty bounds for Gaussian process regression, AAAI 2021.

Conclusions

Goal

Optimize control parameters in distributed MAS, ensuring **safety** and **sample efficiency**.

Contributions

- Implictly model unobserved behavior by introducing **time as latent variable**
- **Time-varying local interpretation** of global static reward function with custom **spatio-temporal kernel**
- **BO algorithm** for parameter tuning in distributed MAS

Future work

- **Confidence intervals** for safety guarantees