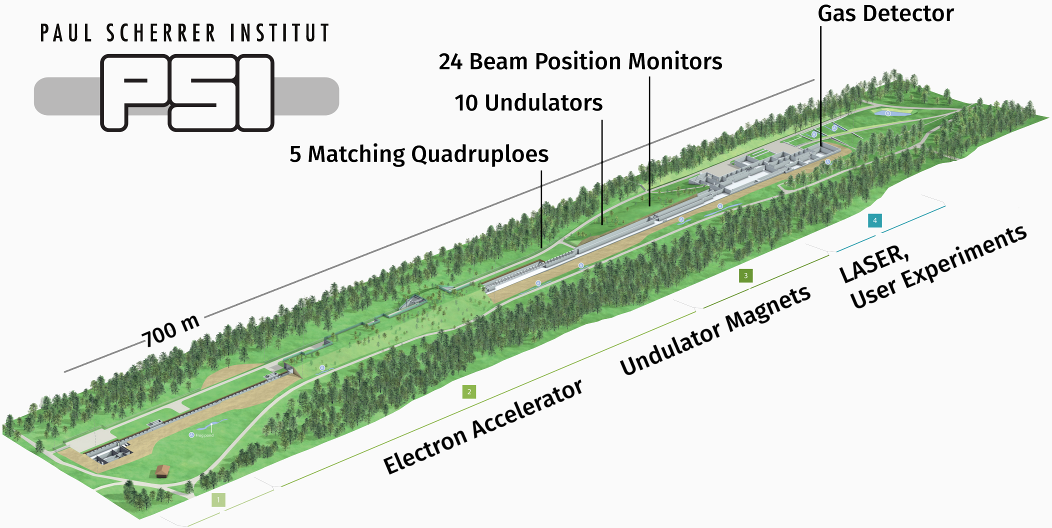
Adaptive and Safe Bayesian Optimization in High Dimensions via One-Dimensional Subspaces

Johannes Kirschner, Mojmir Mutny, Nicole Hiller, Rasmus Ischebeck, Andreas Krause

ICML, June 13th, 2019

ETH zürich

Motivation: Tuning the Swiss Free Electron Laser



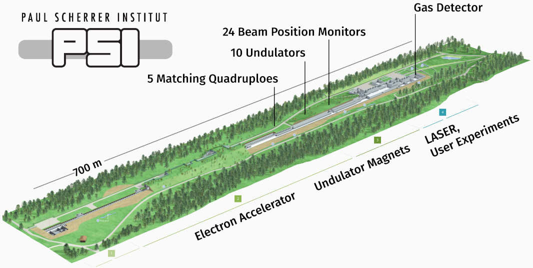
The Optimization Problem

Objective

Safety Constraints

$$\max_{x} f(x) \quad \text{s.t.} \quad g(x) \leq 0$$

PAUL SCHERRER INSTITUT

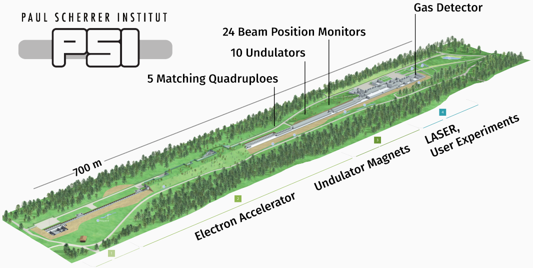


The Optimization Problem

Photon Energy

Electron Losses

$$\max_x f(x) \quad \text{s.t.} \quad g(x) \leq 0$$



Free Electron Laser:

- $\sim$ 50 continuous parameters
- Non-linear, local optima
- **Noisy zero-order** access to f and g
- **Safe optimization:** $g(x_t) \leq 0 \;\; \forall \; x_t$

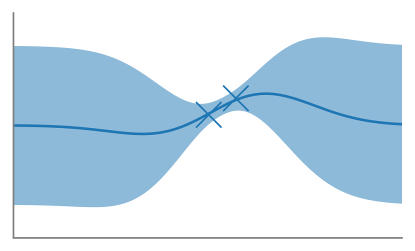
Our Work: Line Bayesian Optimization

$\mathcal{D}_0 = \{\}$

For $t = 1, 2, 3, \ldots$

1 : **Estimate:** $\hat{f}_t, \hat{g}_t | \mathcal{D}_t$

- ▷ Gaussian process / kernel regression



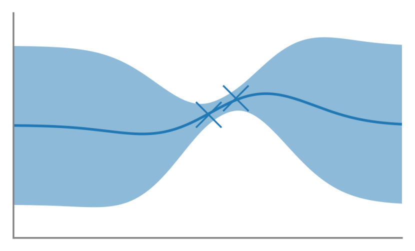
Our Work: Line Bayesian Optimization

$\mathcal{D}_0 = \{\}$

For $t = 1, 2, 3, \ldots$

1 : **Estimate:** $\hat{f}_t, \hat{g}_t | \mathcal{D}_t$

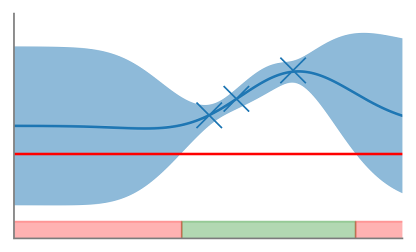
- Gaussian process / kernel regression



2 : **Acquisition**: $x_t = \arg\max_x \alpha(x | \hat{f}_t, \hat{g}_t)$

- balances exploration-exploitation and safety
- augment data $\mathcal{D}_{t+1}$

Our Work: Line Bayesian Optimization



$\mathcal{D}_0 = \{\}$

For $t = 1, 2, 3, \ldots$

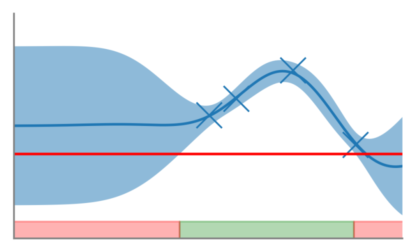
1 : **Estimate:** $\hat{f}_t, \hat{g}_t | \mathcal{D}_t$

- ▷ Gaussian process / kernel regression

2 : **Acquisition**: $x_t = \arg\max_x \alpha(x | \hat{f}_t, \hat{g}_t)$

- ▷ balances exploration-exploitation and safety
- ▷ augment data $\mathcal{D}_{t+1}$

Our Work: Line Bayesian Optimization



$\mathcal{D}_0 = \{\}$

For $t = 1, 2, 3, \ldots$

1 : **Estimate:** $\hat{f}_t, \hat{g}_t | \mathcal{D}_t$

- ▷ Gaussian process / kernel regression

2 : **Acquisition**: $x_t = \arg\max_x \alpha(x | \hat{f}_t, \hat{g}_t)$

- ▷ balances exploration-exploitation and safety
- ▷ augment data $\mathcal{D}_{t+1}$

Our Work: Line Bayesian Optimization

$\mathcal{D}_0 = \{\}$

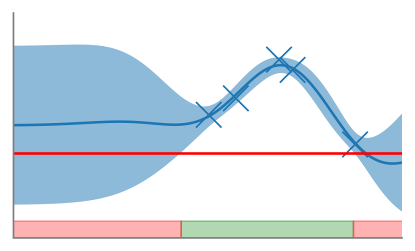
For $t = 1, 2, 3, \ldots$

1 : **Estimate:** $\hat{f}_t, \hat{g}_t | \mathcal{D}_t$

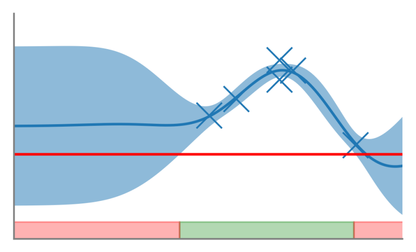
- ▷ Gaussian process / kernel regression

2 : **Acquisition**: $x_t = \arg\max_x \alpha(x | \hat{f}_t, \hat{g}_t)$

- ▷ balances exploration-exploitation and safety
- ▷ augment data $\mathcal{D}_{t+1}$



Our Work: Line Bayesian Optimization



$\mathcal{D}_0 = \{\}$

For $t = 1, 2, 3, \ldots$

1 : **Estimate:** $\hat{f}_t, \hat{g}_t | \mathcal{D}_t$

- Gaussian process / kernel regression

2 : **Acquisition**: $x_t = \arg\max_x \alpha(x | \hat{f}_t, \hat{g}_t)$

- balances exploration-exploitation and safety
- augment data $\mathcal{D}_{t+1}$

Our Work: Line Bayesian Optimization

$\mathcal{D}_0 = \{\}$

For $t = 1, 2, 3, \ldots$

1: **Estimate:** $\hat{f}_t, \hat{g}_t | \mathcal{D}_t$

- Gaussian process / kernel regression

2: **Acquisition**: $x_t = \arg\max_x \alpha(x | \hat{f}_t, \hat{g}_t)$

- balances exploration-exploitation and safety
- augment data $\mathcal{D}_{t+1}$

Difficult optimization problem

Our Work: Line Bayesian Optimization

$\mathcal{D}_0 = \{\}$

For $t = 1, 2, 3, \ldots$

Restrict to 1d subspace $\mathcal{L}$

1 : **Estimate:** $\hat{f}_t, \hat{g}_t | \mathcal{D}_t$

- Gaussian process / kernel regression

2 : **Acquisition on $\mathcal{L}$:** $x_t = \arg\max_{x \in \mathcal{L}} \alpha(x | \hat{f}_t, \hat{g}_t)$

- balances exploration-exploitation and safety
- augment data $\mathcal{D}_{t+1}$

Our Work: Line Bayesian Optimization

$\mathcal{D}_0 = \{\}$

For $t = 1, 2, 3, \ldots$

Restrict to 1d subspace $\mathcal{L}$

1: **Estimate:** $\hat{f}_t, \hat{g}_t | \mathcal{D}_t$

- Gaussian process / kernel regression

2: **Acquisition on $\mathcal{L}$:** $x_t = \arg\max_{x \in \mathcal{L}} \alpha(x | \hat{f}_t, \hat{g}_t)$

- balances exploration-exploitation and safety
- augment data $\mathcal{D}_{t+1}$

3: **If** error$(\mathcal{L}) < \epsilon$:

New 1d subspace $\mathcal{L}$ at best point $\hat{x}_t$ in *random* direction

Our Work: Line Bayesian Optimization

$\mathcal{D}_0 = \{\}$

For $t = 1, 2, 3, \ldots$

1: **Estimate:** $\hat{f}_t, \hat{g}_t | \mathcal{D}_t$

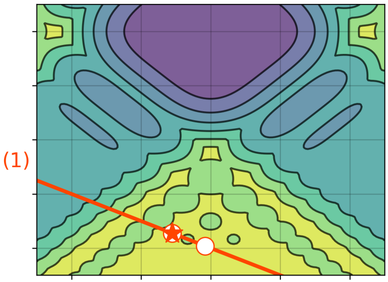
- ▷ Gaussian process / kernel regression

2: **Acquisition on** $\mathcal{L}$: $x_t = \arg\max_{x \in \mathcal{L}} \alpha(x | \hat{f}_t, \hat{g}_t)$

- ▷ balances exploration-exploitation and safety
- ▷ augment data $\mathcal{D}_{t+1}$

3: **If** error$(\mathcal{L}) < \epsilon$:

New 1d subspace $\mathcal{L}$ at best point $\hat{x}_t$ in *random* direction



Our Work: Line Bayesian Optimization

$\mathcal{D}_0 = \{\}$

For $t = 1, 2, 3, \ldots$

1 : **Estimate:** $\hat{f}_t, \hat{g}_t | \mathcal{D}_t$

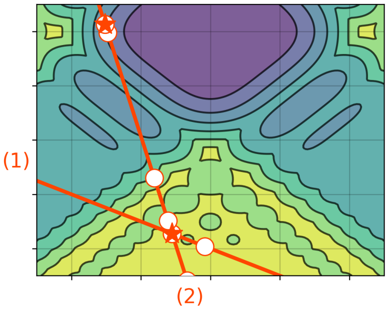
- ▷ Gaussian process / kernel regression

2 : **Acquisition on** $\mathcal{L}$: $x_t = \arg\max_{x \in \mathcal{L}} \alpha(x | \hat{f}_t, \hat{g}_t)$

- ▷ balances exploration-exploitation and safety
- ▷ augment data $\mathcal{D}_{t+1}$

3 : **If** error($\mathcal{L}$) $< \epsilon$:

New 1d subspace $\mathcal{L}$ at best point $\hat{x}_t$ in *random* direction



Our Work: Line Bayesian Optimization

$\mathcal{D}_0 = \{\}$

For $t = 1, 2, 3, \ldots$

1 : **Estimate:** $\hat{f}_t, \hat{g}_t | \mathcal{D}_t$

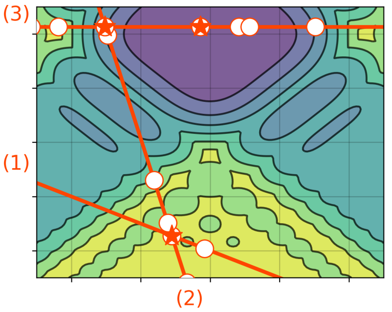
- ▷ Gaussian process / kernel regression

2 : **Acquisition on** $\mathcal{L}$: $x_t = \arg\max_{x \in \mathcal{L}} \alpha(x | \hat{f}_t, \hat{g}_t)$

- ▷ balances exploration-exploitation and safety
- ▷ augment data $\mathcal{D}_{t+1}$

3 : **If** error($\mathcal{L}$) $< \epsilon$:

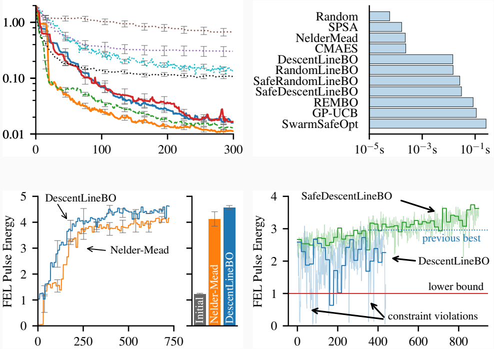
New 1d subspace $\mathcal{L}$ at best point $\hat{x}_t$ in *random* direction



Our Work: Line Bayesian Optimization

Contributions

1) *Global* convergence
2) Adapts to invariant subspaces
3) *Fast local convergence* if the function is smooth & strongly convex
4) *Safety* via Safe-BO
5) Experiments on SwissFEL
6) *User feedback* via slice plots



For details come to our poster! (Pacific Ballroom #147)