Sequential- and Parallel- Constrained Max-value Entropy Search via Information Lower Bound (ICML2022)

Shion Takeno[1, 2], Tomoyuki Tamura[1],
Kazuki Shitara[3,4], and Masayuki Karasuyama[1]

[1]Nagoya Institute of Technology
[2]RIKEN AIP
[3]Osaka University
[4]Japan Fine Ceramics Center

Constrained optimization for black-box functions

- Black-box optimization with unknown constraints:

$$\max_{\boldsymbol{x}\in\mathcal{X}} f(\boldsymbol{x}),\ \text{s.t. } g_c(\boldsymbol{x}) \geq z_c \text{ for } c = 1, \ldots, C,$$

Constrained optimization for black-box functions

- Black-box optimization with unknown constraints:

$$\max_{\boldsymbol{x}\in\mathcal{X}} f(\boldsymbol{x}),\ \text{s.t. } g_c(\boldsymbol{x}) \geq z_c \text{ for } c = 1, \ldots, C,$$

Drug discovery:

f: Medicinal effect

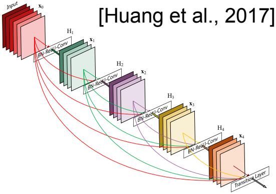
g: Side effects

Materials design:

f: Ion-conductivity

g: Safety

AutoML:



[Huang et al., 2017]

f: Accuracy

g: Fairness

Constrained optimization for black-box functions

- Black-box optimization with unknown constraints:

$$\max_{\boldsymbol{x} \in \mathcal{X}} f(\boldsymbol{x}), \text{ s.t. } g_c(\boldsymbol{x}) \geq z_c \text{ for } c = 1, \ldots, C,$$

Drug discovery:

f: Medicinal effect

g: Side effects

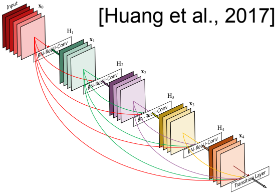
Materials design:

f: Ion-conductivity

g: Safety

AutoML:

[Huang et al., 2017]



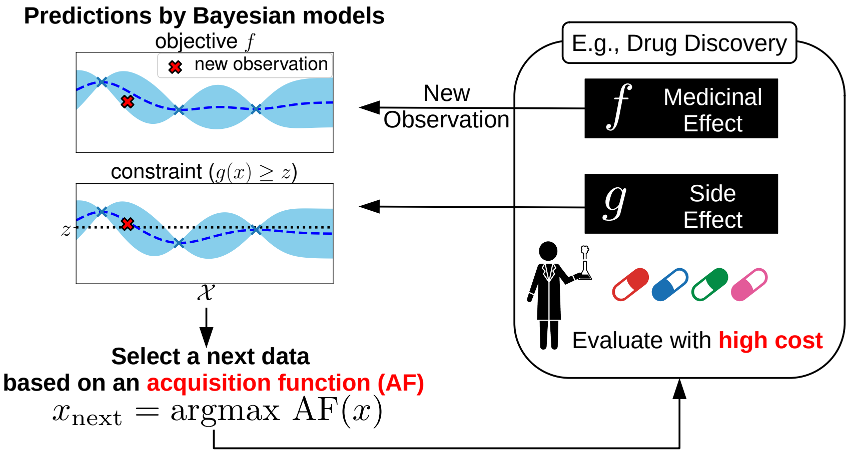
f: Accuracy

g: Fairness

Evaluation cost for f and $g_1, \ldots, g_C$ is often expensive. ☹

Constrained Bayesian optimization

- Constrained Bayesian optimization (CBO) aims for sample-efficient optimization.



CBO based on Max-value entropy search (MES)

- MES for unconstrained problem [Wang and Jegelka, 2017]

$$\mathrm{AF}(\boldsymbol{x}) = \mathrm{MI}(f_*; f(\boldsymbol{x})), \text{ where } f_* = \max_{\boldsymbol{x}} f(\boldsymbol{x}).$$

CBO based on Max-value entropy search (MES)

- MES for unconstrained problem [Wang and Jegelka, 2017]

$$\mathrm{AF}(\boldsymbol{x}) = \mathrm{MI}(f_*; f(\boldsymbol{x})), \text{ where } f_* = \max_{\boldsymbol{x}} f(\boldsymbol{x}).$$

- However, for constrained problems, we revealed that
  - optimal value f_* may not exist

CBO based on Max-value entropy search (MES)

- MES for unconstrained problem [Wang and Jegelka, 2017]

$$\mathrm{AF}(\boldsymbol{x}) = \mathrm{MI}(f_*; f(\boldsymbol{x})), \text{ where } f_* = \max_{\boldsymbol{x}} f(\boldsymbol{x}).$$

- However, for constrained problems, we revealed that
  - optimal value f_* may not exist
  - conventional mutual information (MI) approximation can be negative

CBO based on Max-value entropy search (MES)

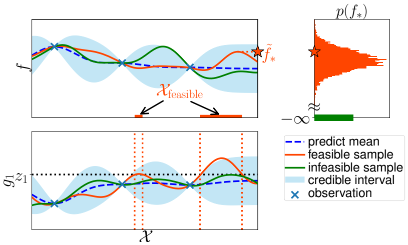
- MES for unconstrained problem [Wang and Jegelka, 2017]

$$\mathrm{AF}(\boldsymbol{x}) = \mathrm{MI}(f_*; f(\boldsymbol{x})), \text{ where } f_* = \max_{\boldsymbol{x}} f(\boldsymbol{x}).$$

- However, for constrained problems, we revealed that
  - optimal value f_* may not exist
  - conventional mutual information (MI) approximation can be negative

our key ideas

- Re-define f_* with infinite penalty of infeasibility
  - integrate uncertainty of feasibility



CBO based on Max-value entropy search (MES)

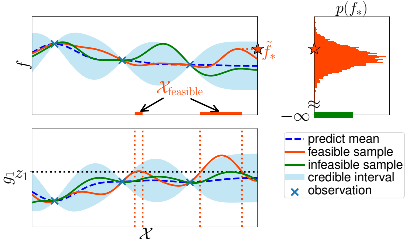
- MES for unconstrained problem [Wang and Jegelka, 2017]

$$\mathrm{AF}(\boldsymbol{x}) = \mathrm{MI}(f_*; f(\boldsymbol{x})), \text{ where } f_* = \max_{\boldsymbol{x}} f(\boldsymbol{x}).$$

- However, for constrained problems, we revealed that
  - optimal value f_* may not exist
  - conventional mutual information (MI) approximation can be negative

our key ideas

- Re-define f_* with infinite penalty of infeasibility
  - integrate uncertainty of feasibility
- AF based on the lower bound of MI
  - easy-to-compute
  - bounded from below by PI (> 0)
  - low estimation variance

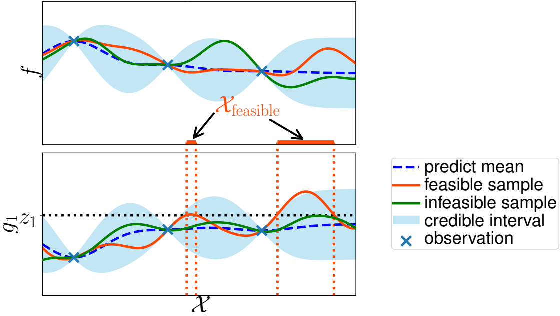


Definition of optimal value f_* for constrained problems

- A straightforward definition: $\max_{\boldsymbol{x} \in \mathcal{X}_{\mathrm{feasible}}} f(\boldsymbol{x})$ [Perrone et al., 2019]
  - $\mathcal{X}_{\mathrm{feasible}}$ is a feasible region.

Definition of optimal value f_* for constrained problems

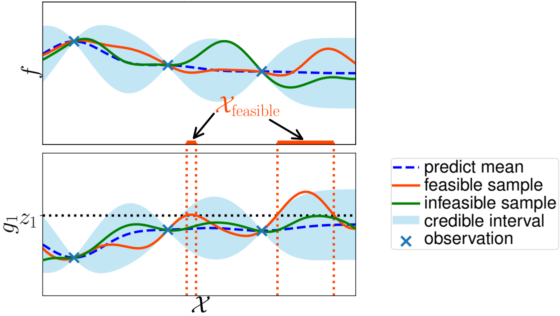
- A straightforward definition: $\max_{\boldsymbol{x} \in \mathcal{X}_{\text{feasible}}} f(\boldsymbol{x})$ [Perrone et al., 2019]
  - $\mathcal{X}_{\text{feasible}}$ is a feasible region.
  - However, $\mathcal{X}_{\text{feasible}}$ **can be empty** as in the following green sample path.



Definition of optimal value f_* for constrained problems

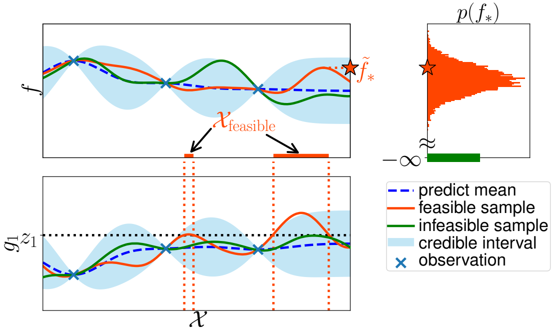
- A straightforward definition: $\max_{\boldsymbol{x} \in \mathcal{X}_{\text{feasible}}} f(\boldsymbol{x})$ [Perrone et al., 2019]
  - $\mathcal{X}_{\text{feasible}}$ is a feasible region.
  - However, $\mathcal{X}_{\text{feasible}}$ **can be empty** as in the following green sample path.
- We set:

$$f_* := \begin{cases} \max_{x \in \mathcal{X}_{\text{feasible}}} f(\boldsymbol{x}), & \text{if } \mathcal{X}_{\text{feasible}} \neq \emptyset, \\ -\infty, & \text{otherwise.} \end{cases}$$



Definition of optimal value f_* for constrained problems

- A straightforward definition: $\max_{\boldsymbol{x} \in \mathcal{X}_{\text{feasible}}} f(\boldsymbol{x})$ [Perrone et al., 2019]
  - $\mathcal{X}_{\text{feasible}}$ is a feasible region.
  - However, $\mathcal{X}_{\text{feasible}}$ **can be empty** as in the following green sample path.
- We set: $$f_* := \begin{cases} \max_{x \in \mathcal{X}_{\text{feasible}}} f(\boldsymbol{x}), & \text{if } \mathcal{X}_{\text{feasible}} \neq \emptyset, \\ -\infty, & \text{otherwise.} \end{cases}$$



- $p(f_*)$ contains the **uncertainty of feasibility**.

Acuqisition function of the proposed method

- We develop Monte Carlo (MC) estimator of lower bound of MI for AF.

Acuqisition function of the proposed method

- We develop Monte Carlo (MC) estimator of lower bound of MI for AF.
- Our AF results in,

$$\mathrm{AF}(\boldsymbol{x}) = -\frac{1}{|\mathcal{F}_*|} \sum_{\tilde{f}_* \in \mathcal{F}_*} \log\Big(1 - \boxed{\Pr(f(\boldsymbol{x}) \geq \tilde{f}_* \text{ and } \forall c, g_c(\boldsymbol{x}) \geq z_c)}\Big),$$

where $\mathcal{F}_*$ is a sample set of f_*.

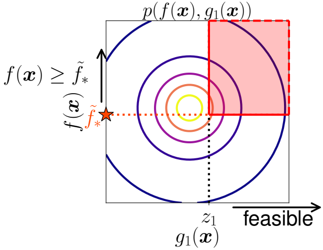
Acuqisition function of the proposed method

- We develop Monte Carlo (MC) estimator of lower bound of MI for AF.
- Our AF results in,

$$\mathrm{AF}(\boldsymbol{x}) = -\frac{1}{|\mathcal{F}_*|} \sum_{\tilde{f}_* \in \mathcal{F}_*} \log\left(1 - \boxed{\Pr(f(\boldsymbol{x}) \geq \tilde{f}_* \text{ and } \forall c, g_c(\boldsymbol{x}) \geq z_c)}\right),$$

where $\mathcal{F}_*$ is a sample set of f_*.

- Red probability can be seen as PI from $\tilde{f}_*$.

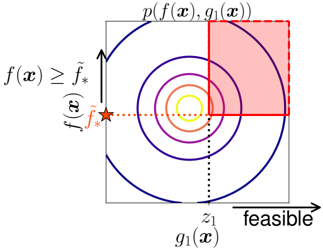


- We develop Monte Carlo (MC) estimator of lower bound of MI for AF.
- Our AF results in,

$$\mathrm{AF}(\boldsymbol{x}) = -\frac{1}{|\mathcal{F}_*|} \sum_{\tilde{f}_* \in \mathcal{F}_*} \log\left(1 - \boxed{\Pr(f(\boldsymbol{x}) \geq \tilde{f}_* \text{ and } \forall c, g_c(\boldsymbol{x}) \geq z_c)}\right),$$

where $\mathcal{F}_*$ is a sample set of f_*.

- Red probability can be seen as $\boxed{\text{PI from } \tilde{f}_*}$.
    - Computed easily by Gaussian CDF.

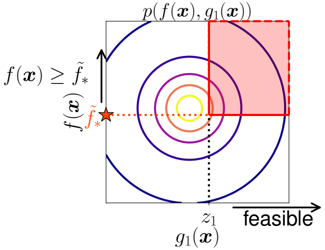


- We develop Monte Carlo (MC) estimator of lower bound of MI for AF.
- Our AF results in,

$$\mathrm{AF}(\boldsymbol{x}) = -\frac{1}{|\mathcal{F}_*|} \sum_{\tilde{f}_* \in \mathcal{F}_*} \log\left(1 - \boxed{\Pr(f(\boldsymbol{x}) \geq \tilde{f}_* \text{ and } \forall c, g_c(\boldsymbol{x}) \geq z_c)}\right),$$

where $\mathcal{F}_*$ is a sample set of f_*.

- Red probability can be seen as $\boxed{\text{PI from } \tilde{f}_*}$.
    - Computed easily by Gaussian CDF.
- Two important properties of our AF:
    - AF is bounded from below by $\boxed{\text{PI}}$ (Remark 4.1);



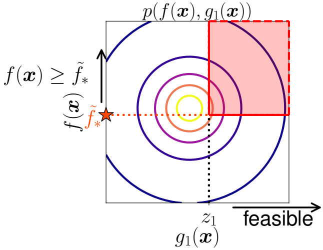
Acuqisition function of the proposed method

- We develop Monte Carlo (MC) estimator of lower bound of MI for AF.
- Our AF results in,

$$\mathrm{AF}(\boldsymbol{x}) = -\frac{1}{|\mathcal{F}_*|} \sum_{\tilde{f}_* \in \mathcal{F}_*} \log\big(1 - \boxed{\Pr(f(\boldsymbol{x}) \geq \tilde{f}_* \text{ and } \forall c, g_c(\boldsymbol{x}) \geq z_c)}\big),$$

where $\mathcal{F}_*$ is a sample set of f_*.

- Red probability can be seen as PI from $\tilde{f}_*$.
  - Computed easily by Gaussian CDF.
- Two important properties of our AF:
  - AF is bounded from below by PI (Remark 4.1);
    - ★ Directly implies **non-negativity**;

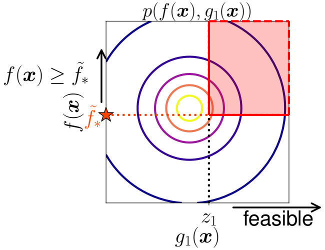


Acuqisition function of the proposed method

- We develop Monte Carlo (MC) estimator of lower bound of MI for AF.
- Our AF results in,

$$\mathrm{AF}(\boldsymbol{x}) = -\frac{1}{|\mathcal{F}_*|} \sum_{\tilde{f}_* \in \mathcal{F}_*} \log\left(1 - \boxed{\Pr(f(\boldsymbol{x}) \geq \tilde{f}_* \text{ and } \forall c, g_c(\boldsymbol{x}) \geq z_c)}\right),$$

  where $\mathcal{F}_*$ is a sample set of f_*.
- Red probability can be seen as $\boxed{\text{PI from } \tilde{f}_*}$.
  - Computed easily by Gaussian CDF.
- Two important properties of our AF:
  - AF is bounded from below by $\boxed{\text{PI}}$ (Remark 4.1);
    - ★ Directly implies **non-negativity**;
  - low estimation variance of MC estimation (Theorem 4.1).



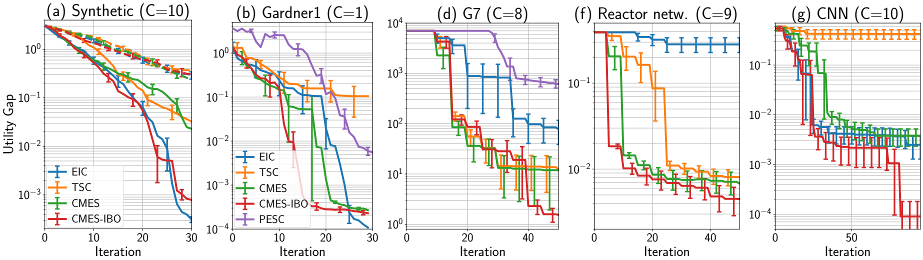


Figure: The solid line and error bar shows the mean and standard error, respectively.

- We evaluate following utility gap:

$$\mathrm{UG}_t := \begin{cases} f_* - f(\hat{\boldsymbol{x}}_t) & \text{if } \hat{\boldsymbol{x}}_t \text{ is feasible,} \\ f_* - \min f(\boldsymbol{x}) & \text{otherwise} \end{cases}$$

  - $\hat{\boldsymbol{x}}_t := \mathrm{argmax}_{\boldsymbol{x} \in \mathcal{X}} \mu_t^{(f)}(\boldsymbol{x}), \text{s.t. } \forall c,\ \Pr(g_c(\boldsymbol{x}) \geq z_c) \geq \sqrt[C]{0.95}$
    - ★ $\mu_t^{(f)}(\boldsymbol{x})$ is a predicted mean of the objective f at iteration t.

- Proposed CMES-IBO (red line) shows superior performance.

References I

Valerio Perrone, Iaroslav Shcherbatyi, Rodolphe Jenatton, Cedric Archambeau, and Matthias Seeger. Constrained Bayesian optimization with max-value entropy search. *arXiv:1910.07003*, 2019.

Zi Wang and Stefanie Jegelka. Max-value entropy search for efficient Bayesian optimization. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70, pages 3627–3635. PMLR, 2017.