Causal Bayesian Optimization

Virginia Aglietti

Research Scientist @ DeepMind

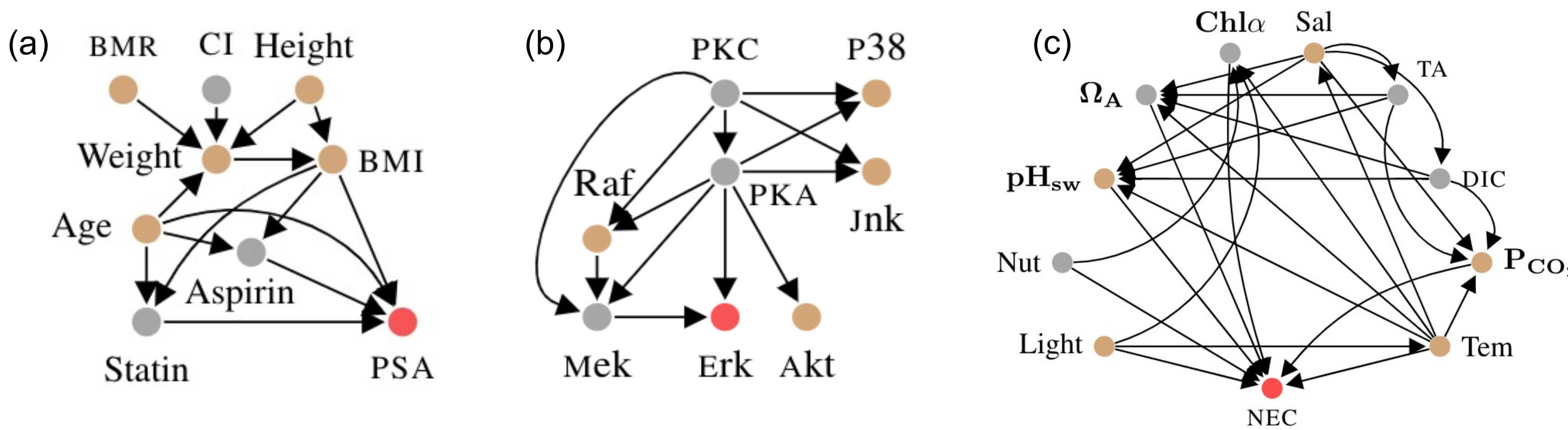
Outline of the Tutorial

- Quick Overview of the BO Framework and GPs
- Summary of advances in GPs and Acquisition Functions
- Bayesian Optimization over Discrete/Hybrid Spaces

- High-Dimensional Bayesian Optimization
- BoTorch Hands-on Demonstration

- Causal Bayesian Optimization
- Outstanding Challenges in BO

Applications



- Intervenable / Manipulative
- Non intervenable / Non manipulative
- Target node

a) Healthcare: Minimize the prostate specific antigen (PSA) by setting the dosage of statin and aspirin or changing calories intake (CI) (or a combination of all these interventions).

b) Genomics: Minimizing the extracellular signal-regulated kinase (Erk) by perturbing the protein kinases Mek, PKA and PKC (or a combination of these perturbations).

c) Ecology: Maximize the net ecosystem calcification (NEC) rate by intervening on the level of nutrients (Nut), the total alkalinity (TA), the dissolved inorganic carbon (DIC) and others.

Every time we want to **optimize a variable that is part of a system of interconnected nodes** that can be represented by a causal graph.

Setting and goal

- A **known** causal graph (Directed Acyclic Graph - DAG).
- **Observe the system** in a non-perturbed state and collect observational data from all (non hidden) nodes.
- **Run experiments** (in reality or in simulation). Every experiment corresponds to a function evaluation in BO where we fix the value of the inputs to a chosen level.
- Every experiment (every function evaluation) has a cost that depends on the number and type of nodes in which we intervene.

Goal: Efficiently find the optimal intervention (optimal intervention set and level) to perform.

Exploiting all available source of information, e.g. the causal graph

Intervention optimizing a target node in a graph.

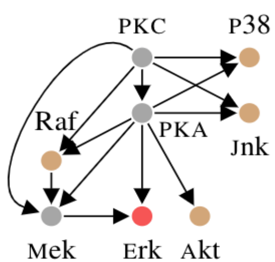
Setting and goal

- A **known** causal graph (Directed Acyclic Graph - DAG).
- **Observe the system** in a non-perturbed state and collect observational data from all (non hidden) nodes.
- **Run experiments** (in reality or in simulation). Every experiment corresponds to a function evaluation in BO where we fix the value of the inputs to a chosen level.
- Every experiment (every function evaluation) has a cost that depends on the number and type of nodes in which we intervene.

Goal: Efficiently find the optimal intervention (optimal intervention set and level) to perform.

Exploiting all available source of information, e.g. the causal graph

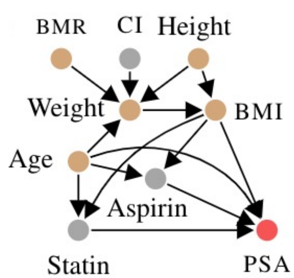
Intervention optimizing a target node in a graph.



Example 1.

(Intervention set) In order to minimize Erk should we perturb Mek? Or should we perturb PKA? Or both?

(Intervention level) If the optimal perturbation is Mek, what level should we set this kinase to?



Example 2.

(Intervention set) In order to minimize PSA should we intervene on the dosage of Statin? Or should we intervene on the CI? Or both?

(Intervention level) If the optimal perturbation is CI, what is be the optimal level?

Causal Global Optimization

- *Target* variable $Y \in \boldsymbol{V}$
- *Intervenable* variables $\boldsymbol{I} \subseteq \boldsymbol{V} \backslash Y$
- Interventional domain $D(\boldsymbol{X}) := \times_{X \in \boldsymbol{X}} D(X)$ for each $\boldsymbol{X} \subseteq \boldsymbol{I}$

$$\boxed{\boldsymbol{X}^*, \boldsymbol{x}^*} = \underset{\boxed{\substack{\boldsymbol{X} \in \mathcal{P}_{\boldsymbol{I}}, \\ \boldsymbol{x} \in D(\boldsymbol{X})}}}{\arg\min} \boxed{\mathbb{E}_{p(Y \mid \mathrm{do}(\boldsymbol{X}=\boldsymbol{x}))}[Y]}$$

- Optimal intervention set
- Optimal intervention level

Exploring all possible interventions in a casual graph

Optimizing the expected value of the target with respect to the interventional distribution

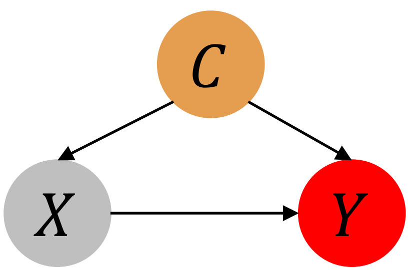
The distribution of Y when we intervene in the system and fix the value(s) of X to x.

Causal Global Optimization

$$\boldsymbol{X}^*, \boldsymbol{x}^* = \underset{\substack{\boldsymbol{X} \in \mathcal{P}_{\boldsymbol{I}}, \\ \boldsymbol{x} \in D(\boldsymbol{X})}}{\arg\min} \boxed{\mathbb{E}_{p(Y \mid \mathrm{do}(\boldsymbol{X}=\boldsymbol{x}))}[Y]}$$

Optimizing the expected value of the target with respect to the interventional distribution

Observed universe



$$\begin{aligned} C &= f_c(U_c) \\ X &= f_x(C, U_x) \quad \Leftrightarrow \quad p(X, C, Y) \\ Y &= f_y(X, C, U_y) \end{aligned}$$

Structural Causal Model

$$\langle \boldsymbol{U}, \boldsymbol{V}, \boldsymbol{F}, p(\boldsymbol{U}) \rangle$$

- $\boldsymbol{U} = \{U_1, \ldots, U_{K'}\}$ is a set of unobserved variables with distribution $p(\boldsymbol{U}) = \prod_k p(U_k)$.
- $\boldsymbol{F} = \{f_{V_1}, \ldots, f_{V_K}\}$ is a set of deterministic functions such that $V_k = f_{V_k}(\mathrm{pa}(V_k), \boldsymbol{U}_k)$ for each $V_k \in \boldsymbol{V}$.

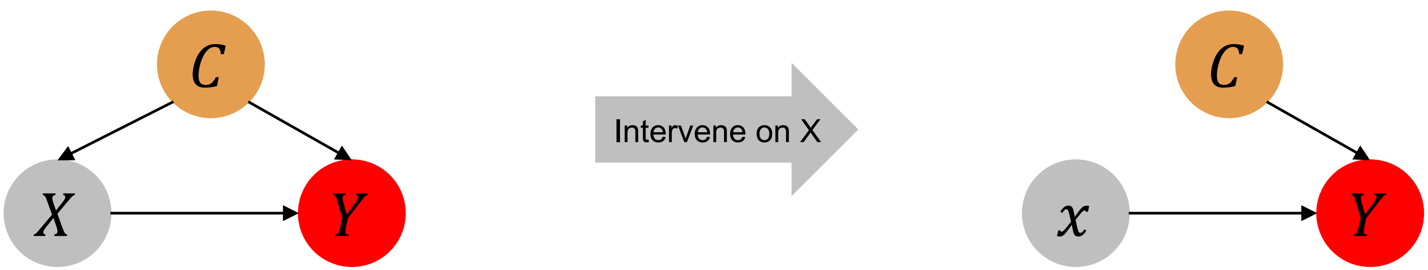
Causal Global Optimization

$$\boldsymbol{X}^*, \boldsymbol{x}^* = \arg\min_{\substack{\boldsymbol{X}\in\mathcal{P}_{\boldsymbol{I}},\\ \boldsymbol{x}\in D(\boldsymbol{X})}} \boxed{\mathbb{E}_{p(Y|\mathrm{do}(\boldsymbol{X}=\boldsymbol{x}))}[Y]}$$

Optimizing the expected value of the target with respect to the interventional distribution

Observed universe

Post-intervention universe

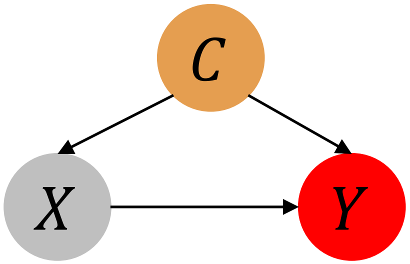


$$\begin{aligned} C &= f_c(U_c) \\ X &= f_x(C, U_x) \quad \Leftrightarrow \quad p(X, C, Y) \\ Y &= f_y(X, C, U_y) \end{aligned}$$

$$\begin{aligned} C &= f_c(U_c) \\ X &= x \quad \Leftrightarrow \quad p(C, Y | do(X = x)) \\ Y &= f_y(x, C, U_y) \end{aligned}$$

Causal Global Optimization

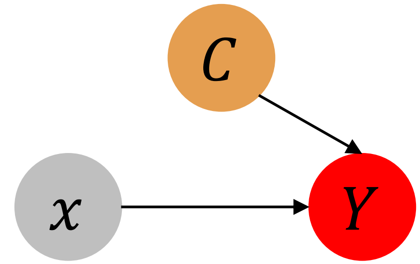
Observed universe



$$\begin{aligned} C &= f_c(U_c) \\ X &= f_x(C, U_x) \\ Y &= f_y(X, C, U_y) \end{aligned} \Leftrightarrow p(X, C, Y)$$

Intervene on X

Post-intervention universe

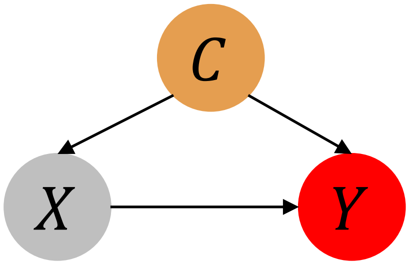


$$\begin{aligned} C &= f_c(U_c) \\ X &= x \\ Y &= f_y(x, C, U_y) \end{aligned} \Leftrightarrow p(C, Y | do(X = x))$$

How to do **inference in the post-intervention universe**?

Causal Global Optimization

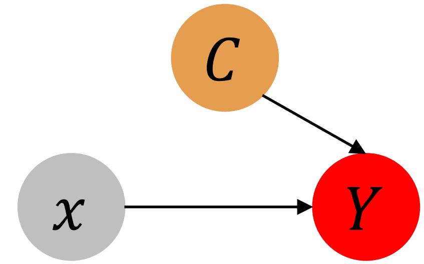
Observed universe



$$\begin{aligned} C &= f_c(U_c) \\ X &= f_x(C, U_x) \\ Y &= f_y(X, C, U_y) \end{aligned} \iff p(X, C, Y)$$

Intervene on X

Post-intervention universe



$$\begin{aligned} C &= f_c(U_c) \\ X &= x \\ Y &= f_y(x, C, U_y) \end{aligned} \iff p(C, Y | do(X = x))$$

$$p(Y|\mathrm{do}(X=1)) = \int p(Y|X=1, C=c)\, p(C=c)dc$$

How to do **inference in the post-intervention universe**?

Observe,
collect observational data and approximate the interventional distribution with the **do-calculus**

Intervene,
collect interventional data

DeepMind

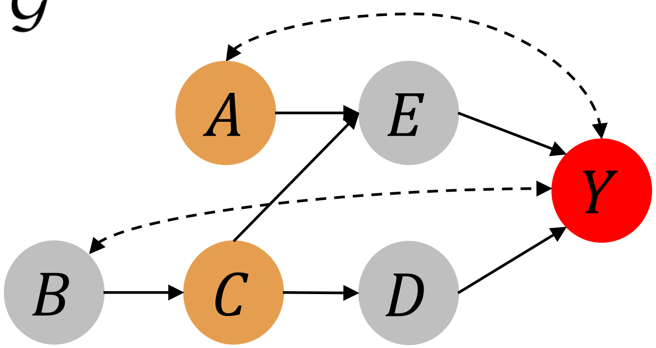
NEURAL INFORMATION PROCESSING SYSTEMS

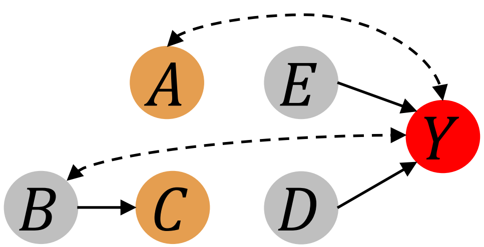
Global vs Causal Global Optimization

Causal Global Optimization

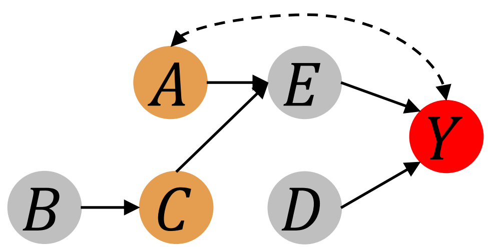
$$\boldsymbol{X}^*, \boldsymbol{x}^* = \underset{\substack{\boldsymbol{X} \in \mathcal{P}_{\boldsymbol{I}}, \\ \boldsymbol{x} \in D(\boldsymbol{X})}}{\arg\min}\, \mathbb{E}_{p(Y \mid \mathrm{do}(\boldsymbol{X}=\boldsymbol{x}))}[Y]$$

$\mathcal{G}$





$\mathcal{G}_{\overline{D,E}}$



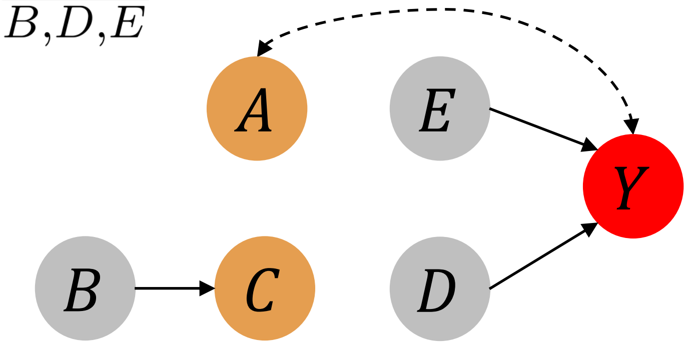
$\mathcal{G}_{\overline{B,D}}$

...

Global Optimization

$$\boldsymbol{x}^* = \underset{\boldsymbol{x} \in D(\boldsymbol{I})}{\arg\min}\, \mathbb{E}_{p(Y \mid \mathrm{do}(\boldsymbol{I}=\boldsymbol{x}))}[Y]$$

$\mathcal{G}_{\overline{B,D,E}}$



Global vs Causal Global Optimization

Causal Global Optimization

$$\boldsymbol{X}^*, \boldsymbol{x}^* = \underset{\substack{\boldsymbol{X} \in \mathcal{P}_{\boldsymbol{I}}, \\ \boldsymbol{x} \in D(\boldsymbol{X})}}{\arg\min}\, \mathbb{E}_{p(Y|\mathrm{do}(\boldsymbol{X}=\boldsymbol{x}))}[Y]$$

Global Optimization

$$\boldsymbol{x}^* = \underset{\boldsymbol{x} \in D(\boldsymbol{I})}{\arg\min}\, \mathbb{E}_{p(Y|\mathrm{do}(\boldsymbol{I}=\boldsymbol{x}))}[Y]$$

- Target function is explicitly unknown and multimodal
- Evaluations are perturbed by noise
- Evaluations are expensive

Bayesian Optimization

Global vs Causal Global Optimization

Causal Global Optimization

$$\boldsymbol{X}^*, \boldsymbol{x}^* = \underset{\substack{\boldsymbol{X} \in \mathcal{P}_{\boldsymbol{I}}, \\ \boldsymbol{x} \in D(\boldsymbol{X})}}{\arg\min}\, \mathbb{E}_{p(Y|\mathrm{do}(\boldsymbol{X}=\boldsymbol{x}))}[Y]$$

- Target function is explicitly unknown and multimodal
- Evaluations are perturbed by noise
- Evaluations are expensive

+ Causal Graph

Causal Bayesian Optimization

Global Optimization

$$\boldsymbol{x}^* = \underset{\boldsymbol{x} \in D(\boldsymbol{I})}{\arg\min}\, \mathbb{E}_{p(Y|\mathrm{do}(\boldsymbol{I}=\boldsymbol{x}))}[Y]$$

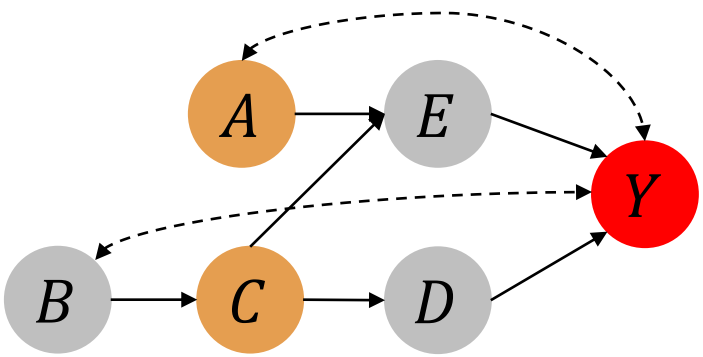
- Target function is explicitly unknown and multimodal
- Evaluations are perturbed by noise
- Evaluations are expensive

Bayesian Optimization

Global vs Causal Global Optimization

Causal Global Optimization

$$\boldsymbol{X}^*, \boldsymbol{x}^* = \underset{\substack{\boldsymbol{X} \in \mathcal{P}_{\boldsymbol{I}}, \\ \boldsymbol{x} \in D(\boldsymbol{X})}}{\arg\min}\, \mathbb{E}_{p(Y \mid \mathrm{do}(\boldsymbol{X}=\boldsymbol{x}))}[Y]$$



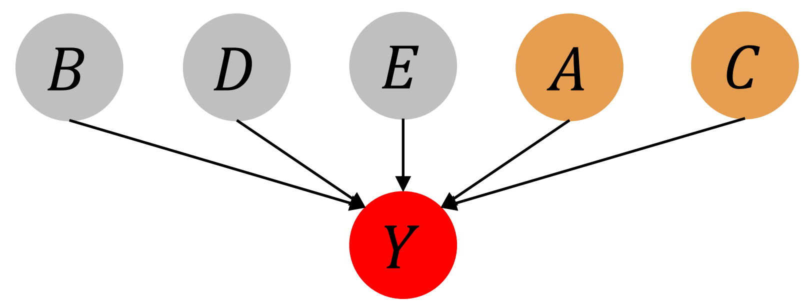
- Target function is explicitly unknown and multimodal
- Evaluations are perturbed by noise
- Evaluations are expensive

+ Causal Graph

Causal Bayesian Optimization

Global Optimization

$$\boldsymbol{x}^* = \underset{\boldsymbol{x} \in D(\boldsymbol{I})}{\arg\min}\, \mathbb{E}_{p(Y \mid \mathrm{do}(\boldsymbol{I}=\boldsymbol{x}))}[Y]$$



When there is not causal structure among the inputs and the output[1] the causal global optimization problem reduces to a global optimization problem. CBO is not needed.

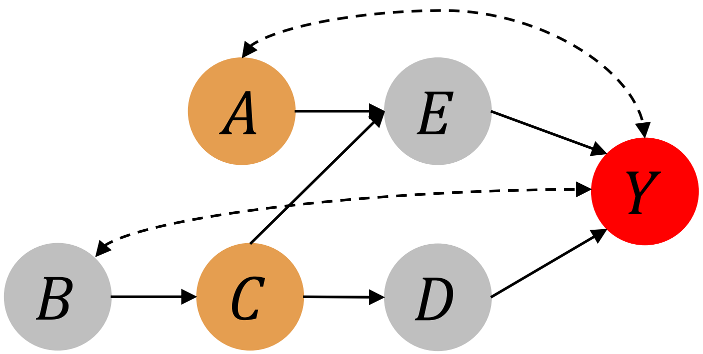
[1]and the observational domains equal the interventional domains D(**X**) for every intervention set **X** among the MISs.

[2]This is also true when there are no hidden confounders and the observational domains equal the interventional domains D(**X**) for every intervention **X** set among the MISs.

Global vs Causal Global Optimization

Causal Global Optimization

$$\boldsymbol{X}^*, \boldsymbol{x}^* = \arg\min_{\substack{\boldsymbol{X}\in\mathcal{P}_{\boldsymbol{I}},\\ \boldsymbol{x}\in D(\boldsymbol{X})}} \mathbb{E}_{p(Y|\mathrm{do}(\boldsymbol{X}=\boldsymbol{x}))}[Y]$$



- Target function is explicitly unknown and multimodal
- Evaluations are perturbed by noise
- Evaluations are expensive

+ Causal Graph

Causal Bayesian Optimization

1. **Search space:** Limit the sets to explore by identifying interventions worth exploring;
2. **Objective function:** Construct a surrogate model incorporating observational and interventional data;
3. **Acquisition function:** Extend the expected improvement acquisition function to explore different intervention sets;
4. **Observation/Intervention trade off** Allow the agent to observe or intervene.

BO dimensions

CBO is related to settings where actions or arms correspond to interventions on an arbitrary causal graph and there exists complex links between the agent's decisions and the received rewards, for instance Causal Bandits [1, 2] and Causal RL [3, 4].

Search space

Continuous variables (+ discrete number of sets). When all variables are discrete and the interventional domains are set equal to the observational domains this problem reduces to a causal bandit problem [1].

[1] Lee, Sanghack, and Elias Bareinboim. "Structural causal bandits: where to intervene?." *Advances in Neural Information Processing Systems 31* 31 (2018).

[2] Lattimore, F., Lattimore, T., & Reid, M. D. (2016). Causal bandits: Learning good interventions via causal inference. *Advances in Neural Information Processing Systems*, *29*.

[3] Zhang, J., & Bareinboim, E. (2016). *Markov decision processes with unobserved confounders: A causal approach*. Technical report, Technical Report R-23, Purdue AI Lab.

[4] Zhang, Junzhe. "Designing optimal dynamic treatment regimes: A causal reinforcement learning approach." *International Conference on Machine Learning*. PMLR, 2020.

BO dimensions

Search space

Continuous variables (+ discrete number of sets). When all variables are discrete and the interventional domains are set equal to the observational domains this problem reduces to a causal bandit problem [1].

Number of objective

Single objective. We have only *one* target variable in the graph that we wish to optimize.

Number of fidelities

Single fidelity. We incorporate different data types in the prior parameters for the surrogate models.

Constraints

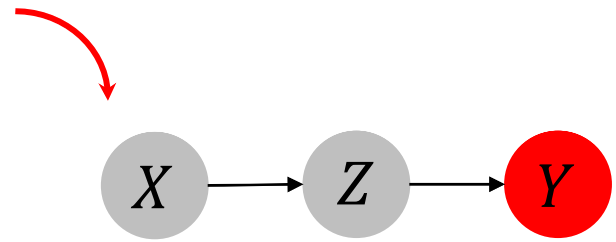
Both *constrained and unconstrained* settings. In some applications e.g. healthcare satisfying the constraints is very important.

CBO - Reducing the search space

$$\boldsymbol{X}^*, \boldsymbol{x}^* = \underset{\substack{\boldsymbol{X} \in \mathcal{P}_{\boldsymbol{I}}, \\ \boldsymbol{x} \in D(\boldsymbol{X})}}{\arg\min}\, \mathbb{E}_{p(Y \mid \mathrm{do}(\boldsymbol{X}=\boldsymbol{x}))}[Y]$$

1. **Search space:** Limit the sets to explore by identifying interventions worth exploring

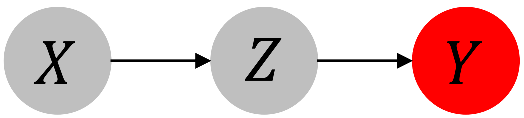
Definition 3.1. Minimal Intervention set (MIS). Given $\langle \mathcal{G}, \mathbf{Y}, \mathbf{X}, \mathbf{C} \rangle$, a set of variables $\mathbf{X}_s \in \mathcal{P}(\mathbf{X})$ is said to be a MIS if there is no $\mathbf{X}'_s \subset \mathbf{X}_s$ such that $\mathbb{E}[Y|\mathrm{do}(\mathbf{X}_s = \mathbf{x}_s)] = \mathbb{E}[Y|\mathrm{do}(\mathbf{X}'_s = \mathbf{x}'_s)]$.



$$\mathbb{E}[Y|\mathrm{do}(X = x), \mathrm{do}(Z = z)] = \mathbb{E}[Y|\mathrm{do}(Z = z)]$$

$$\mathbb{M}_{Y,\mathcal{G}} \subseteq \mathcal{P}_{\mathbf{I}}$$

Example – Reducing the search space



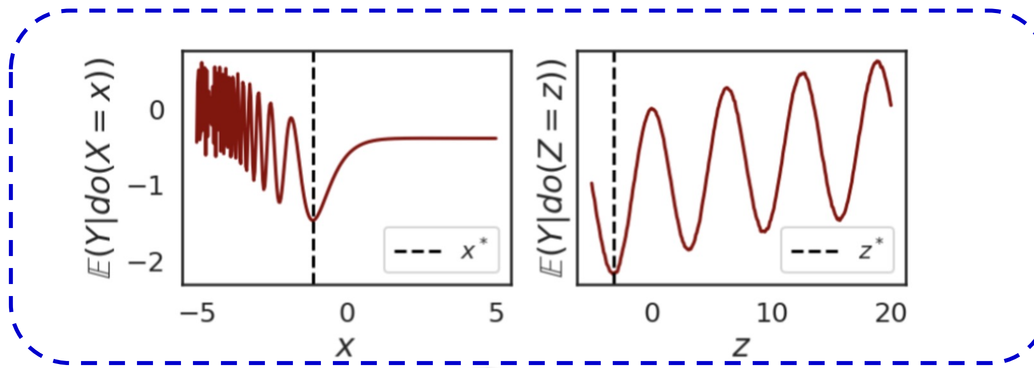
$$X = \epsilon_X$$

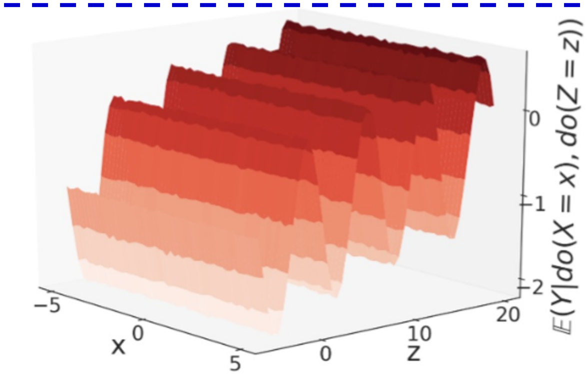
$$Z = \exp(-X) + \epsilon_Z$$

$$Y = \cos(Z) - \exp(-\frac{Z}{20}) + \epsilon_Y$$

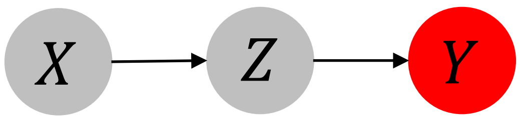
$$\boxed{\begin{aligned} \mathbb{M}_{\mathcal{G},Y} &= \{\emptyset, \{X\}, \{Z\}\} \\ \mathbb{P}_{\mathcal{G},Y} &= \{\{Z\}\} \\ \mathbb{B}_{\mathcal{G},Y} &= \{\{X, Z\}\} \end{aligned}}$$

Sets worth intervening on based on the causal graph structure.





Example – Reducing the search space



$$X = \epsilon_X$$

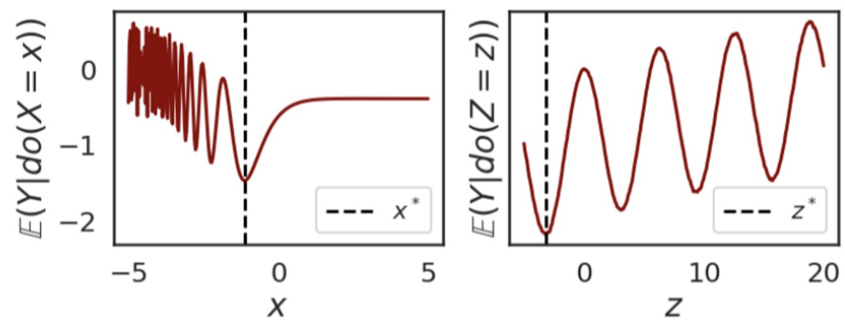
$$Z = \exp(-X) + \epsilon_Z$$

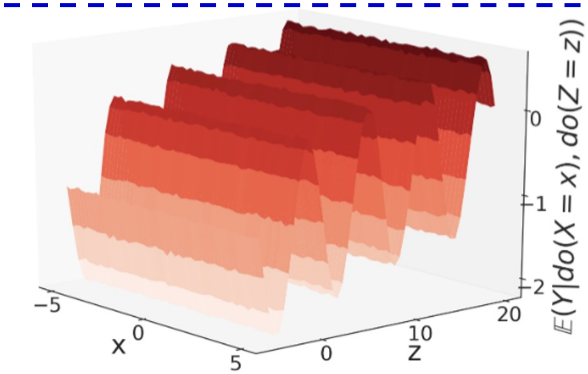
$$Y = \cos(Z) - \exp(-\frac{Z}{20}) + \epsilon_Y$$

$$\boxed{\begin{aligned} \mathbb{M}_{\mathcal{G},Y} &= \{\varnothing, \{X\}, \{Z\}\} \\ \mathbb{P}_{\mathcal{G},Y} &= \{\{Z\}\} \\ \mathbb{B}_{\mathcal{G},Y} &= \{\{X, Z\}\} \end{aligned}}$$

Knowing the causal graph allows to reason about the effective dimensionality of the problem, which is called the *causal intrinsic dimensionality.*

Sets worth intervening on based on the causal graph structure.





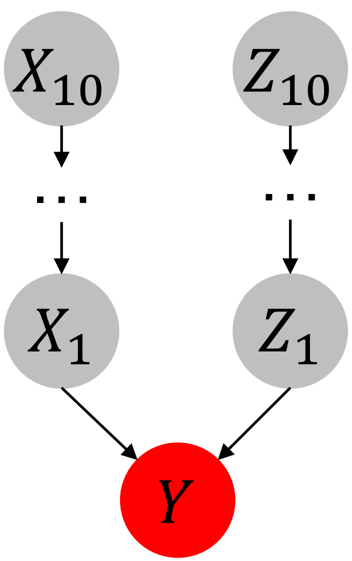
DeepMind

Example – Reducing the search space

Definition 2.1. The causal intrinsic dimensionality of a causal function $\mathbb{E}_{P(Y|\text{do}(\mathbf{X}=\mathbf{x}))}[Y]$ is given by the number of parents of the target variable, that is $|Pa(Y)|$.

Knowing the causal graph allows to reason about the effective dimensionality of the problem, which is called the *causal intrinsic dimensionality.*

$$Y \perp (X_2, \ldots, X_{10}, Z_2, \ldots, Z_{10}) | X_1, X_2$$



$$f_{\mathbf{X}}(x_1, \ldots x_{10}, z_1, \ldots z_{10}) = \mathbb{E}[Y|\text{do}(\mathbf{X} = x_1, \ldots x_{10}, z_1, \ldots z_{10})]$$

$$\mathbf{X} = \{X_1, \ldots, X_{10}, Z_1, \ldots, Z_{10}\}$$

We don't need to consider interventions on three variables. The input space of the surrogate models is either one or two.

Avoid high dimensional optimization which we know to be problematic for BO.

CBO - Surrogate models

$$\boldsymbol{X}^*, \boldsymbol{x}^* = \arg\min_{\substack{\boldsymbol{X}\in\mathcal{P}_{\boldsymbol{I}},\\ \boldsymbol{x}\in D(\boldsymbol{X})}} \boxed{\mathbb{E}_{p(Y|\mathrm{do}(\boldsymbol{X}=\boldsymbol{x}))}[Y]}$$

2. Objective function: Construct a surrogate model for all sets in the restricted search space by incorporating observational and interventional data

For every intervention set in $\mathbb{M}_{Y,\mathcal{G}}$ we model:

$$f_{\mathbf{X}}(\mathbf{x}) \sim \mathcal{GP}(m_{\mathbf{X}}(\mathbf{x}), K_{\mathbf{X}}(\mathbf{x}, \mathbf{x}'))$$

$$m_{\mathbf{X}}(\mathbf{x}) = \mathbb{E}_{\hat{p}(Y|\mathrm{do}(\mathbf{X}=\mathbf{x}))}[Y]$$

$$K_{\mathbf{X}}(\mathbf{x}, \mathbf{x}') = K_{\mathrm{RBF}}(\mathbf{x}, \mathbf{x}') + \sigma_{\mathbf{X}}(\mathbf{x})\sigma_{\mathbf{X}}(\mathbf{x}')$$

CBO - Surrogate models

$$\boldsymbol{X}^*, \boldsymbol{x}^* = \arg\min_{\substack{\boldsymbol{X} \in \mathcal{P}_{\boldsymbol{I}}, \\ \boldsymbol{x} \in D(\boldsymbol{X})}} \boxed{\mathbb{E}_{p(Y \mid \mathrm{do}(\boldsymbol{X}=\boldsymbol{x}))}[Y]}$$

2. Objective function: Construct a surrogate model for all sets in the restricted search space by incorporating observational and interventional data

For every intervention set in $\mathbb{M}_{Y,\mathcal{G}}$ we model:

$$f_{\mathbf{X}}(\mathbf{x}) \sim \mathcal{GP}(m_{\mathbf{X}}(\mathbf{x}), K_{\mathbf{X}}(\mathbf{x}, \mathbf{x}'))$$

$$m_{\mathbf{X}}(\mathbf{x}) = \mathbb{E}_{\hat{p}(Y \mid \mathrm{do}(\mathbf{X}=\mathbf{x}))}[Y]$$

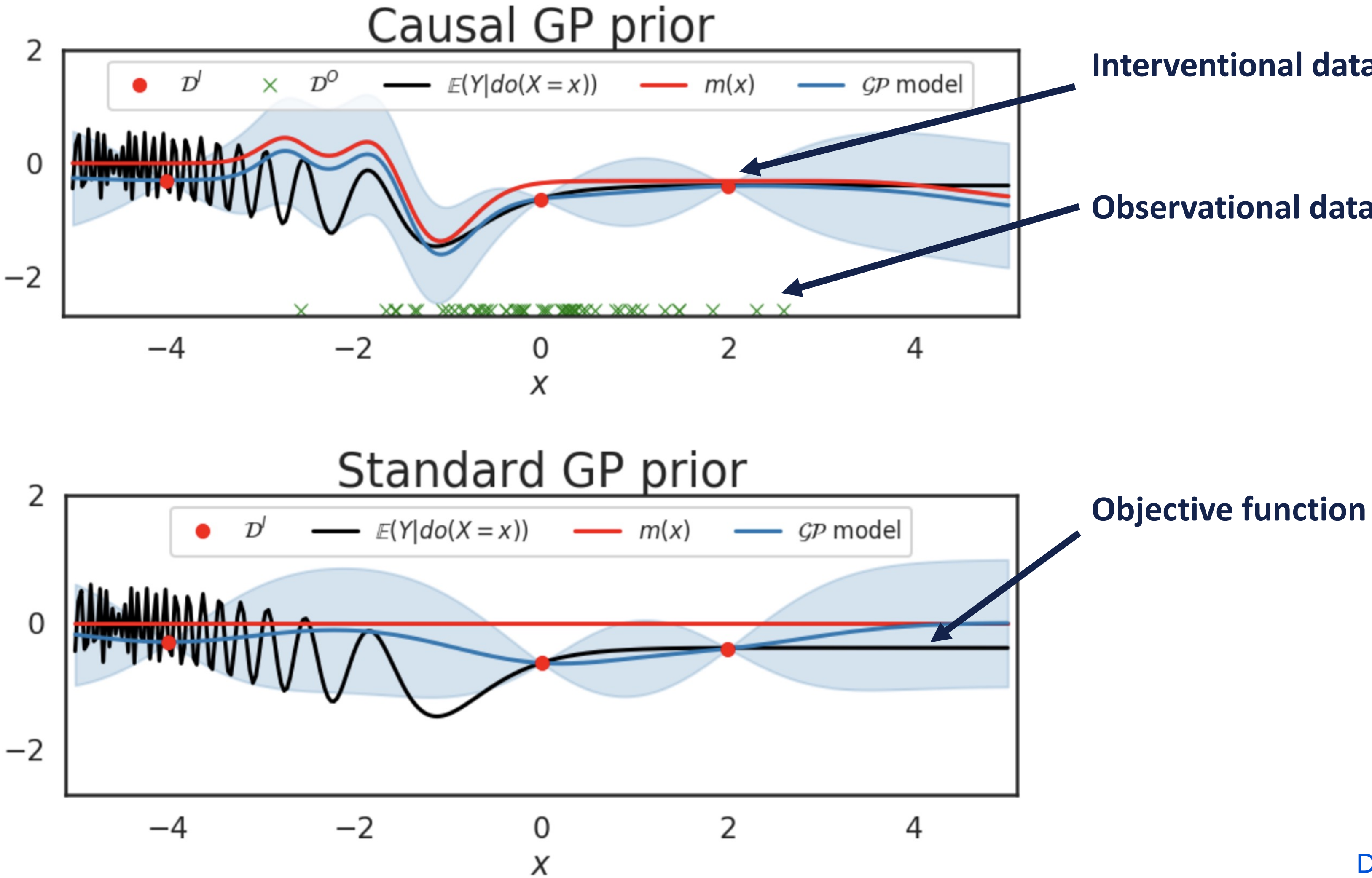
$$K_{\mathbf{X}}(\mathbf{x}, \mathbf{x}') = K_{\mathrm{RBF}}(\mathbf{x}, \mathbf{x}') + \sigma_{\mathbf{X}}(\mathbf{x})\sigma_{\mathbf{X}}(\mathbf{x}')$$

Approximated interventional distributions computed using observational data and the do-calculus.

The choice of the kernel is application specific!

$$\sqrt{\mathbb{V}_{\hat{p}(Y \mid \mathrm{do}(\mathbf{X}=\mathbf{x}))}[Y]}$$

Example – Surrogate models



CBO - Acquisition function

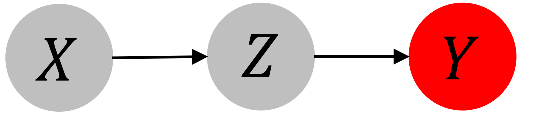
The cost structure is application specific!

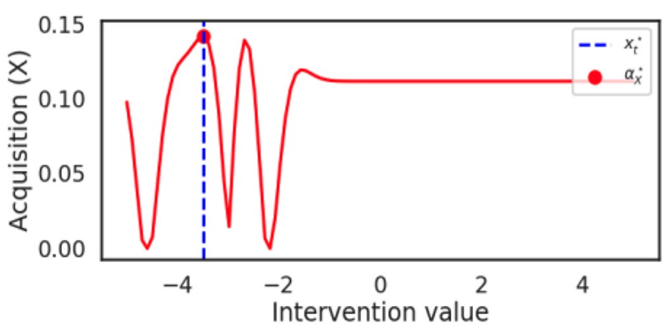
$$\mathrm{EI}_{\mathbf{X}}(\mathbf{x}) = \mathbb{E}_{p(y_{\mathbf{X}}|\mathcal{D}^I)}[\max(y_{\mathbf{X}}(\mathbf{x}) - y^\star, 0)]/\mathrm{Cost}(\mathbf{X}, \mathbf{x})$$

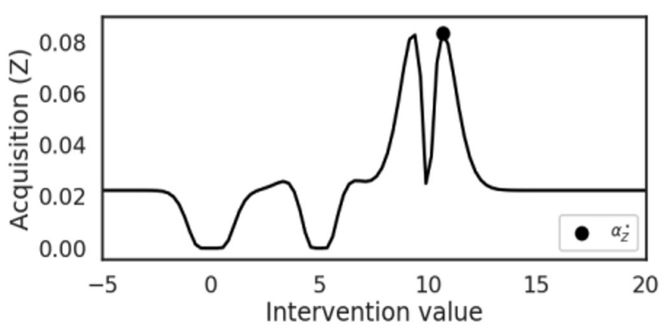
$$y_{\mathbf{X}}(\mathbf{x}) = f_{\mathbf{X}}(\mathbf{x}) + \epsilon_{\mathbf{X}}$$
$$\epsilon_{\mathbf{X}} \sim \mathcal{N}(0, \sigma^2_{\mathbf{X}})$$

$$y^\star = \max_{\mathbf{X} \in \mathbb{M}_{Y,\mathcal{G}}} \mathbf{y}_{\mathcal{D}^I_{\mathbf{X}}}$$

Optimize EI for every intervention set and select the set giving the highest expected improvement with respect to the optimum collected across sets.

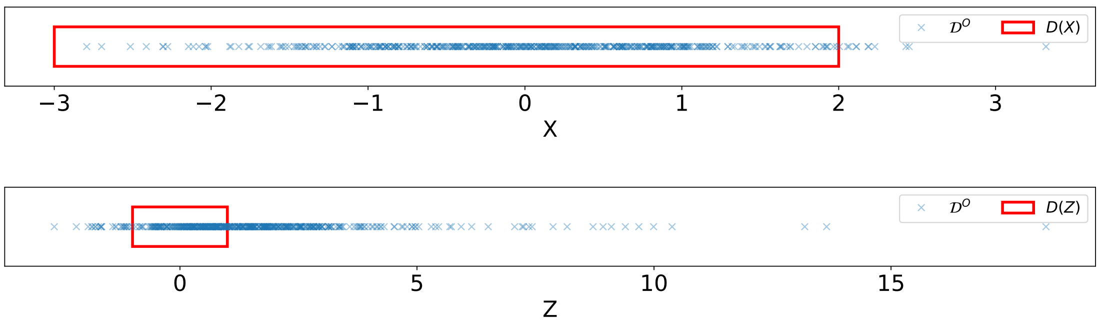


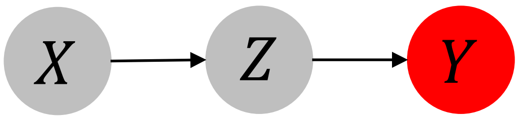


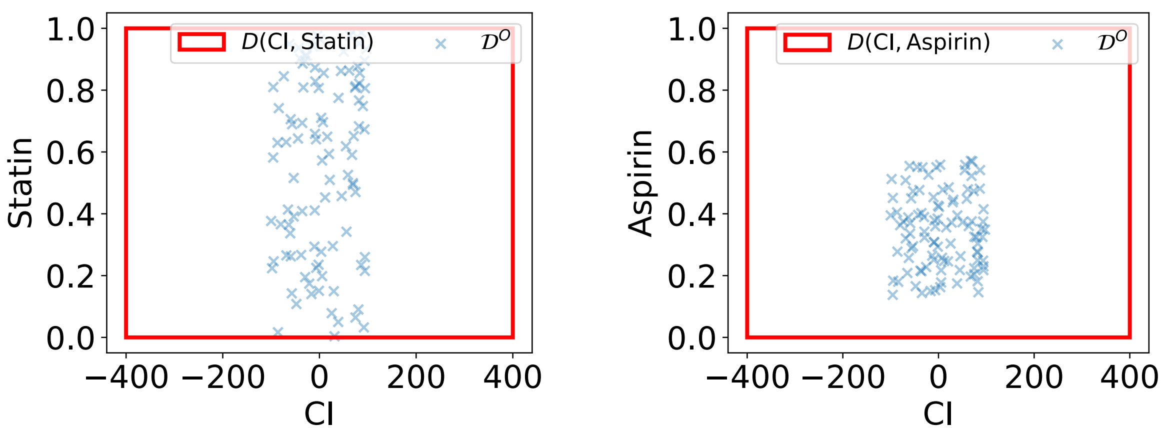


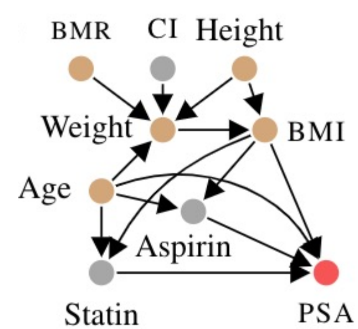
CBO - Observation Intervention trade off

$$\epsilon = \frac{\text{Vol}(\mathcal{C}(\mathcal{D}^O))}{\text{Vol}(\times_{X \in \mathbf{x}}(D(X)))} \times \frac{N}{N_{\text{max}}}$$









CBO algorithm

Algorithm: Causal Bayesian Optimization

Data: $\mathcal{D}^{\mathrm{O}}$, $\mathcal{D}^{\mathrm{I}}$, $\mathcal{G}$, **es**, number of steps T

Result: $\mathbf{X}_s^\star, \mathbf{x}_s^\star, \hat{\mathbb{E}}[\mathbf{Y}^\star|\mathrm{do}(\mathbf{X}_s^\star = \mathbf{x}_s^\star)]$

Initialise: Set $\mathcal{D}_0^{\mathrm{I}} = \mathcal{D}^{\mathrm{I}}$ and $\mathcal{D}_0^{\mathrm{O}} = \mathcal{D}^{\mathrm{O}}$

for t=1, ..., T **do**

Compute ϵ and sample $u \sim \mathcal{U}(0, 1)$

if $\epsilon > u$ **then**

(Observe)

1. Observe new observations $(\mathbf{x}_t, c_t, \mathbf{y}_t)$.
2. Augment $\mathcal{D}^{\mathrm{O}} = \mathcal{D}^{\mathrm{O}} \cup \{(\mathbf{x}_t, c_t, \mathbf{y}_t,)\}$.
3. Update prior of the causal GP.

end

else

(Intervene)

1. Compute $EI^s(\mathbf{x})$ for each element $s \in$ **es**.
2. Obtain the optimal interventional set-value pair $(s^\star, \alpha^\star)$.
3. Intervene on the system.
4. Update posterior of the interventional GP.

end

end

BO loop

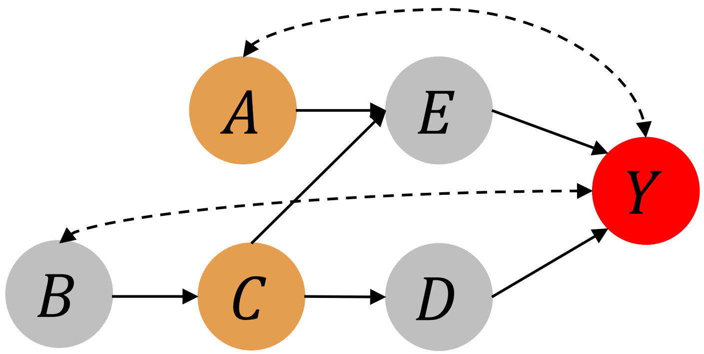
Update prior parameters

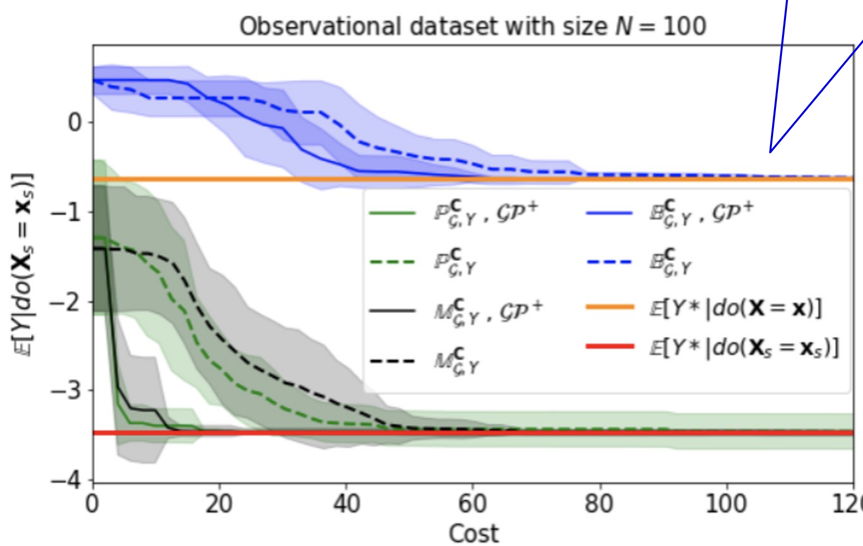
Update surrogate models

DeepMind

Simulation Results

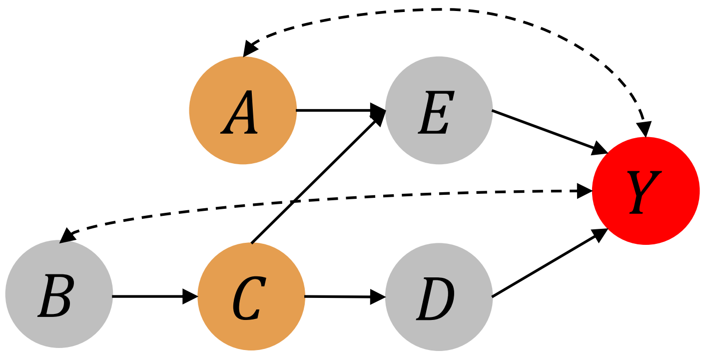
BO is slower and identifies a suboptimal intervention

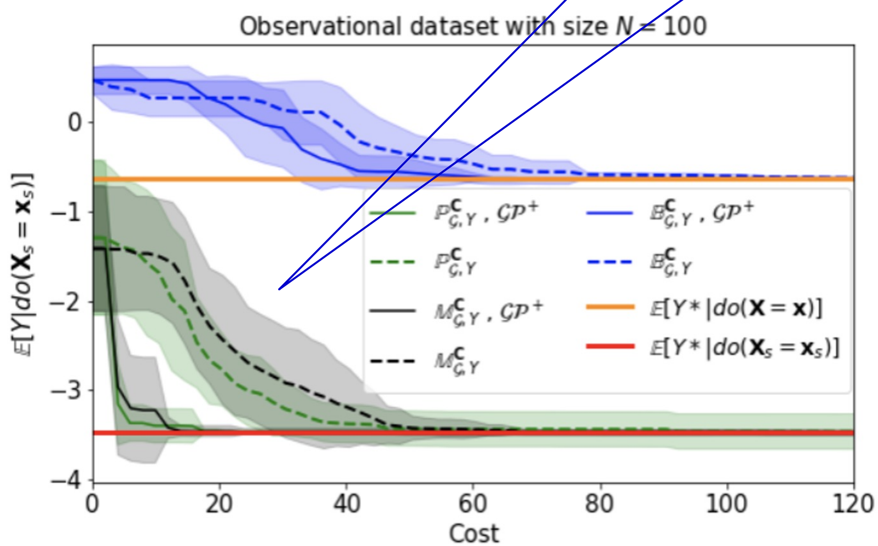




Simulation Results

CBO achieves the best result when using the causal GP model

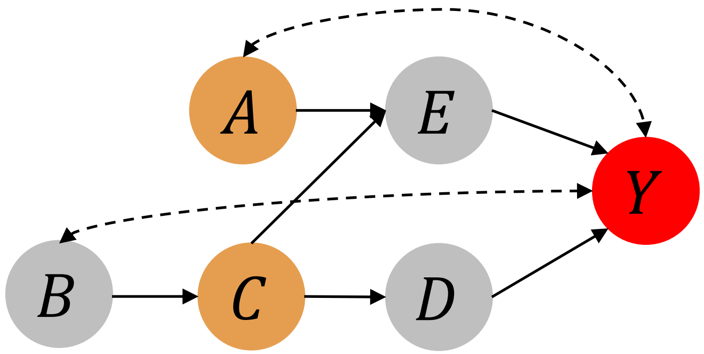


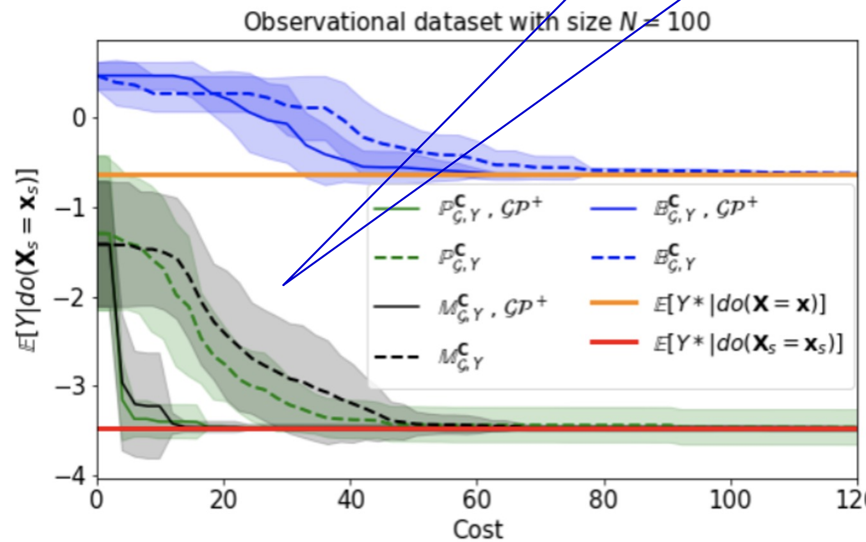


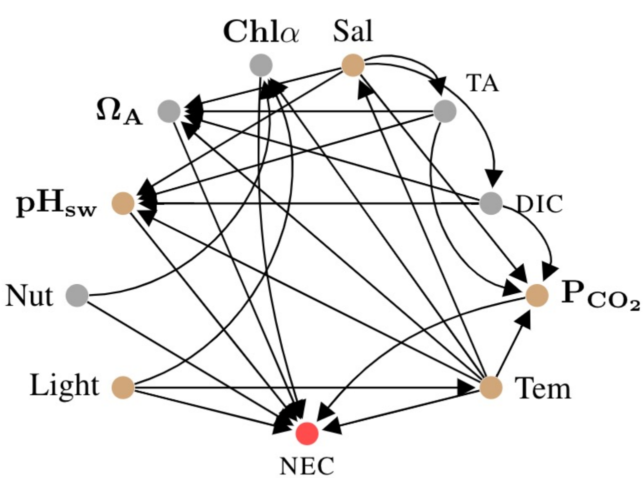
DeepMind

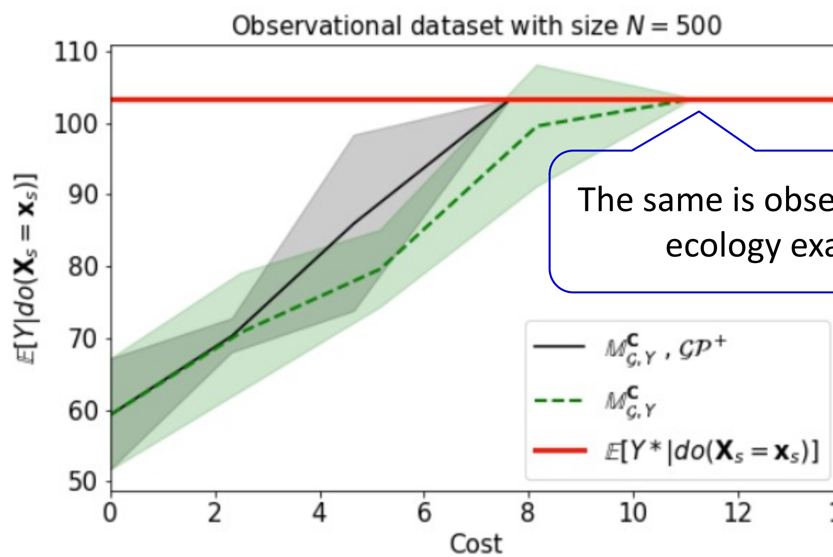
Simulation Results

CBO achieves the best result when using the causal GP model







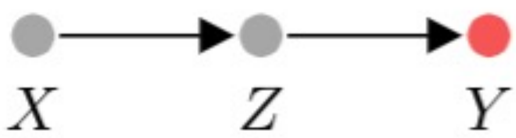


The same is observed for the ecology example.

DeepMind

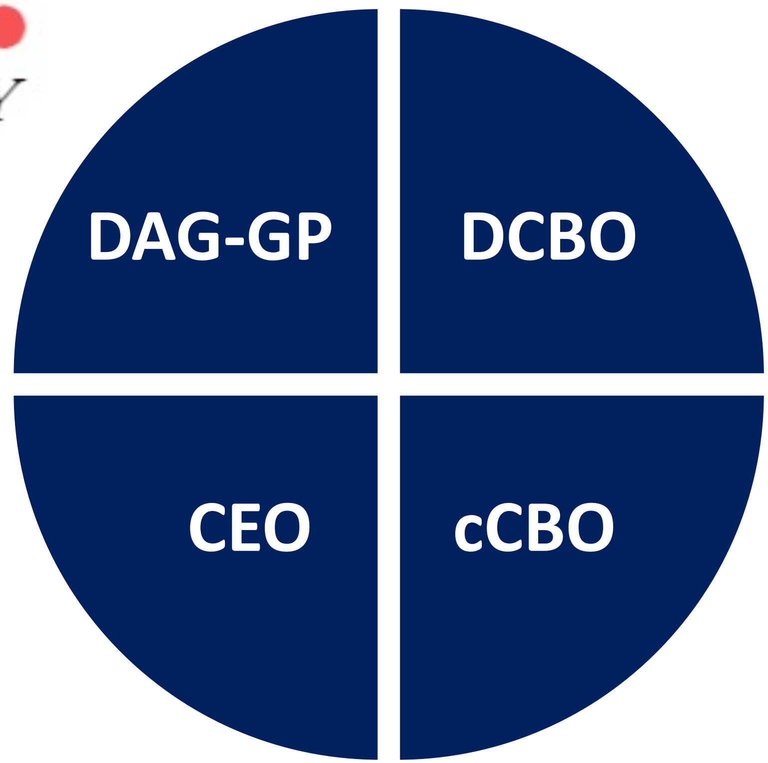
CBO limitations and advancements

DAG-GP surrogate model



$$\mathbb{E}[Y|\mathrm{do}\,(X=x)]$$

$$= \int \mathbb{E}[Y|\mathrm{do}\,(Z=z)]p(z|\mathrm{do}\,(X=x))\mathrm{d}z$$

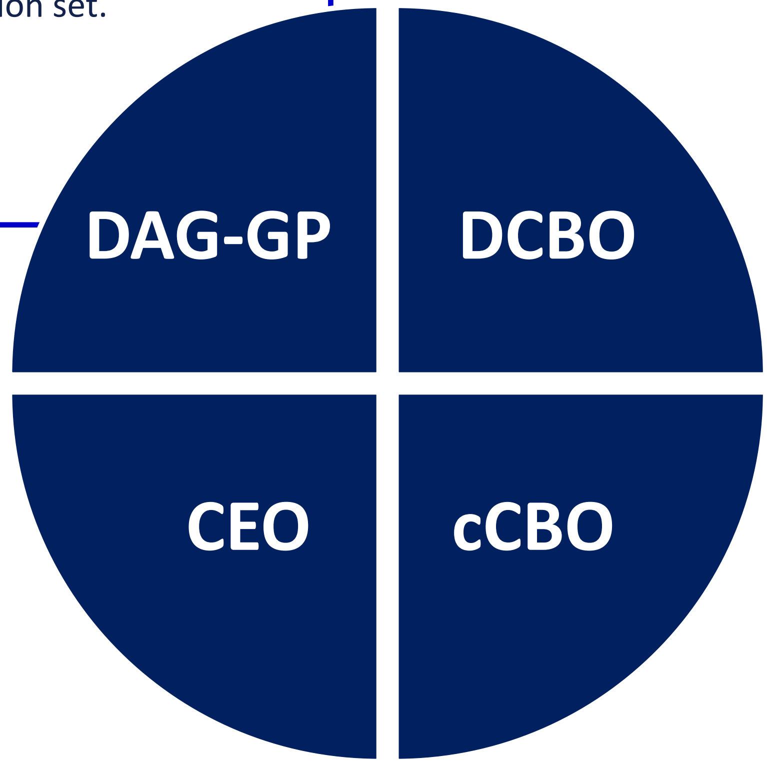


DeepMind

CBO limitations and advancements

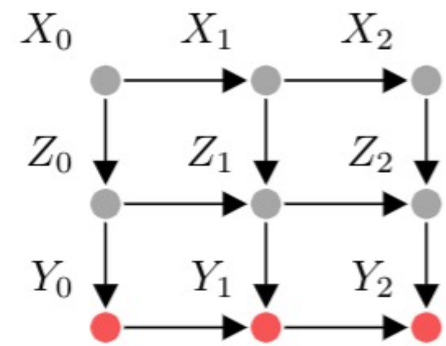
- *Modify the surrogate models* to incorporate the correlations across the objective functions obtained for different intervention set.

Aglietti, V., Damoulas, T., Álvarez, M., & González, J. (2020). Multi-task causal learning with gaussian processes. *Advances in Neural Information Processing Systems*, *33*, 6293-6304.



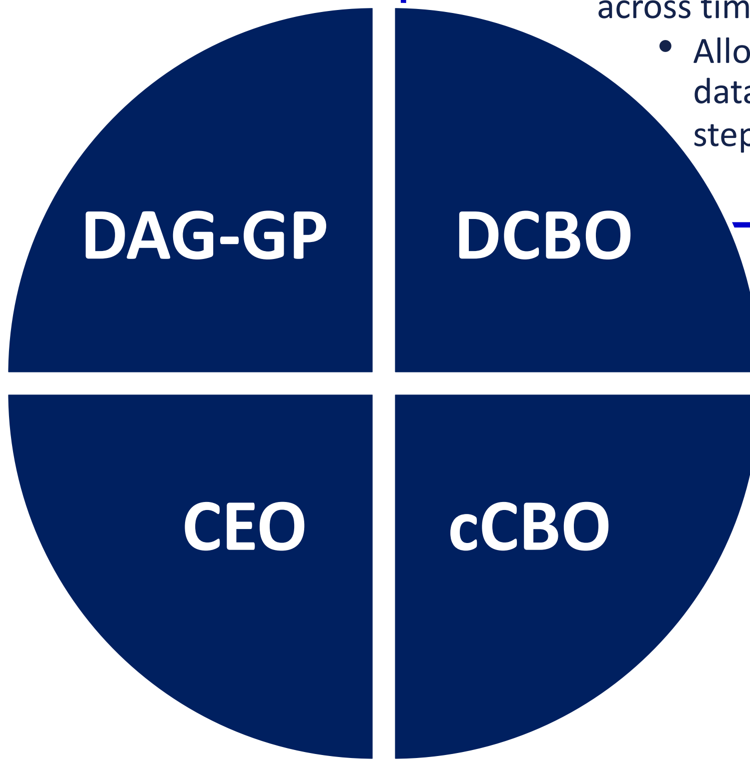
CBO limitations and advancements

Dynamic CBO



DeepMind

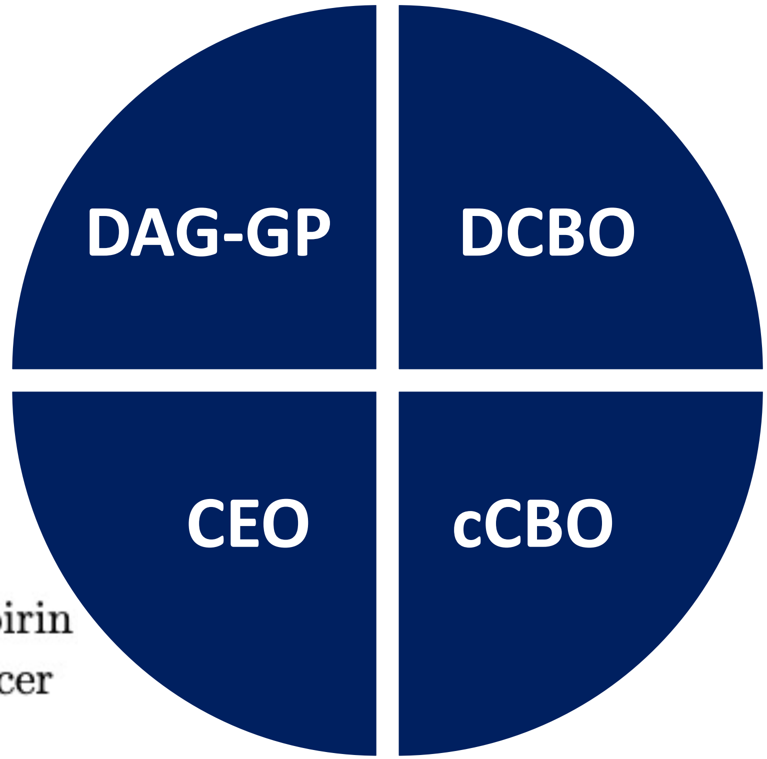
CBO limitations and advancements

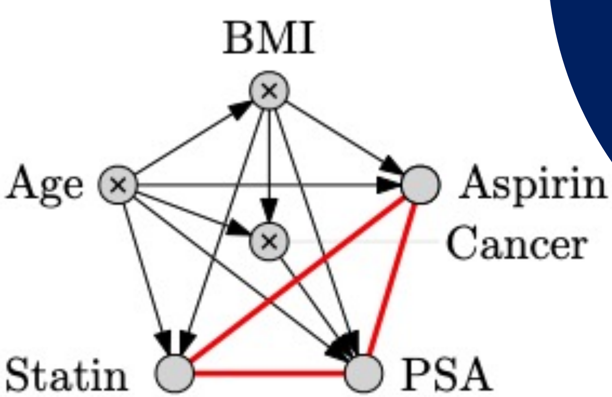


- *Modify the surrogate models* to incorporate the recursion existing among causal effects across time steps in the prior parameters.
  - Allow reusing interventional datapoints collected at previous time steps.

Aglietti, V., Dhir, N., González, J., & Damoulas, T. (2021). Dynamic Causal Bayesian Optimization. *Advances in Neural Information Processing Systems*, *34*, 10549-10560.

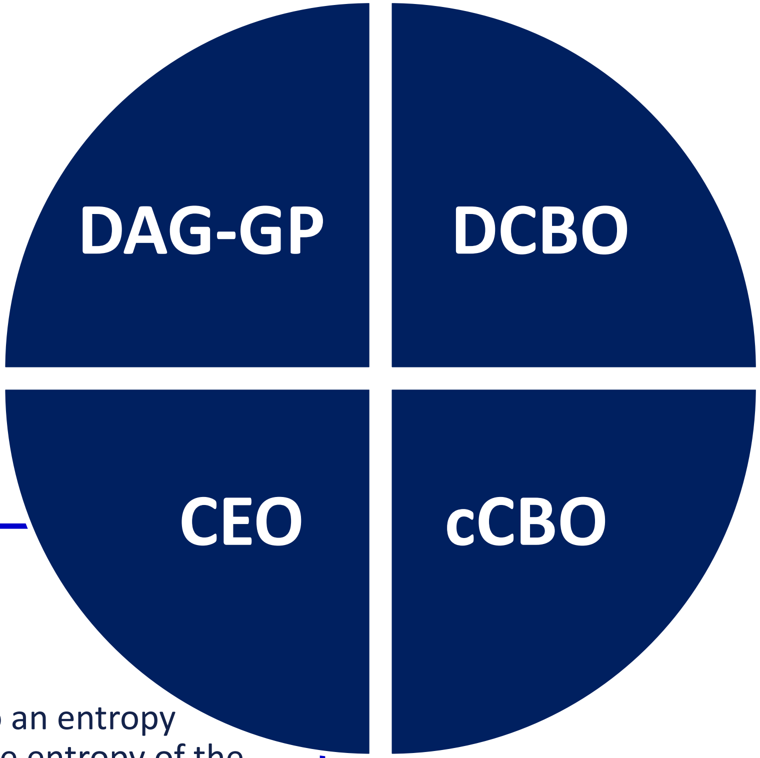
CBO limitations and advancements





Causal Entropy Optimization

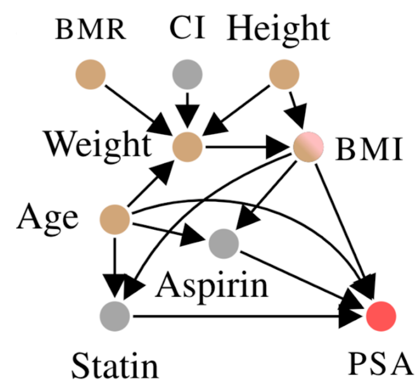
CBO limitations and advancements



Branchini N., Aglietti, V., Dhir, N., & Damoulas, T. (2022). Causal Entropy Optimization. *arXiv preprint*

- *Modify the surrogate models* to account for the uncertainty in the graph structure.
- *Modify the acquisition function* to an entropy based one. We aim at reducing the entropy of the joint distribution of the optimum and the graph structure.

CBO limitations and advancements



Constrained CBO

DeepMind

CBO limitations and advancements

Aglietti, V., Malek, A., Ktena, I. & Chiappa, S. (2022). Constrained Causal Bayesian Optiimization. *arXiv preprint.*

- *Reduce the search space* to account for constraints.
- *Modify the surrogate models* to incorporate correlations across target and constraints functions.
- *Modify the acquisition function* to a constrained expected improvement.

Outstanding Challenges in BO

- **High-dimensional BO**
  - Need more efficient approaches for high-dimensional spaces
  - How can we deal with large causal graphs?
- **BO over Combinatorial Structures**
  - How to combine domain knowledge, kernels, and (geometric) deep learning to build effective surrogate models?
  - How can we incorporate prior information when causal effects are not identifiable? How do we reduce the search space when the graph is partially known?
  - Effective methods to select large and diverse batches?
- **BO over Nested Function Pipelines**
  - Relatively less explored problem
- **BO with Resource Constraints**
  - Real-world experiments need resources and setup time. In CBO, how can we incorporate cost assumptions in the reduction of the search space?
  - Critical for BO deployment in science and engineering labs

DeepMind