Supplementary Material for "Causal Bayesian Optimization"

Virginia Aglietti
University of Warwick
The Alan Turing Institute
V.Aglietti@warwick.ac.uk

Xiaoyu Lu
Amazon
Cambridge, UK
luxiaoyu@amazon.com

Andrei Paleyes
Amazon
Cambridge, UK
paleyes@amazon.com

Javier González
Amazon
Cambridge, UK
gojav@amazon.com

1 Derivations of *do*-calculus for the synthetic experiment

1.1 $Do(B = b)$

$$\begin{aligned}
p(y|do(B=b)) &= \int p(y|c, do(B=b))p(c|B=b)dc \\
&= \int p(y|do(C=c), do(B=b))p(C=c|B=b)dc \quad (Y \perp\!\!\!\perp C|B \;\; in \; \mathcal{G}_{\bar{B}\underline{C}}) \\
&= \int p(y|do(C=c))p(c|B=b)dc \quad (Y \perp\!\!\!\perp B|C \;\; in \; \mathcal{G}_{\bar{B},\bar{C}}) \\
&= \int p(y|b', do(C=c))p(b'|do(C=c))p(c|B=b)db'dc \\
&= \int p(y|b', C=c)p(b')p(c|B=b)db'dc \quad (Y \perp\!\!\!\perp C|B \;\; in \; \mathcal{G}_{\bar{B},\underline{C}})
\end{aligned}$$

1.2 $Do(D = d)$

$$\begin{aligned}
p(y|do(D=d)) &= \int p(y|c, do(D=d))p(c|do(D=d))db \\
&= \int p(y|c, D=d)p(c)dc \quad (Y \perp\!\!\!\perp D|C \;\; in \; \mathcal{G}_{\underline{D}})
\end{aligned}$$

1.3 $Do(E = e)$

$$\begin{aligned}
p(y|do(E=e)) &= \int p(y|a, c, do(E=e))p(a, c|do(E=e))dadc \\
&= \int p(y|a, c, E=e)p(a)p(c)dadc \quad (Y \perp\!\!\!\perp E|A, C \;\; in \; \mathcal{G}_{\underline{E}})
\end{aligned}$$

1.4 $Do(B = b, D = d)$

$$\begin{aligned}
p(y|do(B=b), do(D=d)) &= \int p(y|do(B=b), c, do(D=d))p(c|do(B=b), do(D=d))dc \\
&= \int p(y|do(B=b), do(C=c), do(D=d))p(c|B=b)dc \quad (Y \perp\!\!\!\perp C|B, D \;\; in \; \mathcal{G}_{\underline{C}\bar{B}\bar{D}}) \\
&= \int p(y|do(C=c), do(D=d))p(c|B=b)dc \quad (Y \perp\!\!\!\perp B|C, D \;\; in \; \mathcal{G}_{\bar{B}\bar{C}\bar{D}}) \\
&= \int p(y|b', do(C=c), do(D=d))p(b'|do(C=c), do(D=d))p(c|B=b)dcdb' \\
&= \int p(y|b', C=c, do(D=d))p(b')p(c|B=b)dcdb' \quad (Y \perp\!\!\!\perp C|B, D \;\; in \; \mathcal{G}_{\bar{B}\bar{D}\underline{C}}) \\
&= \int p(y|b', C=c, D=d)p(b')p(c|B=b)dcdb' \quad (Y \perp\!\!\!\perp D|B, C \;\; in \; \mathcal{G}_{\underline{D}})
\end{aligned}$$

1.5 $Do(B = b, E = e)$

$$
\begin{aligned}
p(y|do(B=b), do(E=e)) &= \int p(y|do(B=b), c, do(E=e))p(c|B=b)dc \\
&= \int p(y|do(B=b), do(C=c), do(E=e))p(c|B=b)dc \quad (Y \perp\!\!\!\perp C|B,E \ in\ \mathcal{G}_{\bar{B}\bar{E}\underline{C}}) \\
&= \int p(y|do(C=c), do(E=e))p(c|B=b)dc \quad (Y \perp\!\!\!\perp B|C,E \ in\ \mathcal{G}_{\bar{C}\bar{E}\bar{B}}) \\
&= \int p(y|do(C=c), do(E=e), b')p(b'|do(C=c), do(E=e))p(c|B=b)db'dc \\
&= \int p(y|C=c, do(E=e), b')p(b')p(c|B=b)db'dc \quad (Y \perp\!\!\!\perp C|B,E \ in\ \mathcal{G}_{\bar{E}\underline{C}}) \\
&= \int p(y|a, C=c, do(E=e), b')p(a|C=c, do(E=e), b')p(b')p(c|B=b)db'dcda \\
&= \int p(y|a, b', C=c, E=e)p(a)p(b')p(c|B=b)db'dcda \quad (Y \perp\!\!\!\perp E|A,B,C \ in\ \mathcal{G}_{\underline{E}})
\end{aligned}
$$

1.6 $Do(D = d, E = e)$

$$
\begin{aligned}
p(y|do(D=d), do(E=e)) &= \int p(y|a, c, do(D=d), do(E=e))p(a, c|do(D=d), do(E=e))dadc \\
&= \int p(y|a, c, D=d, E=e)p(a)p(c)dadc \quad (Y \perp\!\!\!\perp (D,E)|A,C \ in\ \mathcal{G}_{\underline{D,E}})
\end{aligned}
$$

1.7 $Do(B = b, D = d, E = e)$

$$
p(y|do(B=b), do(D=d), do(E=e)) = p(y|do(D=d), do(E=e)) \quad (Y \perp\!\!\!\perp B|D,E \ in\ \mathcal{G}_{\bar{D},\bar{E},\bar{B}})
$$

2 SEM for the synthetic experiment

The SEM for the synthetic example is:

$$
\begin{aligned}
U_1 &= \epsilon_{YA} \\
U_2 &= \epsilon_{YB} \\
F &= \epsilon_F \\
A &= F^2 + U_1 + \epsilon_A \\
B &= U_2 + \epsilon_B \\
C &= \exp(-B) + \epsilon_C \\
D &= \exp(-C)/10. + +\epsilon_D \\
E &= \cos(A) + C/10 + \epsilon_E \\
Y &= \cos(D) + \sin(E) + U_1 + U_2\epsilon_y
\end{aligned}
$$

3 Cost configurations

Denote by $Co(\mathbf{X}, \mathbf{x})$ the cost of intervening on node $\mathbf{X}$ at the value $\mathbf{x}$. For the toy example and the real-data examples we consider fix unit cost across nodes. For the synthetic example we consider three possible cost

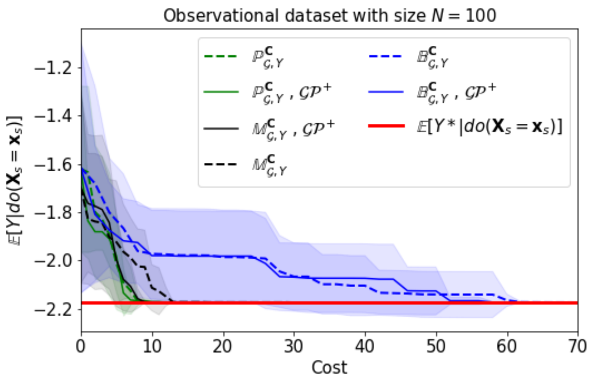

Figure 1: Toy example. Convergence of CBO and standard BO across different initializations of $\mathcal{D}^I$. The red line gives the optimal Y^* when intervening on sets in $\mathbb{M}^{\mathbf{C}}_{\mathcal{G},Y}$, $\mathbb{P}^{\mathbf{C}}_{\mathcal{G},Y}$ or $\mathbb{B}^{\mathbf{C}}_{\mathcal{G},Y}$. Solid lines give CBO results when using the causal GP model which is denoted by $\mathcal{GP}^+$. Dotted line correspond to CBO with a standard GP prior model $p(f(\mathbf{x}_s)) = \mathcal{GP}(0, k_{\text{RBF}}(\mathbf{x}_s, \mathbf{x}'_s))$. Shaded areas are $\pm$ standard deviation.

configurations: equal fix costs across nodes, different fix costs across nodes and variable costs across nodes. These are set to:

1. Fix equal costs: $Co(B, b) = Co(D, d) = Co(E, e) = Co(F, f) = 1$.
2. Fix different costs: $Co(B, b) = 10$, $Co(D, d) = 5$, $Co(E, e) = 20$ and $Co(F, f) = 3$.
3. Variable costs: $Co(B, b) = 10 + |b|$, $Co(D, d) = 5 + |d|$, $Co(E, e) = 20 + |e|$ and $Co(F, f) = 3 + |f|$.

4 Additional synthetic results

In Fig. 1 we show the results for the toy experiment across different initialization of $\mathcal{D}^I$.

In Fig. 2 we show the results for the synthetic experiment across different cost structures and values of N.

5 Example in Healthcare

The DAG describing the causal relationships between statin drugs and PSA (Thompson, 2019; Ferro et al., 2015) is given in Fig. 3. The SEM for this example is:

$$
\begin{aligned}
age &= \mathcal{U}(55, 75) \\
bmi &= \mathcal{N}(27.0 - 0.01 \times age, 0.7) \\
aspirin &= \sigma(-8.0 + 0.10 \times age + 0.03 \times bmi) \\
statin &= \sigma(-13.0 + 0.10 \times age + 0.20 \times bmi) \\
cancer &= \sigma(2.2 - 0.05 \times age + 0.01 \times bmi - 0.04 \times statin + 0.02 \times aspirin) \\
Y &= \mathcal{N}(6.8 + 0.04 \times age - 0.15 \times bmi - 0.60 \times statin + 0.55 \times aspirin + 1.00 \times cancer, 0.4)
\end{aligned}
$$

where $\mathcal{U}(a, b)$ denotes a uniform random variable with parameters a and b, $\mathcal{N}(m, s)$ represents a normal random variable with mean m and standard deviation s and σ denotes the sigmoidal function computed as $\sigma(x) = \frac{1}{1+e^{-x}}$.

6 Example in Ecology

The DAG describing the causal relationships between a set of environmental variables and NEC (Courtney et al., 2017) is given in Fig. 4. The variables included in the DAG are:

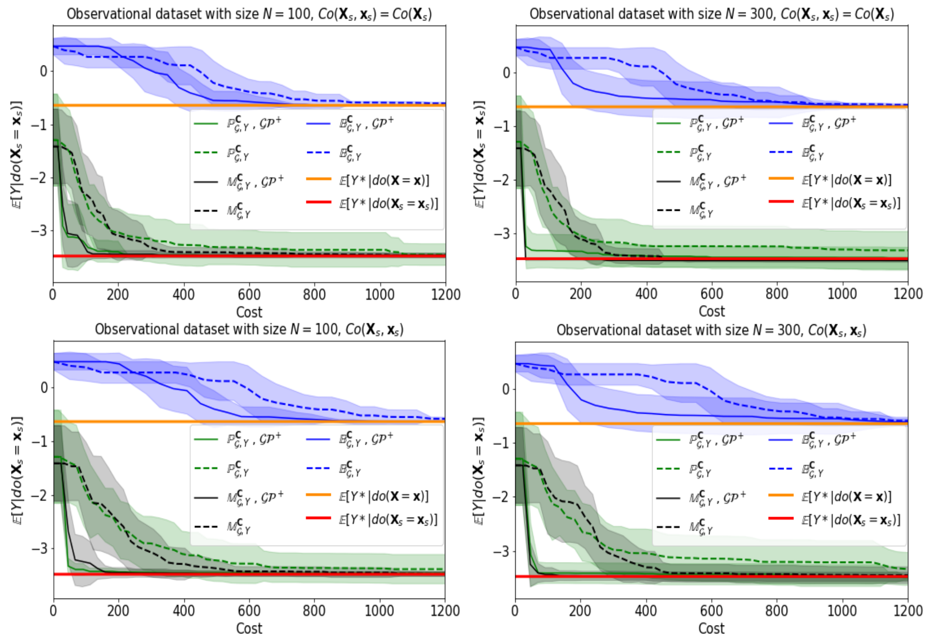

Figure 2: Synthetic example. Convergence of CBO and standard BO. The orange line gives the optimal Y^* when intervening on $\mathbb{B}^{\mathbf{C}}_{\mathcal{G},Y}$. The red line gives the optimal Y^* when intervening on sets in $\mathbb{M}^{\mathbf{C}}_{\mathcal{G},Y}$ or $\mathbb{P}^{\mathbf{C}}_{\mathcal{G},Y}$. Solid lines give CBO results when using the causal GP model which is denoted by $\mathcal{GP}^+$. Dotted line correspond to CBO with a standard GP prior model. *Upper left*: option (2) in §3, $N = 100$. *lower left*: option (3) in §3, $N = 100$. *Upper right*: option (2) in §3, $N = 300$. *Lower right*: option (3) in §3, $N = 300$.

- $Chl\alpha$: sea surface chlorophyll a;
- Sal: sea surface salinity;
- TA: seawater total alkalinity;
- DIC: seawater dissolved inorganic carbon;
- P_{CO_2}: seawater P_{CO_2};
- Tem: bottom temperature;
- NEC: net ecosystem calcification;
- Light: bottom light levels;
- Nut: PC1 of NH4, NiO2+NiO3, SiO4;
- pH_{SW}: seawater pH;
- Ω_A: seawater saturation with respect to aragonite.

See Andersson (2018) for more details.

References

Andersson, A., B. N. (2018). In situ measurements used for coral and reef-scale calcification structural equation modeling including environmental and chemical measurements, and coral calcification rates in bermuda from 2010 to 2012 (beacon project). *Biological and Chemical Oceanography Data Management Office (BCO-DMO). Dataset version 2018-03-02. http://lod.bco-dmo.org/id/dataset/720788.*

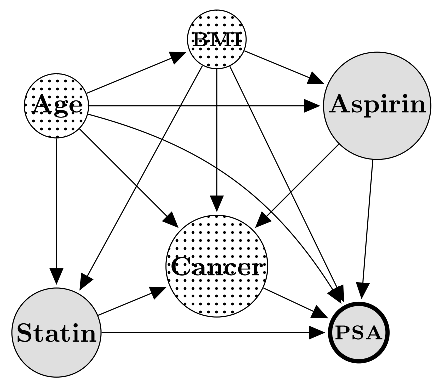

Figure 3: Causal graph of PSA level. Shaded nodes represent variables which can be intervened and dotted nodes represent non-manipulative variables. The target variable PSA is denoted with a thick shaded node.

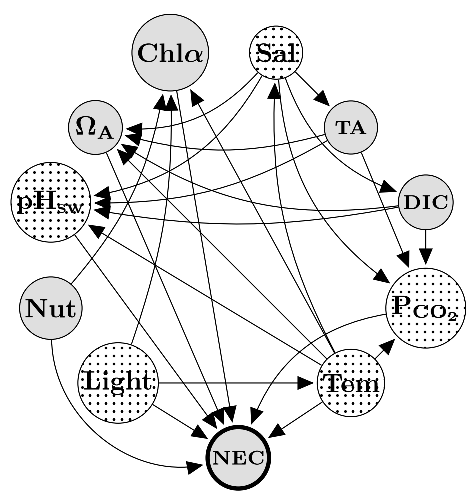

Figure 4: DAG of NEC level. Shaded nodes represent manipulative variables. Dotted nodes represent non-manipulative variables. The target variable NEC is denoted with a thick shaded node.

Courtney, T. A., Lebrato, M., Bates, N. R., Collins, A., De Putron, S. J., Garley, R., Johnson, R., Molinero, J.-C., Noyes, T. J., Sabine, C. L., et al. (2017). Environmental controls on modern scleractinian coral and reef-scale calcification. *Science advances*, 3(11):e1701356.

Ferro, A., Pina, F., Severo, M., Dias, P., Botelho, F., and Lunet, N. (2015). Use of statins and serum levels of prostate specific antigen. *Acta Urológica Portuguesa*, 32(2):71–77.

Thompson, C. (2019). Causal graph analysis with the causalgraph procedure. `https://www.sas.com/content/dam/SAS/support/en/sas-global-forum-proceedings/2019/2998-2019.pdf`.