Supplementary Material for: Differentiable Causal Backdoor Discovery

A Proofs

Lemma 1. If $W \perp\!\!\!\perp Y \mid Z^\star \cup \{X\}$, then there exists some scalar $\phi_X(Z^\star)$ such that $W \perp\!\!\!\perp Y \mid \{\phi_X(Z^\star), X\}$.

Proof. We will discuss the case for binary Y, as the proof for linear-Gaussian models follows the same idea. Let the structural equation for Y be given by $f_y(x, \pi_Z, \pi_U)$, where π_Z and π_U are the observed and unobserved parents of Y in the corresponding causal graph. The conditional distribution of Y is given by

$$
\begin{aligned}
p(y \mid x, \mathbf{z}^\star) = \\
p(f_y(x, \pi_Z, \pi_U) = 1 \mid x, \mathbf{z}^\star)^y \times \\
(1 - p(f_y(x, \pi_Z, \pi_U) = 1 \mid x, \mathbf{z}^\star))^{1-y}.
\end{aligned}
$$

By assumption, $f_y(\cdot)$ is functionally independent of W. Now we just have to show that the random variable $f_y(x, \pi_Z, \pi_U)$ is conditionally independent of W given X and $Z^\star$. Since $W \perp\!\!\!\perp Y \mid Z^\star \cup \{X\}$, it cannot be the case that W and $\pi_{\backslash Z^\star, X}$, the parents of Y not in $Z^\star \cup \{X\}$, are conditionally dependent given $Z^\star \cup \{X\}$. We define $\phi_X(\mathbf{z}^\star)$ as $p(f_y(x, \pi_Z, \pi_U) = 1 \mid x, \mathbf{z}^\star)$ for each possible realization of X. Given X, we can fully reconstruct from $\phi_X(\mathbf{z}^\star)$ a conditional distribution of Y that makes information about W irrelevant. □

Theorem 1. If $W \not\perp\!\!\!\perp Y \mid Z^\star \cup \{X\}$, and

$$
\begin{aligned}
\sum_{\mathbf{z}^\star \in \Phi^f_{x\mathbf{z}^\star}} p(y \mid w, x, \mathbf{z}^\star) \frac{p(\mathbf{z}^\star \mid w, x)}{\Pr(Z^\star \in \Phi^f_{x\mathbf{z}^\star} \mid w, x)} \neq \\
\sum_{\mathbf{z}^\star \in \Phi^f_{x\mathbf{z}^\star}} p(y \mid x, \mathbf{z}^\star) \frac{p(\mathbf{z}^\star \mid x)}{\Pr(Z^\star \in \Phi^f_{x\mathbf{z}^\star} \mid x)},
\end{aligned}
\tag{6}
$$

for some value f in the range of $\phi_x(\cdot)$, then $W \not\perp\!\!\!\perp Y \mid \{\phi_X(Z^*), X\}$.

Proof. Assume, contrary to the hypothesis, that $W \perp\!\!\!\perp Y \mid \{\phi_X(Z^\star), X\}$. Then

$$
\begin{aligned}
p(y \mid w, x, \phi_x(Z^\star) = f) = p(y \mid x, \phi_x(Z^\star) = f) \Rightarrow \\
\sum_{\mathbf{z}^\star} p(y \mid w, x, \phi_x(Z^\star) = f, \mathbf{z}^*) p(\mathbf{z}^* \mid w, x, \phi_x(Z^\star) = f) = \\
\sum_{\mathbf{z}^\star} p(y \mid x, \phi_x(Z^\star) = f, \mathbf{z}^\star) p(\mathbf{z}^\star \mid x, \phi_x(Z^\star) = f) \Rightarrow \\
\sum_{\mathbf{z}^\star} p(y \mid w, x, \mathbf{z}^\star) p(\mathbf{z}^\star \mid w, x, \phi_x(Z^\star) = f) = \\
\sum_{\mathbf{z}^\star} p(y \mid x, \mathbf{z}^\star) p(\mathbf{z}^* \mid x, \phi_x(Z^\star) = f) \Rightarrow \\
\sum_{\mathbf{z}^\star \in \Phi^f_{x\mathbf{z}^\star}} p(y \mid w, x, \mathbf{z}^\star) \frac{p(\mathbf{z}^\star \mid w, x)}{\Pr(Z^\star \in \Phi^f_{x\mathbf{z}^\star} \mid w, x)} = \\
\sum_{\mathbf{z}^\star \in \Phi^f_{x\mathbf{z}^\star}} p(y \mid x, \mathbf{z}^\star) \frac{p(\mathbf{z}^\star \mid x)}{\Pr(Z^\star \in \Phi^f_{x\mathbf{z}^\star} \mid x)},
\end{aligned}
$$

which contradicts the hypothesis. □