# Condition Number Bounds for Causal Inference (Supplementary Material)

**immediate**

# A SUPPLEMENTARY MATERIAL

## A.1 UPPER BOUNDS ON CONDITION NUMBERS

In this section, we give a proof of Lemma 3.

*Proof of Lemma 3.* For items 1 and 2 of the lemma, we will use the characterization of $\kappa(f, P)$ from Lemma 9. Note that when $f(P) > 0$, $\kappa(f, P) = \kappa(1/f, P)$, since

$$\kappa(1/f, P) = \max_{a \in \mathcal{A}(P)} \frac{\langle \nabla(1/f)(P), a \circ P \rangle}{1/f(P)} = \max_{a \in \mathcal{A}(P)} \frac{\langle -\nabla f(P), a \circ P \rangle}{f(P)} = \kappa(f, P),$$

where the first equality is from the chain rule while the second follows from the fact that $a \in \mathcal{A}(P)$ if and only if $-a \in \mathcal{A}(P)$. Thus, for proving item 1 of Lemma 3, it is sufficient to consider the case of the product $fg$. For this we have

$$\begin{aligned}
\kappa(fg, P) &= \max_{a \in \mathcal{A}(P)} \frac{\langle \nabla(fg)(P), a \circ P \rangle}{f(P) \cdot g(P)} \\
&= \max_{a \in \mathcal{A}(P)} \left\{ \frac{\langle \nabla f(P), a \circ P \rangle}{f(P)} + \frac{\langle \nabla g(P), a \circ P \rangle}{g(P)} \right\} \\
&\leq \max_{a \in \mathcal{A}(P)} \frac{\langle \nabla f(P), a \circ P \rangle}{f(P)} + \max_{a \in \mathcal{A}(P)} \frac{\langle \nabla g(P), a \circ P \rangle}{g(P)} \\
&= \kappa(f, P) + \kappa(g, P),
\end{aligned}$$

where the first equality is from Lemma 9 and the second is from the product rule for differentiation. This proves item 1 of Lemma 3 for products of functions, and as discussed above, the claim about quotients follows from this. We now turn to item 2. For this, we need the following standard fact: if $a, b, c, d$ are real numbers with $c, d > 0$ then

$$\frac{a+b}{c+d} \leq \max\left[\frac{a}{c}, \frac{b}{d}\right]. \tag{1}$$

We now have

$$\begin{aligned}
\kappa(f+g, P) &= \max_{a \in \mathcal{A}(P)} \frac{\langle \nabla f(P), a \circ P \rangle + \langle \nabla g(P), a \circ P \rangle}{f(P) + g(P)} \\
&\leq \max_{a \in \mathcal{A}(P)} \max\left\{ \frac{\langle \nabla f(P), a \circ P \rangle}{f(P)}, \frac{\langle \nabla g(P), a \circ P \rangle}{g(P)} \right\} \\
&= \max\left\{\kappa(f, P), \kappa(g, P)\right\},
\end{aligned}$$

where the first equality is from Lemma 9 and the first inequality uses eq. (1). This establishes item 2 of Lemma 3.

Finally, we prove item 3. For this case, it would be more convenient to directly use the definition of $\kappa$ (eq. (3) in Definition 1). We use the notation established in Definition 1. Let $\delta > 0$ be arbitrary. By the definitions of $\kappa(h, P)$ and $\kappa(f, h(P))$ respectively, there exist positive $\epsilon_1 = \epsilon_1(\delta)$ and $\epsilon_2 = \epsilon_2(\delta)$ such that

1. whenever $\mathrm{Rel}(P, \tilde{P}) \leq \epsilon_1$, it holds that

$$\mathrm{Rel}(h(P), h(\tilde{P})) \leq (\kappa(h, P) + \delta) \cdot \mathrm{Rel}(P, \tilde{P}), \text{ and} \tag{2}$$

2. whenever $\mathrm{Rel}(h(P), h(\tilde{P})) \leq \epsilon_2$, it holds that

$$\mathrm{Rel}(f(h(P)), f(h(\tilde{P}))) \leq (\kappa(f, h(P)) + \delta) \cdot \mathrm{Rel}(h(P), h(\tilde{P})). \tag{3}$$

Now choose $\epsilon = \epsilon(\delta) := \min\left\{\epsilon_1, \frac{\epsilon_2}{\kappa(h,P)+\delta}\right\}$. Then, combining eqs. (2) and (3), we get that for all $\tilde{P} \in B_\epsilon(P)$, we have

$$\mathrm{Rel}(f(h(P)), f(h(\tilde{P}))) \leq (\kappa(h, P) + \delta) \cdot (\kappa(f, h(P)) + \delta) \cdot \mathrm{Rel}(P, \tilde{P}). \tag{4}$$

To recap, we have argued that for any $\delta > 0$, we can choose $\epsilon > 0$ such that for all $\tilde{P} \in B_\epsilon(P)$, eq. (4) holds. The definition of $\kappa(f \circ h, P)$, according to eq. (3) in Definition 1, then implies the claim of item 3 of Lemma 3.

$$\kappa(f \circ h, P) \leq \kappa(f, h(P)) \cdot \kappa(h, P).$$

This completes the proof of the lemma. □

### A.2 A BRIEF DESCRIPTION OF THE NOTATION USED IN THE ID ALGORITHM

We give here a brief glossary of the notation used in the **ID** algorithm of Shpitser and Pearl [2006], reproduced in the main paper as algorithm 1.

The notations $\mathrm{An}(\mathbf{Y}) = \mathrm{An}(\mathbf{Y})_G$ and $\mathrm{Pa}(\mathbf{Y}) = \mathrm{Pa}(\mathbf{Y})_G$ are already defined in the main paper: they correspond respectively to the set of ancestors (including $\mathbf{Y}$) of $\mathbf{Y}$, and the set of parents of $\mathbf{Y}$, respectively. Given a set of vertices $\mathbf{X}$ in $G$, the graph $G_{\overline{\mathbf{X}}}$ is obtained by removing all incoming edges into vertices in $\mathbf{X}$. The set $\mathrm{An}(\mathbf{Y})_{G_{\overline{\mathbf{X}}}}$ appearing on line 3 therefore is the set of those vertices which have a directed path to $\mathbf{Y}$ that does not have any vertex in $\mathbf{X}$ as an intermediate vertex. The set $\mathbf{W}$ on line 3 thus consists of those observed vertices of $G \setminus \mathbf{X}$ for which all directed paths to any vertex in $\mathbf{Y}$ must pass through some vertex in $\mathbf{X}$.

On lines 4-6, $C(H)$, for a causal graph $H$, denotes the set of all maximum C-components in $H$. The notation $\pi$ on lines 6 and 7 is for a topological ordering of the observed nodes in the causal DAG $G$, and $V_i$ and $V_\pi^{(i-1)}$ denote, respectively, the $i$th vertex, and the set of the first $i - 1$ vertices, in the order induced by $\pi$. As before, capital letters have been used to denote vertices (or sets of vertices) in the causal DAG, while the corresponding small letters are used to denote assignments of values to the corresponding vertices.

For more detailed discussion of algorithm 1, we refer to Shpitser and Pearl [2006].