# Efficient Inference of Actual Causality via SAT Solving - Proof of Lemma 1

Amjad Ibrahim, Simon Rehwald, and Alexander Pretschner
{ibrahim, rehwald, pretschn}@in.tum.de

Department of Informatics, Technical University of Munich, Germany

## A Appendix: Proofs

**Lemma 1.** *In a binary model, if $\boldsymbol{X} = \boldsymbol{x}$ is a cause of $\varphi$, according to HP [1] definition, then every $\boldsymbol{x}'$ in the definition of AC2 always satisfies $\forall i.x'_i = \neg x_i$.*

*Proof.* We will use the following notation: $\overrightarrow{X}_{(n)}$ stands for a vector of length $n$, $X_1, \ldots, X_n$; and $\overrightarrow{X}_{(n)} = \overrightarrow{x}_{(n)}$ stands for $X_1 = x_1, \ldots, X_n = x_n$. Let $\overrightarrow{X}_{(n)} = \overrightarrow{x}_{(n)}$ be a cause for $\varphi$ in some model $M$.

1. AC1 yields
$$(M, \overrightarrow{u}) \models (\overrightarrow{X}_{(n)} = \overrightarrow{x}_{(n)}) \wedge (M, \overrightarrow{u}) \models \varphi. \tag{1}$$
2. Assume that the lemma does not hold. Then there is some index $k$ such that $x'_k = x_k$ and AC2 hold. Because we are free to choose the ordering of the variables, let us set $k = n$ wlog. We may then rewrite AC2 as follows:
$$\exists \overrightarrow{W}, \overrightarrow{w}, \overrightarrow{x}'_{(n)} \bullet (M, \overrightarrow{u}) \models (\overrightarrow{W} = \overrightarrow{w}) \implies (M, \overrightarrow{u}) \models \\ [\overrightarrow{X}_{(n-1)} \leftarrow \overrightarrow{x}'_{(n-1)}, X_n \leftarrow x_n, \overrightarrow{W} \leftarrow \overrightarrow{w}]\neg\varphi. \tag{2}$$
3. We will show that equations 1 and 2 give rise to a smaller cause, namely $\overrightarrow{X}_{(n-1)} = \overrightarrow{x}_{(n-1)}$. We need to show that $\overrightarrow{X}_{(n-1)} = \overrightarrow{x}_{(n-1)}$ satisfy AC1 and AC2, as stated by equations 3 and 4 below. This violates the minimality requirement of AC3 for $\overrightarrow{X}_{(n)} = \overrightarrow{x}_{(n)}$.
$$(M, \overrightarrow{u}) \models (\overrightarrow{X}_{(n-1)} = \overrightarrow{x}_{(n-1)}) \wedge (M, \overrightarrow{u}) \models \varphi \tag{3}$$
states AC1 for a potential "smaller" cause $\overrightarrow{X}_{(n-1)}$. Similarly,
$$\exists \overrightarrow{W}^*, \overrightarrow{w}^*, \overrightarrow{x}'^*_{(n-1)} \bullet (M, \overrightarrow{u}) \models (\overrightarrow{W}^* = \overrightarrow{w}^*) \\ \implies (M, \overrightarrow{u}) \models [\overrightarrow{X}_{(n-1)} \leftarrow \overrightarrow{x}'^*_{(n-1)}, \overrightarrow{W}^* \leftarrow \overrightarrow{w}^*]\neg\varphi \tag{4}$$
formulates AC2 for this potential smaller cause $\overrightarrow{X}_{(n-1)}$.

4. Let $\Psi$ denote the structural equations that define $M$. Let $\Psi'$ be $\Psi$ without the equations that define the variables $\overrightarrow{X}_{(n)}$ and $\overrightarrow{W}$; and let $\Psi''$ be $\Psi$ without the equations that define the variables $\overrightarrow{X}_{(n-1)}$ and $\overrightarrow{W}$. Clearly, $\Psi'' \implies \Psi'$. We can turn equation 1 into a propositional formula, namely

$$E_1 := \left(\Psi \wedge \overrightarrow{X}_{(n-1)} = \overrightarrow{x}_{(n-1)} \wedge X_n = x_n\right) \wedge \varphi. \tag{5}$$

Similarly, equation 3 is reformulated as

$$E_2 := \left(\Psi \wedge \overrightarrow{X}_{(n-1)} = \overrightarrow{x}_{(n-1)}\right) \wedge \varphi. \tag{6}$$

Because equation 2 holds, we fix some $\overrightarrow{W}$, $\overrightarrow{w}$, $\overrightarrow{x}'_{(n)}$ that make it true and rewrite this equation as

$$E_3 := \left(\Psi' \wedge \overrightarrow{X}_{(n-1)} = \overrightarrow{x}'_{(n-1)} \wedge X_n = x_n \wedge \overrightarrow{W} = \overrightarrow{w}\right) \implies \neg\varphi. \tag{7}$$

Finally, in equation 4, we use exactly these values to also fix $\overrightarrow{W}^* = \overrightarrow{W}$, $\overrightarrow{w}^* = \overrightarrow{w}$, and $\overrightarrow{x}'^*_{(n-1)} = \overrightarrow{x}'_{(n-1)}$, and reformulate this equation as

$$E_4 := \left(\Psi'' \wedge \overrightarrow{X}_{(n-1)} = \overrightarrow{x}'_{(n-1)} \wedge \overrightarrow{W} = \overrightarrow{w}\right) \implies \neg\varphi. \tag{8}$$

It is then a matter of equivalence transformations to show that

$$(\Psi'' \implies \Psi') \implies \left((E_1 \wedge E2) \implies (E_3 \wedge E_4)\right) \tag{9}$$

is a tautology, which proves the lemma.

## References

1. Halpern, J.Y.: A modification of the halpern-pearl definition of causality. In: Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015, Buenos Aires, Argentina, July 25-31, 2015. pp. 3022–3033 (2015), http://ijcai.org/Abstract/15/427