Supplement to "Causal Discovery with Linear Non-Gaussian Models under Measurement Error: Structural Identifiability Results"

This supplementary material provides the proofs which are omitted in the paper. The equation numbers in this material are consistent with those in the paper.

A.1: Proof of Proposition 3

Proof. Both $R_{j\leftarrow i}$ and $\tilde{X}_i^*$ are linear mixtures of independent, non-Gaussian variables $\tilde{E}_i$. According to the Darmois-Skitovich theorem (Kagan et al., 1973), $R_{j\leftarrow i}$ and $\tilde{X}_i^*$ are statistically independent if and only if for any k, at most one of the kth entries of their coefficient vectors, $\alpha_{j\leftarrow i}$ and $\mathbf{A}_{i\cdot}^{\mathrm{NL}}$, is non-zero, which is equivalent to the condition (11). □

A.2: Proof of Proposition 4

Proof. In the constructed ordered group decomposition, each group has one and only one non-leaf node. Just consider the non-leaf nodes in the ordered group decomposition. Combining Lemma 1 in (Shimizu et al., 2011b) and Proposition 3, one can see that the discovered causal ordering among them must be correct. □

A.3: Proof of Proposition 5

Proof. In each ordered group, there is a single non-leaf node, and all the others are leaf-nodes. Denote by $\tilde{X}_q^{*(k)}$ the qth node in the ordered group $g^{(k)}$. Denoted by $\tilde{X}_{\mathrm{NL}}^{*(k)}$ the only non-leaf node in the ordered group $g^{(k)}$. Denote by $\mathrm{PA}(\tilde{X}_{\mathrm{NL}}^{*(k)})$ the set of direct causes of $\tilde{X}_{\mathrm{NL}}^{*(k)}$ in $\tilde{G}^*$.

First consider the case where leaf node O has a parent which is not a parent of the non-leaf node in $g^{(k)}$ (the former part of Assumption A2). Let us regress each variable $\tilde{X}_q^{*(k)}$ in this group on all variables $\tilde{X}_i^*$ in the first $(k-1)$ ordered groups; in this regression task, all predictors are causally earlier than $\tilde{X}_q^{*(k)}$ because of the identifiable causal ordering among the ordered groups. Although the realizations of variables $\tilde{X}_i^*$ are unknown, such regression models can be estimated from the estimated matrix $\hat{\mathbf{A}}^{\mathrm{NL}}$, as done in Section 4.1, or by analyzing the estimated covariance matrix of $\tilde{\mathbf{X}}^*$, which is $\hat{\mathbf{A}}^{\mathrm{NL}}\hat{\mathbf{A}}^{NL\intercal}$ (we have assumed that $\mathbb{V}ar(\tilde{E}_i^{\mathrm{NL}}) = 1$ without loss of generality). There are two possible cases to consider.

i) Under Assumption A0, for the non-leaf node $\tilde{X}_{\mathrm{NL}}^{*(k)}$ in the kth group, $\mathrm{PA}(\tilde{X}_{\mathrm{NL}}^{*(k)})$ provides a minimal set of predictors with non-zero coefficients. (Such a minimal set of predictors may not be unique because of possible deterministic relations among $\tilde{X}_i^*$; however, there does not exist any smaller subset of $g^{(1)} \cup g^{(2)} \cup g^{(k-1)}$ which can predict $\tilde{X}_{\mathrm{NL}}^{*(k)}$ equally well.)

ii) Then consider a leaf node in this group, $\tilde{X}_{q'}^{*(k)}$, $\tilde{X}_{q'}^{*(k)} \neq \tilde{X}_{\mathrm{NL}}^{*(k)}$. Recall that when regressing $\tilde{X}_{q'}^{*(k)}$ on the variables in causally earlier ordered groups, $\tilde{X}_{\mathrm{NL}}^{*(k)}$ is not among the predictors because it is also in the kth group. First note that each node in $\mathrm{PA}(\tilde{X}_{\mathrm{NL}}^{*(k)})$ is always d-connected to $\tilde{X}_{q'}^{*(k)}$ given any variable set that does not include $\tilde{X}_{\mathrm{NL}}^{*(k)}$. As a consequence of the nondeterministically faithfulness assumption for $\tilde{G}^*$, in the regression model for $\tilde{X}_{q'}^{*(k)}$, $\mathrm{PA}(\tilde{X}_{\mathrm{NL}}^{*(k)}) \cup \mathrm{O}$ provides a minimal set of predictors with non-zero coefficients for O. Furthermore, under assumption A2, O has at least a direct cause that is not in $\mathrm{PA}(\tilde{X}_{\mathrm{NL}}^{*(k)})$. Therefore, in the regression model for O, the set of predictors with non-zero coefficients is a proper superset of $\mathrm{PA}(\tilde{X}_{\mathrm{NL}}^{*(k)})$ (the former has more elements).

That is, when regressing variables $\tilde{X}_q^{*(k)}$ in the considered group on variables in earlier groups, the non-leaf node, as well as possibly some of the leaf nodes, always has a smaller number of predictors with non-zero coefficients, compared to the regression model for leaf node O. Hence we can determine O as a leaf node.

Then let us consider the case where leaf node O has a parent which is not a parent of some other leaf node in $g^{(k)}$, say, node O' (the latter part of Assumption A2). Let us regress any variable in $g^{(k)}$, denoted by $\tilde{X}_q^{*(k)}$, on variables in $g^{(1)} \cup g^{(2)} \cup g^{(k-1)} \cup \{Q\}$, where $Q \in g^{(k)}$ and $Q \neq \tilde{X}_q^{*(k)}$. Consider two possible situations.

i) Suppose Q is $\tilde{X}_{\mathrm{NL}}^{*(k)}$, the non-leaf node in the group. Then $\mathrm{PA}(\tilde{X}_q^{*(k)})$ provides a minimal set of predictors with non-zero coefficients–because of assumption A0, when regressing $\tilde{X}_q^{*(k)}$ on this variable set, all coefficients are nonzero. (Again, such a minimal set of predictors may not be unique because of possible deterministic relations among $\tilde{X}_i^*$; however, there does not exist any smaller subset of $g^{(1)} \cup g^{(2)} \cup g^{(k-1)} \cup \{Q\}$ which determines $\tilde{X}_q^{*(k)}$.)

ii) Suppose Q is a leaf node in the group. If $\tilde{X}_q^{*(k)}$ is the non-leaf node, then $\mathrm{PA}(Q) \cup \{Q\} \setminus \{\tilde{X}_{\mathrm{NL}}^{*(k)}\}$ provides a minimal set of predictors with non-zero coefficients. Otherwise, $\mathrm{PA}(\tilde{X}_q^{*(k)}) \cup \{Q\} \cup \mathrm{PA}(Q) \setminus \{\tilde{X}_{\mathrm{NL}}^{*(k)}\}$ provides such a minimal set of predictors.

The cardinality of $\mathrm{PA}(\tilde{X}_q^{*(k)}) \cup \{Q\} \cup \mathrm{PA}(Q) \setminus \{\tilde{X}_{\mathrm{NL}}^{*(k)}\}$ is bigger than or equal to that of $\mathrm{PA}(\tilde{X}_q^{*(k)})$. Now let $\tilde{X}_q^{*(k)}$ be O' and Q be O. According to the latter part of Assumption A2, O has at least a parent that is not a parent of O'. Therefore, The cardinality of $\mathrm{PA}(O') \cup \{O\} \cup \mathrm{PA}(O) \setminus \{\tilde{X}_{\mathrm{NL}}^{*(k)}\}$ is strictly bigger than that of $\mathrm{PA}(O')$. This shows the asymmetry between O and the non-leaf node (as well as possibly some other leaf nodes). So O can be detected as a leaf node.

□

A.4: Proof of Proposition 6

Proof. We note that the ordered group decomposition can be correctly identified from the values of $\mathbf{X}$ as $N \to \infty$, as implied by Proposition 4. Denote by W the non-leaf node in $g^{(k)}$. Let us first find a subset of the nodes causally following $g^{(k)}$ in which each node, denoted by S, is always non-deterministically dependent on at least one of the nodes in $g^{(k)}$ conditional on any subset of the remaining variables in $g^{(1)} \cup g^{(2)} \cup ... \cup g^{(k)}$. Denoted by $\mathbf{S}$ this set of nodes. If Assumptions A0 and A3 holds, $\mathbf{S}$ is not empty.

We then show that the non-leaf node W is always non-deterministically dependent on every node in $\mathbf{S}$ given any subset of $g^{(1)} \cup g^{(2)} \cup ... \cup g^{(k)} \setminus \{W\}$. Suppose this is not the case, i.e., there is $S \in \mathbf{S}$ which is non-deterministically independent from W given a subset of $g^{(1)} \cup g^{(2)} \cup ... \cup g^{(k)} \setminus \{W\}$. Denote by $\mathbf{R}_1$ this subset. If $\mathbf{R}_1$ contains any leaf nodes in $\tilde{G}^*$, let us remove those leaf nodes from $\mathbf{R}_1$ and denote by $\mathbf{R}_1'$ the resulting variable set. Further note that S and W are still de-separated by $\mathbf{R}_1'$. Then U', a leaf node in $g^{(k)}$, is always d-separated from S given $\mathbf{R}_1' \cup (\mathrm{PA}(U') \setminus W)$. Since all nodes in $\mathbf{R}_1' \cup (\mathrm{PA}(U') \setminus W)$ are non-leaf nodes, W can not be represented as their linear combination; thus U' is not their deterministic function. Furthermore, S is not a deterministic function of nodes in $\mathbf{R}_1' \cup (\mathrm{PA}(U') \setminus W)$ either; otherwise, according to the construction procedure of the ordered group decomposition, S will belong to $g^{(1)} \cup g^{(2)} \cup ... \cup g^{(k)}$ because all elements of $\mathbf{R}_1' \cup (\mathrm{PA}(U') \setminus W)$ belong to it. Hence any leaf node U' in $g^{(k)}$ will be non-deterministically independent from S conditional on some subset of $g^{(1)} \cup g^{(2)} \cup ... \cup g^{(k)} \setminus \{U'\}$, so for any node in $g^{(k)}$, there exists some subset of $g^{(1)} \cup g^{(2)} \cup ... \cup g^{(k)}$ given which it is non-deterministically conditionally independent from S. That is, $S \notin \mathbf{S}$, leading to a contradiction.

Next, we show that for leaf node U in $g^{(k)}$, there exists at least one element of $\mathbf{S}$ which is non-deterministically conditionally independent from U given a subset of $g^{(1)} \cup g^{(2)} \cup ... \cup g^{(k)} \setminus \{U\}$. Denote by V one of the nodes that causally follow $g^{(k)}$ and satisfy the two conditions in assumption A3. Because of condition 2), $V \in \mathbf{S}$. Condition 1) states that V and leaf node U are d-separated by a subset of $g^{(1)} \cup g^{(1)} \cup ... \cup g^{(k)} \setminus \{U\}$ that does not include all parents of U. Denote by $\mathbf{R}_2$ this variable set. If $\mathbf{R}_2$ contains any leaf nodes in $\tilde{G}^*$, remove them from $\mathbf{R}_2$ and denote by $\mathbf{R}_2'$ the resulting variable set. V and U are still de-separated by $\mathbf{R}_2'$, but all elements of $\mathbf{R}_2'$ are non-leaf nodes. Because all non-leaf nodes in $\tilde{G}^*$ are linearly independent, the parents of U that are not in $\mathbf{R}_2'$ can not be written as linear combinations of the elements of $\mathbf{R}_2'$. Therefore, U is not a deterministic function of $\mathbf{R}_2'$. Moreover, V is not a deterministic function of $\mathbf{R}_2'$ either, because otherwise V will not in a group causally following $g^{(k)}$. This means that leaf node U is non-deterministically independent from V, as an element of $\mathbf{S}$, given $\mathbf{R}_2'$.

That is, we can distinguish between any leaf node U and the non-leaf node W in the same ordered group by checking non-deterministic conditional independence relationships between $\tilde{X}_i^*$ in the above way.

□

A.5: Proof of Proposition 7

Proof. First note that under the assumptions in the proposition, the ordered group decomposition is identifiable, and all leaf nodes are asymptotically identifiable. The causal ordering among the variables $\tilde{X}_i^*$ is then fully known. The causal graph $\tilde{G}$ can then be readily estimated by regression: for a leaf node, its direct causes are those non-leaf nodes that determine it; for a non-leaf node, we can regress it on all non-leaf nodes that causally precede it according to the causal ordering, and under Assumption A0, those predictors with non-zero linear coefficients are its parents. □