1 Proofs

We first define the following lemma, which we will be using in the later proofs.

Lemma 3. Each constraint \( l_i \) in Lemma 1 can be rewritten in the form of

\[
\rho_{z_i y} W_i \Psi = q_{i1} \theta_1 + \cdots + q_{in} \theta_n,
\]

(1)
such that \( \Psi \) is a function on correlations among variables in \( M \), and each \( q_{il} \) for all \( i = 1, \ldots, n' \) and \( l = 1, \ldots, n \) satisfies the following conditions.

1. If \( \theta_l \) is a directed edge, then \( q_{il} = \sum_{j=0}^{n} b_{ij} a_{ij} l \), where \( a_{ij} l \) and \( b_{ij} \) are defined the same way as \cite{Brito and Pearl, 2012} Equation (10).
2. If \( \theta_l \) is a bidirected edge, then \( q_{il} = b_{i0} \).

The proof of Lemma 3 is given in Section 1.4.

1.1 Proof of Lemma 1

Proof. Given Lemma 3, and \cite{Brito and Pearl, 2012} Section 7.4, we have all those coefficients are functions on the correlations of variables in \( M \).

Note that the functions are not necessarily polynomials, since from the proof of Lemma 3, \( \phi_i \) is a polynomial on correlations, while \( \rho_{z_i y} W_i \) is \( \phi_i \) divided by some functions on the correlations, which results in an arbitrary function.

1.2 Proof of Lemma 2

We prove Lemma 2 together with Theorem 1.

1.3 Proof of Theorem 1

Proof. To prove there exists a full-rank set of \( N = n' + n_k + n_e \) linear constraints on \( E \), we first construct a set of constraints, \( L \), such that \( |L| = N \). Then we prove each of the \( N \) constraints is linear, and finally we show that the set is full-rank.

**Constructing the \( N \) constraints:** We first construct the first \( n' \) constraints. Given a partial-instrumental set \( Z \) for \( E \) on \( E' \), w.l.o.g, denote \( Z = \{z_1, \ldots, z_{n'}\} \), \( E = \{e_1, \ldots, e_n\} \), \( E' = \{e_1, \ldots, e_{n'}\} \). Also denote the triples in the definition of a basic-partial-instrumental set as \((z_1, W_1, p_1) \ldots (z_{n'}, W_{n'}, p_{n'})\). Since each \( p_i \) is a path from \( z_i \) to \( Ta(e_i) \), we can say each \( z_i \) matches to an edge \( e_i \in E' \). From Lemma 1, we can create \( l_i \), which is matched to \( z_i \) and \( e_i \). See Lemma 3.

The left-hand side expression from Equation (1) and \( q_{i1}, \ldots, q_{in} \) can all be calculated from the data. Hence, the first \( n' \) linear equations we construct for \( L \) are Equation (1) for \( i = 1, \ldots, n' \).

Next, we construct the next group of \( n_e \) constraints in \( L \). For each \( j = 1, \ldots, n_e \), we write the \( j \)-th constraint in \( E_e \) as

\[
0 = d_j e_{j1}^e + e_{j2}^e,
\]

(2)

where \( d_j \) is a constant, and \( e_{j1}^e \) and \( e_{j2}^e \) are the two edges involved in this equality constraint. W.l.o.g, we assume for each \( j \), in the \( j \)-th constraint, the first edge, \( e_{j1}^e \), is selected for the selection defined in the theorem.
Finally, we construct the remaining \( n_k \) of the constraints in \( \mathcal{L} \). For each \( h = 1, \ldots, n_k \), the \( h \)-th edge in \( E_k \) is \( e_h^k \), and we have a constraint
\[
\lambda_h = e_h^k,
\]
where \( \lambda_h \) is the known value of \( e_h^k \).

**Constructing a matrix of the constraints** Now that we have a set of \( N = n' + n_e + n_k \) constraints, in order to prove that they are linearly independent, we want to construct a matrix, and prove the matrix is full-row-rank. We first construct an ordering of the edges involved in those constraints.

The first \( n' \) edges are the edges in \( E' \), in the order of \( e_1, \ldots, e_{n'} \). Since there exists a way to non-repetitively select one edge from each equality constraint that certain conditions are satisfied, let the selected edges, \( E_s \), be the next \( n_e \) edges, with the ordering the same as the ordering of the equality constraints in \( \mathcal{L} \). Denote those edges as \( \{e_1^{e_1}, \ldots, e_{n_e}^{e_1}\} \), and the edges that are paired with those edges as \( \{e_2^{e_2}, \ldots, e_{n_e-2}^{e_2}\} \). Next, the last \( n_k \) edges are those in \( E_k \), with the ordering the same as the constraints in \( \mathcal{L} \). Finally, any edges in \( (E \cup E_k) \setminus (E' \cup E_s \cup E_k) \) can be of any order in the end. We can construct this order because as specified in the theorem, \( E', E_s \), and \( E_k \) do not share any element.

Given the ordering of the edges, we can construct a matrix, where each term in the matrix is the coefficient in front of an edge in a constraint. Each row is one constraint in \( \mathcal{L} \), in order, and each column is one edge, in the order we just specified. So we have an \( N \times |E| \) matrix. To prove this matrix is full-row-rank, it suffices to prove the \( N \times N \) sub-matrix containing the first \( N \) columns of the original matrix is full-rank. Below we give what the sub-matrix looks like (the first row in parentheses is used to indicate the edges for the matrix, and is not part of the matrix.)

\[
\begin{pmatrix}
(e_1 & e_2 & \ldots & e_{n'} & e_1^{e_1} & \ldots & e_{n_e}^{e_1} & \ldots & e_1^k & \ldots & e_k^k) \\
q_{11} & q_{12} & \ldots & q_{1n'} & U & \ldots & U & U & \ldots & U \\
\vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots \\
q_{n'1} & q_{n'2} & \ldots & q_{n'n'} & U & \ldots & U & U & \ldots & U \\
0 & \ldots & \ldots & 0 & d_1 & \ldots & \ldots & 1 & \ldots & 0 \\
\vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots \\
0 & \ldots & \ldots & \ldots & 0 & d_{n_e} & \ldots & \ldots & 0 \\
\vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots \\
0 & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & 0 & 1
\end{pmatrix}
\]

\( U \) denotes “unknown”, which might be zero (if the edge corresponding to that column is not in \( E \), or is in \( E \) but not in the constraint corresponding to that row,) or non-zero (if the edge corresponding to that column is in \( E \) and is in the constraint corresponding to that row.)

**Proof that the matrix is full-rank** To prove this matrix is full-rank, we simply have to prove that the determinant does not vanish. The determinant of an \( N \times N \) matrix can be calculated using the Leibniz formula, which is summing up the product of \( N \) entries corresponding to all possible permutations of the set \( \{1, 2, \ldots, N\} \). Hence, we only have to prove that the product we get by selecting the first permutation, i.e., \( \{1, 2, \ldots, N\} \), cannot be canceled by any other products. In other words, we only have to prove that the product of the diagonal of the matrix has a term that cannot be canceled out by any other term from the expression of the determinant.

We define a term, \( T^* \) to be
\[
T^* = \prod_{j=1}^{n'} T(p_j) \prod_{i=1}^{n_e} d_i,
\]
where \( T(p_j) \) is the product of the edge coefficients along the path \( p_j \). \( T^* \) must exist in the product of the diagonal, since \( \prod_{j=1}^{n'} T(p_j) \) exists in the product of the first \( n' \) entries from the diagonal (Lemma 3), and \( \prod_{i=1}^{n_e} d_i \) is the product of the rest of the diagonal entries.
Suppose that $T^*$ appears at least twice in the expression of the determinant. We first prove that $T^*$ must come from selecting the diagonal terms of the matrix.

Note that each selection must select one entry from each row and each column, from the Leibniz formula. We must select the diagonal for the last $n_k$ entries, since if a non-diagonal entry was selected, that entry must be 0, and the whole product would be 0.

Next, we must also select the diagonal for the middle $n_e$ entries, $d_1, \ldots, d_{n_e}$. We prove this argument by proving that if we do not select the diagonal, then we cannot reproduce the product of the $n_e$ diagonal entries no matter what edges we select, which means any term in our selection cannot cancel out $T^*$. Suppose this is not true, i.e., even if we do not select the diagonal entries for the middle $n_e$ rows, we can still get the product somewhere else.

Recall that for each $j = 1, \ldots, n_e$, $d_j e_{j1}^e + e_{j2}^e = 0$. Since this equality constraint should comply with the actual values of the edges $e_{j1}^e, e_{j2}^e$ in the model $M$, we have for each $j$,

$$d_j = -\frac{e_{j2}^e}{e_{j1}^e}. \quad (5)$$

Denote the product of the diagonal entries for the middle $n_e$ rows as $T_m$, then

$$T_m = (-1)^{n_e} \prod_{j=1}^{n_e} \frac{e_{j2}^e}{e_{j1}^e}. \quad (6)$$

Terms cancel out if we have the same edge with one occurrence on the numerator and one occurrence on the denominator. So we might end up having a simplified expression,

$$T_m = (-1)^{n_e} \prod_{i} \frac{e_{i1}^e}{e_{i2}^e}. \quad (7)$$

Note that $T_m$ cannot be $(-1)^{n_e}$, where all edges cancel out. We next examine where those edges might appear in the matrix. First note that the terms in the first $n'$ rows do not contain any edge in $Inc(y)$ (Lemma 3).

$b_{ij}$ are polynomials on the correlations among $z_i, W_{i_1}, \ldots, W_{i_k}$ and $a_{i,j} = \rho_{W_{i}, z_i}$. All edges in $Inc(y)$ have a head $y$, which means no edge in $Inc(y)$ can appear in the correlations among the non-descendants of $y$ (this can be seen from Wrights’ rules.) So $q_{i,l}$, which is made up of correlations among $W_{i_j}, x_i, z_i$ (all non-descendants of $y$) does not contain any edge in $Inc(y)$.

Hence, the edges in $T_m$ cannot be canceled out by anything in the first $n'$ rows, which means $T^*$ will contain $T_m$ as it is.

Suppose we select only a subset of $n'_e$ the diagonal entries for the middle $n_e$ rows. For each row where the diagonal is not selected, 1 must be selected (otherwise we will have to select 0, and the product will be 0.) So we end up having the product of the selected entries from the middle $n_e$ rows, $T'_m$ as

$$T'_m = (-1)^{n'_e} \prod_{i} \frac{e_{i1}^{e'}}{e_{i2}^{e'}}. \quad (8)$$

$T_m$ and $T'_m$ cannot be equal to each other. Otherwise, we produce a constraint on those edges by equating $T_m$ and $T'_m$. However, given that the equality constraints are linearly independent, the values of those edges should vary independently and should not comply to any constraint. In other words, they are equal only when the constraint is satisfied, which has Lebesgue measure 0, so we assume that is not the case. Thus, to cancel out $T^*$, we must select the diagonal of the middle $n_e$ rows.

We have proved that for the last $n_e + n_k$ rows, we must select the entries on the diagonal. For the first $n'$ rows, we can only select from the first $n'$ columns, since we can only select one entry from each column, and the last $n_e + n_k$ columns already have entries been selected. Therefore, we only need to analyze the top left $n' \times n'$ submatrix. The problem reduces to proving the term

$$t^* = \prod_{j=1}^{n'} T(p_j) \quad (9)$$
exists only once in the determinant of this submatrix. We first prove that $t^*$ appears only once in the product of the diagonal entries. We use the same proof strategy as in [Brito and Pearl, 2012] Proof of Lemma 8. To get $t^*$, we need to select one term from each diagonal entry such that the product of those terms gives $t^*$. From [Brito and Pearl, 2012] Proof of Lemma 8, for $q_{ij}$ where column $j$ is a directed edge, if we select the second or the third term of $q_{ij}$ in [Brito and Pearl, 2012] Equation (11), then it must bring in a term that is not in $t^*$, or causes the product to contain a term in $t^*$ twice. Hence, for those $q_{ij}$ entries, we can only select from the first term in Equation in [Brito and Pearl, 2012] Equation (11). After eliminating those terms from consideration, the remaining terms in the product of the $n'$ diagonal terms are given by

$$
t^* \prod_{i \text{ directed}} (1 + \hat{b}_{i0}) \prod_{k \text{ bidirected}} (1 + \hat{b}_{k0}) = t^* \prod_j (1 + \hat{b}_{j0}), \tag{10}
$$

From [Brito and Pearl, 2012], $\hat{b}_{j0}$ are polynomials on correlations among $W_i$, and they do not have any constant terms. As a result, $t^*$ appears only once in Equation (11), and thus appears only once in the product of the diagonal entries.

What remains to prove is that $t^*$ does not appear in the product of another selection of entries, which is different from selecting all the diagonals. For the columns that correspond to bidirected edges, we have to select the diagonal terms, since those are the only terms in those column that are non-zero. We generate a submatrix by removing those columns corresponding to bidirected edges and those rows with the same row numbers as those column numbers. This submatrix is a square matrix, and all columns correspond to directed edges. This reduces to the proof of Theorem 1 from [Brito and Pearl, 2012], where they proved that no matter which selection we have, the term $\prod_{j \text{ directed}} T(p_j)$ can never be canceled.

To sum up, we showed that one can never find another term in the determinant that can cancel out a term, $T^*$, which is also in the determinant. Hence, the $N \times N$ sub-matrix is full-rank, and the $N \times E$ matrix is full-row-rank.

Finally, when $N = |E|$, we have a full-rank set of $N$ linear equations on $N$ edges, so we can solve for all of the edges. \square

### 1.4 Proof of Lemma 3

**Proof.** From Lemma 1 in [Brito and Pearl, 2012], denoting $W_i = \{W_{i_1}, \ldots, W_{i_k}\}$ (we assume $W_i$ contains $k$ single variables), we have

$$
\rho_{z, y_i W_i} = \frac{\phi_i(z_i, y, W_{i_1}, \ldots, W_{i_k})}{\psi_i(z_i, W_{i_1}, \ldots, W_{i_k})}, \tag{12}
$$

where $\phi$ is linear on the correlations $\rho_{z, y_i}, \rho_{W_{i_1}, y}, \ldots, \rho_{W_{i_k}, y}$, and the square of each of the $\psi$ functions is a polynomial on correlations among the variables it takes. We can write

$$
\phi_i = b_{i0} \rho_{z, y_i} + b_{i1} \rho_{W_{i_1}, y} + \cdots + b_{in} \rho_{W_{i_k}, y}. \tag{13}
$$

We only need to prove that $\phi_i$ is linear on the edges $e_1, \ldots, e_n$ and does not contain any constant term. Since $\rho_{z, y_i W_i}$ vanishes in $G_{E \cap D \cup \{e_i\}}$, from the definition of a partial-instrumental set, $\phi(z_i, y, W_{i_1}, \ldots, W_{i_k})$ must also vanish in $G_{E \cap D \cup \{e_i\}}$. For all bidirected edges in $Inc(y)$, we can treat them as two directed sub-edges connected at the tails. Hence, [Brito and Pearl, 2012]’s Lemmas 6 and 7 apply. Let $e_i'$ be the same as $e_j$ if $e_j$ is directed, and the sub-edge pointing to $y$ if $e_j$ is bidirected, and we immediately have that $\phi_i$ is linear on the edges $e_1', \ldots, e_n'$ and does not contain any constant term. If $e_j$ is bidirected, $\phi_i$ being linear on $e_i'$ is equivalent to that $\phi_i$ is linear on $e_j$. From Lemma 7, we have that all edges not in $E \cap D \cup \{e_j\}$ have coefficient 0. Hence, either $e_{z, y_i}$ is the only bidirected edge in the constraint $l_i$, or there exists no bidirected edge in $l_i$.

$\rho_{z, y_i W_i}$ can be written in the form of $\rho_{z, y_i W_i} = c_{i1} e_1 + \cdots + c_{in} e_n$. We then apply the results from Section 7.4 in [Brito and Pearl, 2012] and we have for each $j$ where $e_j$ is a directed edge, $c_{ij}$ is a function of the correlations of variables in $M$. 
If there does not exist a bidirected edge among \( \theta_1, \ldots, \theta_n \), then the lemma is evident from the result from [Brito and Pearl, 2012]. If there exists a bidirected edge, w.l.o.g, assume \( \theta_n \) is the bidirected edge. Now we examine every \( q_{ij} \).

First we can decompose \( \theta \) into two directed edges, one pointing to \( y \) and one does not include \( y \). Let the decomposition be \( \theta_n = \alpha \beta \), where \( \alpha \) is the edge pointing to \( y \). We can thus write

\[
\rho_{z,y} W_i = q_{i1} \theta_1 + \cdots + q_{i(n-1)} \theta_{n-1} + q_{in} \beta \alpha.
\]

Now we have a linear equation on directed edges \( \theta_1, \ldots, \theta_{n-1}, \alpha \). Hence, the results from [Brito and Pearl, 2012] applies, and we know for \( j \) where \( \theta_j \) is a directed edge, \( q_{ij} \) is the same as the way defined in [Brito and Pearl, 2012].

The coefficient of \( \alpha \) can also be regarded as \( \sum_{j=0}^{n} b_{ij} a_{ij,n} \). Recall the definition in [Brito and Pearl, 2012], each \( a_{ij,n} \) is the sum of paths from \( z_i \) or \( W_i \) to \( y \) passing through \( \theta_n \), but not including \( \theta_n \). From Definition 4, each \( W_{ij} \) is non-descendant of \( z_i \), so any unblocked path from \( W_{ij} \) to \( z_i \) must have an arrowhead at \( z_i \), which makes \( z_i \) a collider (also named as “sink” or “convergent”) between \( W_{ij} \) and \( y \), and blocks the path between \( z_i \) and \( y \). Since no paths from other \( z_i \) or \( W_i \) can pass through the bidirected edge, the only non-zero \( a_{ij,n} \) is \( a_{i0,n} \), which is the sum of paths from \( z_i \) to \( y \) through \( \theta_n \) but not including \( \theta_n \), which is equal to 1. The corresponding multiplier is \( b_{i0} \). Since the index of the bidirected edge among \( \theta_1, \ldots, \theta_n \) does not matter, we assumed the bidirected edge is of index \( n \) for the convenience of discussion. Now we can replace \( n \) with \( l \) and we have the coefficient

\[
q_{il} = b_{i0} \cdot 1 = b_{i0}.
\]

### 2 Discussion on the Example in Section 7.1

In Figure 3 left, if the equality constraint is instead \( \lambda_{ux} = \lambda_{uw} \), then the equality constraint in the latent projection DAG is \( \varepsilon_{xy} = \varepsilon_{wy} \). \( \{w, x\} \) form a partial-instrumental set for \( \{\varepsilon_{wy}, \varepsilon_{xy}, \lambda_{xy}\} \) on \( \{\varepsilon_{wy}, \lambda_{xy}\} \). Together with the equality constraint, we can solve for all edges.

If the equality constraint is instead \( \lambda_{ux} = \lambda_{uy} \), then the equality constraint in the latent projection DAG is \( \varepsilon_{xw} = \varepsilon_{wy} \). \( \varepsilon_{xw} \) is identified (\( \varepsilon_{xw} = \rho_{xw} \)). Then with the equality constraint, \( \varepsilon_{wy} \) is identified. \( \varepsilon_{xy} = \varepsilon_{wy} \). \( \{w, x\} \) form a partial-instrumental set for \( \{\varepsilon_{wy}, \varepsilon_{xy}, \lambda_{xy}\} \) on \( \{\varepsilon_{xy}, \lambda_{xy}\} \). Together with the value of \( \varepsilon_{wy} \), we can solve for all edges.

### References