Technical Proof for ANOCE: Analysis of Causal Effects with Multiple Mediators via Constrained Structural Learning

1 Connection to Literature

We establish the connection between our proposed method to the literature from three different angles. First, we show that the individual mediation effect defined in Chakrabortty et al. [2] can be decomposed into our defined DM and IM when the LSEM assumption holds. Next, we give an equivalent definition of the DM through a type of special edge (last edge) in the causal graph. Lastly, we prove that the proposed DM is consistent with the interventional effect via a particular mediator defined in Vansteelandt and Daniel [5] under the LSEM.

1.1 From Individual Mediation Viewpoint

Chakrabortty et al. [2] defined the individual mediation effect under the LSEM as follows.

Definition 1.1 *[2] Individual mediation effect for M_i:*

$$\begin{aligned}\eta_i =&\Big[E\{M_i|do(A=a+1)\}-E\{M_i|do(A=a)\}\Big]\\ &\times\Big[E\{Y|do(M_i=m_i+1)\}-E\{Y|do(M_i=m_i)\}\Big].\end{aligned} \tag{1}$$

In the following theorem, we show that the summation of η_i is strictly larger than the IE if the mediators are not parallel. The proof is given in Section 2.2.

Theorem 1.1 *If there exists at least one directed path $\pi^* \in \{\pi_{AY}(\mathcal{G})\}$ such that the length of π^* is larger than 2, and the element in B is nonnegative, then*

$$\sum \eta_i > IE. \tag{2}$$

Remark 1.1 *From the above theorem, it is clear that the mediator effect defined in Chakrabortty et al. [2] is not appropriate for interpreting the decomposition of the indirect effect, when there exists interaction among mediators (a common situation as described in the introduction). Here, we keep the condition that the element in B is nonnegative, as the multiple count mediation effects in Chakrabortty et al. [2] may cancel out in some cases and their summation would equal to the IE by chance.*

Inspired by the proof of Theorem 1.1, the mediator effect η_i can be decomposed into two parts, the natural direct and indirect effect for i-th mediator, as shown in the following corollary.

Corollary 1.1 *Under assumptions (A1-A3) and Model 1, we have*

$$\eta_i = DM_i + IM_i. \tag{3}$$

Remark 1.2 *Corollary 1.1 together with the definition of η_i in Chakrabortty et al. [2] provides a feasible way to numerically calculate the natural indirect effect IM_i. Specifically, by deleting the mediator M_i in the causal graph, the reduced treatment effect corresponds to η_i, then $IM_i = \eta_i - DM_i$, where the explicit expression of the DM_i is provided in Theorem 3.1. See more implementation details in Section 4.*

1.2 From Graphical Perspective

Next, we give the definition of the edge-specific effect following Avin et al. [1]. Suppose a directed edge of interest as $X_i \to X_j$ in a weighted DAG $\mathcal{G}$. Define a new weighted DAG $\mathcal{G}'_{i,j}$ by deleting the directed edge $X_i \to X_j$ in $\mathcal{G}$, i.e., $\mathcal{G}'_{i,j} \equiv \mathcal{G} \setminus (X_i \to X_j)$.

Definition 1.2 *[1] Edge-specific effect:*

$$ET(X_i, X_j) = TE_{\mathcal{G}} - TE_{\mathcal{G}'_{i,j}}, \tag{4}$$

where $TE_{\mathcal{G}}$ means the total effect in graph $\mathcal{G}$.

We next give an equivalent definition of our proposed DM from a graphical perspective. Let the edge in $\mathcal{G}$ that starts with i-th mediator and ends with node Y, i.e., $M_i \to Y$, as the i-th last edge. Denote the graph $\mathcal{G}$ deleting the i-th last edge $(M_i \to Y)$ as $\mathcal{G}'_i$. We define the ith last edge-specific effect as

Definition 1.3 *Last edge-specific effect for M_i:*

$$LE_i = \begin{cases} TE_{\mathcal{G}} - TE_{\mathcal{G}'_i}, & \text{if there exists edge } M_i \to Y \text{ in } \mathcal{G}; \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

By Theorem 3.1, we have $(I_p - B_M^\top)^{-1}\boldsymbol{\alpha}$ is the causal effect of A on M. Let $\zeta \equiv (I_p - B_M^\top)^{-1}\boldsymbol{\alpha}$, with its i-th element $\zeta_i \equiv \{(I_p - B_M^\top)^{-1}\boldsymbol{\alpha}\}_i$. Next, we show that the i-th last edge-specific effect can be presented as $\boldsymbol{\beta}_i\zeta_i$ under the LSEM in the following theorem, where $\boldsymbol{\beta}_i$ is the i-th element of the vector $\boldsymbol{\beta}$ and corresponds to the weight of the edge $M_i \to Y$. The proof can be found in Section 2.3.

Theorem 1.2 *Under assumptions (A1-A3) and Model 1, we have*

$$LE_i = \boldsymbol{\beta}_i\zeta_i. \tag{6}$$

Based on Theorem 3.1, the natural direct effect of M_i on Y can be expressed as $DM_i = \boldsymbol{\beta}_i\zeta_i$. Thus, with the result of Theorem 1.2, it is easy to show the following corollary.

Corollary 1.2 *Under assumptions (A1-A3) and Model 1, the natural direct effect of M_i is equal to the i-th last edge-specific effect:*

$$LE_i = DM_i = \boldsymbol{\beta}_i \zeta_i. \tag{7}$$

Remark 1.3 *Here, both definitions describe the direct impact of one mediator M_i on the outcome. The natural direct effect of a particular mediator M_i can be understood as the influence when removing the direct edge between M_i and Y. Thus, we have the equivalence between two definitions.*

Then, we can decompose the total natural indirect effect into p last edge-specific effects or p DMs as the following additive form, based on Theorem 3.2 and Corollary 1.2.

Corollary 1.3 *Under assumptions (A1-A3) and Model 1, we have*

$$IE = \boldsymbol{\beta}^\top \zeta = \sum_{i=1}^{p} \boldsymbol{\beta}_i \zeta_i = \sum_{i=1}^{p} DM_i = \sum_{i=1}^{p} LE_i. \tag{8}$$

In fact, based on the uniqueness of each last edge, the natural indirect effect can be decomposed into p last edge-specific effect regardless of the LSEM setting through the graphical perspective. We give the following intuitive conclusion. The proof can be found in Section 2.4.

Theorem 1.3 *The IE can be decomposed through LEs as:*

$$IE = \sum_{i=1}^{p} LE_i. \tag{9}$$

Remark 1.4 *One can view the last edge-specific effect as the generalized definition of the natural direct effect for mediator without the LSEM assumption.*

1.3 From Interventional Effect Level

Finally, we show the consistency of our defined DM to the interventional effect via a particular mediator defined in Vansteelandt and Daniel [5] under the LSEM.

Definition 1.4 *[5] Under assumptions (A1-A3), the interventional effect via M_i is*

$$\begin{aligned} \xi_i = \sum_{m_1 \in \mathbb{M}_1} \cdots \sum_{m_p \in \mathbb{M}_p} \Bigg[& E(Y|A=a, M_i=m_i, \Omega_i=o_i) P(\Omega_i=o_i|A=a) \\ & \times \Big\{ P(M_i=m_i|A=a+1) - P(M_i=m_i|A=a) \Big\} \Bigg], \end{aligned} \tag{10}$$

where $\mathbb{M}_i$ is the support of M_i, $o_i = [m_1, \cdots, m_{i-1}, m_{i+1}, \cdots, m_p]$, $P(M = m|A = a)$ is the probability of $M = m$ when setting $A = a$.

Theorem 1.4 *Under assumptions (A1-A3) and Model 1, we have*

$$DM_i = \xi_i,$$

Remark 1.5 *The proof can be found in Section 2.5. Based on Definition 3.2 and Equation 10, both the proposed DM and the effect defined in Vansteelandt and Daniel [5] contain the information of the causal effect of A on the mediator M_i, i.e., $P(M_i = m_i|A = a + 1) - P(M_i = m_i|A = a)$.*

2 Technical Proofs

2.1 Proof of Theorem 3.1

Proof 2.1 *In this proof, we will give the explicit expressions of causal effects defined under the LSEM. First, Equation 3 is equivalent to*

$$\begin{cases} A \equiv \epsilon_A, \\ M = \boldsymbol{\alpha} A + B_M^\top M + \epsilon_M, \\ Y = \gamma A + \boldsymbol{\beta}^\top M + \epsilon_Y. \end{cases} \tag{11}$$

Based on $M = \boldsymbol{\alpha} A + B_M^\top M + \epsilon_M$*, by moving* $B_M^\top M$ *to the left-hand side, we have*

$$(I_p - B_M^\top) M = \boldsymbol{\alpha} A + \epsilon_M.$$

Suppose the mediators are sorted in the topological order (a series of elementary transformation of the matrix), then the matrix $B_M^\top$ *is strictly upper triangular with the diagonal element as 0. Thus, we have* $I_p - B_M^\top$ *is invertible, then* $I_p - B_M^\top$ *under its original order should be also invertible (any invertible matrix after elementary transformation is still invertible).*

Therefore, we can rewrite M *as a purely function of* A *plus the error term as follows.*

$$M = (I_p - B_M^\top)^{-1} \boldsymbol{\alpha} A + (I_p - B_M^\top)^{-1} \epsilon_M. \tag{12}$$

Then we replace mediators in Equation 11 with Equation 12 and obtain

$$\begin{cases} A &\equiv \epsilon_A, \\ M &= (I_p - B_M^\top)^{-1} \boldsymbol{\alpha} A + (I_p - B_M^\top)^{-1} \epsilon_M, \\ Y &= \gamma A + \boldsymbol{\beta}^\top M + \epsilon_Y \\ &= \gamma A + \{\boldsymbol{\beta}^\top (I_p - B_M^\top)^{-1} \boldsymbol{\alpha}\} A + \{\boldsymbol{\beta}^\top (I_p - B_M^\top)^{-1} \epsilon_M + \epsilon_Y\}. \end{cases} \tag{13}$$

Next, we show how to get the explicit expressions of $E\{Y|do(A=a)\}$ *under the LSEM. Following the results in Rosenbaum and Rubin [4], under the assumption (A2), we have* $P\{M|do(A=a)\} = P(M|A=a)$*, and thus,*

$$E\{M|do(A=a)\} = E(M|A=a).$$

Similarly, we can get $E\{Y|do(A=a)\} = E(Y|A=a)$ *under the assumption (A1), and* $E\{Y|do(A=a, M=m)\} = E(Y|A=a, M=m)$ *under the assumption (A3).*

Based on above results and Equation 13, we have

$$\begin{aligned} & E\{Y|do(A=a)\} = E\{Y|A=a\} \\ =& E\{\gamma A + \boldsymbol{\beta}^\top M + \epsilon_Y | A=a\} \\ =& \gamma a + \boldsymbol{\beta}^\top E\{M|A=a\} \\ =& \gamma a + \boldsymbol{\beta}^\top E\{(I_p - B_M^\top)^{-1} \boldsymbol{\alpha} A + (I_p - B_M^\top)^{-1} \epsilon_M | A=a\} \\ =& \gamma a + \boldsymbol{\beta}^\top (I_p - B_M^\top)^{-1} \boldsymbol{\alpha} a, \end{aligned} \tag{14}$$

where the first '=' is held under the assumption (A1), the second and forth '=' are given by Equation 13 that $Y = \gamma A + \boldsymbol{\beta}^\top M + \epsilon_Y$ *and* $M = (I_p - B_M^\top)^{-1}\boldsymbol{\alpha}A + (I_p - B_M^\top)^{-1}\epsilon_M$.

Following the same calculation procedure of $E\{Y|do(A=a)\}$*, we next give the natural direct effect under assumptions (A1-A3) and Model 1 as*

$$\begin{aligned} DE &= E\{Y|do(A=a+1, M=m^{(a)})\} - E\{Y|do(A=a)\} \\ &= \{\gamma(a+1) + \boldsymbol{\beta}^\top m^{(a)}\} - \{\gamma a + \boldsymbol{\beta}^\top m^{(a)}\} \\ &= \gamma, \end{aligned}$$

where the first '=' is given by the definition of the DE.

Similarly, the natural indirect effect is

$$\begin{aligned} IE &= E\{Y|do(A=a, M=m^{(a+1)})\} - E\{Y|do(A=a)\} \\ &= \{\gamma a + \boldsymbol{\beta}^\top m^{(a+1)}\} - \{\gamma a + \boldsymbol{\beta}^\top m^{(a)}\} \\ &= \boldsymbol{\beta}^\top (I_p - B_M^\top)^{-1}\boldsymbol{\alpha}(a+1) - \boldsymbol{\beta}^\top (I_p - B_M^\top)^{-1}\boldsymbol{\alpha}a \\ &= \boldsymbol{\beta}^\top (I_p - B_M^\top)^{-1}\boldsymbol{\alpha}. \end{aligned}$$

Thus, the total effect of A *on* Y *is*

$$TE = E\{Y|do(A=a+1)\} - E\{Y|do(A=a)\} = DE + IE = \gamma + \boldsymbol{\beta}^\top (I_p - B_M^\top)^{-1}\boldsymbol{\alpha}.$$

Finally, we give the expression for the natural direct effect of M_i *on* Y *under the LSEM. Based on the assumption (A2) and Equation 12, we have*

$$\begin{aligned} & E\{M_i|do(A=a+1)\} - E\{M_i|do(A=a)\} \\ =& E\{M_i|A=a+1\} - E\{M_i|A=a\} \\ =& \{(I_p - B_M^\top)^{-1}\boldsymbol{\alpha}\}_i(a+1) - \{(I_p - B_M^\top)^{-1}\boldsymbol{\alpha}\}_i a \\ =& \{(I_p - B_M^\top)^{-1}\boldsymbol{\alpha}\}_i, \end{aligned} \tag{15}$$

where $\{(I_p - B_M^\top)^{-1}\boldsymbol{\alpha}\}_i$ *is the* i*-th element of the vector* $(I_p - B_M^\top)^{-1}\boldsymbol{\alpha}$.

Then, based on $Y = \gamma A + \boldsymbol{\beta}^\top M + \epsilon_Y$ *and the assumption (A3), we have,*

$$\begin{aligned} & E\{Y|do(A=a, M_i = m_i^{(a)} + 1, \Omega_i = o_i^{(a)})\} - E\{Y|do(A=a)\} \\ =& E\{Y|A=a, M_i = m_i^{(a)} + 1, \Omega_i = o_i^{(a)}\} - E\{Y|A=a\} \\ =& \gamma a + \boldsymbol{\beta}^\top \begin{bmatrix} m_1^{(a)} \\ \vdots \\ m_i^{(a)} + 1 \\ \vdots \\ m_p^{(a)} \end{bmatrix} - \gamma a - \boldsymbol{\beta}^\top \begin{bmatrix} m_1^{(a)} \\ \vdots \\ m_i^{(a)} \\ \vdots \\ m_p^{(a)} \end{bmatrix} = \boldsymbol{\beta}^\top \boldsymbol{1}_i = \boldsymbol{\beta}_i, \end{aligned}$$

where $\boldsymbol{1}_i$ *is a* $p \times 1$ *vector with the* i*-th element as 1 while others qual to 0, and* $\boldsymbol{\beta}_i$ *is the* i*-th element of the vector* $\boldsymbol{\beta}$.

Thus, we have

$$\begin{aligned}
DM_i =&\Big[E\{M_i|do(A=a+1)\} - E\{M_i|do(A=a)\}\Big] \\
&\times \Big[E\{Y|do(A=a, M_i=m_i^{(a)}+1, \Omega_i=o_i^{(a)})\} - E\{Y|do(A=a)\}\Big], \\
=&\{(I_p - B_M^\top)^{-1}\boldsymbol{\alpha}\}_i \times \boldsymbol{\beta}_i \\
=&\boldsymbol{\beta}_i\{(I_p - B_M^\top)^{-1}\boldsymbol{\alpha}\}_i. \qquad \square
\end{aligned}$$

2.2 Proof of Theorem 1.1

Proof 2.2 *1. If there is no directed path $\pi^* \in \{\pi_{AY}(\mathcal{G})\}$ such that the length of π^* is larger than 2, i.e., the length of $\pi^* \in \{\pi_{AY}(\mathcal{G})\}$ is either 1 or 2. Here, the path with length 1 corresponds to $A \to Y$, and paths with length 2 are $A \to M_i \to Y$ with possibly $i = 1, \cdots, p$. Thus, there is no interaction among mediators.*

*By the definition of the LSEM, we have $B_M = \boldsymbol{0}_{p\times p}$, where $\boldsymbol{0}_{p\times p}$ is a $p\times p$ zero matrix. Following the path method (the causal effect of X_i on X_j along a directed path from $X_i \to X_j$ in $\mathcal{G}$ can be calculated by multiplying all edge weights along the path) illustrated in Wright [6] and Nandy et al. [3], we could obtain $\sum \eta_i = \sum_i \boldsymbol{\beta}_i\boldsymbol{\alpha}_i = IE$. (See a toy example provided in section **??** to illustrate how to use the path method to manually compute the causal effects.)*

2. If there exists at least one directed path $\pi^* \in \{\pi_{AY}(\mathcal{G})\}$ such that the length of π^* is larger than 2, and the element in B is nonnegative, we have $B_M \neq \boldsymbol{0}_{p\times p}$. Without loss of generality, suppose there exists $M_i \in M$ with a set of directed path that starts with A, contains M_i, then goes through other mediators, and ends with Y, denoted each path in such set as $\pi_{i,j} = \{A \to \cdots \to M_i \cdots \to \cdots \to Y\}$ for $j = 1, \cdots, n_i$, where n_i is the size of such path set for M_i, and the weights of edges in $\pi_{i,j}$ is positive. Note the set $\{\pi_{i,j}\}$ excludes the paths end with $M_i \to Y$.

Let $e_{\pi_{i,j}}$ denote the causal effect of A on Y through directed path $\pi_{i,j}$. Based on the path method in Wright [6] and Nandy et al. [3] with the definition of IM_i, we have its theoretical form as

$$IM_i = \sum_{j=1}^{n_i} e_{\pi_{i,j}}. \tag{16}$$

By Equation 15 and the definition of η_i, we have its first multiplier as

$$E\{M_i|do(A=a+1)\} - E\{M_i|do(A=a)\} = \{(I_p - B_M^\top)^{-1}\boldsymbol{\alpha}\}_i,$$

which is also the first multiplier in both DM_i and IM_i.

And the second multiplier of η_i can be expressed as

$$\begin{aligned}
&E\{Y|do(M_i=m_i+1)\} - E\{Y|do(M_i=m_i)\} \\
=&E\{Y|do(M_i=m_i^{(a)}+1)\} - E\{Y|do(M_i=m_i^{(a)})\} \\
=&E\{Y|do(A=a, M_i=m_i^{(a)}+1)\} - E\{Y|do(A=a, M_i=m_i^{(a)})\}, \\
=&E\{Y|do(A=a, M_i=m_i^{(a)}+1)\} - E\{Y|do(A=a)\},
\end{aligned}$$

where $m_i^{(a)}$ is the value of M_i when setting $do(A=a)$. Here, the first '=' is valid since m_i can be arbitrary number, and the second and third '=' are based on the equivalent interventions.

Based on the technique of plus and minus the same term, we decompose the second multiplier of η_i into two parts as follows

$$\begin{aligned}
&E\{Y|do(A=a,M_i=m_i^{(a)}+1)\}-E\{Y|do(A=a)\}\\
=&\underbrace{\Big[E\{Y|do(A=a,M_i=m_i^{(a)}+1,\Omega_i=o_i^{(a)})\}-E\{Y|do(A=a)\}\Big]}_{\textit{the second multiplier of } DM_i}\\
&+\underbrace{\Big[E\{Y|do(A=a,M_i=m_i^{(a)}+1)\}-E\{Y|do(A=a,M_i=m_i^{(a)}+1,\Omega_i=o_i^{(a)})\}\Big]}_{\textit{the second multiplier of } IM_i}
\end{aligned}\tag{17}$$

where $\Omega_i = M \setminus M_i$ is the sets of mediators except M_i, and $o_i^{(a)}$ is the value of Ω_i when setting $do(A=a)$. Here, the first term in the above equation corresponds to the second multiplier of DM_i, while the second term is the second multiplier of IM_i.

Thus, the summation of η_i is

$$\begin{aligned}
\sum\eta_i=&\sum_i\Bigg\{\Big[E\{M_i|do(A=a+1)\}-E\{M_i|do(A=a)\}\Big]\\
&\times\Big[E\{Y|do(M_i=m_i+1)\}-E\{Y|do(M_i=m_i)\}\Big]\Bigg\}\\
=&\sum_i\{DM_i+IM_i\}=\sum_i DM_i+\sum_i IM_i=IE+\sum_i\sum_{j=1}^{n_i}e_{\pi_{i,j}},
\end{aligned}$$

where the first '=' is from Definition 1, the second '=' is given by Equation 17 and Definition 3.2 and 3.3, and the last '=' comes from Theorem 3.2 and the theoretical form of IM in Equation 16.

Here, we have $e_{\pi_{i,j}}>0$ since the weights of edges in $\pi_{i,j}$ is positive based on the path method in Wright [6] and Nandy et al. [3]. Then, $\sum_i\sum_{j=1}^{n_i}e_{\pi_{i,j}}$ is also strictly larger than 0. Therefore, we have

$$\sum\eta_i>IE. \qquad \square$$

2.3 Proof of Theorem 1.2

Proof 2.3 *1. If there doesn't exist edge $M_i\to Y$ in $\mathcal{G}$, then by definition we have $\boldsymbol{\beta}_i=0$. Thus, $LE_i=\boldsymbol{\beta}_i\zeta_i=0$.*

2. If there exists edge $M_i\to Y$ in $\mathcal{G}$. Suppose there is a directed path set with size m_i associated to the edge $M_i\to Y$, where each directed path $\tilde{\pi}_{i,j}$ starts with node A and ends with $M_i\to Y$, denoted as $\tilde{\pi}_{i,j}=\{A\to\cdots\to\cdots\to M_i\to Y\}$ for $j=1,\cdots,m_i$.

Let $e_{\tilde{\pi}_{i,j}}$ denote the causal effect of A on Y through directed path $\tilde{\pi}_{i,j}$, $e_{\tilde{\pi}_{i,j}}^{(A,M_i)}$ be the causal effect of A on M_i through directed path $\tilde{\pi}_{i,j}$, and $e^{(M_i,Y)}$ is the causal effect of M_i on Y through

directed edge $M_i \to Y$. Following the path method in Wright [6] and Nandy et al. [3], we have $e_{\tilde{\pi}_{i,j}} = e^{(A,M_i)}_{\tilde{\pi}_{i,j}} e^{(M_i,Y)}$.

Then the i-th last edge-specific effect is equal to the summation of the effect through each path $\tilde{\pi}_{i,j}$, i.e.,

$$LE_i = \sum_{j=1}^{n_i} e_{\tilde{\pi}_{i,j}} = \sum_{j=1}^{n_i} e^{(A,M_i)}_{\tilde{\pi}_{i,j}} e^{(M_i,Y)} = e^{(M_i,Y)} \sum_{j=1}^{n_i} e^{(A,M_i)}_{\tilde{\pi}_{i,j}}.$$

Here, by the similar argument based on the path method, we have $e^{(M_i,Y)} = \boldsymbol{\beta}_i$ and $\sum_{j=1}^{n_i} e^{(A,M_i)}_{\tilde{\pi}_{i,j}}$ as the total causal effect of A on M_i.

Recall that $\zeta_i \equiv \{(I_p - B_M^\top)^{-1}\boldsymbol{\alpha}\}_i$ is the causal effect of A on M_i. Therefore, the i-th LE is the product of the causal effect of A on M_i and the causal effect of M_i on Y, i.e.,

$$LE_i = \boldsymbol{\beta}_i \zeta_i. \qquad \square$$

2.4 Proof of Theorem 1.3

Proof 2.4 *Given a general DAG $\mathcal{G}$ with nodes $\{A, M, Y\}$, let the union of all directed paths that contain the i-th last edge as $\tau_i = \{\pi : A \to \cdots \to M_i \to Y\}, i = 1, \cdots p$. Here, we have $\tau_i = \{\tilde{\pi}_{i,j}\}_{1 \le j \le m_j}$ established in Section 2.3. It is clear that the union set of τ_i in $\mathcal{G}$ is equal to the set of all directed paths start with A and end with node Y (except $A \to Y$) in $\mathcal{G}$ as*

$$\bigcup_i \tau_i = \{\pi_{AY}(\mathcal{G})\} \setminus \{A \to Y\}.$$

Also, based on the uniqueness of each last edge, τ_i is pairwise disjoint, i.e.

$$\tau_i \bigcap \tau_j = \varnothing, \qquad \forall i \neq j.$$

Since the IE is defined as the total causal effect of A on Y that goes through mediators, we have the IE equal to the causal effect that goes through the set $\{\pi_{AY}(\mathcal{G})\} \setminus \{A \to Y\}$, i.e., the IE equal to the causal effect that goes through set $\bigcup_i \tau_i$. Based on the mutual disjoint property of τ_i, we have the IE is exactly the summation of the causal effect through τ_i. Lastly, from the definition of LE_i, we have

$$IE = \sum_{i=1}^{p} LE_i. \qquad \square$$

2.5 Proof of Theorem 1.4

Proof 2.5 *The proof of the consistency of our defined DM to the interventional effect ξ_i can be completed based on Equation 3 under assumptions (A1-A3) and Model 1.*

Recall the definition in Equation 10, we have

$$\begin{aligned}\xi_i = \sum_{m_1\in\mathbb{M}_1}\cdots\sum_{m_p\in\mathbb{M}_p}\Bigg[&E(Y|A=a,M_i=m_i,\Omega_i=o_i)P(\Omega_i=o_i|A=a)\\ &\times\Big\{P(M_i=m_i|A=a+1)-P(M_i=m_i|A=a)\Big\}\Bigg].\\ =\sum_{m_1\in\mathbb{M}_1}\cdots\sum_{m_p\in\mathbb{M}_p}\Big\{&E(Y|A=a,M_i=m_i,\Omega_i=o_i)P(\Omega_i=o_i|A=a)P(M_i=m_i|A=a+1)\\ &-E(Y|A=a,M_i=m_i,\Omega_i=o_i)P(\Omega_i=o_i|A=a)P(M_i=m_i|A=a)\Big\}.\end{aligned}$$

Given $A=a$, the value of M_i is $m_i^{(a)}$ and Ω_i takes $o_i^{(a)}$; while when setting $A=a+1$, the value of M_i is $m_{a+1}^{(i)}$. Therefore, we have $P(M_i=m_i|A=a)=1$ if $m_i=m_i^{(a)}$ otherwise is 0, and $P(\Omega_i=o_i|A=a)=1$ if $o_i=o_i^{(a)}$ otherwise is 0.

Under assumptions (A1-A3), we have

$$\xi_i = E(Y|A=a,M_i=m_i^{(a+1)},\Omega_i=o_i^{(a)})-E(Y|A=a,M_i=m_i^{(a)},\Omega_i=o_i^{(a)}).$$

Then, based on the LSEM that $Y=\gamma A+\boldsymbol{\beta}^\top M+\epsilon_Y$, we can further obtain that

$$\xi_i = \gamma a+\boldsymbol{\beta}^\top\begin{bmatrix}m_1^{(a)}\\ \vdots\\ m_i^{(a+1)}\\ \vdots\\ m_p^{(a)}\end{bmatrix}-\gamma a-\boldsymbol{\beta}^\top\begin{bmatrix}m_1^{(a)}\\ \vdots\\ m_i^{(a)}\\ \vdots\\ m_p^{(a)}\end{bmatrix}=\boldsymbol{\beta}_i\{m_i^{(a+1)}-m_i^{(a)}\}.$$

From Equation 12, we have

$$\xi_i=\boldsymbol{\beta}_i\Big[\{(I_p-B_M^\top)^{-1}\boldsymbol{\alpha}\}_i(a+1)-\{(I_p-B_M^\top)^{-1}\boldsymbol{\alpha}\}_i a\Big]=\boldsymbol{\beta}_i\{(I_p-B_M^\top)^{-1}\boldsymbol{\alpha}\}_i.$$

Thus, under assumptions (A1-A3) and Model 1, we have

$$DM_i=\xi_i. \qquad \square$$

References

[1] Avin, C., Shpitser, I. and Pearl, J. [2005], 'Identifiability of path-specific effects'.

[2] Chakrabortty, A., Nandy, P. and Li, H. [2018], 'Inference for individual mediation effects and interventional effects in sparse high-dimensional causal graphical models', *arXiv preprint arXiv:1809.10652* .

[3] Nandy, P., Maathuis, M. H., Richardson, T. S. et al. [2017], 'Estimating the effect of joint interventions from observational data in sparse high-dimensional settings', *The Annals of Statistics* **45**(2), 647–674.

[4] Rosenbaum, P. R. and Rubin, D. B. [1983], 'The central role of the propensity score in observational studies for causal effects', *Biometrika* **70**(1), 41–55.

[5] Vansteelandt, S. and Daniel, R. M. [2017], 'Interventional effects for mediation analysis with multiple mediators', *Epidemiology (Cambridge, Mass.)* **28**(2), 258.

[6] Wright, S. [1921], 'Correlation and causation', *Journal of agricultural research* **20**(7), 557–585.