Unfolding Latent Tree Structures using 4th Order Tensors
Appendix

8. Properties and Notations used

Nuclear and Frobenius norms:

- Let σ_i be the singular values of A. Then
$$\|A\|_* = \sum_i \sigma_i\,, \quad \|A\|_F^2 = \sum_i \sigma_i^2 \quad \text{and} \quad \|A\|_F \le \|A\|_*\,. \tag{13}$$
- (Nuclear and Frobenius norms are unitarily invariant) For any orthogonal Q we have
$$\begin{aligned} \|A\|_* &= \|QA\|_* = \|AQ\|_*\,, \\ \|A\|_F &= \|QA\|_F = \|AQ\|_F\,. \end{aligned} \tag{14}$$
-
$$\|AB\|_* \le \|A\|_F\|B\|_F \le \|A\|_*\|B\|_*\,. \tag{15}$$
- Let σ_i be the singular values of X and $\tilde{\sigma}_i$ be the singular values of $\tilde{X} = X + E$. Then
$$\|\mathrm{diag}(\tilde{\sigma}_i - \sigma_i)\|_* \le \|\tilde{X} - X\|_*\,. \tag{16}$$

Kronecker and Khatri-Rao products:

$$(A \otimes B)^\top = A^\top \otimes B^\top \tag{17}$$
$$(A + B) \otimes C = A \otimes C + B \otimes C \tag{18}$$
$$AB \otimes CD = (A \otimes C)(B \otimes D) \tag{19}$$
$$AB \odot CD = (A \otimes C)(B \odot D) \tag{20}$$
$$\|A \otimes B\|_F = \|A\|_F \|B\|_F$$
$$\mathrm{rank}(A \otimes B) = \mathrm{rank}(A)\,\mathrm{rank}(B)$$

Tensor operations:

We use the following tensor-matrix products of a tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ with matrices $M^{(n)} \in \mathbb{R}^{J_n \times I_n}$, $n = 1, 2, 3$:
$$\begin{aligned} &\textit{mode-1 product:} \quad (\mathcal{A} \bullet_1 M^{(1)})_{j_1 i_2 i_3} = \sum\nolimits_{i_1=1}^{I_1} a_{i_1 i_2 i_3} m^{(1)}_{j_1 i_1}\,, \\ &\textit{mode-2 product:} \quad (\mathcal{A} \bullet_2 M^{(2)})_{i_1 j_2 i_3} = \sum\nolimits_{i_2=1}^{I_2} a_{i_1 i_2 i_3} m^{(2)}_{j_2 i_2}\,, \\ &\textit{mode-3 product:} \quad (\mathcal{A} \bullet_3 M^{(3)})_{i_1 i_2 j_3} = \sum\nolimits_{i_3=1}^{I_3} a_{i_1 i_2 i_3} m^{(3)}_{j_3 i_3}\,, \end{aligned}$$
where $1 \le i_n \le I_n$, $1 \le j_n \le J_n$. These products can be considered as a generalization of the left and right multiplication of a matrix A with a matrix M. The mode-1 product signifies multiplying the columns (mode-1 vectors) of $\mathcal{A}$ with the rows of $M^{(1)}$ and similarly for the other tensor-matrix products.

The *contracted product* $\mathcal{C}$ of two tensors $\mathcal{A} \in \mathbb{R}^{I \times J \times M}$ and $\mathcal{B} \in \mathbb{R}^{K \times L \times M}$ along their third modes is a 4th order tensor denoted by $\mathcal{C} = \langle \mathcal{A}, \mathcal{B} \rangle_3$. $\mathcal{C} \in \mathbb{R}^{I \times J \times K \times L}$ and its entries $\mathcal{C}(i, j, k, l)$, $1 \le i \le I$; $1 \le j \le J$; $1 \le k \le K$; $1 \le l \le L$ are defined as
$$\mathcal{C}(i, j, k, l) = \sum\nolimits_{m=1}^{M} a_{ijm}\, b_{klm}.$$
It can be interpreted as taking inner products of the mode-3 vectors of $\mathcal{A}$ and $\mathcal{B}$ and storing the results in $\mathcal{C}$.

The 3 different *reshapings* A, B and C (2)–(4) of the tensor $\mathcal{P}$ contain exactly the same entires as $\mathcal{P}$ but in different order.

- A corresponds to the grouping $\{\{1,2\},\{3,4\}\}$ of the variables. The rows of A correspond to dimensions 1 and 2 of $\mathcal{P}$, and its columns to dimensions 3 and 4. Suppose all observed variables take values from $\{1,\ldots,n\}$, then entry of A at $(x_1+n(x_2-1))$-th row and $(x_3+n(x_4-1))$-th column is equal to $\mathcal{P}(x_1,x_2,x_3,x_4)$;
- B corresponds to the grouping $\{\{1,3\},\{2,4\}\}$, and its entry at $(x_1+n(x_3-1))$-th row and $(x_2+n(x_4-1))$-th column is equal to $\mathcal{P}(x_1,x_2,x_3,x_4)$;
- C corresponds to the grouping $\{\{1,4\},\{2,3\}\}$, and its entry at $(x_1+n(x_4-1))$-th row and $(x_2+n(x_3-1))$-th column is equal to $\mathcal{P}(x_1,x_2,x_3,x_4)$.

9. Matrix Representations A, B, C of $\mathcal{P}$

From $\mathcal{P}$ to A, B, C:

Let $X\in\mathbb{R}^{m\times k}$, $Y\in\mathbb{R}^{k\times l}$, $Z\in\mathbb{R}^{n\times l}$, $X=(x_1,\ldots,x_k)$ and $Z=(z_1,\ldots,z_l)$. A useful property that we will use in our derivations is the following

$$X\,Y\,Z^\top=\sum_{i,j}x_i\,y_{ij}\,z_j^\top. \tag{21}$$

We can derive the formula for A starting from the element-wise formula (1)

$$\mathcal{P}(x_1,x_2,x_3,x_4)=\sum_{h,g}P(x_1|h)P(x_2|h)P(h,g)P(x_3|g)P(x_4|g)$$

and placing all entries in the matrix A in the correct order. Note that given h and g we only need one column of each $P_{1|H}$, $P_{2|H}$, $P_{3|G}$ and $P_{4|G}$, which we will denote by $(P_{1|H})_h$, $(P_{2|H})_h$, $(P_{3|G})_g$ and $(P_{4|G})_g$. In order to obtain a matrix such that X_1 and X_2 are mapped to rows and X_3 and X_4 are mapped to columns, we need to map all possible products of single element of $(P_{1|H})_h$ and single element of $(P_{2|H})_h$ to rows and and similarly, we need to map all possible products of single element of $(P_{3|G})_g$ and single element of $(P_{4|G})_g$ to columns. This can be done using Khatri-Rao products in the following way

$$\begin{aligned}A\;&=\;\sum_{h,g}\Big((P_{2|H})_h\odot(P_{1|H})_h\Big)\,(P_{HG})_{hg}\,\Big((P_{4|G})_g\odot(P_{3|G})_g\Big)^\top\\&\overset{(21)}{=}\;\Big(P_{2|H}\odot P_{1|H}\Big)\;P_{HG}\;\Big(P_{4|G}\odot P_{3|G}\Big)^\top.\end{aligned}$$

The matrix B is unfolding of $\mathcal{P}$, such that the rows of B correspond to X_1 and X_3 and the columns of B correspond to X_2 and X_4. We have

$$\begin{aligned}B\;&=\;\sum_{h,g}\Big((P_{3|G})_g\odot(P_{1|H})_h\Big)\,(P_{HG})_{hg}\,\Big((P_{4|G})_g\odot(P_{2|H})_h\Big)^\top\\
&\overset{(17)}{=}\;\sum_{h,g}\Big((P_{3|G})_g\otimes(P_{1|H})_h\Big)\,(P_{HG})_{hg}\,\Big((P_{4|G})_g^\top\otimes(P_{2|H})_h^\top\Big)\\
&\overset{(19)}{=}\;\sum_{h,g}(P_{HG})_{hg}\,\Big((P_{3|G})_g(P_{4|G})_g^\top\Big)\otimes\Big((P_{1|H})_h(P_{2|H})_h^\top\Big)\\
&\overset{(18)}{=}\;\sum_{h}\Big(\sum_{g}(P_{HG})_{hg}(P_{3|G})_g(P_{4|G})_g^\top\Big)\otimes\Big((P_{1|H})_h(P_{2|H})_h^\top\Big)\\
&\overset{(21)}{=}\;\sum_{h}\Big(P_{3|G}\;\mathrm{diag}((P_{HG})_h)\,P_{4|G}^\top\Big)\otimes\Big((P_{1|H})_h(P_{2|H})_h^\top\Big)\\
&\overset{(19)}{=}\;\sum_{h}\Big(P_{3|G}\otimes(P_{1|H})_h\Big)\,\mathrm{diag}((P_{HG})_h)\,\Big(P_{4|G}^\top\otimes(P_{2|H})_h^\top\Big)\\
&\overset{\text{block}-(21)}{=}\;\Big(P_{3|G}\otimes P_{1|H}\Big)\;\mathrm{diag}(P_{HG}(:))\;\Big(P_{4|G}^\top\otimes P_{2|H}^\top\Big)\\
&\overset{(17)}{=}\;\Big(P_{3|G}\otimes P_{1|H}\Big)\;\mathrm{diag}(P_{HG}(:))\;\Big(P_{4|G}\otimes P_{2|H}\Big)^\top.\end{aligned}$$

The expression for C is derived in a similar way.

Other representations of A, B, C:

Using the properties in Section 8 and the formulas (5)–(7) for the matrix unfoldings A, B and C, we can derive the following additional formulas,

$$
\begin{aligned}
A &= \left(P_{2|H}\odot P_{1|H}\right)\ P_{HG}\ \left(P_{4|G}\odot P_{3|G}\right)^\top\\
&= \left(I_n\, P_{2|H}\odot P_{1|H}\, I_H\right)\ P_{HG}\ \left(I_n\, P_{4|G}\odot P_{3|G}\, I_G\right)^\top\\
&\overset{(20)}{=} \left(I_n\otimes P_{1|H}\right)\ \left(P_{2|H}\odot I_H\right)\ P_{HG}\ \left(P_{4|G}\odot I_G\right)^\top\ \left(I_n\otimes P_{3|G}\right)^\top\\
&= \begin{pmatrix} P_{1|H} & & \\ & \ddots & \\ & & P_{1|H}\end{pmatrix}
\begin{pmatrix} p_{2|H}^{(1,1)} & & \\ & p_{2|H}^{(1,2)} & \\ & & \ddots \\ p_{2|H}^{(2,1)} & & \\ \vdots & \ddots & \end{pmatrix}
P_{HG}
\begin{pmatrix} p_{4|G}^{(1,1)} & & \\ & p_{4|G}^{(1,2)} & \\ & & \ddots \\ p_{4|G}^{(2,1)} & & \\ \vdots & \ddots & \end{pmatrix}^\top
\begin{pmatrix} P_{3|G} & & \\ & \ddots & \\ & & P_{3|G}\end{pmatrix}^\top ,
\end{aligned} \tag{22}
$$

$$
\begin{aligned}
B &= \left(P_{3|G}\otimes P_{1|H}\right)\ \mathrm{diag}(P_{HG}(:))\ \left(P_{4|G}\otimes P_{2|H}\right)^\top\\
&= \left(P_{3|G}\, I_G\otimes I_n\, P_{1|H}\right)\ \mathrm{diag}(P_{HG}(:))\ \left(P_{4|G}\, I_G\otimes I_n\, P_{2|H}\right)^\top\\
&\overset{(19),(17)}{=} \left(P_{3|G}\otimes I_n\right)\ \left(I_G\otimes P_{1|H}\right)\ \mathrm{diag}(P_{HG}(:))\ \left(I_G\otimes P_{2|H}\right)^\top\ \left(P_{4|G}\otimes I_n\right)^\top\\
&= \begin{pmatrix} (p_{3|G}^{(1,1)}) & \cdots \\ (p_{3|G}^{(2,1)}) & \\ \vdots & \end{pmatrix}
\begin{pmatrix} P_{1|H} & & \\ & \ddots & \\ & & P_{1|H}\end{pmatrix}
\mathrm{diag}(P_{HG}(:))
\begin{pmatrix} P_{2|H} & & \\ & \ddots & \\ & & P_{2|H}\end{pmatrix}^\top
\begin{pmatrix} (p_{4|G}^{(1,1)}) & \cdots \\ (p_{4|G}^{(2,1)}) & \\ \vdots & \end{pmatrix}^\top ,
\end{aligned} \tag{23}
$$

where $(p^{(i,j)})$ is a diagonal block of size $(n\times n)$ with all diagonal elements equal to $p^{(i,j)}$.

The formula for C can be obtained from the ones for B by swapping the positions of $P_{3|G}$ and $P_{4|G}$.

Rank properties of A, B, C:

In this section we prove the rank properties used in Section 3.2 of the paper.

Lemma. *If $X\in\mathbb{R}^{m\times n}$, $Y\in\mathbb{R}^{n\times k}$, $Z\in\mathbb{R}^{l\times m}$, Y has full row rank, and Z has full column rank, then*

$$
\begin{aligned}
\mathrm{rank}(XY) &= \mathrm{rank}(X),\\
\mathrm{rank}(ZX) &= \mathrm{rank}(X).
\end{aligned}
$$

We assume that all CPTs have full column (or row) rank. Then the first two matrices in (22) also have full column rank. The last two matrices have full row rank. From the lemma, it follows that

$$
\mathrm{rank}(A) = \mathrm{rank}(P_{HG}) = k \tag{24}
$$

Analogously, the first two matrices in (23) have full column rank. The last two matrices have full row rank. From the lemma, it follows that

$$
\mathrm{rank}(B) = \mathrm{nnz}(P_{HG}), \tag{25}
$$

i.e., generically,

$$
\mathrm{rank}(B) = k^2.
$$

10. Algorithms

Algorithm 3 $\mathcal{T}_{next}$ = QuartetTree($\mathcal{T}_1$, $\mathcal{T}_2$, $\mathcal{T}_3$, X_4)

Require: Leaf($\mathcal{T}$): leaves of a tree $\mathcal{T}$;

for $j = 1$ to 3 **do**
  $X_i \leftarrow$ Randomly choose a variable from Leaf($\mathcal{T}_i$)
end for
$i^* \leftarrow$ Quartet(X_1, X_2, X_3, X_4), $\mathcal{T}_{next} \leftarrow \mathcal{T}_{i^*}$

Algorithm 4 $\mathcal{T}$ = Insert($\mathcal{T}, \widetilde{\mathcal{T}}, X_i$)

Require: Left($\mathcal{T}$) and Right($\mathcal{T}$): left and right child branch of the root respectively; $\mathcal{T} + \mathcal{T}'$: return a new tree connecting the root of two trees by an edge and use the root of $\mathcal{T}$ as the new root

if $|$Leaf($\mathcal{T}$)$| = 1$ **then**
  $\mathcal{T} \leftarrow$ Form a tree with root R connecting Leaf($\mathcal{T}$) and X_i.
else
  $\mathcal{T}_{next} \leftarrow$ QuartetTree(Left($\mathcal{T}$), Right($\mathcal{T}$), $\widetilde{\mathcal{T}}$, X_i)
  if $\mathcal{T}_{next}$ = Left($\mathcal{T}$) **then**
    $\mathcal{T} \leftarrow$ Insert($\mathcal{T}_{next}$, Right($\mathcal{T}$) + $\widetilde{\mathcal{T}}$, X_i)
  else if $\mathcal{T}_{next}$ = Right($\mathcal{T}$) **then**
    $\mathcal{T} \leftarrow$ Insert($\mathcal{T}_{next}$, Left($\mathcal{T}$) + $\widetilde{\mathcal{T}}$, X_i)
  end if
end if
$\mathcal{T} \leftarrow \mathcal{T} + \widetilde{\mathcal{T}}$

Algorithm 5 $\mathcal{T}$ = BuildTree($\{X_1, \ldots, X_d\}$)

Randomly choose X_1, X_2, X_3 and X_4
$i^* \leftarrow$ Quartet(X_1, X_2, X_3, X_4)
$\mathcal{T} \leftarrow$ Form a tree with two connecting hidden variables H and G, where H joins X_{i^*} and X_4, while G joins variables in $\{X_1, X_2, X_3\} \setminus \{X_{i^*}\}$
for $i = 5$ to d **do**
  Pick a root R from $\mathcal{T}$ which split it to three branches of equal sizes, and $\mathcal{T}_{next} \leftarrow$ QuartetTree(Left($\mathcal{T}$), Right($\mathcal{T}$), Middle($\mathcal{T}$), X_i)
  if $\mathcal{T}_{next}$ = Left($\mathcal{T}$) **then**
    $\mathcal{T} \leftarrow$ Insert($\mathcal{T}_{next}$, Right($\mathcal{T}$) + Middle($\mathcal{T}$), X_i)
  else if $\mathcal{T}_{next}$ = Right($\mathcal{T}$) **then**
    $\mathcal{T} \leftarrow$ Insert($\mathcal{T}_{next}$, Left($\mathcal{T}$) + Middle($\mathcal{T}$), X_i)
  else if $\mathcal{T}_{next}$ = Middle($\mathcal{T}$) **then**
    $\mathcal{T} \leftarrow$ Insert($\mathcal{T}_{next}$, Right($\mathcal{T}$) + Left($\mathcal{T}$), X_i)
  end if
end for

11. Recovery Conditions for Quartet

Latent variables H and G are independent. In this case, rank$(P_{HG}) = 1$, since $P(h, g) = P(h)P(g)$. Applying the relation in Equation 8, we have that $rank(A) = 1 \ll rank(B)$. Furthermore, since A has only one nonzero singular value, we have $\|A\|_* = \|A\|_F = \|B\|_F \leq \|B\|_*$, since $\|M\|_F \leq \|M\|_*$ for any M. In this case, we know for sure that the nuclear norm quartet test will return the correct topology.

Latent variables H and G are not independent. We analyze this case by treating it as perturbation Δ away from the P_{HG} in the independent case. We want to characterize how large Δ can be while still allowing the nuclear norm quartet test to find the correct latent relation. Suppose $A_\perp$ and $B_\perp$ are the unfolding matrices in the case where H and G are independent. Suppose we add perturbation Δ to P_{HG}, then $A_\perp = \left(P_{2|H} \odot P_{1|H}\right)\ P_{HG}\ \left(P_{4|G} \odot P_{3|G}\right)^\top$ and its perturbed version is $A = \left(P_{2|H} \odot P_{1|H}\right)\ (P_{HG} + \Delta)\ \left(P_{4|G} \odot P_{3|G}\right)^\top$. We want to bound the difference $|\,\|A_\perp\|_* - \|A\|_*|$. We have

$$
\begin{aligned}
|\,\|A_\perp\|_* - \|A\|_*| &= \left|\sum\nolimits_i \sigma_i(A_\perp) - \sum_i \sigma_i(A)\right| \\
&\leq \sum\nolimits_i |\sigma_i(A_\perp) - \sigma_i(A)| \\
&\overset{(16)}{\leq} \|A_\perp - A\|_* \\
&\leq \left\|\left(P_{2|H} \odot P_{1|H}\right)\ \Delta\ \left(P_{4|G} \odot P_{3|G}\right)^\top\right\|_* \\
&\overset{(15)}{\leq} \left\|P_{2|H} \odot P_{1|H}\right\|_F\ \|\Delta\|_F\ \left\|P_{4|G} \odot P_{3|G}\right\|_F \\
&\leq k\,\|\Delta\|_F\,,
\end{aligned}
$$

since $P_{2|H} \odot P_{1|H}$ and $P_{4|G} \odot P_{3|G}$ are CPTs with k columns each, and thus $\left\|P_{2|H} \odot P_{1|H}\right\|_F^2 \leq k$ and $\left\|P_{4|G} \odot P_{3|G}\right\|_F^2 \leq k$.

Analogously, $B_\perp = \left(P_{3|G} \otimes P_{1|H}\right)\ \mathrm{diag}(P_{HG}(:))\ \left(P_{4|G} \otimes P_{2|H}\right)^\top$ and its perturbed version is $B = \left(P_{3|G} \otimes P_{1|H}\right)\ \mathrm{diag}(P_{HG}(:) + \Delta(:))\ \left(P_{4|G} \otimes P_{2|H}\right)^\top$. We want to bound the difference $|\|B_\perp\|_* - \|B\|_*|$. We have

$$
\begin{aligned}
|\|B_\perp\|_* - \|B\|_*| &= \left|\sum\nolimits_i \sigma_i(B_\perp) - \sum_i \sigma_i(B)\right| \\
&\leq \sum\nolimits_i |\sigma_i(B_\perp) - \sigma_i(B)| \\
&\overset{(16)}{\leq} \|B_\perp - B\|_* \\
&\leq \left\|\left(P_{3|G} \otimes P_{1|H}\right)\ \mathrm{diag}(\Delta(:))\ \left(P_{4|G} \otimes P_{2|H}\right)^\top\right\|_* \\
&\overset{(15)}{\leq} \left\|P_{3|G} \otimes P_{1|H}\right\|_F\ \left\|\mathrm{diag}(\Delta(:))\right\|_F\ \left\|P_{4|G} \otimes P_{2|H}\right\|_F \\
&\leq k^2\,\|\mathrm{diag}(\Delta(:))\|_F \\
&= k^2\,\|\Delta\|_F\,,
\end{aligned}
$$

since $P_{3|G}\otimes P_{1|H}$ and $P_{4|G}\otimes P_{2|H}$ are CPTs with k^2 columns, and thus $\left\|P_{3|G}\otimes P_{1|H}\right\|_F^2 \leq k^2$ and $\left\|P_{4|G}\otimes P_{2|H}\right\|_F^2 \leq k^2$.

Therefore, we get the following upper and lower bound:

$$
\begin{aligned}
\|A\|_* &\leq \|A_\perp\|_* + k\ \|\Delta\|_F\,, \\
\|B\|_* &\geq \|B_\perp\|_* + k^2\,\|\Delta\|_F\,.
\end{aligned}
$$

If we require that

$$
\|A_\perp\|_* + k\,\|\Delta\|_F \leq \|B_\perp\|_* + k^2\,\|\Delta\|_F\,,
$$

then we will have $\|A\|_* \leq \|B\|_*$.

We can derive similar condition for the relationship $\|A\|_* \leftrightarrow \|C\|_*$. Let

$$
\theta := \min\{\|B_\perp\|_* - \|A_\perp\|_*,\ \|C_\perp\|_* - \|A_\perp\|_*\}.
$$

We thus obtain an upper bound on the allowed perturbation:

$$\|\Delta\|_F \leq \frac{\theta}{k^2+k}\,. \tag{26}$$

12. Recovery Conditions for Latent Tree

When latent variables H and G are independent, we have that $P_{HG} = P_H P_G^\top$. In this case,

$$\begin{aligned} \|B_\perp\|_* &= \left\|(P_{3|G}\otimes P_{1|H})(\mathrm{diag}(P_G)\otimes\mathrm{diag}(P_H))(P_{4|G}\otimes P_{2|H})^\top\right\|_* \\ &= \left\|(P_{3|G}\,\mathrm{diag}(P_G)P_{4|G}^\top)\otimes(P_{1|H}\,\mathrm{diag}(P_H)P_{2|H}^\top)\right\|_* \\ &= \|P_{34}\otimes P_{12}\|_* \\ &\geq \|P_{34}\otimes P_{12}\|_F \end{aligned} \tag{27}$$

and

$$\begin{aligned} \|A_\perp\|_* &= \left\|(P_{2|H}\odot P_{1|H})\ \ P_H P_G^\top\ \ (P_{4|G}\odot P_{3|G})^\top\right\|_* \\ &= \left\|P_{12}(:)P_{34}(:)^\top\right\|_* \\ &= \left\|P_{12}(:)P_{34}(:)^\top\right\|_F \\ &= \|P_{34}\otimes P_{12}\|_F \end{aligned} \tag{28}$$

and thus

$$\|A_\perp\|_* \ \leq\ \|B_\perp\|_*\,.$$

Suppose now that H and G are not independent and thus we have $P_{HG} = P_H P_G^\top + \Delta$. The goal is to characterize all Δs, such that $\|A\|_* \leq \|B\|_*$ still holds for any quartet. From the above formulas it follows that the upper bound on Δ depends only on pairwise marginal distributions.

Since the perturbed version of $P_H P_G^\top$ remains a joint probability table, all entries of the perturbation matrix Δ have to sum to 0, *i.e.*, $\mathbf{1}^\top\Delta(:) = 0$. We further assume that each column sum and each row sum of Δ is also equal to 0, *i.e.*, $\mathbf{1}^\top\Delta = \mathbf{0}$ and $\Delta\,\mathbf{1} = \mathbf{0}$. In this case, $\mathbf{1}^\top\Delta(:) = 0$ is satisfied automatically.

The recovery conditions for latent trees can be derived in two steps. The first step is to provide recovery conditions for those quartet relations corresponding to a single edge $H - G$ in the tree (Fig. 6, left). In the second step we study quartet relations corresponding to paths $H - M_1 - M_2 - \cdots - M_l - G$ in the tree (Fig. 6, right). We provide a condition under which the recovery condition of such quartets is reduced to the recovery condition on quartets from step 1. That is, we provide a condition under which the perturbation on the path is guaranteed to be smaller than the maximum allowed perturbation on an edge.

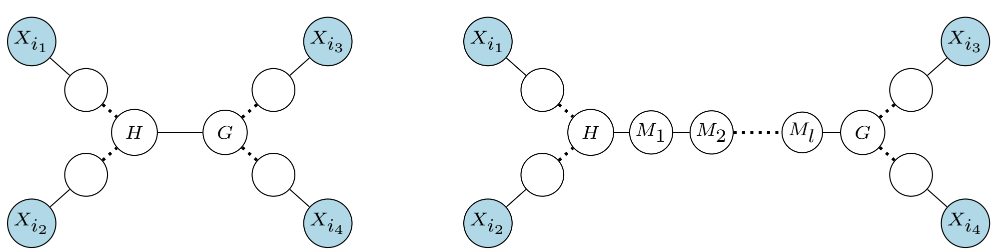

Figure 6. Topologies of quartets corresponding to a single edge $H - G$ and to a path $H - M_1 - M_2 - \cdots - M_l - G$.

Let

$$\delta := \max_{H-G \text{ an edge}} \|\Delta_{HG}\|_F\,.$$

Our goal is to obtain conditions on δ, under which recovery of any quartet relation is guaranteed.

12.1. Quartets Corresponding to a Single Edge

The first step is readily obtained from §11 if we assume that all CPTs (including $P_{X_{i_1}|H}$, $P_{X_{i_2}|H}$, $P_{X_{i_3}|G}$, $P_{X_{i_4}|G}$) have full rank. Let $\theta_{\min} = \min_{\text{quarter } q} \theta_q$. From (26), we have

$$\delta \le \min \frac{\|B_\perp\|_* - \|A_\perp\|_*}{k^2+k} = \frac{\theta_{\min}}{k^2+k}. \tag{29}$$

12.2. Quartets Corresponding to a Path

Path of independent latent variables. For the second step, we start again from the fully factorized case (independent case). The joint probability table P_{HG} of the two end points in a path $H - M_1 - M_2 - \cdots - M_l - G$ is

$$\begin{aligned}
P_{HG} &= P_{H|M_1} P_{M_1|M_2} \cdots P_{M_l|G} P_G \\
&= P_{HM_1} \operatorname{diag}(P_{M_1})^{-1} P_{M_1 M_2} \operatorname{diag}(P_{M_2})^{-1} \cdots \operatorname{diag}(P_{M_l})^{-1} P_{M_l G} \\
&= P_H P_{M_1}^\top \operatorname{diag}(P_{M_1})^{-1} P_{M_1} P_{M_2}^\top \operatorname{diag}(P_{M_2})^{-1} \cdots \operatorname{diag}(P_{M_l})^{-1} P_{M_l} P_G^\top \\
&= P_H (P_{M_1}^\top \operatorname{diag}(P_{M_1})^{-1}) P_{M_1} (P_{M_2}^\top \operatorname{diag}(P_{M_2})^{-1}) \cdots \operatorname{diag}(P_{M_l})^{-1} P_{M_l} P_G^\top \\
&= P_H \mathbf{1}^\top P_{M_1} \mathbf{1}^\top \cdots \mathbf{1}^\top P_{M_l} P_G^\top \\
&= P_H P_G^\top \,,
\end{aligned}$$

where we have used $P_{M_i}^\top \operatorname{diag}(P_{M_i}(:))^{-1} = \mathbf{1}^\top$.

Path of dependent latent variables. Next, we add perturbation matrices to the joint probability tables associated with each edge $M_i - M_j$ in the tree and assume that the resulting joint probability table $P_{M_i M_j} = P_{M_i} P_{M_j}^\top + \Delta_{ij}$ has full rank. Furthermore, we assume that the resulting joint probability table P_{HG} of the two end points in a path $H - M_1 - M_2 \cdots M_l - G$ also has full rank. We have

$$\begin{aligned}
P_{HG} &= P_{H|M_1} P_{M_1|M_2} \cdots P_{M_l|G} P_G \\
&= P_{HM_1} \operatorname{diag}(P_{M_1})^{-1} P_{M_1 M_2} \operatorname{diag}(P_{M_2})^{-1} \cdots \operatorname{diag}(P_{M_l})^{-1} P_{M_l G} \\
&= (P_H P_{M_1}^\top + \Delta_1) \operatorname{diag}(P_{M_1})^{-1} (P_{M_1} P_{M_2}^\top + \Delta_2) \operatorname{diag}(P_{M_2})^{-1} \cdots \operatorname{diag}(P_{M_l})^{-1} (P_{M_l} P_G^\top + \Delta_l) \\
&= P_H P_{M_1}^\top \operatorname{diag}(P_{M_1})^{-1} P_{M_1} P_{M_2}^\top \operatorname{diag}(P_{M_2})^{-1} \cdots \operatorname{diag}(P_{M_l})^{-1} P_{M_l} P_G^\top \\
&\quad + 0 \;\; \text{(terms not involving all the Δs will all be zero)} \\
&\quad + \Delta_1 \operatorname{diag}(P_{M_1})^{-1} \Delta_2 \operatorname{diag}(P_{M_2})^{-1} \cdots \operatorname{diag}(P_{M_l})^{-1} \Delta_l \\
&= P_H P_G^\top + \Delta_1 \operatorname{diag}(P_{M_1})^{-1} \Delta_2 \operatorname{diag}(P_{M_2})^{-1} \cdots \operatorname{diag}(P_{M_l})^{-1} \Delta_l \,.
\end{aligned} \tag{30}$$

The reason why we do not need to perturb the term $\operatorname{diag}(P_{M_i})^{-1}$ is that if $\widetilde{P}_{M_i}$ is the perturbed P_{M_i},

$$\widetilde{P}_{M_i} = \widetilde{P}_{M_i M_j} \mathbf{1} = (P_{M_i} P_{M_j}^\top + \Delta_{ij}) \mathbf{1} = P_{M_i} P_{M_J}^\top \mathbf{1} + \mathbf{0} = P_{M_i} \,,$$

since $\Delta_{ij} \mathbf{1} = \mathbf{0}$. And the reason why terms not involving all the Δs will all be zero is that such terms contain either $\mathbf{1}^\top \Delta = \mathbf{0}^\top$ or $\Delta \mathbf{1} = \mathbf{0}$.

Now, from (30) it follows that the perturbation corresponding to the path $H - M_1 - M_2 - \cdots - M_l - G$ is

$$\Delta := \Delta_1 \operatorname{diag}(P_{M_1})^{-1} \Delta_2 \operatorname{diag}(P_{M_2})^{-1} \cdots \operatorname{diag}(P_{M_l})^{-1} \Delta_l. \tag{31}$$

Bounding the perturbation on the path. We still need to show under which condition Δ from (31) will satisfy $\|\Delta\|_F \le \delta$. Assume that the smallest entry in a marginal distribution of an internal node is bounded from below by $\gamma_{\min}$, *i.e.*,

$$\gamma_{\min} := \min_{\text{hidden node } H} \min_i P_H(i) \,. \tag{32}$$

Then we have

$$\begin{aligned}\|\Delta\|_F &= \left\|\Delta_1 \operatorname{diag}(P_{M_1})^{-1}\Delta_2 \operatorname{diag}(P_{M_2})^{-1}\cdots\Delta_l\right\|_F\\ &\le \left\|\Delta_1 \operatorname{diag}(P_{M_1})^{-1}\right\|_F \left\|\Delta_2 \operatorname{diag}(P_{M_2})^{-1}\right\|_F\cdots\|\Delta_l\|_F\\ &\le \frac{\delta^l}{\gamma_{\min}^{l-1}}\,.\end{aligned}$$

The perturbation Δ on the path $H - M_1 - M_2 \cdots M_l - G$ is bounded by δ if $\frac{\delta^l}{\gamma_{\min}^{l-1}} \le \delta$, *i.e.*, if

$$\delta \le \gamma_{\min}. \tag{33}$$

From (29) and (33) we arrive at the condition for successful quartet test for all quartets

$$\delta \le \min\left\{\frac{\theta_{\min}}{k^2+k}, \gamma_{\min}\right\}.$$

Intuitively, it means that the size of the perturbation δ away from independence can not be too large. In particular, it has to be small compared to the smallest marginal probability $\gamma_{\min}$ of a hidden state; it also has to be small compared to the smallest excessive dependence $\theta_{\min}$.

13. Recovery Conditions in Case of Different Number of Hidden States

13.1. Simple quartets in case of violated (A1)

Let us first revisit the rank conditions in case **(A1)** is violated. Let the (simple) quartet relation of 4 variables, X_1, X_2, X_3 and X_4, be $\{\{1,2\},\{3,4\}\}$, and the connecting hidden variables be H and G (see Fig. 2(b), left). Further, let $\operatorname{rank}(P_{HG}) = k_0$, $\operatorname{rank}(P_{1|H}) = k_1$, $\operatorname{rank}(P_{2|H}) = k_2$, $\operatorname{rank}(P_{3|G}) = k_3$, $\operatorname{rank}(P_{4|G}) = k_4$. Based on equations (5), (6) and (7), we can conclude that

$$\operatorname{rank}(A) \le \operatorname{rank}(P_{HG}) = k_0$$

and generically

$$\begin{aligned}\operatorname{rank}(B) &= \min(k_1k_3, k_2k_4),\\ \operatorname{rank}(C) &= \min(k_1k_4, k_2k_3).\end{aligned}$$

Thus

$$\operatorname{rank}(A) < \min(\operatorname{rank}(B), \operatorname{rank}(C)) \quad \text{if} \quad k_0 < \min(k_1k_3, k_2k_4, k_1k_4, k_2k_3). \tag{34}$$

13.2. General quartets with hidden variables having different number of states

Suppose now that (**A1**) holds but the hidden variables have different number of states. Let the smallest number be $k_{\min}$ and let (**A4**) all hidden variables have less than $k_{\min}^2$ states.

In a general quartet from the tree (as in Fig. 2(a)), each combined CPT corresponding to a path (*e.g.*, $H_i - \ldots - X_{i_1}$) may have low rank, although each CPT of adjacent variables has full rank. Thus we can think of resolving a general quartet relation in case of different number of hidden states as resolving a simple quartet relation in case (**A1**) is violated (see Section 13.1). Generically, the rank of a combined CPT corresponding to a path is equal to the minimal rank of a single CPT corresponding to an edge in the path. Now, under assumption (**A4**), the condition in (34) is satisfied for any quartet form the tree and thus $\operatorname{rank}(A) < \min(\operatorname{rank}(B), \operatorname{rank}(C))$ holds.

Condition (**A4**) is sufficient but not necessary, as it can be seen from the following special case.

Remark 8 *As it is natural to assume that the hidden variables are simpler than the observed ones ($k \le n$), extending this intuition to the relations between hidden variables, it is meaningful to assume that the hidden variables get simpler when we move away from the observed variables. In this case (using the notation from Section 13.1 but this time for general quartets), $k_0 < k_1$ and $k_0 < k_2$ (or otherwise $k_0 < k_3$ and $k_0 < k_4$) and the condition in (34) is satisfied for any quartet form the. Thus* $\operatorname{rank}(A) < \min(\operatorname{rank}(B), \operatorname{rank}(C))$ *holds.*

13.3. Rank vs. nuclear norm condition

This being said, note that the rank condition only motivates our approach but the test is based only on nuclear norm comparisons rather than on rank comparisons. Thus, if $\|A\|_* < \min(\|B\|_*, \|C\|_*)$, then our test will succeed even if $\operatorname{rank}(A) \geq \min(\operatorname{rank}(B), \operatorname{rank}(C))$.

When two hidden variables H and G have different numbers of states, k and g respectively, the condition of Lemma 4 becomes

$$\|\Delta\|_F <= \frac{\theta}{kg + \sqrt{kg}}. \tag{35}$$

In our analysis in Section 12, we have used stronger than necessary assumptions to make the presentation as clear as possible. However, full rank condition for the combined CPTs is not necessary for the correctness of the proofs and they are thus valid for the case of different number of hidden states as well.

Remark 9 *There is a minimal representation of each single CPTs and thus also of the marginal probabilities of each hidden variable. Thus, for $\gamma_{\min}$ in (32) we should have $\gamma_{\min} > 0$ and (**A3**) is still meaningful even in case of different number of hidden states.*

14. Statistical Guarantee for the Quartet Test

Based on the concentration result for nuclear norm in (10), we have that, given m samples, the probability that the finite sample nuclear norm deviates from its true quantity by $\epsilon := \frac{2\sqrt{2\tau}}{\sqrt{m}}$ is bounded

$$\mathbb{P}\left\{\|\widehat{A}\|_* \geq \|A\|_* + \epsilon\right\} \leq 2e^{-\frac{m\epsilon^2}{8}} \quad \text{and} \quad \mathbb{P}\left\{\|\widehat{B}\|_* \leq \|B\|_* - \epsilon\right\} \leq 2e^{-\frac{m\epsilon^2}{8}}, \tag{36}$$

where we have used $\tau = \frac{m\epsilon^2}{8}$. Now we can derive the probability of making an error for individual quartet test. First, let $q = \{\{i_1, i_2\}, \{i_3, i_4\}\}$ and

$$\alpha = \min\left\{\|B(q)\|_* - \|A(q)\|_*, \|C(q)\|_* - \|A(q)\|_*\right\}.$$

Then, for sufficiently large m, we can bound the error probability by

$$\begin{aligned}
&\mathbb{P}\left\{\text{Quartet test returns incorrect result}\right\}\\
&= \mathbb{P}\left\{\|\widehat{A}\|_* \geq \|\widehat{B}\|_* \;\text{ or }\; \|\widehat{A}\|_* \geq \|\widehat{C}\|_*\right\}\\
&\leq \mathbb{P}\left\{\|\widehat{A}\|_* \geq \|\widehat{B}\|_*\right\} + \mathbb{P}\left\{\|\widehat{A}\|_* \geq \|\widehat{C}\|_*\right\} \qquad \text{(union bound)}\\
&= \mathbb{P}\left\{\|\widehat{A}\|_* - \|A\|_* + \|B\|_* - \|\widehat{B}\|_* \geq \|B\|_* - \|A\|_*\right\}\\
&\quad + \mathbb{P}\left\{\|\widehat{A}\|_* - \|A\|_* + \|C\|_* - \|\widehat{C}\|_* \geq \|C\|_* - \|A\|_*\right\}\\
&\leq \mathbb{P}\left\{\|\widehat{A}\|_* - \|A\|_* \geq \frac{\|B\|_* - \|A\|_*}{2}\right\} + \mathbb{P}\left\{\|B\|_* - \|\widehat{B}\|_* \geq \frac{\|B\|_* - \|A\|_*}{2}\right\}\\
&\quad + \mathbb{P}\left\{\|\widehat{A}\|_* - \|A\|_* \geq \frac{\|C\|_* - \|A\|_*}{2}\right\} + \mathbb{P}\left\{\|C\|_* - \|\widehat{C}\|_* \geq \frac{\|C\|_* - \|A\|_*}{2}\right\}\\
&\leq \mathbb{P}\left\{\|\widehat{A}\|_* - \|A\|_* \geq \frac{\alpha}{2}\right\} + \mathbb{P}\left\{\|B\|_* - \|\widehat{B}\|_* \geq \frac{\alpha}{2}\right\}\\
&\quad + \mathbb{P}\left\{\|\widehat{A}\|_* - \|A\|_* \geq \frac{\alpha}{2}\right\} + \mathbb{P}\left\{\|C\|_* - \|\widehat{C}\|_* \geq \frac{\alpha}{2}\right\}\\
&\leq 8e^{-\frac{m\alpha^2}{32}}
\end{aligned}$$

15. Statistical Guarantee for the Tree Building Algorithm

Let $\alpha_q = \min\left\{\|B(q)\|_* - \|A(q)\|_*, \|C(q)\|_* - \|A(q)\|_*\right\}$. We define

$$\alpha_{\min} = \min_{\text{quartet } q} \alpha_q.$$

For a latent tree with d observed variables, the tree building algorithm described in the paper requires $O(d \log d)$ calls to the quartet test procedure. The probability that the tree is constructed incorrectly is bounded by the probability that either one of these quartet tests returns incorrect result. That is

$$\begin{aligned}
&\mathbb{P}\left\{\text{The latent tree is constructed incorrectly}\right\} \\
\leq\ &\mathbb{P}\left\{\text{Either one of the } O(d \log d) \text{ quartet tests returns incorrect result}\right\} \\
\leq\ &c \cdot d \log d \cdot \mathbb{P}\left\{\text{quartet test returns incorrect result}\right\} \qquad \text{(union bound)} \\
\leq\ &8c \cdot d \log d \cdot e^{-\frac{m\alpha^2}{32}},
\end{aligned}$$

which implies that the probability of constructing the tree incorrectly decreases exponentially fast as we increase the number of samples m.