Algebraic Properties of Wyner Common Information Solution under Graphical Constraints

Md Mahmudul Hasan, Shuangqing Wei, Ali Moharrer

Abstract—The Constrained Minimum Determinant Factor Analysis (CMDFA) setting was motivated by Wyner’s common information problem where we seek a latent representation of a given Gaussian vector distribution with the minimum mutual information under certain generative constraints. In this paper, we explore the algebraic structures of the solution space of the CMDFA, when the underlying covariance matrix $\Sigma_x$ has an additional latent graphical constraint, namely, a latent star topology. In particular, sufficient and necessary conditions in terms of the relationships between edge weights of the star graph have shown that the CMDFA problem has either a rank one solution or a rank $n-1$ solution where $n$ is the dimension of the observable vector. Numerical results are provided to demonstrate the difference between the optimal mutual information and that derived under a naive star constraint.

Index Terms—Factor Analysis, MTFA, CMTFA, CMDFA

I. INTRODUCTION

The literature that approached factor analysis can be classified in three major categories [1] [2]. Firstly, algebraic approaches [3] [4], which did not offer scalable algorithms for higher dimensional statistics. Secondly, factor analysis via heuristic local optimization techniques, which offered no provable performance guarantees. The third and final type are the convex optimization based methods such as Constrained Minimum Trace Factor Analysis (CMTFA) [5] and CMDFA [6], Moharrer and Wei derived CMDFA as a variety of a broader class of convex optimization problem defined in [7]. CMDFA decomposes the population covariance matrix of an $n$-dimensional random vector $\vec{X} \in \mathbb{R}^n$, $\Sigma_x$ as $\Sigma_x = (\Sigma_x - D) + D$ such that the matrix $D$ is diagonal, both $D$ and $(\Sigma_x - D)$ are Gramian matrices and $|\Sigma_x - D|$ is minimized where $|\Sigma_x - D|$ denotes the determinant of the matrix $(\Sigma_x - D)$. We next discuss the operational significance of such optimization framework in terms of Wyner’s common information.

Wyner’s common information $C(X_1, X_2)$ characterizes the minimum amount of common randomness needed to approximate the joint density between a pair of random variables $X_1$ and $X_2$ to be $C(X_1, X_2) = \min_{p_{X_1, Y; X_2}} I(X_1, X_2; Y)$, where $I(X_1, X_2; Y)$ is the mutual information between $X_1$, $X_2$ and $Y$, $X_1 - Y - X_2$ indicates the conditional independence between $X_1$ and $X_2$ given $Y$, and the joint density function $p_{X_1, Y; X_2}$ is sought to ensure such conditional independence as well as the given joint density of $X_1$ and $X_2$ [8]. The second interpretation to common information that Wyner provided in his work is more relevant in our setting, which essentially says that the minimum rate of information (bits per channel use) needed to be able to synthesize the joint distribution of two random variables with arbitrarily high accuracy is equal to the common information shared by the variables [9]. A straightforward extension to an $n$ dimensional case is that, to be able to synthesize a jointly distributed random vector $\vec{X} = [X_1, \ldots, X_n]^T$ the required minimum rate of information has to be equal to $C(X_1, \ldots, X_n) = \min \{I(\vec{X}, \vec{Y})\}$ over all $\vec{Y}$ such that $X_i \perp \!\!\!\!\perp X_j | \vec{Y}, \ i \neq j$. Now if the aforementioned $\vec{X}$ is a Gaussian random vector then it can be represented by the factor analysis model $\vec{X} = A\vec{Y} + \vec{Z}$, where $A$ is an $n \times k$ real matrix, $\vec{Y} \in \mathbb{R}^k$, $k < n$ is the vector of latent variables that captures all the dependencies in $\vec{X}$ and $\vec{Z}$ is a Gaussian vector of zero mean and covariance matrix $\Sigma_z$. From optimization point of view $I(\vec{X}, \vec{Y}) = h(\vec{X}) - h(\vec{X} | \vec{Y}) = h(\vec{X}) - h(\vec{Z})$ where $h(\vec{X})$ and $h(\vec{Z})$ are differential entropies of $\vec{X}$ and $\vec{Z}$ respectively, and $h(\vec{X} | \vec{Y})$ is the conditional entropy of $\vec{X}$ given $\vec{Y}$. Hence characterizing the common information between $\vec{X}$ and $\vec{Y}$ [10] [8] [11] would be $\min_{A, \Sigma_z} I(\vec{X}; \vec{Y})$ which is an equivalent problem to $\max_{\Sigma_z} h(\vec{Z})$ hence equivalent to $\min_{\Sigma_z} - \log |\Sigma_z|$, which is essentially finding the minimum determinant solution i.e. CMDFA solution of $\Sigma_z$.

The main contribution of our work is that, we analysed the solution space of CMDFA and recovered the underlying graphical structures. We find the explicit conditions under which the CMDFA solution of $\Sigma_x$ is a rank 1 solution i.e. recovers a star structure. Since star may not always be the optimum solution, we have also shown the existence and uniqueness of a rank $n-1$ CMDFA solution of $\Sigma_x$ which is the only other possible solution. In the bigger picture, from operational point of view, such characterization of the CMDFA solution space helps us synthesize the jointly Gaussian vector $\vec{X}$ with minimum rate of information. At the end we presented some numerical results to show the merit of our rigorous analysis. It is important to remark that our work is not concerned about the algorithm side of the optimization technique, rather, our focus is to characterize and find insights about the underlying solution space.

II. DEFINITIONS AND NOTATIONS

Let $\vec{b}$ be a real $n$ dimensional column vector and $A$ be an $n \times n$ matrix. As in literature in general we denote the $i$th element $\vec{b}_i$ of $\vec{b}$ and the $(i,j)$th element of $A$ as $A_{i,j}$. Here
we define all the vector operations and notations in terms of \( \vec{b} \) and \( A \), that will carry their meaning on other vectors and matrices throughout this paper unless stated otherwise.

Let \( \alpha_{i,*} \) and \( \alpha_{*,i} \) denote the \( i \)th row and \( i \)th column vector of matrix \( A \) respectively. Function \( \lambda_{\min}(A) \) is defined to be the smallest eigenvalue of matrix \( A \). \( N(A) \) stands for the null space of matrix \( A \).

Vectors \( \vec{1} \) and \( \vec{0} \) are the \( n \) dimensional column vectors with each element equal to 1 and 0 respectively. When we write \( \vec{b} \geq 0 \) we mean that each element of the vector \( b(i) \geq 0, 1 \leq i \leq n \). \( \vec{b}^2 \) is the Hadamard product of vector \( \vec{b} \) with itself. \( ||\vec{b}|| \) denotes the \( L_2 \) norm of vector \( \vec{b} \).

Now we define two terms i.e. dominance and non-dominance of a vector which will repeatedly appear throughout the paper. When we talk about the dominance or non-dominance of any vector \( \vec{b} \) we assume that the elements of the vector are sorted in a way such that \( |b_1| \geq |b_2| \geq \cdots \geq |b_n| \). We call vector \( \vec{b} \) dominant and \( b_j \) the dominant element if for the above sorted vector \( |b_1| > \sum_{j \neq 1} |b_j| \) holds. Otherwise \( \vec{b} \) is non-dominant.

### III. Formulation of the Problem

First of all we define the real column vector \( \vec{\alpha} \) as \( \vec{\alpha} = [\alpha_1, \ldots, \alpha_n]^T \in \mathbb{R}^n \) where \( 0 < |\alpha_j| < 1, j = 1, 2, \ldots, n \) and

\[
|\alpha_1| \geq |\alpha_2| \geq \cdots \geq |\alpha_n|
\]

(1)

Let us consider a star structured population covariance matrix \( \Sigma_x \), having all the diagonal components 1 as given by equation (2).

\[
\Sigma_x = \begin{pmatrix}
1 & \alpha_1 \alpha_2 & \cdots & \alpha_1 \alpha_n \\
\alpha_2 \alpha_1 & 1 & \cdots & \alpha_2 \alpha_n \\
\vdots & \vdots & \ddots & \vdots \\
\alpha_n \alpha_1 & \alpha_n \alpha_2 & \cdots & 1
\end{pmatrix}
\]

(2)

The above covariance matrix could be produced by the graphical model given by equation (3).

\[
\vec{X} = \vec{\alpha} \vec{Y} + \vec{Z}
\]

(3)

where

- \( \vec{X} \sim \mathcal{N}(0, \Sigma_x) \) is a jointly Gaussian random vector of conditionally independent Gaussian random variables \( X_1, \ldots, X_n \) given \( \vec{Y} \), where \( \vec{Y} \sim \mathcal{N}(0, 1) \).
- \( \vec{Z} \) is a vector of independent Gaussian random variables \( Z_1, \ldots, Z_n \) with \( Z_j \sim \mathcal{N}(0, 1 - \alpha_j^2) \) \( 1 \leq j \leq n \).

The above graphical model assumes the conditional independence among the observables given the latent variable given by (4) giving rise to a star topology.

\[
p(X_1, X_2, \ldots, X_n|Y) = \Pi_{i=1}^n p(X_i|Y)
\]

(4)

Operationally speaking, we are trying to find the minimum rate of information required to synthesize the jointly distributed Gaussian vector \( \vec{X} \), which is, as we discussed in the introduction, an equivalent problem to minimizing the mutual information between the observable Gaussian random vector \( \vec{X} \) and the latent vector \( \vec{Y} \). It is thus to seek joint distribution between the latent variables and the observable ones such that the differential entropy \( h(\vec{X}|\vec{Y}) \) is maximized. Under the joint Gaussian distribution, it is the same as seeking factorization of \( \Sigma_x \) such that the determinant of the diagonal matrix \( D \) is maximized as in equation (5) under the constraint that both \( (\Sigma_x - D) \) and \( D \) are Gramian matrices i.e. seeking the CMDFA solution of \( \Sigma_x \).

\[
\Sigma_x = (\Sigma_x - D) + D
\]

(5)

\( \Sigma_x \) being produced by the model in (3) would equivalently mean (5) having a rank 1 solution i.e. \( \Sigma_x - D \) being \( \Sigma_{1,ND} \) given by equation (6).

\[
\Sigma_{1,ND} = \begin{pmatrix}
\alpha_1^2 & \alpha_1 \alpha_2 & \cdots & \alpha_1 \alpha_n \\
\alpha_2 \alpha_1 & \alpha_2^2 & \cdots & \alpha_2 \alpha_n \\
\vdots & \vdots & \ddots & \vdots \\
\alpha_n \alpha_1 & \alpha_n \alpha_2 & \cdots & \alpha_n^2
\end{pmatrix}
\]

(6)

But CMDFA solution of \( \Sigma_x \) may not always be rank 1, indicating star may not always be the optimum solution from common information point of view. It remains to be seen if CMDFA solution to \( \Sigma_x \) recovers the graphical model given by (3). Also to be investigated is the exact solution to CMDFA if it fails to recover the underlying star topology. In the rest of the paper, we will present both sufficient and necessary conditions under which the rank of the optimal \( \Sigma_x - D \) and the values of \( D \)'s entries are determined.

### IV. Solutions to CMDFA

In this section we present the detailed analysis of the CMDFA solution space of \( \Sigma_x \). We define the real column vector \( \vec{\theta} \in \mathbb{R}^n \) as \( \vec{\theta} = [\theta_1, \ldots, \theta_n]^T \) where \( \theta_i = \frac{|\alpha_i|}{\sqrt{1-\alpha_i^2}}, 1 \leq i \leq n \).

As we can see, each element in \( \vec{\theta} \) is equal to the square root of the signal to noise ratio (SNR) of the corresponding element of vector \( \vec{\alpha} \). The following order of the elements of \( \vec{\theta} \) is a necessary consequence of our assumption in (1),

\[
\theta_1 \geq \theta_2 \geq \cdots \geq \theta_n
\]

(7)

As we mentioned before, we are interested to find out if CMDFA low rank decomposition of \( \Sigma_x \) produces a rank 1 matrix. Next we analyse the solution space of CMDFA and find explicit conditions for both when the solution is rank 1 and when it is not. To start the proceedings we state Theorem 1 given in [6] that gives the necessary and sufficient condition for \( D^* \) to be the CMDFA solution of the decomposition given in (5).

**Theorem 1.** The matrix \( D^* \) is the CMDFA solution of \( \Sigma_x \) if and only if \( \lambda_{\min}(\Sigma_x - D^*) = 0 \), and there exists \( n \times r \) matrix \( T \) such that \( \vec{L}_{i,*} \in N(\Sigma_x - D^*), 1 \leq i \leq r \) and the \( ||\vec{L}_{i,*}||^2 = \frac{1}{D_{i,i}}, 1 \leq i \leq n \).

In the first of the two subsections of this section, we find the conditions under which CMDFA solution of \( \Sigma_x \) recovers the model given by (3) or equivalently speaking, find conditions under which CMDFA solution of \( \Sigma_x \) is the rank 1 matrix given.
by (6). In the other subsection, we show the detailed analysis on the existence and uniqueness of the CMDFA solution of \( \Sigma_x \), when the solution is not a rank 1 matrix.

### A. CMDFA Non-dominant Case

Here we analyse the conditions under which the CMDFA solution of \( \Sigma_x \) recovers a star structure. Lemma 1 sets the groundwork for the Theorem to follow.

**Lemma 1.** Non-dominance of vector \( \vec{\theta} \) given by (8) is a necessary condition for the existence of such \( n \times r \) matrix \( T \) that \( \vec{t}_{s,i} \in N(\Sigma_{t,N,D}) \), 1 \( \leq i \leq r \) and \( ||\vec{t}_{s,i}||^2 = \frac{1}{1-\alpha_i^2} \), 1 \( \leq j \leq n \).

\[
\theta_i \leq \sum_{i=2}^{n} \theta_i \tag{8}
\]

The proof of Lemma 1 and the associated geometric interpretation is given in [12]. We are now well equipped to state and prove the statement of Theorem 2 that has the main result of this subsection.

**Theorem 2.** CMDFA solution of \( \Sigma_x \) is \( \Sigma_{t,N,D} \) if and only if \( \vec{\theta} \) is non-dominant.

**Proof of Theorem 2:** Now we refer back to the necessary and sufficient condition for CMDFA solution at the beginning of this section given by Theorem 1. Since, \( \Sigma_{t,N,D} \) is rank 1, its minimum eigenvalue is 0. To complete the proof of Theorem 2, we only need to show the existence of matrix \( T \) such that the column vectors of \( T \) are in the null space of \( \Sigma_{t,N,D} \) and the \( L_2 \)-norm square of the \( r \)th row of \( T \) is \( \frac{1}{1-\alpha_i^2} \), 1 \( \leq i \leq n \).

Lemma 1 has already shown that, for the existence of such \( T \) non-dominance given by equation (8) is a necessary condition. Next we show, by constructing such a \( T \) matrix under the assumption of non-dominance of \( \vec{\theta} \), that non-dominance is also a sufficient condition. And that should complete the proof of Theorem 2.

It is straightforward to find the following basis vectors for the null space of \( \Sigma_{t,N,D} \):

\[
\vec{v}_1 = \begin{bmatrix} \frac{\alpha_1}{\alpha_1} \\ \frac{\alpha_1}{1} \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \vec{v}_2 = \begin{bmatrix} \frac{-\alpha_1}{\alpha_1} \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \quad \vec{v}_{n-1} = \begin{bmatrix} \frac{\alpha_1}{\alpha_1} \\ 0 \\ \vdots \\ 1 \end{bmatrix} \tag{9}
\]

We define matrix \( V \) so that its columns span the null space of \( \Sigma_{t,N,D} \):

\[
V = [\vec{v}_1 \, \cdots \, \vec{v}_{n-1}] \tag{10}
\]

where \( c_i = \frac{\vec{v}_i}{\sqrt{1-\alpha_i^2}}, \quad i = 2, \ldots, n \) and \( \vec{v}_i \in \{1,-1\} \).

The columns of \( V \) span the null space of \( \Sigma_{t,N,D} \). To construct our desired matrix \( T \), under the assumption of non-dominance of \( \vec{\theta} \), it will suffice for us to find a diagonal matrix \( B_{n \times n} \) such that the following holds.

\[
T_{n \times n} = V_{n \times n} \cdot B_{n \times n} \Rightarrow TT^T = VBB^TV = V\beta V^T \tag{11}
\]

We require the diagonal matrix \( \beta = BB^T \) to have only non-negative entries and the \( L_2 \)-norm square of the \( r \)th row of \( T \) be \( \frac{1}{1-\alpha_i^2} \). Based on the conditions imposed on the matrix \( T \), using (10) and (11) we have the following \( n \) equations,

\[
\frac{\alpha_1^2}{\alpha_1^2} \beta_{11} + \frac{\alpha_2^2}{\alpha_1^2} \beta_{22} + \cdots + \frac{\alpha_n^2}{\alpha_1^2} \beta_{nn} + \left( \frac{\alpha_2^2}{\alpha_1^2} + \frac{\alpha_3^2}{\alpha_1^2} + \cdots + \frac{\alpha_n^2}{\alpha_1^2} \right) \beta_{nn} = \frac{1}{1-\alpha_1^2} \tag{12}
\]

\[
\beta_{ii} + c_i^2 \beta_{nn} = \frac{1}{1-\alpha_i^2}, \quad i = 1, \ldots, n-1 \tag{13}
\]

Solving, (12) with the help of (13) we get,

\[
\beta_{nn} = \frac{1}{1-\alpha_1^2} - \frac{1}{1-\alpha_2^2} - \cdots - \frac{1}{1-\alpha_n^2} \sum_{i \neq j, i \neq 1, j \neq 1} c_i c_j \alpha_i \alpha_j \tag{14}
\]

It is straightforward to see that, to ensure all the \( \beta_{ii}, 1 \leq i \leq n \) are non-negative, we need \( \beta_{nn} \leq 1 \). We select \( \vec{v}_i, 2 \leq i \leq n \) such that,

\[
c_i \alpha_i = \frac{\vec{v}_i}{\sqrt{1-\alpha_i^2}} = \theta_i, \quad i = 2, \ldots, n \tag{15}
\]

Under such selection of \( \vec{v}_i, 2 \leq i \leq n \), \( \beta_{nn} \) becomes,

\[
\beta_{nn} = \frac{\theta_1^2 - \theta_2^2 - \cdots - \theta_n^2}{\sum_{i \neq j, i \neq 1, j \neq 1} \theta_i \theta_j} \tag{16}
\]

Now, using the non-dominance assumption given in (8), we have

\[
\theta_1^2 \leq \left( \sum_{i=2}^{n} \theta_i \right)^2 \Rightarrow \theta_1^2 - \sum_{i=2}^{n} \theta_i^2 \leq 1 \Rightarrow \beta_{nn} \leq 1
\]

Which means non-dominance of vector \( \vec{\theta} \) is a sufficient condition to construct the kind of \( T \) matrix we are looking for. That completes the proof of Theorem 2.

### B. Dominant Case

Having proved that the non-dominance of vector \( \vec{\theta} \) is a sufficient and necessary condition for CMDFA solution of \( \Sigma_x \) to recover a star structure, we are left with only the dominance case now i.e. \( \theta_1 > \sum_{i=2}^{n} \theta_i \). Under the dominant condition we want to show the existence of a rank \( n-1 \) solution of \( \Sigma_x \). Any solution we find will be unique, because CMDFA is a special type of the broader class of convex optimization problems defined in [7]. We still have to satisfy the same sufficient and necessary condition set by Theorem 1. Like the non-dominant case, for the matrix \( D^* \) to be the CMDFA solution of \( \Sigma_x \) under the dominant case, the minimum eigen value of \( \Sigma_x - D^* \) has to be \( \lambda_{\min}(\Sigma_x - D^*) = 0 \) and the \( L_2 \)-norm square of
the $i$th row of the null space matrix $T$ has to be $\frac{1}{D_{i,i}}$. The only difference with the non-dominant case is that, since our conjecture for the dominant case is an $n-1$ rank solution, the null space matrix $T$ will always be rank 1 i.e. a column vector. Mathematically speaking, we need to show the existence of $0 < a_i < 1, 1 \leq i \leq n$ such that the following orthogonality condition holds.

\[
\begin{bmatrix}
  a_1 & a_1 a_2 & a_1 a_3 & \ldots & a_1 a_n \\
  a_2 a_1 & a_2 & a_2 a_3 & \ldots & a_2 a_n \\
  \vdots & \vdots & \vdots & \ddots & \vdots \\
  a_n a_1 & a_n a_2 & a_n a_3 & \ldots & a_n \\
\end{bmatrix}
\begin{bmatrix}
  a
  \sqrt{1-a_i}
  \vdots \\
  c_i
\end{bmatrix}
= \begin{bmatrix}
  0 \\
  \vdots \\
  0
\end{bmatrix}
\] (17)

where $c_i \in \{-1,1\}$. Once we have such $a_i, 1 \leq i \leq n$ the $i$th diagonal element of the CMDFA solution matrix $D^*$ under the dominant case will be $1-a_i, 1 \leq i \leq n$. The above orthogonality relationship gives us the following $n$ equations.

\[
\frac{a_i c_i}{\sqrt{1-a_i}} + \sum_{j \neq i} \frac{a_i a_j c_j}{\sqrt{1-a_j}} = 0, 1 \leq i \leq n
\] (18)

Let $(i)$ denote the $i$th equation given by (18). Using the linear combination $a_{i+1} \times (i) - a_i \times (i+1), 1 \leq i \leq n$ gives us the following $n-1$ equations given by (19) which in turn implies the expression in (20) for some ratio $\mu$.

\[
a_{i+1} c_i \eta_i - a_i c_{i+1} \eta_{i+1} = 0, \quad 1 \leq i \leq n - 1
\] (19)
\[
\Rightarrow [c_1 \eta_1, \ldots, c_n \eta_n]^T = \mu [a_1, \ldots, a_n]^T
\] (20)

where $\eta_i = \frac{a_i - a_i^2}{\sqrt{1-a_i}}$, $1 \leq i \leq n$. Now plugging the expressions from (20) in any of the $n$ equations given by (18) we get,

\[
\sum_{i=1}^{n} \frac{1}{1 - \frac{a_i}{a_i}} = 1
\] (21)

It will suffice for us to prove the existence of $0 < a_i < 1, 1 \leq i \leq n$ such that (21) holds. From (20) and the definition of $\eta_i$ we see that, to find each $a_i, 1 \leq i \leq n$ we need to solve the following second order polynomial.

\[
a_i^2 + a_i a_i (\mu^2 - 2) + a_i^2 (\mu^2 - \mu^2) = 0, \quad 1 \leq i \leq n
\] (22)

If we solve equation (22) for each $a_i$ we will get a left root and a right root. Our initial conjecture is that the left root for $a_1$ and right roots for $a_2, \ldots, a_n$ that we get solving (22) will give us $0 < a_i < 1, 1 \leq i \leq n$ that satisfy (21). If we can prove that our conjecture is true, then that should be the only possible solution to (21) because of the uniqueness of solution to such convex optimization problems proved in [7]. Plugging in the left root for $a_1$, right roots for $a_2, \ldots, a_n$ in (21) we get us the following equation.

\[
1 + \frac{1}{2} \sum_{i=1}^{n} \frac{\alpha_i^2}{1 - \alpha_i^2} = \frac{|\alpha_1|}{\sqrt{1 - \alpha_1^2}} \frac{1}{\sqrt{1 - \alpha_1^2}} \frac{\alpha_1^2}{4 - \alpha_1^2} + \frac{1}{\mu^2}
\]

\[
- \sum_{i=2}^{n} \frac{|\alpha_i|}{\sqrt{1 - \alpha_i^2}} \frac{1}{\sqrt{1 - \alpha_i^2}} \frac{\alpha_i^2}{4 - \alpha_i^2} + \frac{1}{\mu^2}
\]

(23)

We define

\[
M_i = \frac{1}{\sqrt{1 + \frac{1}{\mu^2}} - \frac{\alpha_i^2}{1 - \alpha_i^2}} = \frac{1}{\sqrt{1 + \frac{1}{\mu^2}}}, \quad i = 1, 2, \ldots, n
\] (24)

Under these newly defined $M_i$s (23) becomes,

\[
\theta_i^2 M_1 - \sum_{i=2}^{n} \theta_i^2 M_i = 1 + \frac{1}{3} \sum_{i=1}^{n} \theta_i^2
\] (25)

And using the definition of $M_i, 1 \leq i \leq n$ given in (24), we get the following cylinders of hyperbolas.

\[
\theta_i^2 M_1^2 - \theta_i^2 M_i^2 = \frac{1}{4} (\theta_i^2 - \theta_i^2), \quad 2 \leq i \leq n
\] (26)

Equations given by (26) imply that for each value of $M_1$ we get a point $[M_1, M_2(M_1), \ldots, M_n(M_1)]$, in the $n$ dimensional space where each $M_i(M_1), 1 \leq i \leq n$ is a function of $M_1$. For the range of values of $(M_1, M_1, \ldots, M_n(M_1))$ all such points together produce an $n$ dimensional space curve. If we project this space curve on any of the two dimensional $(M_1, M_i), 2 \leq i \leq n$ planes we get a hyperbola.

Another important thing to note is that, each equation given by (26) is a cylinder of hyperbolas originated from $(M_1, M_i)$ plane and projected onto $n$ dimensional space. Each point in the space curve represents an intersection point of all $n-1$ cylinders of hyperbolas originated from $(M_1, M_i), 2 \leq i \leq n$ planes. The next theorem has our revised goal at this point summed up.

**Theorem 3.** There exists an intersection point among the plane given by (25) and the $n-1$ cylinders of hyperbolas given by (26), that satisfies $M_i > \frac{1}{2}, 1 \leq i \leq n$.

Proving the above Theorem would mean that, there exists $0 < a_i < 1, 1 \leq i \leq n$ such that (21) holds, which in turn would mean the existence of an $n-1$ rank CMDFA solution under the dominance of vector $\hat{\theta}$. Here we present an outline of the proof and a more detailed version is given in [12].

**Outline of the Proof of Theorem 3:** Let us define the function $G(.)$ of $M_1$ as the inner product between the vectors $[M_1, \ldots, M_n]$ and $[\theta_1^2, \ldots, \theta_n^2]^T$, where each $M_i, 1 \leq i \leq n$ is a function of $M_1$. Which means, $G(M_1) = \theta_1^2 M_1 - \sum_{i=2}^{n} \theta_i^2 M_i(M_1)$. So, our revised goal becomes to find the existence of such $M_1 > \frac{1}{2}$ for which the function value of $G(M_1)$ becomes $G(M_1) = 1 + \frac{1}{3} \sum_{i=1}^{n} \theta_i^2$

Equation (26) dictates that each $M_i(M_1), 2 \leq i \leq n$ is a concave function of $M_1 > \frac{1}{2}$, which makes $G(M_1)$ a convex
Fig. 1. Trend of the function $G(M_1)$ against $M_1$

Fig. 2. Difference of mutual information against $\theta_1$

function of $M_1$ as the sum of convex functions of $M_1$. Using (26) we get,

$$G \left( \frac{1}{2} \right) = \frac{1}{2} \left( \theta_1^2 - \sum_{i=2}^{n} \theta_i^2 \right)$$  \hspace{1cm} (27)

Using (24) we get,

$$\frac{dM_1(M_1)}{dM_1} = \frac{\theta_1^2 M_1}{\theta_i^2 M_i(M_1)} \Rightarrow dG(M_1) = \theta_1^2 \left[ 1 - \sum_{i=2}^{n} \frac{M_i(M_1)}{M_1} \right]$$

which is an increasing function of $M_1 > \frac{1}{2}$, because (26) dictates that $\frac{M_i(M_1)}{M_1}, 2 \leq i \leq n$ are decreasing functions of $M_1 > \frac{1}{2}$. We also have $\frac{dG(M_1)}{dM_1} \bigg|_{M_1=\frac{1}{2}} = -\theta_1^2(n - 2)$ which is a negative value.

Figure 1 has been drawn based on the above functional analysis. It can be seen in Figure 1 that $G(M_1)$ is an increasing function for the values $M_1 > \frac{\bar{M}}{2}$ and $G(M_1) < G \left( \frac{1}{2} \right) < 1 + \frac{1}{2} \sum_{i=1}^{n} \theta_i^2$. Hence, there must exist $M_1^* > \frac{\bar{M}}{2}$ such that $G(M_1^*) = 1 + \frac{1}{2} \sum_{i=1}^{n} \theta_i^2$.

While we can find the exact CMDFA solution $M_1^*$ by any suitable line search algorithm, as part of our groundwork to present our numerical data in [12], we analytically found an upperbound $M_1^{up}$ and a lowerbound $M_1^{low}$ to the actual solution $M_1^*$.

V. NUMERICAL RESULTS

We motivated CMDFA in terms common information which is a function of the minimum mutual information between the observables and the latent factors. Let $I_{\text{star}}, I_{\text{CMDFA}}, I^{up}$ and $I^{low}$ be the corresponding minimum mutual information between the latent variables and the observables for a star solution, CMDFA solution $M_1^*$, the upperbound $M_1^{up}$ and the lowerbound $M_1^{low}$. In general people tend to assume a star topology to find common information, hence any value of mutual information less than $I_{\text{star}}$ works to our advantage. Our numerical results for a 3 dimensional case demonstrated in Figure 2 shows that under the dominant case the star solution is not optimal. $I_{\text{star}} - I^{up}$ is an increasing function of $\theta_1$ indicates that the lower bound of the advantage of CMDFA solution over star increases as vector $\bar{\theta}$ becomes more and more dominant. We numerically calculated $I_{\text{CMDFA}}$ and the curve in Figure 2 gives the actual advantage that CMDFA sution has over star under the dominance of $\bar{\theta}$ whereas $I^{\text{star}} - I^{\text{low}}$ gives an upperbound to the actual advantage of CMDFA over a star topology. The gap between $I_{\text{star}} - I^{\text{low}}$ and $I_{\text{star}} - I^{\text{up}}$ is gradually increasing indicating $I^{\text{up}} - I^{\text{low}}$ is increasing with $\theta_1$. The analytical justification for this numerical data is given in [12].

REFERENCES

[1] Y. Chen, X. Li, and S. Zhang, “Structured latent factor analysis for large-scale data: Identifiability, estimability, and their implications,” arXiv preprint arXiv:1712.08966, 2017.

[2] D. Bertsimas, M. S. Copenhaver, and R. Mazumder, “Certifiably optimal low rank factor analysis,” Journal of Machine Learning Research, vol. 18, no. 29, pp. 1–53, 2017.

[3] A. A. Albert, “The matrices of factor analysis,” Proceedings of the National Academy of Sciences, vol. 30, no. 4, pp. 90–95, 1944.

[4] M. Drton, B. Sturmfels, and S. Sullivant, “Algebraic factor analysis: tetrads, pentads and beyond,” Probability Theory and Related Fields, vol. 138, no. 3–4, pp. 463–493, 2007.

[5] P. Bentler and J. A. Woodward, “Inequalities among lower bounds to reliability: With applications to test construction and factor analysis,” Psychometrika, vol. 45, no. 2, pp. 249–267, 1980.

[6] A. Moharrer and S. Wei, “Algebraic properties of solutions to common information of Gaussian graphical models,” in Communication, Control, and Computing (Allerton), 2017 55th Annual Allerton Conference on. IEEE, 2017.

[7] G. Della Riccia and A. Shapiro, “Minimum rank and minimum trace of covariance matrices,” Psychometrika, vol. 47, no. 4, pp. 443–448, 1982.

[8] A. Wyner, “The common information of two dependent random variables,” IEEE Transactions on Information Theory, vol. 21, no. 2, pp. 163–179, 1975.

[9] P. Cuff, “Communication requirements for generating correlated random variables,” in 2008 IEEE International Symposium on Information Theory. IEEE, 2008, pp. 1393–1397.

[10] G. Xu, W. Liu, and B. Chen, “A lossy source coding interpretation of Wyner’s common information,” IEEE Transactions on Information Theory, vol. 62, no. 2, pp. 754–768, 2016.

[11] S. Satpathy and P. Cuff, “Gaussian secure source coding and Wyner’s common information,” in Information Theory (ISIT), 2015 IEEE International Symposium on. IEEE, 2015, pp. 116–120.

[12] M. M. Hasan, S. Wei, and A. Moharrer, “Algebraic properties of Wyner common information solution under graphical constraints,” Louisiana State University, USA, Tech. Rep., 2019. [Online]. Available: https://arxiv.org/abs/1701.05952