Fifty-Fifth Annual Allerton Conference
Allerton House, UIUC, Illinois, USA
October 3-6, 2017

Algebraic Properties of Solutions to Common Information of Gaussian Vectors under Sparse Graphical Constraints

Ali Moharrer and Shuangqing Wei

Abstract— **We formulate Wyner's common information for random vectors $\mathrm{x} \in \mathbb{R}^n$ with joint Gaussian density. We show that finding the common information of Gaussian vectors is equivalent to maximizing a log-determinant of the additive Gaussian noise covariance matrix. We coin such optimization problem as a constrained minimum determinant factor analysis (CMDFA) problem. The convexity of such problem with necessary and sufficient conditions on CMDFA solution is shown. We study the algebraic properties of CMDFA solution space, through which we study two sparse Gaussian graphical models, namely, *latent Gaussian stars*, and *explicit Gaussian chains*. Interestingly, we show that depending on pairwise covariance values in a Gaussian graphical structure, one may not always end up with the same parameters and structure for CMDFA solution as those found via graphical learning algorithms. In addition, our results suggest that Gaussian chains have little room left for dimension reduction in terms of the number of latent variables required in their CMDFA solutions.**

Keywords: Factor Analysis, Common Information, Gaussian Graphs

I. Introduction

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 $C(X_1, X_2) = \min_{\substack{P_Y \\ X_1-Y-X_2}} I(X_1, X_2; Y)$, where $X_1 - Y - X_2$ represents conditional independence between X_1 and X_2, given Y, where the joint density function is sought to esnure such conditional independence, as well as the given joint density of X_1 and X_2. In other words, one may see Wyner's common information as the *optimal* way of generating random outputs, through which the number of common random bits to produce the desired output is minimized. Han and Verdu, in [1] along the same problem, define the notion of *resolvability* of a given channel, which is defined as the minimal required randomness to generate output statistics in terms of a vanishing total variation distance between the synthesized and prescribed joint densities. Resolvability of a channel is found to be a very intuitive description of common randomness in our settings, since it can be related to channel quality in terms of its noise power, and the noisier the channel the less number of common random bits needed to simulate the output [1]. Also, Cuff in [2] completely characterized the achievable rate regions needed to synthesize a memoryless channel, where he also used the total variation distance metric to show the quality of the proposed scheme.

There are several works that extend the classical bi-variate synthesis problem in Wyner's study to more general scenarios. A lower bound on the generalized Wyner's common information is obtained in [3]. In [4]–[6], the authors aim to define the common information of n dependent random variables, to further address the same question in this setting. In fact, the authors characterize the closed form solution for common information of Gaussian vectors with homogeneous (i.e., equal) pairwise correlation values. They resort to the same scenario as Wyner [7] did, i.e., considering one random variable Y to define such common randomness. Also, in [8] the authors completely characterize the common information between two jointly Gaussian vectors, as a function of certain singular values that are related to both joint and marginal covariance matrices of two Gaussian random vectors. However, they still divide the random vector into two groups, which makes it similar to Wyner's scenario.

In many cases, introducing one latent variable is not enough to capture the entire common information. Here, we motivate the *multi-variate* common information problem. Consider the following motivating example, showing that we need at least two latent random variables to capture common information.

Example 1: Suppose we are given a zero mean Gaussian random vector $\{X_1, X_2, X_3, X_4\}$ forming a Markov chain $X_1 - X_2 - X_3 - X_4$ with corresponding correlation matrix (normalized covariance matrix) $\Sigma = [\rho_{ij}]$, where $\rho_{ij},\ (i,j) \in [1,4]$ is a pairwise correlation between X_i and X_j. Also, to ignore infeasible or trivial cases, we need to have $\rho_{ij} \in (-1,1)$ and non-zero. The correlation space of Gaussian trees is fully characterized in [9]. It is also shown that in order to have a chain the pairwise correlations ρ_{ij} are the product of correlation values for those variables along the path from X_i to X_j [10]. Our objective is to show if it is possible to form a latent star-shape Gaussian graph by adding a single random variable Y given which all four X_i are conditionally independent. We only need to consider jointly Gaussian vector $(\mathbf{X}, Y)$, as it is shown in [6] that jointly Gaussian vector $(\mathbf{X}, Y)$ maximizes the conditional entropy $h(\mathbf{X}|Y)$ hence minimizing the mutual information $I(\mathbf{X}; Y)$. Assuming Y is a zero mean Gaussian random variable with variance σ_y^2, we may write the signal model $[X_1, X_2, X_3, X_4]' = [a_1, a_2, a_3, a_4]'Y + [Z_1, Z_2, Z_3, Z_4]'$, where the vector $A = [a_1, a_2, a_3, a_4]'$ is to be determined by given constraints in the problem and $\{Z_1, Z_2, Z_3, Z_4\}$

A. Moharrer and S. Wei are with the School of Electrical Engineering and Computer Science, Louisiana State University, Baton Rouge, LA 70803, USA. Email: amohar2@lsu.edu, swei@lsu.edu.

This material is based upon work supported in part by the National Science Foundation under Grant No. 1320351 and No.1642991.

are independent zero mean Gaussian noises with variance $\sigma_{z_i}^2$. From such signal model we may easily see that $\Sigma = AA'\sigma_y^2 + \Sigma_z$, where $AA'\sigma_y^2$ is a rank one positive semi-definite matrix and Σ_z is diagonal, with diagonal elements $\sigma_{z_i}^2$. We may move Σ_z to the other part of equation to show the matrix $\Sigma - \Sigma_z$ is a rank one positive semi-definite matrix. Such Hermitian matrix has the following form,

$$\Sigma - \Sigma_z = \begin{pmatrix} t_1 & \rho_{12} & \rho_{12}\rho_{23} & \rho_{12}\rho_{23}\rho_{34} \\ \rho_{12} & t_2 & \rho_{23} & \rho_{23}\rho_{34} \\ \rho_{12}\rho_{23} & \rho_{23} & t_3 & \rho_{34} \\ \rho_{12}\rho_{23}\rho_{34} & \rho_{23}\rho_{34} & \rho_{34} & t_4 \end{pmatrix} \tag{1}$$

where due to the fact that $\Sigma - \Sigma_z$ is positive semi-definite and $\sigma_{z_i}^2$'s are non-negative, each $t_i = 1 - \sigma_{z_i}^2,\ i \in [1, 4]$ is between 0 and 1. Since the matrix in (1) is rank one, hence we may pick one of the rows as a basis for the row space of this matrix. One may see that by choosing either the first or second row as a basis, we end up setting $\rho_{23} = \pm 1$, which we know is an infeasible value. Due to symmetry of chain structure, similar answers will be deduced by setting the third or fourth row as a basis. Hence, overall we reached to a contradictory conclusion: the matrix $\Sigma - \Sigma_z$ cannot be a rank one matrix.

This simple case study shows that depending on the covariance matrix structure, we may need a Gaussian random vector (instead of a single random variable) to capture the common information among variables with certain structures.

Recently, Veld and Gastpar [11] characterized common information problem for this general setting, and formulated the problem as a specific instance of maximum determinant (*maxdet*) [12] problems. They also analytically computed the common information value for a specific set of Gaussian joint densities with *circulant* covariance matrices. Steeg *et. al*, in [13] defined *information sieve*, which is closely related to common information metric to represent deep latent Gaussian structures. They showed that in many applications, such as *Blind Source Separation* (BSS), the intrinsic latent structure consists of more than a single variable, and that a multi-variate notion of common information is necessary to discover all the latent Gaussian sources. In our previous works, [14], [15] we addressed such multi-variate latent structure in a special case of Gaussian trees. We proved that for such cases, the common information restricted to the underlying tree structure is equal to the mutual information between observed variables and the latent variables.

Similar to these works, we in this paper first show that in a Gaussian case the common information problem is equivalent to minimizing the negative of log-determinant function of the additive Gaussian noise covariance matrix, under certain constraints. We show the relation of such problem to the classical *constrained minimum trace factor analysis* (CMTFA) problem [16]–[18], where the objective is to minimize the trace of an additive Gaussian noise covariance matrix. Therefore, we name the common information problem as *constrained minimum determinant factor analysis* (CMDFA). Rather than proposing a numerical algorithm for solving such convex programming problem (which as we discuss, there are certain algorithms for numerically solving maxdet problem), we focus on studying the algebraic features of the solution space of CMDFA problem in general, and specifically for several case studies, where Σ follows certain latent (or explicit) graphical structure.

The paper is organized as follows. In section II we give a proper formulation of CMDFA problem. In section III we show the solution space of CMDFA problem, and study couple of special cases for $n = 3$. Finally, we conclude the paper in section IV.

II. Problem Formulation

We may straightforwardly generalize Wyner's common information into multi-variable setting as follows,

$$C(\mathbf{X}) = \min_{p_{\mathbf{Y}}(\mathbf{y}), p(\mathbf{X}|\mathbf{Y})} I(\mathbf{X};\mathbf{Y}) \quad s.t. \quad p(\mathbf{X}|\mathbf{Y}) = \prod_{i=1}^{n} p(X_i|\mathbf{Y}) \tag{2}$$

where $\mathbf{X} \in \mathbb{R}^n$ is a Gaussian vector with covariance matrix $\Sigma_{\mathbf{x}}$ (without loss of generality we can set $\mu_{\mathbf{x}}$ to a zero vector), and $\mathbf{Y} \in \mathbb{R}^k$ ($k \leq n$) is the auxiliary (latent) random vector (a single random variable, in a special case) capturing common randomness in $\mathbf{X}$. Also, $I(\mathbf{X};\mathbf{Y})$ captures the mutual information between these two vectors. The only constraint is the conditional independence of all $X_i \in \mathbf{X}$ given the latent vector $\mathbf{Y}$.

We know $I(\mathbf{X};\mathbf{Y}) = h(\mathbf{X}) - h(\mathbf{X}|\mathbf{Y})$, with $h(.)$ being the differential entropy, since given $\Sigma_{\mathbf{x}}$, the first term is fixed. The common information problem is equivalent to maximizing the conditional entropy $h(\mathbf{X}|\mathbf{Y})$ with conditional independence constraint. It is shown in [6] that a jointly Gaussian latent vector $\mathbf{Y}$ can maximize such quantity, hence, we can limit the search space of problem to Gaussian $p_{\mathbf{Y}}$'s. Let us define an affine model $\mathbf{X} = A\mathbf{Y} + \mathbf{Z}$, where A is $n \times k$ transition matrix and $Z \in \mathbb{R}^n$ is a zero mean additive Gaussian noise vector, with diagonal covariance matrix D (hence, all $z_i \in \mathbf{Z}$ are independent), where the diagonal elements d_i are the corresponding variances for each z_i. The noise elements are independent of the latent vector $\mathbf{Y}$. We assume that the generative (affine) model's parameters, i.e., the transition matrix A_G and the diagonal covariance matrix D_G are known to us, either a priori or through a specific learning algorithm [19], [20]. Using such affine mapping we also satisfy the conditional independence constraint. As a result, one may show that $I(\mathbf{X};\mathbf{Y}) = \frac{1}{2}\log\frac{|\Sigma_{\mathbf{x}}|}{|D|}$. We may re-write the common information problem in (2) as follows,

$$\min_{D} -\log|D|,\ \ s.t. \quad \begin{cases} D \succ 0 \ and\ diagonal \\ \Sigma_{\mathbf{x}} - D \succeq 0 \end{cases} \tag{3}$$

We coin the optimization problem in (3) as CMDFA. The matrix D has to be positive definite, i.e., all diagonal entries d_i should be positive. Otherwise, if for some i, d_i is zero, then we know $-\log|D| = -\log\prod_{i=1}^{n} d_i \to +\infty$, which maximizes the objective function. The second constraint is due to affine modeling: $\Sigma_{\mathbf{x}} = AA' + D$, where A' is the transpose of A. Therefore, $\Sigma_{\mathbf{x}} - D = AA' \succeq 0$. In particular, the rank of AA' is at most $k \leq n$. It can be easily shown that CMDFA is an instance of general class of optimization problems known as *max-det* problems [12]. Hence, several iterative algorithms proposed in [12] can be used to numerically find the solution for such optimization problem. In fact, a Matlab-based modeling system for convex optimization, known as CVX [21], [22] can be used to solve such problem. Rather, our goal is to study the algebraic properties of CMDFA solution space in general, and for certain Gaussian graphical structures.

From now on, for simplicity and without loss of generality, we assume $\Sigma_{\mathbf{x}}$ to be a correlation matrix, i.e., all $\sigma_{ij} \in \Sigma_{\mathbf{x}}$ are normalized to $\rho_{ij} = \sigma_{ij}/\sqrt{\sigma_{ii}\sigma_{jj}}$. As a result, due to the constraint $\Sigma_{\mathbf{x}} - D \succeq 0$, for all $d_i \in D$ we have $d_i < 1$. This would be fine, since it is shown [12][p. 3] that such problem is invariant under congruence transformations. Hence, once the solution D_1 for the correlation matrix is found, one may propose $D_2 = \Lambda^{1/2} D_1 \Lambda'^{1/2}$ (Λ is a diagonal matrix with $\lambda_i = \sigma_{ii}$) as a solution to the un-normalized CMDFA problem.

III. Main Results

In this section, we first give the necessary and sufficient conditions under which D^* can be the solution to CMDFA problem. The proof procedure is very similar to CMTFA proof proposed in [16], [17]. Then, we aim to characterize the solution in certain cases, where the Gaussian density $p_{\mathbf{X}}$ follows either a latent or explicit Gaussian tree structure.

In Theorem 1 whose proof can be found in Appendix I we characterize the conditions under which D^* is the solution to CMDFA.

Theorem 1: The diagonal positive definite matrix D^* is a solution to CMDFA problem if and only if $|\Sigma - D^*| = 0$ and there exists a Gramian matrix $T = [t_{ij}] \succeq 0$, whose entries satisfy the following condition,

$$\frac{1}{\mathbf{d}^*} = \sum_{i=1}^{n-k} \sum_{j=1}^{n-k} t_{ij} \mathbf{e}_i \mathbf{e}_j^* \tag{4}$$

where $1/\mathbf{d}^* = [1/d_1, ..., 1/d_n]'$ and $\mathbf{e}_i \in N(\Sigma - D^*)$ is a basis vector in null space of $\Sigma - D^*$. The notation $\mathbf{e}_i \mathbf{e}_j^* = [e_{i1}e_{j1}, ..., e_{in}e_{jn}]'$ is used for the Hadamard product of two basis vectors.

Remark 1: The rank deficiency constraint in Theorem 1, suggests the solution D^* to be always on the boundary $|\Sigma - D^*| = 0$. This is a necessary condition for CMDFA solution. Otherwise, assume D^* is such that $\Sigma - D^*$ is a full rank. As a result, all n principal minors of this matrix should be positive. However, we know that each of these principal minors are polynomial functions of d_i^*'s. We may propose another matrix $\tilde{D} = diag(d_1^* + \epsilon_1, ..., d_n^* + \epsilon_n)$, where $\epsilon \geq 0$ for all i. Due to smoothness of such polynomial functions, we can always find at least one ϵ_i (although very small) such that still all principal minors of $\Sigma - \tilde{D}$ remain positive, hence, keeping the matrix positive definite. However, now D^* cannot be CMDFA solution, since apparently $-\log|\tilde{D}| < -\log|D^*|$, a contradiction. Therefore, the rank of $\Sigma - D^*$ should be $k \leq n - 1$.

Remark 2: Then, one may question the existence of CMDFA solution, i.e., whether all postive definite matrices Σ can be decomposed into sum of $AA' + D$, where $AA' \succeq 0$ is rank deficient and $D^* \succ 0$ is diagonal. To show the existence of solution, define $\lambda_{min} > 0$ as the smallest eigenvalue of Σ. Then, considering the matrix $D = \lambda_{min} I \succ 0$, where I is $n \times n$ identity matrix, we know $\Sigma - D$ is both positive semi-definite and rank deficient (its minimum eigenvalue is zero). Hence, for any given matrix Σ the search space of CMDFA problem is nonempty.

In Theorem 2, whose proof can be found in Appendix II we take the same steps as in [17][Theorem 4] to show the uniqueness of CMDFA solution.

Theorem 2: The CMDFA problem has a unique solution.

In the remainder of the paper we go through several cases to study their solution space regarding the CMDFA problem.

A. Star Structure

Suppose in the underlying affine model, the latent vector is a singleton Y, i.e., star structure. This can be modeled as

$$\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} Y + \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_n \end{bmatrix} \tag{5}$$

where $0 \leq a_i < 1,\ i \in [1, n]$. A special case for such model, with $n = 3$ is shown in Figure 1.

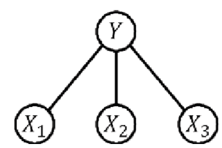

Fig. 1: Star structure with $n = 3$ outputs

Using Theorem 1 we may be able to characterize the solution space of CMDFA problem as follows. Note that the pairwise correlations have the form $\rho_{ij} = a_i a_j,\ i \neq j \in [1, n]$. Basically, each a_i can be seen as an edge weight between a latent factor Y and the corresponding variable X_i.

For all $d_i \in D$ define $t_i = 1 - d_i$, where for CMDFA problem we know $0 \leq t_i < 1$.

In our previous studies [14], [15], we showed that the value of each t_i is given and it is equal to a_i^2. Now, one can easily check that $rank(\Sigma - D) = 1$, also, $Trace(\Sigma - D) = \sum_{i=1}^{n} a_i^2 = \sum \lambda_i > 0$, where λ_i is an eigenvalue of $\Sigma - D$, but since $rank(\Sigma - D) = 1$, hence, $\lambda_2 = ... = \lambda_n = 0$, hence, $\lambda_1 = \sum_{i=1}^{n} a_i^2 > 0$, so $\Sigma - D$ is positive-semidefinite in this case. Also, since $d_i = 1 - a_i^2 > 0$, hence $D \succ 0$. Therefore, such case lies on the feasible solution of CMDFA.

To check if this is indeed the solution of CMDFA, we first need to find the null space of $\Sigma - D$, which has rank $n - 1$. By solving the system of equations $(\Sigma - D).\mathbf{V} = \mathbf{0}$, we deduce that the null space has the form $N(\Sigma - D) =$

$\{\mathbf{V}|\sum_{i=1}^{n} a_i v_i = 0\}$. In other words, the vectors in null space have the form $\mathbf{V} = (v_1, v_2, ..., -\frac{1}{a_n}\sum_{i=1}^{n-1} a_i v_i)$.

One intuitive suggestion for the basis would be choosing the set of linearly independent vectors $e_1 = (1, 0, 0, ..., 0, -\frac{a_1}{a_n}), ..., e_n = (0, 0, ..., 0, 1, -\frac{a_{n-1}}{a_n})$. To find a Gramian matrix $T = [t_{ij}]$ satisfying(4), we obtain the following system of equalities

$$\begin{cases} t_{ii} = \dfrac{1}{1-a_i^2}, \ i \in [1, n-1] \\ 2\sum_{i<j} t_{ij} a_i a_j = \dfrac{a_n^2}{1-a_n^2} - \sum_{i=1}^{n-1} \dfrac{a_i^2}{1-a_i^2}, \ i, j \in [1, n-1] \end{cases} \tag{6}$$

Remark 3: Suppose, all $a_i = a$ are equal. This is the case considered in [4]–[6]. Then, using (6) we obtain $t_{ii} = \frac{1}{1-a^2}$, $i \in [1, n-1]$ and $\sum_{i<j} t_{ij} = -\frac{n-2}{2(1-a^2)}$, $i, j \in [1, n-1]$. Simply putting all t_{ij}'s to be equal, gives $t_{ij} = -\frac{n-2}{n(n-1)(1-a^2)}$, $i \neq j$. In this case $T = [t_{ij}]$ is a strictly digonally dominant matrix, since $|t_{kk}| > \sum_{i \neq k} |t_{ki}| = \frac{(n-2)^2}{n(n-1)(1-a^2)}$. And since all of its diagonal entries are positive, hence T is in fact positive semi-definite. As it can be seen this is a special case that satisfies the system of equalities and shows that CMDFA solution for a such affine model with single hidden variable (i.e., a star), is a star!

One may wonder if the above system of linear equations always holds regardless of given values for a_i's. In other words, does always exist a Gramian matrix T satisfying the following system of equalities? So that the CDMFA solution of a given Gaussian vector, which was generated using a star-generative and latent Gaussian graph, also ends up with a star? Through the following case study we show that this is not the case, even for the smallest star tree with $n = 3$ output variables.

B. CMDFA solution space for $n = 3$

Here, we consider a special case, where the set of output variables is a three dimensional vector $\mathbf{X} = \{X_1, X_2, X_3\}$. As we will, although this seems a small number of variables to begin with, but finding CMDFA solution proves to be a non-trivial task.

1) Star: Rank one matrix: Suppose, $\Sigma - D$ is a rank one matrix, i.e., the latent vector is a singleton Y. Here, we draw an interesting conclusion, that the solution to CMDFA for such affine model, is not necessarily a star. This is shown in Theorem 3, whose proof can be found in Appendix III.

Theorem 3: For $n = 3$, and a rank one $\Sigma - D$ following the affine model in (5), the CMDFA solution is also a star with the same parameters if and only if the following inequality holds,

$$\begin{aligned} S_1 = \{s_1, s_2, s_3 | (s_1 - s_2)^2 + (s_1 - s_3)^2 + (s_2 - s_3)^2 \\ \leq s_1^2 + s_2^2 + s_3^2\} \end{aligned} \tag{7}$$

where $s_i = \frac{a_i^2}{1-a_i^2}$, $i \in [1, 3]$.

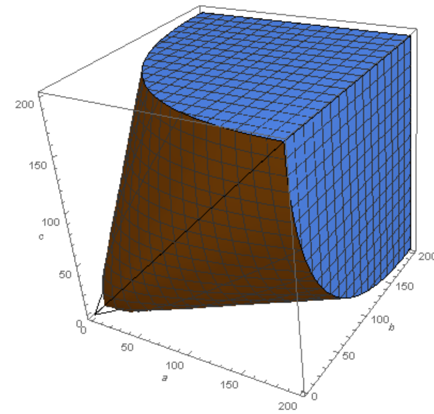

Fig. 2: The feasible region for SNR values s_i which can vary from zero to infinity

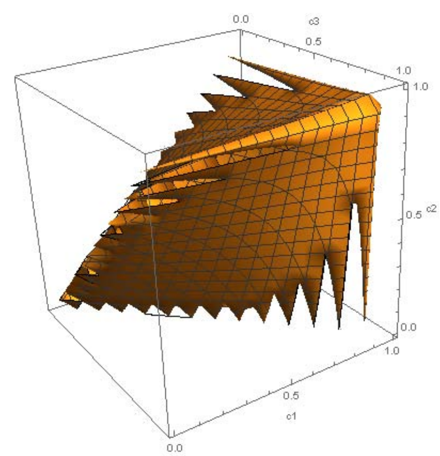

Fig. 3: The feasible region for positive edge weight values a_1, a_2, and a_3 which can vary from zero to one

Hence, in some cases, despite the fact that a latent tree induces a star structure, but the CMDFA solution is not necessarily a star.

Remark 4: s_i's can be seen as Signal to Noise Ratio (SNR) of the three Gaussian channels in the affine model. And the feasible region can be re-written as $(\sqrt{s_1} - \sqrt{s_2})^2 \leq s_3 \leq (\sqrt{s_1} + \sqrt{s_2})^2$. The following figures show the feasible region in both different SNR and (Positive) Edge weights domains.

It is noteworthy to mention that the general case considered for example in [4] is a special case for this region, i.e., the diagonal line inside the region where $s_1 = s_2 = s_3$ or equivalently $|a_1| = |a_2| = |a_3|$.

Using previous remarks, since we know the CMDFA solution is rank-deficient, hence, we have the following corollary.

Corollary 1: For $n = 3$, and a rank one $\Sigma - D$ following the affine model in (5), the CMDFA solution $\Sigma - D^*$ is rank two if and only if all s_i's lie on S_1^c, i.e., the complement of S_1.

Remark 5: In [23] the lower bound on the rank of $\Sigma - D$

is shown to be the total number of eigenvalues $\lambda_i \in \Sigma$ that are greater than 1. Supposing the 2×2 principal submatrix of Σ, its eigenvalues are $\lambda_1' = 1 - |\rho_{12}|$ and $\lambda_2' = 1 + |\rho_{12}|$, where such eigenvalues *interlace* [24] the eigenvalues $\lambda_i,\ i \in \{1,2,3\}$ of Σ and we have the ordering $\lambda_1 \leq \lambda_1' \leq \lambda_2 \leq \lambda_2' \leq \lambda_3$. The lower bound on λ_2 can be made tighter by considering the intrinsic symmetry in Σ structure, hence having a lower bound $\max 1 - |\rho_{ij}|,\ i \neq j$, i.e., $1 - \min |\rho_{12}|, |\rho_{13}|, |\rho_{23}| \leq \lambda_2$. Now, considering the set S_1, we may see that if one of the SNR values s_i dominates the other two, then (by ignoring the other two SNR values) we may not satisfy the inequality in (7). This may happen for example when a single edge-weight a_i is large, i.e., when there is one small correlation value ρ_{ij}, which also will dominate the lower bound on λ_2. As such correlation value decreases, the lower bound becomes closer to 1. Roughly speaking, in this case the search space of rank one matrices $\Sigma - D$ for CMDFA either shrinks or disappears. This might be one reason on having rank two solution on S_1^c in this case.

2) Non-star: Rank two matrix: Suppose this time that in the affine model the matrix $\Sigma - D$ is a rank two matrix, i.e., $\mathbf{Y} = \{Y_1, Y_2\}$. Hence, the row space of such matrix is two dimensional.

In the following Theorem, whose proof can be found in Appendix IV, we characterize the solution space of such matrices for rank two CMDFA solutions.

Theorem 4: For $n = 3$, and a rank two $\Sigma - D$ following the affine model in (5), the CMDFA solution is also a rank two matrix if and only if the following system of non-linear equations holds,

$$\begin{cases} d_3 = 1 + \dfrac{(d_2 - 1)\rho_{13}^2 + 2\rho_{12}\rho_{13}\rho_{23} + (d_1 - 1)\rho_{23}^2}{(1 - d_1)(1 - d_2) - \rho_{12}^2} \\ \dfrac{d_3}{d_1} = \alpha^2 \\ \dfrac{d_3}{d_2} = \beta^2 \end{cases} \tag{8}$$

where the parameters α, β are functions of (d_1, d_2) as follows,

$$\alpha = \frac{\rho_{12}\rho_{23} - \rho_{13}(1 - d_2)}{(1 - d_1)(1 - d_2) - \rho_{12}^2}$$
$$\beta = \frac{\rho_{12}\rho_{13} - \rho_{23}(1 - d_1)}{(1 - d_1)(1 - d_2) - \rho_{12}^2} \tag{9}$$

Similarly, we have the following Corollary,

Corollary 2: For $n = 3$, and a rank two $\Sigma - D$ following the affine model in (5), the CMDFA solution $\Sigma - D$ is rank one if and only if all d_i's lie on S_2^c, where S_2 is the solution set obtained from Theorem 4.

Remark 6: The results in Corollaries 1 and 2 interestingly show the difference between the affine models and CMDFA solutions. We may see that regardless of the rank of generative model, i.e., dimension of latent $\mathbf{Y}$ vector, the CMDFA solution can have either lower or higher dimensions. This shows that, in many situations and depending on the values of the transition matrix A, the generative affine model might not be the optimal one (in terms of achieving minimum number of common randomness) to use in order to generate the Gaussian output vector $\mathbf{X}$. Such generative models are usually learned by a specific learning algorithm. For example, for Gaussian latent trees there are efficient algorithms such as Chow-Liu Recursive Grouping (CLRG) [19] and Neighbor Joining [20] algorithm that can consistently learn both the tree structure and parameters.

Remark 7: While CMDFA problem is similar to CMTFA problem with different cost functions (and accepting zero solutions for d_i's), note that their solution sets are different and for the special case of $n = 3$, they are exclusive. To show this, suppose for CMTFA solution all d_i^*'s are non-zero, then all we need to do is to replace the left hand side of (4) with $[1, 1, 1]'$. Essentially, this is because in CMTFA the objective function to be minimized is negative of $Trace(\Sigma - D) = \sum_{i=1}^{3}(1 - d_i)$. Hence, its gradient with respect to $\mathbf{d}$ becomes $[-1, -1, -1]'$. Now, in CMDFA this means replacing the left hand side of (4) to $1/\mathbf{d}^* = [1, 1, 1]'$, i.e., $d_1^* = d_2^* = d_3^* = 1$. But then $\Sigma - D^*$ obtains the following form:

$$\Sigma - D^* = \begin{pmatrix} 0 & \rho_{12} & \rho_{13} \\ \rho_{12} & 0 & \rho_{23} \\ \rho_{13} & \rho_{23} & 0 \end{pmatrix} \tag{10}$$

In other words $Tr(\Sigma - D^*) = 0$, now if the eigenvalues of $\Sigma - D^*$ are non-zero, then they should have different signs, hence, making $\Sigma - D^*$ non-positive definite, and not Gramian. If they are all zero, then this matrix has a rank zero, and again it violates the Gramian assumption. Note that we assume in CMDFA, the solutions d^* are non-zero, but this cannot be the answer to CMTFA (since then for equivalence of the solutions, we should set all $d_i^* = 1$). Hence, in this case we conclude that the answers for CMTFA and CMDFA are exclusive.

C. Markov Chain

In [14], [15] we showed an operational approach under which any latent Gaussian tree can be efficiently synthesized. We showed that the sources of common randomness can be shrinked into a set of latent variables $\mathbf{X} = \{X_1, ..., X_k\}$ forming a Markov chain structure $X_1 - X_2 - ... - X_k$. Here, we want to show that such chain structure can be efficiently synthesized by a smaller set of variables through an affine Gaussian model. In Theorem 5, whose proof can be found in Appendix V we show these results.

Theorem 5: Supposing a Gaussian vector $\mathbf{X} \in \mathbb{R}^n$, with Σ following a Markov chain structure $X_1 - X_2 - ... - X_n$, the CMDFA solution $\Sigma - D^*$ has rank either $n - 1$ or $n - 2$.

In other words, we can always reduce the number of common random bits required to synthesize chain structure through a latent common variables $\{Y_1, ..., Y_k\}$, where $k \in \{n - 2, n - 1\}$. This also shows that comparing to affine models inducing star structures, such Markov chain structures cannot be significantly made simpler through sum of lower rank and diagonal matrices.

IV. Conclusion

In this paper, we studied the problem of characterizing Wyner's common information for Gaussian vectors following special structures, such as star or a Markov chain. We showed that how such problem can be turned into a specific convex programming problem, which we coined as CMDFA. For a general star Gaussian tree, we obtained the linear system of equations that can be efficiently solved to find the CMDFA solution. For $n=3$ and star Gaussian tree, we showed that interestingly the CMDFA solution can be a rank two matrix. This resulted in computing the general solution space for such case, in which it consists previous solutions as a special case. Finally, for a Gaussian Markov chain we showed that unlike star affine models, these vectors cannot be made as compact such that the lower rank matrix $\Sigma - D$ can be made at most having rank $n-2$, which suggests that there is not much degree of freedom left to further reduce the model complexity for a Gaussian chain structure.

References

[1] T. S. Han and S. Verdú, "Approximation theory of output statistics," *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 752–772, 1993.

[2] P. Cuff, "Distributed channel synthesis," *Information Theory, IEEE Transactions on*, vol. 59, no. 11, pp. 7071–7096, 2013.

[3] Q. Chen, F. Cheng, T. Liu, and R. W. Yeung, "A marginal characterization of entropy functions for conditional mutually independent random variables (with application to wyner's common information)," in *IEEE International Symposium on Information Theory (ISIT)*, 2015, pp. 974–978.

[4] G. Xu and B. Chen, "Information for inference," in *Communication, Control, and Computing (Allerton), 2011 49th Annual Allerton Conference on*. IEEE, 2011, pp. 1516–1520.

[5] P. Yang and B. Chen, "Wyner's common information in gaussian channels," in *IEEE International Symposium on Information Theory (ISIT)*, 2014, pp. 3112–3116.

[6] G. Xu, W. Liu, and B. Chen, "A lossy source coding interpretation of wyner's common information," *Information Theory, IEEE Transactions on*, vol. 62, no. 2, pp. 754–768, 2016.

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

[8] S. Satpathy and P. Cuff, "Gaussian secure source coding and wyner's common information," *arXiv preprint arXiv:1506.00193*, 2015.

[9] N. Shiers, P. Zwiernik, J. A. Aston, and J. Q. Smith, "The correlation space of gaussian latent tree models," *arXiv preprint arXiv:1508.00436*, 2015.

[10] A. Moharrer, S. Wei, G. Amariucai, and J. Deng, "Extractable common randomness from gaussian trees: Topological and algebraic perspectives," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 10, pp. 2312–2321, 2016.

[11] G. J. Op't Veld and M. C. Gastpar, "Caching gaussians: Minimizing total correlation on the gray-wyner network," in *Information Science and Systems (CISS), 2016 Annual Conference on*. IEEE, 2016, pp. 478–483.

[12] L. Vandenberghe, S. Boyd, and S.-P. Wu, "Determinant maximization with linear matrix inequality constraints," *SIAM journal on matrix analysis and applications*, vol. 19, no. 2, pp. 499–533, 1998.

[13] G. V. Steeg, S. Gao, K. Reing, and A. Galstyan, "Sifting common information from many variables," *arXiv preprint arXiv:1606.02307*, 2016.

[14] A. Moharrer, S. Wei, G. T. Amariucai, and J. Deng, "Layered synthesis of latent gaussian trees," *arXiv preprint arXiv:1608.04484*, 2016.

[15] ——, "Synthesis of gaussian trees with correlation sign ambiguity: An information theoretic approach," *arXiv preprint arXiv:1601.06403*, 2016.

[16] A. Shapiro, "Rank-reducibility of a symmetric matrix and sampling theory of minimum trace factor analysis," *Psychometrika*, vol. 47, no. 2, pp. 187–199, 1982.

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

[18] A. Shapiro, "Weighted minimum trace factor analysis," *Psychometrika*, vol. 47, no. 3, pp. 243–264, 1982.

[19] M. J. Choi, V. Y. Tan, A. Anandkumar, and A. S. Willsky, "Learning latent tree graphical models," *The Journal of Machine Learning Research*, vol. 12, pp. 1771–1812, 2011.

[20] N. Saitou and M. Nei, "The neighbor-joining method: a new method for reconstructing phylogenetic trees." *Molecular biology and evolution*, vol. 4, no. 4, pp. 406–425, 1987.

[21] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," http://cvxr.com/cvx, Mar. 2014.

[22] ——, "Graph implementations for nonsmooth convex programs," in *Recent Advances in Learning and Control*, ser. Lecture Notes in Control and Information Sciences, V. Blondel, S. Boyd, and H. Kimura, Eds. Springer-Verlag Limited, 2008, pp. 95–110, http://stanford.edu/~boyd/graph_dcp.html.

[23] L. Guttman, "Some necessary conditions for common-factor analysis," *Psychometrika*, vol. 19, no. 2, pp. 149–161, 1954.

[24] S. Fisk, "A very short proof of cauchy's interlace theorem for eigenvalues of hermitian matrices," *arXiv preprint math/0502408*, 2005.

Appendix I
Proof of Theorem 1

First, we show the convexity of CMDFA problem. We know that the negative log-determinant function is convex. Moreover, the constraint $D \succ 0$ can be written as n linear constraints of the form $d_i > 0,\ i \in [1,n]$, where d_i's are non-zero diagonal entries of D. The constraint $\Sigma - D \succeq 0$ is equivalent to having its non-negative minimum eigenvalue $\lambda(\mathbf{d})$, or equivalently, $-\lambda(\mathbf{d}) \leq 0$. It is proven in [17, Lemma 1] that in fact finding the negative of minimum eigenvalue corresponds to maximizing a set of linear functions, hence making $-\lambda_{min}(\mathbf{d})$ a convex (piecewise linear)function. So, overall CMDFA is a convex optimization problem.

By relying on [17, Theorem 2] (KKT necessary and sufficient conditions) we obtain that $\mathbf{d}^*$ is a solution to the CMDFA problem defined in (3), if and only if $\lambda(\mathbf{d}^*) = 0$, $d_i > 0,\ i \in [1,n]$ and we have the following (with $\alpha \geq 0$)

$$0 \in \nabla(-\log|D|)|_{D=D^*} + \alpha\partial(-\lambda(\mathbf{d}))|_{\mathbf{d}=\mathbf{d}^*} \tag{11}$$

Since all d_i's are non-zero (positive), the gradient $\nabla(-\log|D|)|_{D=D^*}$ can be easily replaced as $\frac{1}{\mathbf{d}^*} = [1/d_1, ..., 1/d_n]'$. Note that the minimum eigenvalue function is piecewise linear, hence, non-smooth. Since such function is maximum over linear functions, it is shown in [17, Lemma 1] that the subdifferential term can be written as $\partial(-\lambda(\mathbf{d})) = \bar{conv}\{\mathbf{x}^2\}$, where $\bar{conv}\{S\}$ denotes the convex hull of set S and the vectors $\mathbf{x}$ are unit eigenvectors of $\Sigma - D$ corresponding to $\lambda(\mathbf{d})$. The term $\mathbf{x}^2$ is the Hadamard product of the vector $\mathbf{x}$ with itself. However, note that as the solution is on the boundary $\lambda(\mathbf{d}^*) = 0$, hence, such eigenvectors in fact correspond to the null space, $N(\Sigma - D)$. Therefore, we rewrite (11) as,

$$0 = -\frac{1}{\mathbf{d}^*} + \Sigma_{i=1}^m \mathbf{x}_i^2 \tag{12}$$

where $\mathbf{x}_i \in N(\Sigma - D)$. Note that m can be arbitrary with $m(m+1)/2 < n$ [16].

Due to [16, Lemma 3.1] one can replace the summation $\Sigma_{i=1}^m \mathbf{x}_i^2$ in (12) with weighted sum of basis vectors $\mathbf{e}_i \in$

$N(\Sigma-D),\ i\in[1,k]$, with the form $\Sigma_{i=1}^{k}\Sigma_{j=1}^{k}t_{ij}\mathbf{e}_i\mathbf{e}_j^*$. This is due to the fact that any vector $\mathbf{x}\in N(\Sigma-D)$ can be also written as a linear combination of the basis vectors $\mathbf{e}_i\in N(\Sigma-D)$, i.e., there exists $(n-k)\times m$ matrix C such that $[\mathbf{x}_1,\mathbf{x_2},...,\mathbf{x}_m]=[\mathbf{e}_1,\mathbf{e}_2,...,\mathbf{e}_{n-k}]C$. Then, the summation in (12) can be replaced with $\Sigma_{i=1}^{k}\Sigma_{j=1}^{k}t_{ij}\mathbf{e}_i\mathbf{e}_j^*$, where $T=[t_{ij}]=CC'$ and hence it has to be Gramian. Hence, we have the desired results in Theorem 1.

Appendix II
Proof of Theorem 2

Similar to [17, Theorem 4], we may assume that the CMDFA solution $\mathbf{d}_1$ is not unique. Hence, there is another solution $\mathbf{d}_2$, where both of them lie on the boundary $\lambda(\mathbf{d}_1)=\lambda(\mathbf{d}_1)=0$. From the convexity of the problem, which is shown in Appendix I, the vector $\mathbf{d}_3=(\mathbf{d}_1+\mathbf{d}_2)/2$ should also be the solution. By [17, Corollary 1] there should be $\mathbf{x}\in N(\Sigma-D_3)$ with non-zero coordinates. We know $N(\Sigma-D_3)\mathbf{x}=0$. Hence, replacing $\mathbf{d}_3=(\mathbf{d}_1+\mathbf{d}_2)/2$, then we obtain $\mathbf{x}'(\Sigma-D_1)\mathbf{x}+\mathbf{x}'(\Sigma-D_2)\mathbf{x}=0$. Both terms in this summation are non-negative, since the matrices $\Sigma-D_i,\ i\in[1,2]$ are Gramian. Therefore, we conclude that $\mathbf{x}'(\Sigma-D_1)\mathbf{x}=\mathbf{x}'(\Sigma-D_2)\mathbf{x}=0$, and we can obtain $(\Sigma-D_1)\mathbf{x}=(\Sigma-D_2)\mathbf{x}=0$. In particular $D_1\mathbf{x}=D_2\mathbf{x}=0$, and since D_1 and D_2 are diagonal we may have the system of equations $d_{i1}x_i=d_{i2}x_i, i\in[1,n]$. Since all the vector $\mathbf{x}$ has non-zero entries, hence the uniqueness condition for CMDFA solutions $D_1=D_2$ holds.

Appendix III
Proof of Theorem 3

Let us find the conditions under which the CMDFA solution is a star. In this case, we have $d_i=1-a_i^2$ as the CMDFA solutions. By the results in Section III-A, the null-space basis are $e_1=(1,0,-a_3/a_1)$ and $e_2=(0,1,-a_3/a_2)$. From (4), for this case we obtain the system of equations, $(1/d_1,1/d_2,1/d_3)=t_{11}e_1^2+t_{22}e_2^2+t_{12}e_1e_2$ that needs to be satisfied for some $T=[t_{ij}]\succeq 0$. solving such system of equations gives us $t_{ii}=\dfrac{a_i^2}{1-a_i^2},\ i\in[1,2]$, and $t_{12}=1/2(s_3-s_1-s_2)$, where $s_i=\dfrac{a_i^2}{1-a_i^2}$. We need to show $T=[t_{ij}]\succeq 0$. Obviously $Trace(T)\geq 0$. For determinant to be non-negative, we need to have $t_{11}t_{22}\geq t_{12}^2$, or $A^2+B^2+C^2-2AB-2AC-2BC\leq 0$. Note that this is not always true, based on the values of s_i, which are the functions of a_i^2. For example, put $s_1=s_2=1$ and $s_3=5$, which corresponds to $a_1^2=a_2^2=1/2$ and $a_3^2=5/6$ (hence, a positive definite matrix Σ), where one may check that the inequality is not satisfied.

Appendix IV
Proof of Theorem 4

Since the row space of $\Sigma-D$ is two-dimensional, we can find non-zero variables α and β such that $\alpha r_1+\beta r_2+r_3=[0,0,0]'$, where r_i is the i-th row of $\Sigma-D$. This, of course is a necessary condition for a rank two matrix, and for sufficiency we need to make sure no r_i and r_j are linearly dependent, since otherwise $\Sigma-D$ becomes a rank one matrix.

By replacing r_i's with their respective vectors, we obtain the following system of equations:

$$\begin{aligned}&\alpha(1-d_1)+\beta\rho_{12}+\rho_{13}=0\\&\alpha\rho_{12}+\beta(1-d_2)+\rho_{23}=0\\&\alpha\rho_{13}+\beta\rho_{23}+(1-d_3)=0\end{aligned}\tag{13}$$

Solving for α, β and d_3, gives us:

$$\begin{aligned}\alpha&=\frac{\rho_{12}\rho_{23}-\rho_{13}(1-d_2)}{(1-d_1)(1-d_2)-\rho_{12}^2}\\\beta&=\frac{\rho_{12}\rho_{13}-\rho_{23}(1-d_1)}{(1-d_1)(1-d_2)-\rho_{12}^2}\\d_3&=1+\frac{(d_2-1)\rho_{13}^2+2\rho_{12}\rho_{13}\rho_{23}+(d_1-1)\rho_{23}^2}{(1-d_1)(1-d_2)-\rho_{12}^2}\end{aligned}\tag{14}$$

Hence, d_3 can be completely determined, via d_1 and d_2.

We know that the null space $N(\Sigma-D)$ is rank one. And a basis vector can be obtained by solving $(\Sigma-D)\mathbf{x}=\mathbf{0}$. After solving, one deduce that the null space has the following form $N(\Sigma-D)=\{(\alpha,\beta,1)^Tx_3:\forall x_3\}$, where it turns out that $p=\alpha$ and $q=\beta$. Hence, the basis is $v=(p,q,1)$, and the normal basis is $e_1=\dfrac{v}{||v||}$

Now, using Theorem 1 we need to satisfy the following equality $te_1^2=(1/d_1 1/d_2 1/d_3)^T$ for $t\geq 0$. Which gives us the system of equations $[1/d_1,1/d_2,1/d_3]'=\dfrac{1}{||v||^2}[t\alpha^2,t\beta^2,t]'$

Replacing the last equality in the first two, gives us $\dfrac{d_3}{d_1}=\alpha^2$ and $\dfrac{d_3}{d_2}=\beta^2$.

Appendix V
Proof of Theorem 5

First, note that since CMDFA solution is rank-deficient so the rank of solution is at most $n-1$. Hence we only need to prove that the rank cannot be less than $n-2$. The proof goes by induction. For the bases case, we may consider the case described in Example 1, where we showed the rank of $\Sigma-D$ cannot be less than two.

We show the matrix $\Sigma_n'=\Sigma_n-D$, with Σ_n corresponding to a Markov chain $X_1-X_2-....-X_{n-1}-X_n$, has rank at least $n-2$; assuming for all $\Sigma_{n-1}'=\Sigma_{n-1}-D$, with Σ_{n-1} regarding to smaller Markov chains $X_{i_1}-X_{i_2}-...-X_{i_{n-1}}$ for $i_1\neq...\neq i_{n-1}\in[1,n-1]$ have ranks at least $n-3$. In other words, if we sum out (drop) a variable $X_i\in[1,n]$ from the Markov chain, we obtain a length $n-1$ Markov chain with Σ_{n-1}, with rank of Σ_{n-1}' at least least $n-3$.

Without loss of generality, we may assume the Gramian matrix Σ'_n has the following generic form,

$$\Sigma'_n = \begin{pmatrix} & & & & \rho_{1,n} \\ & & & & \rho_{2,n} \\ & & \Sigma'_{n-1} & & \vdots \\ & & & & \rho_{n-1,n} \\ \rho_{1,n} & \rho_{2,n} & \dots & \rho_{n-1,n} & t_n \end{pmatrix} \tag{15}$$

Obviously if rank of Σ'_{n-1} is at least $n-2$, then we are done, since the first $n-1$ rows, at least span a $n-2$ dimensional space (adding a new dimension, i.e., the last column, does not reduce the row space dimension). Therefore, we assume Σ'_{n-1} has rank $n-3$.

Consider the first $n-3$ (linearly independent) rows $r_1, ..., r_{n-3}$, and form a linear combination of these rows with row r_n: $\alpha_1 r_1 + ... + \alpha_{n-3} r_{n-3} + \alpha_n r_n$. We are interested to see whether r_n can be written as linear combination of the first $n-3$ rows, and note that $\alpha_n \neq 0$ (since then we have a contradictory conclusion of linear dependence of first $n-3$ rows). Hence, we may ignore α_n and write $\alpha_1 r_1 + ... + \alpha_{n-3} r_{n-3} = r_n$. Extracting the summation elements for the last three columns gives us the following equations,

$$\begin{aligned} \sum_{i=1}^{n-3} \alpha_i \rho_{i,n-2} &= \rho_{n-2,n} \\ \sum_{i=1}^{n-3} \alpha_i \rho_{i,n-1} &= \rho_{n-1,n} \\ \sum_{i=1}^{n-3} \alpha_i \rho_{i,n} &= t_n \end{aligned} \tag{16}$$

Due to Markov chain property, we know $\rho_{i,j} = \prod_{(k,l)\in path(i,j)} \rho_{k,l}$, i.e., the pairwise correlation $\rho_{i,j}$ can be computed as the product of all $\rho_{k,l}$, where (x_k, x_l) pairs are the edges on the path between x_i and x_j. Now, we may multiply the first and second equations by $\rho_{n-2.n}$ and $\rho_{n-1,n}$ and re-write the equations as follows,

$$\begin{aligned} \sum_{i=1}^{n-3} \alpha_i \rho_{i,n} &= \rho^2_{n-2,n} \\ \sum_{i=1}^{n-3} \alpha_i \rho_{i,n} &= \rho^2_{n-1,n} \\ \sum_{i=1}^{n-3} \alpha_i \rho_{i,n} &= t_n \end{aligned} \tag{17}$$

The left hand side on all equations is equal, hence we have $\rho^2_{n-2,n} = \rho^2_{n-1,n}$, which reduces to $\rho^2_{n-2,n-1} = 1$, i.e., a rank-deficient Markov chain with rank $n-1$, a contradiction (since we started with a rank n Markov chain).

This shows the linear independence of r_n with first $n-3$ rows, i.e., the set of vectors $(r_1, ..., r_{n-3}, r_n)$ spans an $n-2$ dimensional space, i.e., the rank of Σ'_n is at least $n-2$.