An improved GraDe method for blind separation of graph signals

Mohammad Sadeghi, Massoud Babaie-Zadeh, *Senior Member, IEEE*

***Abstract*—For blind source separation (BSS) of Gaussian graph signals, an algorithm called GraDe (graph decorrelation) has been introduced in [1], [2]. In the current paper, it is shown that GraDe does not achieve a good performance for some types of graphs. This is attributed to the estimation of covariance/autocovariance matrices using signal samples, which may not be reliable. To address this weakness, an improvement based on the spectral representation of the signals is proposed, focusing on removing the impact of the outlier eigenvalues. Numerical simulations show that the proposed method outperforms the original GraDe algorithm.**

***Index Terms*—Graph Signal Processing, Graph Decorralation, Blind Source Separation**

I. Introduction

GRAPHS are mathematical tools that can be used to model the relationship of data in complicated data structures. Examples include social networks, brain connections and sensor networks [3]. In order to deal with this type of data structures, graph signal processing (GSP) [4] tools have been developed, in which many ideas from classical signal processing, e.g. filtering, frequency analysis and sampling, are extended to graph signals [4]. In GSP frameworks, data elements are entries associated with graph nodes, and edges specify the dependence of these entries. In some cases, the structure of the graph is a priori known, but in some others, it should be learned from a data set that lies on the graph.

On the other hand, a subject in traditional signal processing is blind source separation (BSS) [5]. In linear BSS, there are a set of observed signals that are linear mixtures of some statistically independent source signals, where neither the source signals nor the mixing system is known (hence the term "Blind"), and the goal is to retrieve the source signals only from these observed mixed signals. When the sources are not Gaussian, instantaneous independence of the separated signals or maximizing their non-Gaussianity are used to separate the sources [5]. However, when the sources are Gaussian, utilization of these properties does not yield to source separation, and temporal properties have also to be exploited. For example, in [6] the time coherence of the source signals has been used.

BSS of traditional signals can be extended to graph signals. It is applicable when there are some source signals that lie on one or several graphs, and not the exact sources but some mixtures of them are available. For example, in gene expression analysis, regulatory network and gene expression data (as the graph and the graph signals) can be the inputs of a graph source separation algorithm to extract the sources, which are interpreted as distinct biological processes [1], [7]. As another example, classical BSS has been already used for the analysis of functional magnetic resonance imaging (fMRI) data (see e.g. [8], [9]). By exploiting the brain connectivity network (see e.g. [10], [11]), the problem can be changed to a graph source separation framework.

The idea of using graph information in BSS of graph signals, up to our best knowledge, has been introduced in [1], in which, a graph decorrelation (GraDe) algorithm has been developed. For this purpose, the concepts of auto-correlation and cross-correlation in time series have been extended to graph signals. Then, since the cross-correlation between the latent sources on the graph is assumed to be zero, GraDe tries to find an unmixing system that decorrelates the output signals. In [2], GraDe has been generalized to the case in which each source lies on a different graph. In addition, when sources are non-Gaussian, GraDe has been combined with other BSS algorithms like JADE [12] and FastICA [13], [14], which exploit the non-Gaussianity of the sources for their separation, and the resulting algorithms are called Graph Jade and Graph FastICA [2]. Moreover, the Cramér-Rao bound for the estimation of mixing and unmixing matrices has been derived when the sources are Gaussian graph moving average (GMA) signals [2]. In [15] a method has been proposed to jointly learn the underlying graphs and separate the graph signals, when the structure of the graphs is not a priori known. Unlike GraDe, this method uses many graph signals of every source, because learning graphs requires a data set of graph signals.

To our knowledge, GraDe is the only algorithm that can be used to separate Gaussian signals based on the information given by graph. In this paper, we take a precise look at GraDe and show that it contains a weakness, which results in poor separation in some cases. This weakness mathematically arises from the use of sample covariance/autocovariance matrices of the graph signals as estimations for their expected values, which may not necessarily be reliable. Then, based on the spectral domain representation of the graph signals, we propose a new separation method, which improves GraDe when the sources are Gaussian GMA signals. The proposed method uses the spectral representation of the graph signals, aiming to eliminate the impact of the outlier eigenvalues.

The paper is organized as follows. In Section II, GraDe algorithm is reviewed from a critical point of view and a

This work has been partially supported by Research Office of Sharif University of Technology.

M. Sadeghi and M. Babaie–Zadeh are with Department of Electrical Engineering, Sharif University of Technology, Tehran, Iran, (e-mail: mo.sadeghi@sharif.edu, mbzadeh@yahoo.com).

numerical example is given to illustrate the weakness of this algorithm. In Section III, two solutions to improve GraDe are proposed. Section IV is devoted to numerical study of our method and its comparison with GraDe.

II. A Critical Review on GraDe

In this section, we review GraDe and illustrate how it estimates the covariance/autocovariance matrices of the sources using the sample signals. We show that this estimation is not reliable in the general case of graph structures and signals. Then, we provide a numerical example to confirm the limitations attributed to GraDe.

A. Graph decorrelation

Here, the assumptions about the sources and the mixing model are the same as in [2]. A weighted graph $\mathcal{G} = (\mathcal{N}, \mathcal{E}, \mathcal{W})$ is defined with a set of N nodes ($\mathcal{N}$), a set of edges $\mathcal{E} \subseteq \mathcal{N} \times \mathcal{N}$ indicating the connections between the nodes, and a set of weights ($\mathcal{W}$) corresponding to the edges. The weights are non-negative numbers. The graph can be represented by an $N \times N$ matrix $\mathbf{W}$ known as the adjacency matrix. The ij-th entry of the adjacency matrix represents the weight between the nodes i and j. If there is no edge between these nodes, the entry is zero. For simplicity, the graphs are assumed to be undirected with no self-loops. Therefore, the adjacency matrix is symmetric, and its diagonal entries are zero. A graph signal is a real vector $\mathbf{x} \in \mathbb{R}^N$, where the i-th entry, x_i, represents a value assigned to the i-th node of $\mathcal{G}$.

Consider a graph $\mathcal{G}$ with N nodes and an adjacency matrix $\mathbf{W}$. Let $\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_P \in \mathbb{R}^N$ represent P (unknown) source signals that lie on the graph $\mathcal{G}$. This implies that the i-th entry of $\mathbf{z}_p$ corresponds to the i-th node of $\mathcal{G}$ for $p = 1, 2, \ldots, P$. The graph provides prior knowledge regarding the structure of the signals and captures the relationships between the entries of each source signal. It is important to note that each signal is assumed to be a realization of a random vector, where the entries are not necessarily identically distributed. The probability distribution of the entries of the source signals may vary depending on the graph structure and the underlying signal-graph model, such as the Gaussian GMA model that will be explained in Section II-B. Furthermore, the sources are assumed to be independent of each other, which is a key assumption in the context of BSS that enables their separation.

Let $\mathbf{Z} = [\mathbf{z}_1, \mathbf{z}_2, ..., \mathbf{z}_P]^T$ be the matrix of the P aforementioned latent sources. Thus, $\mathbf{Z}$ is a random matrix where each row represents an independent source signal associated with the graph, and the i-th column of $\mathbf{Z}$ corresponds to the vector of values of the sources at the i-th node of the graph. It is assumed that the sources have zero mean ($\mathbb{E}\{\mathbf{z}_p\} = \mathbf{0}$), and that $\mathbb{E}\{\|\mathbf{z}_p\|^2\} = N$ for $p = 1, 2, ..., P$. Therefore, due to the independence of the sources, the covariance matrix of the sources, $\mathbf{S}^0(\mathbf{Z}) \triangleq N^{-1}\mathbb{E}\{\mathbf{Z}\mathbf{Z}^T\}$, is an identity matrix. As introduced in [1], corresponding to the time shift in temporal signals, the graph shift of the graph signal $\mathbf{z}_p$ is defined as $\mathbf{W}\mathbf{z}_p$. This transformation shifts the entries of the graph signal along the direction specified by the edges of the graph. Using this definition, $\mathbf{Z}\mathbf{W}^k = \{\mathbf{z}_1^T\mathbf{W}^k, \mathbf{z}_2^T\mathbf{W}^k, ..., \mathbf{z}_P^T\mathbf{W}^k\}$ is the matrix of the k-shifted version of the sources for a positive integer k, and the k-th autocovariance matrix of the sources is defined as

$$\mathbf{S}^k(\mathbf{Z}, \mathbf{W}) \triangleq N^{-1}\mathbb{E}\left\{\mathbf{Z}\mathbf{W}^k\mathbf{Z}^T\right\}. \tag{1}$$

We refer to the matrices $\hat{\mathbf{S}}^0(\mathbf{Z}) \triangleq N^{-1}(\mathbf{Z}\mathbf{Z}^T)$ and $\hat{\mathbf{S}}^k(\mathbf{Z}, \mathbf{W}) \triangleq (N)^{-1}(\mathbf{Z}\mathbf{W}^k\mathbf{Z}^T)$ as the sample covariance matrix and the k-th sample autocovariance matrix, respectively.

It is assumed that the observed signals are a linear mixture of the latent sources. The matrix of observed signals $\mathbf{X} \in \mathbb{R}^{P\times N}$ can be modeled as the product of the source matrix and an invertible unknown mixing matrix $\mathbf{\Omega} \in \mathbb{R}^{P\times P}$:

$$\mathbf{X} = \mathbf{\Omega}\mathbf{Z}. \tag{2}$$

This implies that each row of $\mathbf{X}$ is a linear combination of the sources with coefficients specified by the corresponding row of $\mathbf{\Omega}$. The goal of source separation is to estimate the latent sources given the observations and the underlying graph structure. In GraDe, a minimum of P observed signals is required to perform the source separation [2]. If the number of observations exceeds the number of sources, a dimension reduction technique like principal component analysis (PCA) can be used to reduce the dimension of the observed signals to P (see e.g., [5]). Therefore, we assume that $\mathbf{\Omega}$ is a square matrix.

In most BSS algorithms, the first step involves whitening the observed data [5]. This is achieved through a linear transformation of the data by a matrix $\mathbf{H}$, such that the transformed data has unit covariance [5]. Specifically, the whitened data $\widetilde{\mathbf{X}}$ is obtained as $\tilde{\mathbf{X}} = \mathbf{H}\mathbf{X}$, satisfying $N^{-1}\mathbb{E}\{\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T\} = \mathbf{I}_P$, where $\mathbf{I}_P$ is the $P \times P$ identity matrix. The inverse square root of the covariance matrix, $(\mathbf{S}^0(\mathbf{X}))^{-1/2} = (N^{-1}\mathbb{E}\mathbf{X}\mathbf{X}^T)^{-1/2} = (\mathbf{\Omega}\mathbf{\Omega}^T)^{-1/2}$, can be used as a whitening matrix. This choice can be justified as

$$\begin{aligned} N^{-1}\mathbb{E}&\left\{(\mathbf{\Omega}\mathbf{\Omega}^T)^{-1/2}\mathbf{X}\mathbf{X}^T(\mathbf{\Omega}\mathbf{\Omega}^T)^{-T/2}\right\} \\ &= N^{-1}(\mathbf{\Omega}\mathbf{\Omega}^T)^{-1/2}\mathbf{\Omega}\mathbb{E}\{\mathbf{Z}\mathbf{Z}^T\}\mathbf{\Omega}^T(\mathbf{\Omega}\mathbf{\Omega}^T)^{-1/2} \\ &= N^{-1}(\mathbf{\Omega}\mathbf{\Omega}^T)^{-1/2}\mathbf{\Omega}(N\mathbf{I}_P)\mathbf{\Omega}^T(\mathbf{\Omega}\mathbf{\Omega}^T)^{-1/2} \\ &= \mathbf{I}_P. \end{aligned} \tag{3}$$

Note that the square root of a positive semidefinite matrix like $\mathbf{\Omega}\mathbf{\Omega}^T$ is not unique, so a candidate can be obtained, e.g. by Cholesky decomposition [16]. In GraDe, such an inverse square root of the sample covariance matrix, $(\hat{\mathbf{S}}^0(\mathbf{X}))^{-1/2}$, is used as the whitening matrix [2]. Therefore, the whitened observation matrix is obtained as

$$\tilde{\mathbf{X}} = (\hat{\mathbf{S}}^0(\mathbf{X}))^{-1/2}\mathbf{X}. \tag{4}$$

In other words, in GraDe, $\hat{\mathbf{S}}^0(\mathbf{X})$ is considered as an estimate for $\mathbf{S}^0(\mathbf{X}) = \mathbf{\Omega}\mathbf{\Omega}^T$. Similarly, $N^{-1}\mathbf{X}\mathbf{X}^T$ is considered as an estimate for $N^{-1}\mathbb{E}\{\mathbf{X}\mathbf{X}^T\}$. However, one may wonder whether $N^{-1}\mathbf{X}\mathbf{X}^T$ is a good estimate for its expected value.

The answer is generally no. Let us represent the observation matrix by its column vectors: $\mathbf{X} = [\boldsymbol{\chi}_1, \boldsymbol{\chi}_2, ..., \boldsymbol{\chi}_N]$. The entries of the random vector $\boldsymbol{\chi}_n \in \mathbb{R}^P$ are each a linear combination of the values of the latent sources on the n-th node of the graph, for $n = 1, 2, ..., N$. As mentioned earlier, a source

signal can have different probability distributions on different nodes of the graph. Moreover, the values of the sources on different nodes are not statistically independent (actually the graph structure is intended to represent their dependence). Therefore, the random vectors $\boldsymbol{\chi}_1, \boldsymbol{\chi}_2, ..., \boldsymbol{\chi}_N$ generally have different probability distributions and are not independent of each other. By expressing the sample covariance matrix of the observations as

$$N^{-1}\mathbf{X}\mathbf{X}^T = N^{-1}\sum_{n=1}^{N}\boldsymbol{\chi}_n\boldsymbol{\chi}_n^T, \tag{5}$$

we observe that it represents an average of random matrices, with each matrix corresponding to a node of the graph. These matrices possess distinct probability distributions and are not independent of each other. On the other hand, the sample covariance matrix $N^{-1}\mathbb{E}\{\mathbf{X}\mathbf{X}^T\}$ represents the average of their expected values. Therefore, for no reason, the sample covariance matrix can be considered a good estimate for the covariance matrix of the observed data in general. This is the first weakness that we attribute to GraDe.

It is worth noting that traditional BSS methods also estimate the covariance matrix based on time samples of the signals, but they make certain assumptions about the sources to ensure the validity of this estimation. For example, in [6], it is assumed that the sources are either deterministic autocovariance-ergodic sequences or jointly stationary processes with diagonal autocovariance matrices. Similarly, the JADE algorithm [12] assumes jointly stationary processes and consistent estimation of the covariance matrix of the observed signals. However, applying the same assumptions to graph signals limits the generality of graph structures. For instance, the assumption that the sources have the same probability distributions on different nodes of the graph is similar to the stationarity assumption in time signals. Nevertheless, this assumption may not hold for many graph structures and graph signals.

If the whitening process is performed well and $\mathbf{U}^*$ is a matrix such that $\mathbf{U}^*\tilde{\mathbf{X}} = \mathbf{Z}$, it follows that $\mathbf{U}^*$ must be an orthogonal matrix. This is because both $\mathbf{Z}$ and $\tilde{\mathbf{X}}$ have identity covariance matrices. As explained in [2], the next step in GraDe is to find an orthogonal matrix $\mathbf{U}$ such that each row of $\mathbf{U}\tilde{\mathbf{X}}$ is uncorrelated with all other rows of its graph shifted versions, $\mathbf{U}\tilde{\mathbf{X}}\mathbf{W}^k$. To achieve this decorrelation, the GraDe algorithm aims to make the sample covariance matrices $\hat{\mathbf{S}}^k(\mathbf{U}\tilde{\mathbf{X}}, \mathbf{W}) = N^{-1}(\mathbf{U}\tilde{\mathbf{X}}\mathbf{W}^k\tilde{\mathbf{X}}^T\mathbf{U}^T)$ as diagonal as possible, where $k = 1, 2, ..., K$. The objective function to be maximized is defined as

$$\sum_{k=1}^{K}\left\|\mathrm{diag}(\hat{\mathbf{S}}^k(\mathbf{U}\tilde{\mathbf{X}}, \mathbf{W}))\right\|_2^2 = \sum_{k=1}^{K}\left\|\mathrm{diag}(\mathbf{U}\hat{\mathbf{S}}^k(\tilde{\mathbf{X}}, \mathbf{W})\mathbf{U}^T)\right\|_2^2, \tag{6}$$

where the diag operator returns the vector composed of the diagonal entries of its argument. Maximizing this objective function leads to the diagonalization of the sample autocovariance matrices. This is because multiplying a matrix by an orthogonal matrix does not alter its Frobenius norm. Thus, maximizing the sum of squares of the diagonal entries of the sample autocovariance matrices with respect to the orthogonal matrix $\mathbf{U}$ corresponds to minimizing the sum of squares of the off-diagonal entries.

Since the sources are assumed to be mutually independent and have zero means, for two distinct sources $\mathbf{z}_i$ and $\mathbf{z}_j$, we have $\mathbb{E}\{\mathbf{z}_i^T\mathbf{z}_j\} = 0$ and $\mathbb{E}\{\mathbf{z}_i^T\mathbf{W}^k\mathbf{z}_j\} = \mathbb{E}\{\mathbf{z}_i^T\}\mathbf{W}^k\mathbb{E}\{\mathbf{z}_j\} = 0$. Note that $\mathbb{E}\{\mathbf{z}_i^T\mathbf{W}^k\mathbf{z}_j\}$ is the element in the i-th row and j-th column of the matrix $\mathbb{E}\{\mathbf{Z}\mathbf{W}^k\mathbf{Z}^T\}$. Consequently, the off-diagonal entries of the k-th autocovariance matrix $\mathbf{S}^k(\mathbf{Z}, \mathbf{W})$ are zero, making it diagonal. If $\mathbf{U}^*$ is an orthogonal matrix such that $\mathbf{U}^*\tilde{\mathbf{X}} = \mathbf{Z}$, then the matrix $\mathbf{S}^k(\mathbf{U}^*\tilde{\mathbf{X}}, \mathbf{W})$ will be diagonal. However, GraDe aims to find an orthogonal matrix $\mathbf{U}$ that makes the sample autocovariance matrices $\hat{\mathbf{S}}^k(\mathbf{U}\tilde{\mathbf{X}}, \mathbf{W})$ as diagonal as possible, rather than their expected values. In this case, $\hat{\mathbf{S}}^k(\mathbf{U}\tilde{\mathbf{X}}, \mathbf{W})$ is considered as an estimate for $\mathbf{S}^k(\mathbf{U}^*\tilde{\mathbf{X}}, \mathbf{W})$. This estimation may not be valid in general, for the same reasons mentioned earlier regarding the estimation of the covariance matrix from the samples in the whitening process. Consequently, the matrices $\hat{\mathbf{S}}^k(\mathbf{U}^*\tilde{\mathbf{X}}, \mathbf{W})$ may not be close to diagonal, and seeking an orthogonal matrix $\mathbf{U}$ that maximizes the diagonality of the matrices $\hat{\mathbf{S}}^k(\mathbf{U}\tilde{\mathbf{X}}, \mathbf{W})$ may not yield an accurate estimation of $\mathbf{U}^*$.

Note that the weaknesses mentioned about GraDe are for the general case. It is possible that for some specific graph structures and related graph signals the estimations that are made in this method are fine and the separation is performed well.

B. A numerical example

In the previous section, it is explained that GraDe uses the sample covariance and autocovariance matrices $\mathbf{Z}\mathbf{Z}^T$ and $\mathbf{Z}\mathbf{W}^k\mathbf{Z}^T$ as an estimate for their expected values $\mathbb{E}\{\mathbf{Z}\mathbf{Z}^T\}$ and $\mathbb{E}\{\mathbf{Z}\mathbf{W}^k\mathbf{Z}^T\}$. Here a numerical example is provided to illustrate how inaccurate these estimations can be. In this example, the graph is generated by Barabási-Albert (BA) model [17] with total $N = 5000$ nodes, beginning with $m_0 = 100$ initial nodes and connecting every new node to $m = 40$ existing nodes. Weights take only 1 and 0 values. It is assumed that there are $P = 4$ independent sources that are created by Gaussian GMA model. This model is widely used in [1], [2] and [15] to generate graph signals. In this model, the sources can be written as

$$\mathbf{z}_p = \mathbf{y}_p + \sum_{l=1}^{L}\theta_{l,p}\mathbf{W}^l\mathbf{y}_p, \quad p = 1, ..., P, \tag{7}$$

where $\mathbf{y}_p$ is a vector of Gaussian independent and identically distributed (i.i.d.) random variables with zero mean and unit variance. The p-th source $\mathbf{z}_p$ is equal to the sum of $\mathbf{y}_p$ and its graph shifted versions multiplied by the weighting coefficients $\theta_{1,p}, ..., \theta_{L,p}$. L indicates the order of GMA model. The vector $\mathbf{y}_p$ is generated independently for $p = 1, 2, ..., P$, and hence the sources are independent. In our simulation, Gaussian GMA model of order 3, i.e. GMA(3), is used to generate the sources, and for all of them the coefficients are $\theta_1 = 0.05$, $\theta_2 = 0.0015$ and $\theta_3 = 10^{-5}$. These coefficients are chosen such that the mean of the variances of the entries of the shifted and scaled versions of $\mathbf{y}_p$ are in the same scale as of $\mathbf{y}_p$. The source matrix $\mathbf{Z} = [\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3, \mathbf{z}_4]^T$ is generated this way. Fig. 1

$$10^{-4}\mathbf{Z}\mathbf{Z}^T = \begin{bmatrix} 1.04 & -0.44 & 0.38 & 0.99 \\ -0.44 & 1.36 & -0.55 & -1.42 \\ 0.38 & -0.55 & 1.18 & 1.19 \\ 0.99 & -1.42 & 1.19 & 3.85 \end{bmatrix}$$

(a)

$$10^{-6}\mathbf{Z}\mathbf{W}\mathbf{Z}^T = \begin{bmatrix} 0.48 & -0.63 & 0.53 & 1.40 \\ -0.63 & 0.98 & -0.78 & -2.06 \\ 0.53 & -0.78 & 0.70 & 1.72 \\ 1.40 & -2.06 & 1.72 & 4.60 \end{bmatrix}$$

(b)

Fig. 1. Sample covariance and autocovariance matrices for one realization of 4 sources by BA graph and GMA(3) model (a) $\mathbf{Z}\mathbf{Z}^T/10^4$ (b) $\mathbf{Z}\mathbf{W}\mathbf{Z}^T/10^6$

shows the values of $\mathbf{Z}\mathbf{Z}^T$ and $\mathbf{Z}\mathbf{W}\mathbf{Z}^T$ for one realization of the sources. It is seen that these matrices are not diagonal, and their values change in each realization of the sources, but obviously their expected values are diagonal, because of the independence of the sources. It illustrates the inaccuracy of the estimation of the covariance and autocovariance matrices using the samples.

Suppose now that each source has been independently generated M times. The goal is to obtain the average of the sample covariance and autocovariance matrices of these independent set of sources to evaluate how their diagonality changes as M increases. Averaged sample covariance and autocovariance matrices can be written as

$$\mathbf{C}_M^0 = \frac{1}{MN}\sum_{m=1}^{M}\mathbf{Z}_m\mathbf{Z}_m^T\ ,\ \mathbf{C}_M^1 = \frac{1}{MN}\sum_{m=1}^{M}\mathbf{Z}_m\mathbf{W}\mathbf{Z}_m^T, \quad (8)$$

where $\mathbf{Z}_m$ denotes the matrix of the m-th realization of the sources. A metric is required in order to measure the diagonality of these matrices. For this purpose, a modified Frobenius measure, as defined in [18], is employed. For a square matrix $\mathbf{A}$ this measure is defined as

$$\mathcal{D}(\mathbf{A}) = \frac{1}{2}\left\|(\mathrm{Diag}(\mathbf{A}))^{-1/2}\mathbf{A}(\mathrm{Diag}(\mathbf{A}))^{-1/2} - \mathbf{I}\right\|_F^2, \quad (9)$$

where $\mathrm{Diag}(\mathbf{A})$ represents a diagonal matrix with the diagonal elements identical to those of $\mathbf{A}$. When $\mathbf{A}$ is a diagonal matrix, $\mathcal{D}(\mathbf{A})$ will be zero. Conversely, larger values of the diagonality measure indicate that the sum of squared off-diagonal entries of the normalized version of $\mathbf{A}$ is significant.

Fig. 2 illustrates that the averaged matrices $\mathbf{C}_M^0$ and $\mathbf{C}_M^1$ become more diagonal as M increases. That is because they become more close to the covariance and autocovariance matrices $\mathbb{E}\{\mathbf{Z}\mathbf{Z}^T\}$ and $\mathbb{E}\{\mathbf{Z}\mathbf{W}^k\mathbf{Z}^T\}$. This example shows that having only one version of the graph sources, as is used in GraDe, may not be enough for the estimation of the covariance and autocovariance matrices.

III. Solutions to the Drawbacks of GraDe

In this section, two approaches to improve GraDe are discussed. The first one changes the assumptions about the sources and requires more data. The second one is a proposed method based on the spectral domain representation of the signals, and its assumptions are the same as of GraDe.

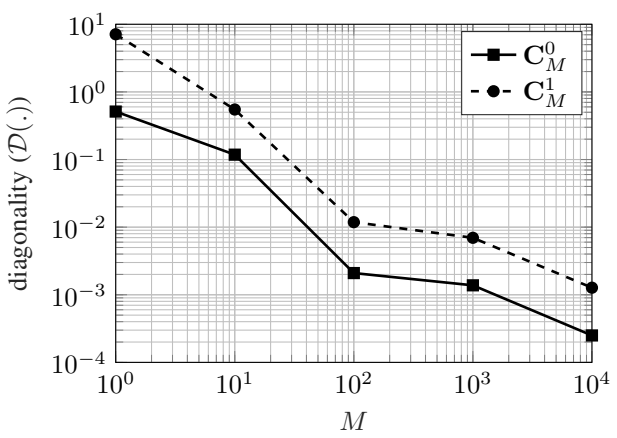

Fig. 2. Diagonality of averaged sample covariance and autocovariance matrices according to number of averaged matrices, M.

A. Averaging over independent realizations of the graph signals

Assume that not a single realization of the latent sources, but rather M independent realizations of them are available. So, there will exist M observation matrices each of which is a mixture of one realization of the latent sources, that is,

$$\mathbf{X}_m = \mathbf{\Omega}\mathbf{Z}_m \quad m = 1, 2, ..., M. \quad (10)$$

The matrices $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_m$ are independent of each other, and have the same probability distributions. So the average of the sample covariance matrices ($M^{-1}N^{-1}\sum_{m=1}^{M}\mathbf{X}_m\mathbf{X}_m^T$) can be used as an estimate for the covariance matrix $N^{-1}\mathbb{E}\{\mathbf{X}\mathbf{X}^T\}$ in the whitening process. Then after the whitening, the average of the sample autocovariance matrices ($M^{-1}N^{-1}\sum_{m=1}^{M}\tilde{\mathbf{X}}_m\mathbf{W}^k\tilde{\mathbf{X}}_m^T$) can be used as an estimate for the autocovariance matrices $N^{-1}\mathbb{E}\{\tilde{\mathbf{X}}\mathbf{W}^k\tilde{\mathbf{X}}^t\}$. As the number of the averaged matrices in averaging process increases, the estimations becomes more accurate. In [15] the same approach is taken and it is assumed that several independent realizations of the signals are available, in which every realization is called a window, and the averaging is performed over these windows.

Assuming that there exist several realizations of data, limits the applicability of the method. In addition, assuming that independent realizations of the sources are mixed by the same mixing matrix is not probably realistic in some applications. For example, in the brain mapping with fMRI, every measured signal at each time course is a new linear mixture of the source signals [8]. So recording more measurements does not lead to the model presented in (10), but rather leads to the model in (2) with a tall mixing matrix, which is an overdetermined BSS problem.

B. Proposed method based on the spectral representation

In the second approach, only a single realization of the sources will be used to separate them, as in GraDe. As mentioned in Section II-A, the weakness of GraDe is that the sample covariance and autocovariance matrices may not be a good estimate for their expected values, so they may not be even nearly diagonal. If a transformation is performed on the

signals such that the sample covariance and autocovariance matrices of the transformed sources are more diagonal, we will show that the transformed signals can be used to yield a better separation result. Here the goal is to find such a transformation, and the spectral representation of the signals makes the problem more clear.

1) Spectral representation of GMA signals: In this approach, the sources are assumed to be Gaussian GMA signals with unknown coefficients, and are generated as specified in (7). The adjacency matrix $\mathbf{W}$ is symmetric, so its eigen decomposition can be written as

$$\mathbf{W} = \mathbf{V}\mathbf{D}\mathbf{V}^T, \tag{11}$$

where $\mathbf{V}$ is the matrix of the eigenvectors of $\mathbf{W}$ and is orthonormal, and $\mathbf{D}$ is a diagonal matrix whose diagonal entries are the corresponding eigenvalues. By plugging (11) into (7),

$$\mathbf{z}_p = \mathbf{V}(\mathbf{I}_N + \sum_{l=1}^{L} \theta_{l,p}\mathbf{D}^l)\mathbf{V}^T\mathbf{y}_p = \mathbf{V}\mathbf{D}^{(p)}\hat{\mathbf{y}}_p, \tag{12}$$

for $p = 1, ..., P$. The vector $\hat{\mathbf{y}}_p = \mathbf{V}^T\mathbf{y}_p$ is the graph Fourier transform of $\mathbf{y}_p$ and

$$\mathbf{D}^{(p)} \triangleq \mathbf{I}_N + \sum_{l=1}^{L} \theta_{l,p}\mathbf{D}^l \tag{13}$$

is a diagonal matrix corresponding to the p-th source. The entries of $\hat{\mathbf{y}}_p$ are i.i.d. random variables, because they are jointly Gaussian and have the identity covariance matrix ($\mathbb{E}\{\hat{\mathbf{y}}_p\hat{\mathbf{y}}_p^T\} = \mathbf{V}^T\mathbf{V} = \mathbf{I}_N$). For different i and j, the vectors $\hat{\mathbf{y}}_i$ and $\hat{\mathbf{y}}_j$ are independent of each other. By using (12), the ij-th entry of the sample covariance matrix $N^{-1}\mathbf{Z}\mathbf{Z}^T$ can be written as

$$\begin{aligned} N^{-1}\left[\mathbf{Z}\mathbf{Z}^T\right]_{ij} &= N^{-1}\mathbf{z}_i^T\mathbf{z}_j = N^{-1}\hat{\mathbf{y}}_i^T\mathbf{D}^{(i)}\mathbf{D}^{(j)}\hat{\mathbf{y}}_j \\ &= N^{-1}\sum_{t=1}^{N} D_{tt}^{(i)}D_{tt}^{(j)}\hat{y}_{i,t}\hat{y}_{j,t}, \quad i,j \in \{1,...,P\}, \end{aligned} \tag{14}$$

where $D_{tt}^{(i)}$ denotes the t-th diagonal entry of $\mathbf{D}^{(i)}$ and $\hat{y}_{i,t}$ denotes the t-th entry of the vector $\hat{\mathbf{y}}_i$. In the same manner, the ij-th entry of the sample autocovariance matrices can be written as

$$\begin{aligned} N^{-1}\left[\mathbf{Z}\mathbf{W}^k\mathbf{Z}^T\right]_{ij} &= N^{-1}\mathbf{z}_i^T\mathbf{W}^k\mathbf{z}_j = N^{-1}\hat{\mathbf{y}}_i^T\mathbf{D}^{(i)}\mathbf{D}^k\mathbf{D}^{(j)}\hat{\mathbf{y}}_j \\ = N^{-1}\sum_{t=1}^{N} D_{tt}^{(i)}D_{tt}^kD_{tt}^{(j)}\hat{y}_{i,t}\hat{y}_{j,t}, &\quad i,j \in \{1,...,P\}. \end{aligned} \tag{15}$$

It can be seen that each entry of the sample covariance and autocovariance matrices can be written as the weighted sum of N i.i.d. random variables. Now, the goal is to find out how the diagonality of the matrices is related to the weights of these i.i.d. random variables. First, the mean and variance of the entries of these matrices are derived. The random variables $\hat{y}_{i,t}$ have standard normal distributions and are independent of each other, so the mean and variance of the i-th diagonal entry of the sample autocovariance matrices can be obtained as

$$\mathbb{E}\{N^{-1}\left[\mathbf{Z}\mathbf{W}^k\mathbf{Z}^T\right]_{ii}\} = N^{-1}\sum_{t=1}^{N} D_{tt}^{(i)}D_{tt}^kD_{tt}^{(i)}, \tag{16}$$

$$\mathrm{Var}\{N^{-1}\left[\mathbf{Z}\mathbf{W}^k\mathbf{Z}^T\right]_{ii}\} = 2N^{-2}\sum_{t=1}^{N}(D_{tt}^{(i)}D_{tt}^kD_{tt}^{(i)})^2, \tag{17}$$

and for each off-diagonal entry of the sample autocovariance matrix ($i \neq j$), the mean and variance can be obtained as

$$\mathbb{E}\{N^{-1}\left(\mathbf{Z}\mathbf{W}^k\mathbf{Z}^T\right)_{ij}\} = 0, \tag{18}$$

$$\mathrm{Var}\{N^{-1}\left(\mathbf{Z}\mathbf{W}^k\mathbf{Z}^T\right)_{ij}\} = N^{-2}\sum_{t=1}^{N}(D_{tt}^{(i)}D_{tt}^kD_{tt}^{(j)})^2. \tag{19}$$

The same expressions can be obtained for the mean and variance of the entries of the covariance matrix by putting $k = 0$.

2) Diagonality of sample matrices: According to (16) and (18), the diagonal entries of covariance/autocovariance matrices have a non-zero mean, whereas the off-diagonal entries have a zero mean. Consequently, the expected value of these random matrices is completely diagonal. However, the diagonality of individual samples of these random matrices is to be examined. The diagonality measure (9) for the sample autocovariance matrices can be expressed in an alternative form as

$$\mathcal{D}(N^{-1}\mathbf{Z}\mathbf{W}^k\mathbf{Z}^T) = \sum_{i=1}^{P}\sum_{j=1}^{i-1}\frac{(\mathbf{z}_i^T\mathbf{W}^k\mathbf{z}_j)^2}{(\mathbf{z}_i^T\mathbf{W}^k\mathbf{z}_i)(\mathbf{z}_j^T\mathbf{W}^k\mathbf{z}_j)}. \tag{20}$$

We define a matrix to be "nearly diagonal" if its diagonality measure is less than a specified threshold, say $\mathcal{D}^*$. The probability that a sample autocovariance/covariance matrix is nearly diagonal depends on the probability distribution of the entries of the corresponding random matrix. However, by considering that the expectations of the numerators in the summation (20) represent the variance of the off-diagonal entries, and the expectations of the denominators correspond to the product of the means of the two corresponding diagonal entries, it can be intuitively inferred that as the standard deviations of the off-diagonal entries decrease relative to the mean of the corresponding diagonal entries, the probability of the sample matrix being nearly diagonal increases. Thus, we will use this ratio to clarify the proposed method in the remainder of this section. The relationship between the expected value of the diagonality measure in equation (20) and the ratio of moments (the standard deviations of the off-diagonal entries to the mean of the corresponding diagonal entries) will be numerically examined in Section IV-B.

3) The effect of eigenvalues on diagonality: To illustrate the impact of eigenvalues on the diagonality of sample covariance/autocovariance matrices, we consider two simple cases. In the first case, it is assumed that the eigenvalues are equal, resulting in equal diagonal entries for matrices $\mathbf{D}$ and $\mathbf{D}^{(p)}$ ($p = 1, ..., P$). Referring to equations (16) and (19), the mean of the p-th diagonal entry of the k-th autocovariance matrix becomes $D_{11}^{(p)}D_{11}^kD_{11}^{(p)}$, while the variance of the ij-th off-diagonal entry becomes $N^{-1}(D_{11}^{(i)}D_{11}^kD_{11}^{(j)})^2$. Consequently, the ratio of the standard deviation of the off-diagonal entries to the mean of the diagonal entries in this case is on the order of $O(N^{-1/2})$.

In the second case, it is assumed that only one eigenvalue (say the first one) has a non-zero value. According to (13), in

this scenario, only the first diagonal entry of $\mathbf{D}^{(p)}$ is non-zero, while the remaining diagonal entries are zero. Hence, in (16) and (19), the first term of the summations has a non-zero value. The mean of the p-th diagonal entry of the k-th autocovariance matrix becomes $N^{-1}D_{11}^{(p)}D_{11}^{k}D_{11}^{(p)}$, and the variance of the ij-th off-diagonal entry is $N^{-2}(D_{11}^{(i)}D_{11}^{k}D_{11}^{(j)})^2$, with a standard deviation of $N^{-1}D_{11}^{(i)}D_{11}^{k}D_{11}^{(j)}$. In this case, the ratio of the standard deviation of the off-diagonal entries to the mean of the diagonal entries is on the order of $O(1)$. By comparing these two cases, it is observed that the mentioned ratio in the first case is lower than that in the second case by approximately $N^{-1/2}$. As explained above, this difference implies a higher probability for the sample covariance/autocovariance matrices in the second case to be nearly diagonal. It is important to note that these are simplified cases that do not occur in real graphs, but they serve to illustrate the impact of the eigenvalues on the diagonality of the sample matrices.

4) Outliers in eigenvalues: Here, we aim to derive a lower bound for the ratio of the standard deviation of the off-diagonal entries to the mean of the diagonal entries by assuming the presence of a subset of eigenvalues with significantly large magnitudes compared to the remaining eigenvalues. Let $\mathcal{N}$ denote the set of indices for all N eigenvalues, and $\mathcal{L}$ represent the set of indices for the T largest eigenvalues. Based on (13), each weight in the summations (16) and (19) is a polynomial function of an eigenvalue, with the degree of the polynomial depending on the order of the GMA model and k. Hence, a large magnitude of an eigenvalue can result in a large magnitude of the corresponding weight in the summations.

To formulate this, let us assume that in the summation corresponding to the mean of each diagonal element in (16), we have

$$\left|\sum_{t\in\mathcal{L}} D_{tt}^{(i)}D_{tt}^{k}D_{tt}^{(i)}\right| \geq \Gamma\left|D_{nn}^{(i)}D_{nn}^{k}D_{nn}^{(i)}\right|\ ,\ \forall n\in\mathcal{N}\backslash\mathcal{L}, \quad (21)$$

where $\Gamma \gg 1$ is the largest value for which the above inequality holds, indicating the difference in magnitude between the weights in $\mathcal{L}$ and the remaining weights.

Under this assumption, we can bound the ratio of the standard deviation of the ij-th off-diagonal element to the mean of the p-th diagonal element as

$$\begin{aligned}
&\frac{N^{-1}\sqrt{\sum_{t=1}^{N}(D_{tt}^{(i)}D_{tt}^{k}D_{tt}^{(j)})^2}}{N^{-1}\sum_{t=1}^{N}D_{tt}^{(p)}D_{tt}^{k}D_{tt}^{(p)}}\\
&\geq \frac{\sqrt{\sum_{t\in\mathcal{L}}(D_{tt}^{(i)}D_{tt}^{k}D_{tt}^{(j)})^2}}{|\sum_{t\in\mathcal{L}}D_{tt}^{(p)}D_{tt}^{k}D_{tt}^{(p)}| + \sum_{t\in\mathcal{N}\backslash\mathcal{L}}|D_{tt}^{(p)}D_{tt}^{k}D_{tt}^{(p)}|}\\
&\geq \frac{\sqrt{\sum_{t\in\mathcal{L}}(D_{tt}^{(i)}D_{tt}^{k}D_{tt}^{(j)})^2}}{|\sum_{t\in\mathcal{L}}D_{tt}^{(p)}D_{tt}^{k}D_{tt}^{(p)}| + \frac{N-T}{\Gamma}|\sum_{t\in\mathcal{L}}D_{tt}^{(p)}D_{tt}^{k}D_{tt}^{(p)}|}\\
&= \frac{1}{1+\frac{N-T}{\Gamma}}\frac{\sqrt{\sum_{t\in\mathcal{L}}(D_{tt}^{(i)}D_{tt}^{k}D_{tt}^{(j)})^2}}{|\sum_{t\in\mathcal{L}}D_{tt}^{(p)}D_{tt}^{k}D_{tt}^{(p)}|} = \frac{1}{1+\frac{N-T}{\Gamma}}O(T^{-1}).
\end{aligned} \quad (22)$$

If Γ is sufficiently large compared to $N-T$, the lower bound becomes $O(T^{-1})$, indicating that it depends on the cardinality of $\mathcal{L}$ while the contribution of the remaining eigenvalues is limited. Thus, if there exists a set of outliers among the eigenvalues with a small cardinality (T) and significantly larger magnitudes compared to the remaining eigenvalues (resulting in a large Γ), according to the lower bound, they will restrict the ratio of the standard deviation of the off-diagonal entries to the mean of the diagonal entries to be $O(T^{-1})$.

5) Eigenvalue distribution of real graphs: Figure 3 illustrates the absolute magnitude of the eigenvalues (sorted in decreasing order) of the adjacency matrix for five different graph models. All graphs have $N = 5000$ nodes, and the parameters used to generate the graphs will be described in Section IV. From the plot, it is observed that the Barabási-Albert and Erdös-Rényi [19] graphs exhibit one eigenvalue that is significantly larger than the others. In the random sensor [20] and geometric graph models, a few eigenvalues with large magnitudes dominate the remaining eigenvalues. Similarly, in the stochastic block graph model [21], three eigenvalues have greater magnitudes compared to the others.

These observations align with previous studies in the literature on spectra of random graphs. For instance, in [22] and [23], it is demonstrated that for an Erdös-Rényi graph with an average degree of $d = pN$ and the eigenvalues $\lambda_1 \geq ... \geq \lambda_N$, the largest eigenvalue λ_1 satisfies $\lambda_1 \geq d$, while the magnitude of the second-largest eigenvalue, $\lambda = \max(\lambda_2, |\lambda_N|)$, scales as $O(\sqrt{d})$. This indicates a significant gap between the magnitude of the largest eigenvalue and the remaining eigenvalues. In a stochastic block graph with M blocks, [24] shows that the M largest eigenvalues are extremal and distinct from the rest of the eigenvalues in terms of magnitude, while the remaining eigenvalues form a continuous part in the spectrum of the adjacency matrix. Additionally, numerical experiments conducted in [25] demonstrate that in Barabási-Albert graphs, the largest eigenvalue λ_1 scales with N as $\lambda_1 \sim N^{1/4}$, and the difference between the magnitudes of the two largest eigenvalues scales as $\sim N^{-0.43}$.

6) Improved separation method: As explained above, these eigenvalues with significantly large magnitudes can act as the outliers and reduce the diagonality of the sample covariance/autocovariance matrices. An idea that comes to mind is to omit the terms corresponding with the extremely large eigenvalues in the summations (14) and (15) to make the resulting matrices more diagonal. This idea can be formulated in a general form. We define the weighted sample covariance matrix of the sources as $N^{-1}\mathbf{Z}\mathbf{V}\mathbf{Q}^{(0)}\mathbf{V}^T\mathbf{Z}^T$ and the k-th weighted sample autocovariance matrix as $N^{-1}\mathbf{Z}\mathbf{V}\mathbf{Q}^{(k)}\mathbf{V}^T\mathbf{Z}^T$, where $\mathbf{V}$ is the matrix of the eigenvectors and $\mathbf{Q}^{(k)}$ is the diagonal matrix of the weights for $k = 0, 1, ..., K$. If we put $\mathbf{Q}^{(0)} = \mathbf{I}_N$ and $\mathbf{Q}^{(k)} = \mathbf{D}^k$, then these matrices will be the same as the sample covariance and autocovariance matrices that are used in GraDe. Same as (14), the ij-th entry of the weighted sample covariance/autocovariance matrices can be written as

$$\begin{aligned}
&N^{-1}\left[\mathbf{Z}\mathbf{V}\mathbf{Q}^{(k)}\mathbf{V}^T\mathbf{Z}^T\right]_{ij} = N^{-1}\mathbf{z}_i^T\mathbf{V}\mathbf{Q}^{(k)}\mathbf{V}^T\mathbf{z}_j\\
&= N^{-1}\sum_{t=1}^{N} D_{tt}^{(i)}D_{tt}^{(j)}Q_{tt}^{(k)}\hat{y}_{i,t}\hat{y}_{j,t}, \quad i,j\in\{1,...,P\}.
\end{aligned} \quad (23)$$

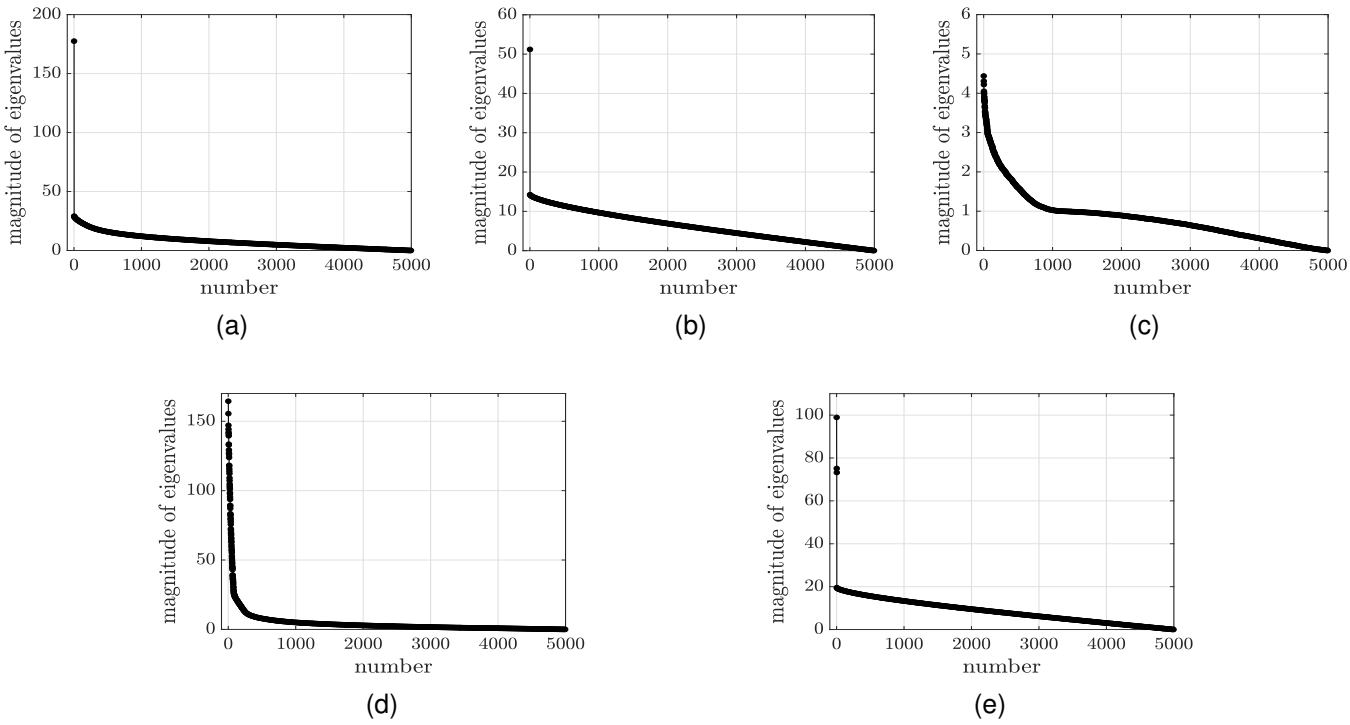

Fig. 3. Amplitude of the eigenvalues of the adjacency matrices of graph models sorted from large to small. (a) Barabási-Albert (b) Erdös-Rényi (c) random sensor (d) geometric (e) stochastic block

If the weight $Q_{tt}^{(k)}$ is set to zero, then the effect of the t-th eigenvalue is eliminated in the summation. Let $\mathcal{L}$ be the set of the indexes of the eigenvalues whose effects are to be eliminated, because they have large magnitudes and act as outliers. For example, for the graph corresponding to Fig. 3e, $\mathcal{L}$ contains the indexes of the first three large eigenvalues. Then, for $k = 0, 1, ..., K$, the weights are set as

$$Q_{tt}^{(k)} = \begin{cases} 0 & \text{if } t \in \mathcal{L} \\ D_{tt}^{k} & \text{otherwise} \end{cases}. \tag{24}$$

Note that this formulation allows for the selection of various weight distributions. However, here our objective is to eliminate the effect of the outlier eigenvalues. Therefore, in (24), the remaining weights, except for those corresponding to the outlier eigenvalues, remain the same as in the basic GraDe. It is now expected that $N^{-1}\mathbf{Z}\mathbf{V}\mathbf{Q}^{(0)}\mathbf{V}^T\mathbf{Z}^T$ and $N^{-1}\mathbf{Z}\mathbf{V}\mathbf{Q}^{(k)}\mathbf{V}^T\mathbf{Z}^T$ are more diagonal than $N^{-1}\mathbf{Z}\mathbf{Z}^T$ and $N^{-1}\mathbf{Z}\mathbf{W}^k\mathbf{Z}^T$.

If the matrices $N^{-1}\mathbf{Z}\mathbf{V}\mathbf{Q}^{(k)}\mathbf{V}^T\mathbf{Z}^T$ are nearly diagonal for $k = 0, 1, ..., K$ and are almost equal to some diagonal matrices $\mathbf{\Lambda}_k$, by assuming model (2) for the mixing system, we have

$$N^{-1}\mathbf{X}\mathbf{V}\mathbf{Q}^{(k)}\mathbf{V}^T\mathbf{X}^T \approx \mathbf{\Omega}\mathbf{\Lambda}_k\mathbf{\Omega}^T, \quad k = 0, 1, ..., K. \tag{25}$$

By defining $\mathbf{\Omega}' \triangleq \mathbf{\Omega}\mathbf{\Lambda}_0^{1/2}$ and $\mathbf{H} \triangleq \left(\mathbf{\Omega}'\mathbf{\Omega}'^T\right)^{-1/2}$, we have

$$\begin{aligned} N^{-1}\mathbf{H}\mathbf{X}\mathbf{V}\mathbf{Q}^{(0)}\mathbf{V}^T\mathbf{X}^T\mathbf{H}^T &\approx \mathbf{I}_P, \\ N^{-1}\mathbf{H}\mathbf{X}\mathbf{V}\mathbf{Q}^{(1)}\mathbf{V}^T\mathbf{X}^T\mathbf{H}^T &\approx \mathbf{H}\mathbf{\Omega}'\mathbf{\Lambda}_1'\mathbf{\Omega}'^T\mathbf{H}^T, \\ &\vdots \\ N^{-1}\mathbf{H}\mathbf{X}\mathbf{V}\mathbf{Q}^{(K)}\mathbf{V}^T\mathbf{X}^T\mathbf{H}^T &\approx \mathbf{H}\mathbf{\Omega}'\mathbf{\Lambda}_K'\mathbf{\Omega}'^T\mathbf{H}^T, \end{aligned} \tag{26}$$

where $\mathbf{\Lambda}_k' = \mathbf{\Lambda}_k\mathbf{\Lambda}_0^{-1}$ for $k = 1, ..., K$. It is noted that the matrices $\mathbf{\Lambda}_k'$ are also diagonal for $k = 1, ..., K$, and moreover, the matrix $\mathbf{U} \triangleq \mathbf{H}\mathbf{\Omega}'$ is orthogonal, because according to the definition of $\mathbf{H}$ we have $\mathbf{U}\mathbf{U}^T = \mathbf{I}_P$. The matrix $\mathbf{H}$ is not available and an estimate of it is required. According to the first approximate equality of (25), it can be seen that $\hat{\mathbf{H}} \triangleq (N^{-1}\mathbf{X}\mathbf{V}\mathbf{Q}^{(0)}\mathbf{V}^T\mathbf{X}^T)^{-1/2}$ can be used as an estimate for $\mathbf{H}$. We define $\tilde{\mathbf{X}} \triangleq \hat{\mathbf{H}}\mathbf{X}$, so we can write

$$\begin{aligned} N^{-1}\tilde{\mathbf{X}}\mathbf{V}\mathbf{Q}^{(0)}\mathbf{V}^T\tilde{\mathbf{X}}^T &\approx \mathbf{I}_P, \\ N^{-1}\tilde{\mathbf{X}}\mathbf{V}\mathbf{Q}^{(1)}\mathbf{V}^T\tilde{\mathbf{X}}^T &\approx \mathbf{U}\mathbf{\Lambda}_1'\mathbf{U}^T, \\ &\vdots \\ N^{-1}\tilde{\mathbf{X}}\mathbf{V}\mathbf{Q}^{(K)}\mathbf{V}^T\tilde{\mathbf{X}}^T &\approx \mathbf{U}\mathbf{\Lambda}_K'\mathbf{U}^T. \end{aligned} \tag{27}$$

Let $\mathbf{S}_k \triangleq N^{-1}\tilde{\mathbf{X}}\mathbf{V}\mathbf{Q}^{(k)}\mathbf{V}^T\tilde{\mathbf{X}}^T$. Finding the orthogonal matrix $\mathbf{U}$ to make the matrices $\mathbf{U}^T\mathbf{S}_k\mathbf{U}$ as diagonal as possible for $k = 1, ..., K$ is a joint diagonalization (JD) problem. We use

$$\tilde{\mathbf{S}}_k \triangleq \frac{\mathbf{S}_k}{\|\mathbf{S}_k\|_F} \tag{28}$$

as the normalized version of $\mathbf{S}_k$, where $\|.\|_F$ denotes the Frobenius norm. So the objective function to be maximized can be defined as

$$f(\mathbf{U}) = \sum_{k=1}^{K} \left\| \mathrm{diag}(\mathbf{U}^T\tilde{\mathbf{S}}_k\mathbf{U}) \right\|_2^2. \tag{29}$$

The method in [26] performs the joint diagonalization based on the Givens rotations, and minimizing its criterion is equivalent to maximizing the objective function $f(\mathbf{U})$. This JD algorithm is used to find an orthogonal matrix $\hat{\mathbf{U}}$, which is an estimate

for $\mathbf{U}$. Then, according to the definition of $\mathbf{U}$, $\mathbf{\Omega}'$ and $\tilde{\mathbf{X}}$, the matrix $\hat{\mathbf{Z}} = \hat{\mathbf{U}}^{-1}\tilde{\mathbf{X}}$ is an estimate for $\mathbf{\Lambda}_0^{-1/2}\mathbf{Z}$, which is an scaled version of the graph source signals.

The final proposed graph source separation algorithm (called improved GraDe) is summarized in Alg. 1.

Algorithm 1 Improved GraDe

Input: Adjacency matrix $\mathbf{W}$, Observation matrix $\mathbf{X}$, K.
Output: Estimate of graph source signals $\hat{\mathbf{Z}}$.

1: Obtain eigenvalues of $\mathbf{W}$, put the indices of the "outlier" eigenvalues in $\mathcal{L}$.
2: Specify $\mathbf{Q}^{(k)}$ for $k = 0, ..., K$, for example use (24).
3: $\hat{\mathbf{H}} \leftarrow (N^{-1}\mathbf{X}\mathbf{V}\mathbf{Q}^{(0)}\mathbf{V}^T\mathbf{X}^T)^{-1/2}$, $\tilde{\mathbf{X}} \leftarrow \hat{\mathbf{H}}\mathbf{X}$
4: $\mathbf{S}_k \leftarrow N^{-1}\tilde{\mathbf{X}}\mathbf{V}\mathbf{Q}^{(k)}\mathbf{V}^T\tilde{\mathbf{X}}^T$ for $k = 1, ..., K$
5: $\tilde{\mathbf{S}}_k \leftarrow \mathbf{S}_k / \|\mathbf{S}_k\|_F$ for $k = 1, ..., K$
6: Use joint diagonalization algorithm in [26] to obtain orthogonal matrix $\hat{\mathbf{U}}$, which maximizes (29).
7: $\hat{\mathbf{Z}} \leftarrow \hat{\mathbf{U}}^{-1}\tilde{\mathbf{X}}$

IV. Numerical Results

In this section, the proposed method is first compared with the original GraDe algorithm. Subsequently, a numerical analysis is conducted on the diagonality measure in Equation 20 and its relationship with the ratio of moments of the autocovariance matrices.

A. Comparison with GraDe

Five different graph models are used to evaluate the proposed method in Section III-B, Barabási-Albert, Erdös-Rényi, random sensor, geometric and stochastic block[1]. Each result reported in this section is average of 100 repetitions, in each of which graphs and signals are generated anew. In the first simulation all the graphs have $N = 5000$ nodes. To generate Barabási-Albert graph $m_0 = 100$ initial nodes are used, in which each of initial nodes is connected to at least $m = 40$ other nodes, and every new node is connected to exactly $m = 40$ existing nodes. In Erdös-Rényi graph the probability of the existence of the edge between two nodes is set to 0.01. To generate geometric graph, a unit square is considered and the location of every node is chosen randomly inside the square. Nodes whose Euclidean distance is less than the neighborhood radius of 0.1 are then connected. Stochastic block model consists of three blocks, in which the probability of edge existence between the nodes of each block is considered 0.05, and this probability between the nodes of two different blocks is considered 0.005. It is assumed that there are $P = 4$ Gaussian GMA(2) sources. For $p = 1, ..., P$, the first parameter is set as $\theta_{1,p} = 0.2p\epsilon_1$ to make this parameter different for every source, and the second parameter $\theta_{2,p}$ is chosen randomly between 0 and $0.5\epsilon_2$ at each repetition. ϵ_1 and ϵ_2 are the normalization coefficients that depend on the graph adjacency matrix. According to (7), the parameter ϵ_k is chosen such that $\mathbb{E}\{\|\epsilon_k\mathbf{W}^k\mathbf{y}_p\|^2\}$ is equal to $\mathbb{E}\{\|\mathbf{y}_p\|^2\} = N$. So by using (11) we have $\epsilon_k = (N/\text{diag}(\mathbf{D}^{2k}))^{1/2}$. At each repetition the entries of the mixing matrix $\mathbf{\Omega}$ are generated from a standard normal distribution. According to the Fig. 3, which shows the magnitude of the eigenvalues of five mentioned graphs, for Barabási-Albert and Erdös-Rényi graph the index of the first large eigenvalue is put in $\mathcal{L}$, and for stochastic block graph the indexes of the first three large eigenvalues is put in $\mathcal{L}$. For sensor network and geometric graphs the optimum number of ignored eigenvalues is obtained from the results of the simulations. For geometric graph the first 250 and for sensor network graph the first 400 large eigenvalues are ignored by putting their indexes in $\mathcal{L}$.

In order to measure the quality of the separated signals, signal to distortion ratio is defined as SDR =

[1]The simulations in this section are done in Matlab-R2017b on a Windows 7 operating system, with a 2.2 GHz Core i7 processor and 8 GB DDR3 memory.

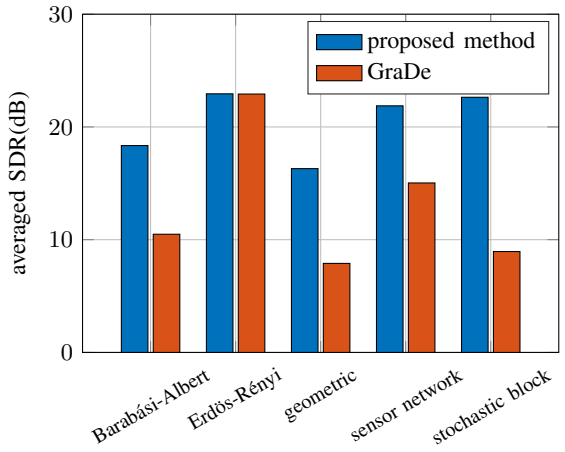

Fig. 4. Averaged signal to distortion ratio of $P = 4$ separated signals for five different graph models and two separation algorithm: proposed method and GraDe. Sources are Gaussian GMA(2) with $N = 5000$ nodes.

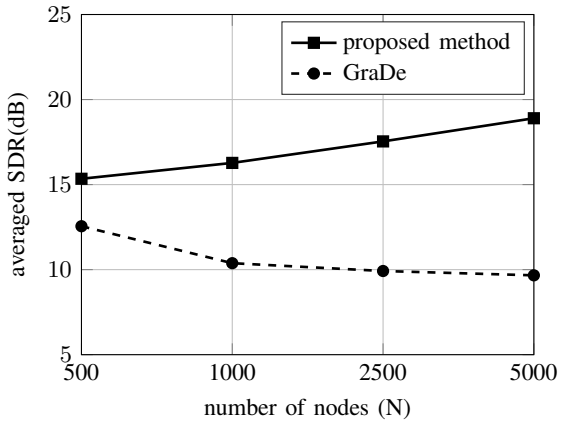

Fig. 5. Comparing the separation quality between GraDe and the proposed method for different number of nodes with $P = 4$ Gaussian GMA(2) sources on Barabási-Albert graph.

TABLE I
Required Time for a Single Run of Two Algorithms

N	Proposed Method	GraDe
500	0.049s	0.004s
1000	0.2543s	0.005s
2500	2.803s	0.009s
5000	22.316s	0.037s

$1/P \sum_{p=1}^{P} 10\log(\|\mathbf{z}_p\|^2 / \|\mathbf{z}_p - \alpha_p \hat{\mathbf{x}}_p\|^2)$, where $\mathbf{z}_p$ is the p-th latent source and $\hat{\mathbf{x}}_p$ is the separated signal corresponding to the p-th source. α_p is a coefficient that compensates the scale difference between the source and the separated signal and is calculated as $\alpha_p = \mathbf{z}_p^T \mathbf{z}_p / \mathbf{z}_p^T \hat{\mathbf{x}}_p$. In each simulation, SDR is averaged over 100 repetitions. The results of the proposed method is compared with GraDe for mentioned graph models in Fig. 4. In both algorithms, only $K = 1$ sample autocovariance matrix is used for the diagonalization. For most graph models, the proposed method yields better SDR than GraDe, that is because the matrices used in the joint diagonalization process of the proposed method are more diagonal than of GraDe. In Erdös-Rényi graph, the entries of the eigenvector corresponding to the first large eigenvalue are almost equal, so ignoring the first large eigenvalue is equivalent to removing the DC component of the signals, which is performed in GraDe as the first stage. So for Erdös-Rényi graph, the two methods are equivalent, and result in the same SDR for the separated signals. In the second simulation, Barabási-Albert graph is used and the sources and the mixing matrix and the graph is generated as in the first simulation, and the number of the graph nodes varies from 500 to 5000. As showed in Fig. 5 by increasing the number of nodes, the SDR of the separated signals decreases in GraDe and increases in proposed method.

Table I presents a comparison of the time required for a single run of two algorithms for different numbers of nodes. The reported times are obtained by averaging over 100 runs of the algorithms. Note that only the separation time is considered, excluding the time for graph and signal generation. The increased time required for the proposed method can be attributed to the need to calculate the eigen decomposition of the adjacency matrix, which is not required in GraDe.

B. Examining the diagonality measure

The diagonality measure, as defined in (20), quantifies the sum of squares of off-diagonal elements, each normalized by dividing by the corresponding diagonal elements. However, in order to present the concept of the proposed method, we opted for an indirect measure: the ratio of the standard deviation of the off-diagonal entries to the mean of the diagonal entries. To explore the relationship between the diagonality measure and the ratio of the moments, two numerical experiments are conducted. These experiments involve averaging over 1000 Monte Carlo realizations to obtain the expectation and the standard deviation of the diagonality measure, and the moments of the autocovariance matrix entries. In the first experiment, geometric graphs with varying neighborhood radii ranging from 0.01 to 0.7 are utilized. The number of the graph nodes is set to $N = 5000$ and the generation process for these geometric graphs remains consistent with the previous experiments. The reason for employing geometric graphs with different neighborhood radii is to yield distinct values for the diagonality measure, as will be demonstrated in the forthcoming results. For each graph, $P = 4$ Gaussian GMA(2) sources are generated. The first parameter of the model is set as $\theta_{1,p} = 0.2p\epsilon_1$ for $p = 1, ..., P$, and the second parameter is set as $\theta_{2,p} = 0.5\epsilon_2$. The expectation and the standard deviation of the diagonality measure of the first order autocovariance matrix ($\mathcal{D}(\mathbf{ZWZ}^T)$) and the moments of the entries of the autocovariance matrix are plotted against the neighborhood radius in Fig. 6. The standard deviation of the ij-th off-diagonal entry is denoted by σ_{ij}, and the mean of the i-th diagonal entry is represented as m_{ii}. An increase in the neighborhood radius results in an augmentation of the expectation of the diagonality measure. This is accompanied by an increase in the ratio of the standard deviation of the off-diagonal elements to the mean of the corresponding diagonal elements, illustrating a direct relationship between the diagonality measure and the mentioned ratio of moments.

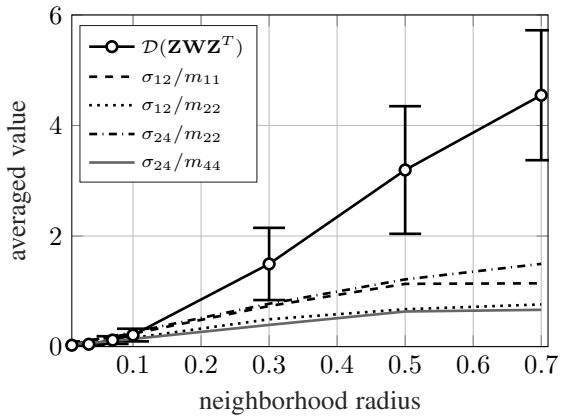

Fig. 6. The expected value of the diagonality measure of the first order autocovariance matrix shows a direct relationship with the ratio of the standard deviation of the off-diagonal entries to the mean of the corresponding diagonal entries. Geometric graph with varying neighborhood radius are employed.

In the second experiment, to generate varying values of the diagonality measure for the autocovariance matrix, we artificially create the graph adjacency matrix. The eigenvectors of the adjacency matrix are chosen to be the same as those of a geometric graph with a neighborhood radius of 0.1, and the eigenvalues are determined as

$$D_{tt} = \exp(-t/\rho) \ , \ \ t = 1, 2, ..., N. \tag{30}$$

The parameter ρ controls the decay rate of the eigenvalues. Smaller values of ρ lead to a faster decay of the eigenvalues, making the initial eigenvalues more significant in terms of magnitude compared to the rest. So, as argued in Section III-B4, a smaller value of ρ results in a larger ratio of the standard deviation of the off-diagonal entries to the mean of the corresponding diagonal entries. This is confirmed by the numerical results presented in Fig. 7. Furthermore, it is obvious that the expected value of the diagonality measure shows a direct relationships with the ratio of the moments.

V. Conclusion

In this paper, we first reviewed GraDe algorithm from a critical viewpoint, then we discussed two solutions to improve GraDe. The second one, which is the proposed one, is a method based on the spectral representation of the graph signals, and is more practical than the first one, because it does

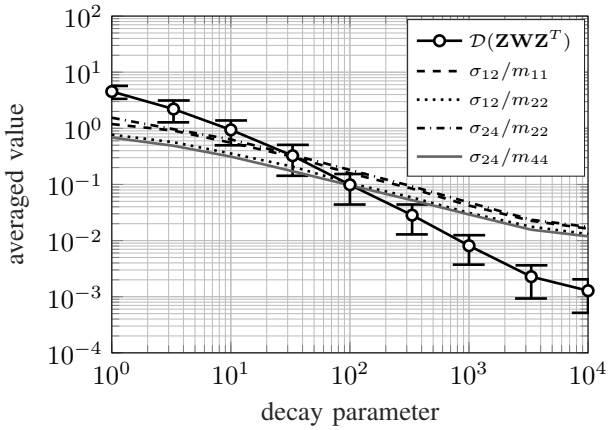

Fig. 7. The standard deviation and the expected value of the diagonality measure of the first order autocovariance matrix, and the ratio of moments of the autocovariance matrix entries plotted against the decay parameter of the eigenvalues of the adjacency matrix.

not require additional data. Numerical simulations demonstrate that the proposed method achieves better performance in the separation of the mixed Gaussian GMA sources.

REFERENCES

[1] F. Blöchl, A. Kowarsch, and F. J. Theis, "Second-order source separation based on prior knowledge realized in a graph model," in *Latent Variable Analysis and Signal Separation, LNCS*, vol. 6365. Springer, 2010, pp. 434–441.

[2] J. Miettinen, E. Nitzan, S. A. Vorobyov, and E. Ollila, "Graph signal processing meets blind source separation," *IEEE Trans. Signal Process.*, vol. 69, pp. 2585–2599, 2021.

[3] I. Jabłoński, "Graph signal processing in applications to sensor networks, smart grids, and smart cities," *IEEE Sensors J.*, vol. 17, no. 23, pp. 7659–7666, Dec. 2017.

[4] A. Ortega, "Graph signal processing: Overview, challenges, and applications," *Proc. IEEE*, vol. 106, no. 5, pp. 808–828, May 2018.

[5] P. Comon and C. Jutten, *Handbook of Blind Source Separation: Independent component analysis and applications*. Academic press, 2010.

[6] A. Belouchrani, K. Abed-Meraim, J. . Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *IEEE Trans. Signal Process.*, vol. 45, no. 2, pp. 434–444, Feb. 1997.

[7] A. Kowarsch, F. Blöchl, S. Bohl, M. Saile, N. Gretz, U. Klingmüller, and F. J. Theis, "Knowledge-based matrix factorization temporally resolves the cellular responses to il-6 stimulation," *BMC Bioinf.*, vol. 11, pp. 1–14, 2010.

[8] M. J. McKeown, S. Makeig, G. G. Brown, T.-P. Jung, S. S. Kindermann, A. J. Bell, and T. J. Sejnowski, "Analysis of fmri data by blind separation into independent spatial components," *Hum. Brain Mapping*, vol. 6, no. 3, pp. 160–188, 1998.

[9] M. J. McKeown and T. J. Sejnowski, "Independent component analysis of fmri data: examining the assumptions," *Hum. Brain Mapping*, vol. 6, no. 5-6, pp. 368–372, 1998.

[10] E. Bullmore and O. Sporns, "Complex brain networks: graph theoretical analysis of structural and functional systems," *Nat. Rev. Neurosci.*, vol. 10, no. 3, pp. 186–198, 2009.

[11] K. Li, L. Guo, J. Nie, G. Li, and T. Liu, "Review of methods for functional brain connectivity detection using fmri," *Comput. Med. Imag. Graph.*, vol. 33, no. 2, pp. 131–139, 2009.

[12] J.-F. Cardoso and A. Souloumiac, "Blind beamforming for non-gaussian signals," in *IEE proceedings F (IEE Proc. F Radar Signal Process.)*, vol. 140, no. 6. IET, 1993, pp. 362–370.

[13] A. Hyvärinen and E. Oja, "A fast fixed-point algorithm for independent component analysis," *Neural Comput.*, vol. 9, no. 7, pp. 1483–1492, 1997.

[14] A. Hyvarinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans. Neural Netw.*, vol. 10, no. 3, pp. 626–634, 1999.

[15] A. Einizade, S. H. Sardouie, and M. B. Shamsollahi, "Simultaneous graph learning and blind separation of graph signal sources," *IEEE Signal Process. Lett.*, vol. 28, pp. 1495–1499, 2021.

[16] N. J. Higham, *Functions of matrices: theory and computation*. SIAM, 2008.

[17] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.

[18] K. Alyani, M. Congedo, and M. Moakher, "Diagonality measures of hermitian positive-definite matrices with application to the approximate joint diagonalization problem," *Linear Algebra Appl.*, vol. 528, pp. 290–320, 2017.

[19] P. Erdős and A. Rényi, "On random graphs i," *Publ. Math. Debrecen*, vol. 6, no. 290-297, p. 18, 1959.

[20] M. Girvan and M. E. Newman, "Community structure in social and biological networks," *Proc. Nat. Acad. Sci.*, vol. 99, no. 12, pp. 7821–7826, 2002.

[21] P. W. Holland, K. B. Laskey, and S. Leinhardt, "Stochastic blockmodels: First steps," *Soc. Netw.*, vol. 5, no. 2, pp. 109–137, 1983.

[22] U. Feige and E. Ofek, "Spectral techniques applied to sparse random graphs," *Random Struct. Algorithms*, vol. 27, no. 2, pp. 251–275, 2005.

[23] Z. Füredi and J. Komlós, "The eigenvalues of random symmetric matrices," *Combinatorica*, vol. 1, pp. 233–241, 1981.

[24] K. Avrachenkov, L. Cottatellucci, and A. Kadavankandy, "Spectral properties of random matrices for stochastic block model," in *Proc. 4th Int. Workshop Phys.-Inspired Paradigms Wireless Commun. Netw.* IEEE, 2015, pp. 537–544.

[25] K.-I. Goh, B. Kahng, and D. Kim, "Spectra and eigenvectors of scale-free networks," *Phys. Rev. E*, vol. 64, no. 5, p. 051903, 2001.

[26] D. B. Clarkson, "Remark AS R74: A least squares version of algorithm AS 211: The FG diagonalization algorithm," *Appl. Statist.*, vol. 37, no. 2, pp. 317–321, 1988.