Journal of the American Statistical Association

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/uasa20

Generalized Liquid Association Analysis for Multimodal Data Integration

Lexin Li, Jing Zeng & Xin Zhang

To cite this article: Lexin Li, Jing Zeng & Xin Zhang (2023) Generalized Liquid Association Analysis for Multimodal Data Integration, Journal of the American Statistical Association, 118:543, 1984-1996, DOI: 10.1080/01621459.2021.2024437

To link to this article: https://doi.org/10.1080/01621459.2021.2024437

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2023, VOL. 118, NO. 543, 1984–1996: Theory and Methods
https://doi.org/10.1080/01621459.2021.2024437

Generalized Liquid Association Analysis for Multimodal Data Integration

Lexin Li[a], Jing Zeng*[b], and Xin Zhang[b]

[a]University of California at Berkeley, Berkeley, CA; [b]Florida State University, Tallahassee, FL

ABSTRACT

Multimodal data are now prevailing in scientific research. One of the central questions in multimodal integrative analysis is to understand how two data modalities associate and interact with each other given another modality or demographic variables. The problem can be formulated as studying the associations among three sets of random variables, a question that has received relatively less attention in the literature. In this article, we propose a novel generalized liquid association analysis method, which offers a new and unique angle to this important class of problems of studying three-way associations. We extend the notion of liquid association from the univariate setting to the sparse, multivariate, and high-dimensional setting. We establish a population dimension reduction model, transform the problem to sparse Tucker decomposition of a three-way tensor, and develop a higher-order orthogonal iteration algorithm for parameter estimation. We derive the nonasymptotic error bound and asymptotic consistency of the proposed estimator, while allowing the variable dimensions to be larger than and diverge with the sample size. We demonstrate the efficacy of the method through both simulations and a multimodal neuroimaging application for Alzheimer's disease research. Supplementary materials for this article are available online.

ARTICLE HISTORY
Received April 2021
Accepted December 2021

KEYWORDS
Liquid association; Multimodal neuroimaging; Sufficient dimension reduction; Tensor analysis; Tucker tensor decomposition

1. Introduction

1.1. Motivation and Problem Formulation

Multimodal data are now prevailing in scientific research, where different types of data are acquired for a common set of experimental subjects. One example is multi-omics, where different genetic information such as gene expressions, copy number alternations, and methylation changes are jointly collected for the same biological samples (Richardson, Tseng, and Sun 2016). Another example is multimodal neuroimaging, where distinct brain characteristics including brain structure, function, and chemical constituents are simultaneously measured for the same study subjects (Uludag and Roebroeck 2014). Integrative analysis of multimodal data aggregates such diverse and often complementary information, and consolidates knowledge across multiple data modalities.

In this article, we aim to address one of the questions of central interest in multimodal integrative analysis, that is, to understand how different data modalities associate and interact with each other given other modalities or covariates. This problem is of a broad scientific interest; for instance, it is useful to understand how gene expressions and microRNA levels interact given the severity of ovarian cancer and other demographical variables (Cai, Cai, and Zhang 2016), or how gene expressions and comparative genomic hybridization measures interact given the progression of breast cancer and demographics (Mai and Zhang 2019). Our motivation is a multimodal positron emission tomography (PET) study for Alzheimer's disease (AD). Amyloid-beta and tau are two hallmark proteins of AD, both of which can be measured in vivo by PET imaging using different nuclear tracers. The two proteins are closely associated in terms of spatial patterns of their accumulations, and such association patterns are believed to be affected by the subject's age (Braak and Braak 1991). Nevertheless, their specific age-dependent regional associations remain unclear. The data we study involve 81 elderly subjects, each receiving two PET scans that measure the depositions of amyloid-beta and tau, respectively. Each PET modality is represented by a vector of protein measurements at a set of brain regions of interest, with 60 regions for amyloid-beta, and 26 for tau. Our goal is to find how and where in the brain the associations of the two proteins are the most contrastive as the subject's age varies.

This problem can be formulated statistically as studying the associations of two sets of random variables $\boldsymbol{X} \in \mathbb{R}^{p_1}$ and $\boldsymbol{Y} \in \mathbb{R}^{p_2}$ conditional on the third set of random variables $\boldsymbol{Z} \in \mathbb{R}^{p_3}$. In our motivation example, $\boldsymbol{X}$ denotes the amyloid-beta PET imaging with $p_1 = 60$, $\boldsymbol{Y}$ denotes the tau PET imaging with $p_2 = 26$, and $\boldsymbol{Z}$ denotes the subject's age with $p_3 = 1$. Meanwhile, in plenty of multimodal applications, $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}$ can all be high-dimensional, and their dimensions can be even larger than the sample size. For instance, in imaging genetics (Nathoo, Kong, and Zhu 2017), $\boldsymbol{X}, \boldsymbol{Y}$ can represent different imaging modalities, whose dimensions can be in hundreds, and $\boldsymbol{Z}$ can denote the genetic information, whose dimension can be in tens of thousands or more. In high-dimensional data analysis, it is common to postulate that the data information can be sufficiently captured by some low-dimensional representations, and most often, some linear combinations of the originally

CONTACT Lexin Li lexinli@berkeley.edu University of California, Berkeley, Division of Biostatistics, Berkeley, CA 94720.
*Jing Zeng is also affiliated with International Institute of Finance, School of Management, University of Science and Technology of China
Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JASA.

high-dimensional variables (Cook 2007). Adopting this view, our question can be formulated as seeking linear combinations of $\boldsymbol{X}$ and linear combinations of $\boldsymbol{Y}$ whose conditional associations given $\boldsymbol{Z}$ are the most contrastive. In other words, we seek linear combinations of $\boldsymbol{X}$ and $\boldsymbol{Y}$ that change the most as $\boldsymbol{Z}$ changes.

1.2. Related Work

There has been a rich statistical literature studying the associations between two sets of multivariate variables $\boldsymbol{X}$ and $\boldsymbol{Y}$. A well studied and commonly used family of methods are canonical correlation analysis (CCA) and its variants (Witten, Tibshirani, and Hastie 2009; Gao et al. 2015; Li and Gaynanova 2018; Shu, Wang, and Zhu 2019; Mai and Zhang 2019, among others). CCA explores the symmetric relations between $\boldsymbol{X}$ and $\boldsymbol{Y}$, and looks for pairs of linear combinations that are most correlated. This goal, however, is different from ours, as the highly correlated linear combinations of $\boldsymbol{X}$ and $\boldsymbol{Y}$ are not necessarily the ones that are the most contrastive. For instance, a pair of linear combinations of $\boldsymbol{X}$ and $\boldsymbol{Y}$ can be highly correlated, while this correlation remains a constant as the value of $\boldsymbol{Z}$ varies, and as such they are not the target of our problem. We later numerically compare our method with CCA to further demonstrate their differences. Another popular family of methods are sufficient dimension reduction (SDR), which looks for linear combinations of $\boldsymbol{X}$ that capture full regression information of $\boldsymbol{Y}$ given $\boldsymbol{X}$; see Li (2018) for a review of this topic. Later we show that our proposed method is connected to several SDR methods, including principal Hessian directions (Li 1992; Cook 1998), and partial and groupwise sufficient dimension reduction (Chiaromonte, Cook, and Li 2002; Li, Li, and Zhu 2010). However, the goals of the two are utterly different. Whereas SDR studies asymmetric relations of $\boldsymbol{Y}$ conditioning on $\boldsymbol{X}$, we seek symmetric relations between $\boldsymbol{X}$ and $\boldsymbol{Y}$ conditioning on the third set of variables $\boldsymbol{Z}$, in that the roles of $\boldsymbol{X}$ and $\boldsymbol{Y}$ are interchangeable, but not with the role of $\boldsymbol{Z}$.

Compared to the setting of two sets of variables, there have been much fewer statistical methods studying the associations among three sets of multivariate variables in the form of $\boldsymbol{X}$ and $\boldsymbol{Y}$ given $\boldsymbol{Z}$. In his groundbreaking work, Li (2002) proposed a novel three-way interaction metric, termed *liquid association*, that measures the extent to which the association of a pair of random variables depends on the value of a third variable. He showed that this metric is particularly useful in discovering co-expressed gene pairs that are regulated by another gene. However, Li (2002) only considered the scenario where all three variables X, Y, Z are one-dimensional. Li et al. (2004) extended the notion of liquid association to the scenario of a multivariate $\boldsymbol{X}$ and a scalar Z, and sought two linear combinations $\boldsymbol{\gamma}_1^\top \boldsymbol{X}$ and $\boldsymbol{\gamma}_2^\top \boldsymbol{X}$ such that $\text{corr}(\boldsymbol{\gamma}_1^\top \boldsymbol{X}, \boldsymbol{\gamma}_2^\top \boldsymbol{X}|Z)$ varies the most with Z. Ho et al. (2011) and Yu (2018) developed some modified versions of liquid association, but still focused on the one-dimensional X, Y, Z scenario. Relatedly, Chen, Xie, and Li (2011) proposed a bivariate conditional normal model to identify the variables that regulate the co-expressions between two genes. That corresponds to the scenario with a scalar X, a scalar Y and a multivariate $\boldsymbol{Z}$. Abid et al. (2018) proposed contrastive principal component analysis for a multivariate $\boldsymbol{X}$ and a binary scalar Z, which sought linear combinations of $\boldsymbol{X}$ that have the largest changes in the conditional variance given $Z = 0$ versus $Z = 1$. Moreover, Lock et al. (2013); Li and Jung (2017) developed a class of matrix and tensor factorization methods, which aimed to decompose the multimodal data into the components that capture joint variation shared across modalities, and the components that characterize modality-specific variation. Their goal is again different from ours, as their methods did not target the conditional distribution of $\boldsymbol{X}, \boldsymbol{Y}$ given $\boldsymbol{Z}$. Finally, Xia et al. (2019) analyzed a similar dataset as our motivation example, but tackled a totally different problem. They studied hypothesis testing of covariance between the two multivariate PET measurements, and worked on the residuals after regressing out the age effect, which involves no conditioning of any third set of variables.

1.3. Proposal and Contributions

In this article, we study the three-way association among multivariate $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}$, and seek a set of linear combinations of $\boldsymbol{X}$ and $\boldsymbol{Y}$ that has a varying association as $\boldsymbol{Z}$ varies. We generalize the notion of liquid association of Li (2002) from the univariate case to the multivariate case, and develop a population dimension reduction framework for three-way association analysis. Our extension is far from trivial though, resulting in a completely different estimation method and the associated asymptotic theory. For the estimation, we transform the liquid association analysis to the problem of sparse Tucker decomposition of a three-way tensor, and introduce sparsity for the linear combinations to improve the interpretability. We then develop a higher-order orthogonal iteration algorithm for parameter estimation, and establish its algorithmic convergence. For the theory, we establish a population model that is essential for the study of statistical properties. We derive the error bound and consistency, while allowing the variable dimensions p_1, p_2, p_3 to be larger than and to diverge to infinity along with the sample size n. As a result, our proposal makes some useful contributions from both the scientific and statistical perspectives.

Scientifically, characterizing the associations between different modalities given other modalities or covariates is of crucial importance for multimodal integrative analysis. However, there is almost no existing statistical solution available for this type of problem, especially when the modalities involved are high-dimensional. Our proposal offers a unique angle for this important problem. As an illustration, for our multimodal PET study, understanding the patterns between amyloid-beta and tau with respect to age would offer insight about how pathological proteins of Alzheimer's disease interact in the aging brains.

Statistically, our proposal of generalized liquid association analysis makes a useful addition to the toolbox of association analysis of more than two sets of variables. Moreover, our method involves sparse tensor decomposition, which is itself of independent interest. Tensor data analysis is gaining increasing attention in recent years (Kolda and Bader 2009; Zhou, Li, and Zhu 2013; Sun et al. 2017; Bi, Qu, and Shen 2018; Tang, Bi, and Qu 2019; Zhang and Han 2019; Hao, Zhang, and Cheng 2020, among others); see also Bi et al. (2021) for a review of tensor analysis in statistics. Nevertheless, our proposal differs in several ways. In particular, our sparse tensor decomposition

algorithm is related to some recent singular value decomposition (SVD) type solutions for matrix denoising (Yang, Ma, and Buja 2016) and tensor denoising (Zhang and Han 2019), in that they share a similar iterative hard thresholding SVD scheme. However, our algorithm is tailored to the tensor parameter estimation with more flexible initialization and tuning. As a result, our theoretical analysis differs considerably from the denoising problems. Whereas both Yang, Ma, and Buja (2016) and Zhang and Han (2019) achieved the minimax optimal estimation for their denoising problems, we establish the dimension reduction subspace recovery consistency, variable selection consistency, and tensor parameter estimation consistency. Our rate of convergence matches the optimal rate in previous works, and all the consistency results are established in the ultrahigh dimensional setting of $s\log(p) = o(n)$, where $s = s_1s_2s_3$ and $p = p_1p_2p_3$ are the products of the number of nonzero entries and dimensions, respectively, of $\boldsymbol{X}$, $\boldsymbol{Y}$ and $\boldsymbol{Z}$. Our theoretical development is highly nontrivial, and may be of independent interest for future research involving tensor parameter estimation in a statistical model with iid data. In a sense, our work further broadens the scope of higher-order sparse SVD and tensor analysis.

The rest of the article is organized as follows. Section 2 develops the concept of generalized liquid association, and the corresponding population model of generalized liquid association analysis. Section 3 introduces the estimation algorithm, and Section 4 establishes the theoretical guarantees. Section 5 presents the simulations, and Section 6 revisits the multimodal PET study. Section 7 concludes the paper with a discussion, and the supplementary material collects all technical proofs and additional results.

2. Generalized Liquid Association Analysis

We first generalize the concept of liquid association from the univariate case to the multivariate case. The conceptual generalization itself is straightforward. Nevertheless, it motivates us to develop a dimension reduction model, along with an optimization formulation, that connects to the problem of tensor decomposition. We show that the solution to this optimization problem is closely related to the low-dimensional representation in the generalized liquid association analysis that we seek. Our method, in a sense, provides a new dimension reduction framework for three-way association analysis.

2.1. Generalized Liquid Association

We begin with a brief review of the concept of liquid association (LA) proposed by Li (2002) for the univariate case. We then extend this notion to the multivariate case.

Suppose X, Y, Z are random variables with mean zero and variance one. Define $g(z) = \mathrm{E}(XY|Z = z) : \mathbb{R} \mapsto \mathbb{R}$. Li (2002) defined the liquid association of X and Y given Z as,

$$\mathrm{LA}(X, Y|Z) = \mathrm{E}\left\{\frac{dg(Z)}{dZ}\right\} = \mathrm{E}\left\{\frac{d}{dZ}\mathrm{E}(XY|Z)\right\},$$

When Z follows a standard normal distribution, by Stein's Lemma (Stein 1981), we have,

$$\mathrm{LA}(X, Y|Z) = \mathrm{E}\left\{g(Z)Z\right\} = \mathrm{E}(XYZ).$$

Intuitively, $\mathrm{LA}(X, Y|Z)$ characterizes the change of the association of X and Y conditioning on Z through $g(z)$, and the normality condition connects this quantity with the simple unconditional expectation $\mathrm{E}(XYZ)$. In practice, the univariate Z is transformed to standard normal using the normal score transformation, and the LA measure is estimated by the sample mean $\mathrm{E}(XYZ)$. Li et al. (2004) considered an extension of LA to a multivariate $\boldsymbol{X} \in \mathbb{R}^{p_1}$ and a scalar Z, by looking for two linear combinations, such that $\mathrm{LA}(\boldsymbol{\gamma}_1^\top \boldsymbol{X}, \boldsymbol{\gamma}_2^\top \boldsymbol{X}|Z) = \boldsymbol{\gamma}_1^\top \mathrm{E}(\boldsymbol{X}\boldsymbol{X}^\top Z)\boldsymbol{\gamma}_2$ is maximized. It has a close-form solution that $\boldsymbol{\gamma}_1 = (\boldsymbol{v}_1 + \boldsymbol{v}_p)/\sqrt{2}$, $\boldsymbol{\gamma}_2 = (\boldsymbol{v}_1 - \boldsymbol{v}_p)/\sqrt{2}$, where $\boldsymbol{v}_1$ and $\boldsymbol{v}_p$ are the eigenvectors of the matrix $\mathrm{E}(\boldsymbol{X}\boldsymbol{X}^\top Z) \in \mathbb{R}^{p\times p}$ with the largest and smallest eigenvalues.

We next extend the concept of liquid association to the multivariate case, where $\boldsymbol{X} \in \mathbb{R}^{p_1}$, $\boldsymbol{Y} \in \mathbb{R}^{p_2}$, and $\boldsymbol{Z} \in \mathbb{R}^{p_3}$. Without loss of generality, suppose each variable entry in $\boldsymbol{X}$, $\boldsymbol{Y}$, and $\boldsymbol{Z}$ are standardized with mean zero and variance one. Define

$$\boldsymbol{g}(\boldsymbol{z}) = \mathrm{E}(\boldsymbol{X}\boldsymbol{Y}^\top|\boldsymbol{Z} = \boldsymbol{z}) : \mathbb{R}^{p_3} \mapsto \mathbb{R}^{p_1\times p_2}.$$

We introduce the generalized liquid association measure, which is a three-way tensor.

Definition 1. The generalized liquid association (GLA) of $\boldsymbol{X}$ and $\boldsymbol{Y}$ with respect to $\boldsymbol{Z}$ is,

$$\boldsymbol{\Phi} = \mathrm{GLA}(\boldsymbol{X}, \boldsymbol{Y}|\boldsymbol{Z}) = \mathrm{E}\left\{\frac{d}{d\boldsymbol{Z}}\boldsymbol{g}(\boldsymbol{Z})\right\} \in \mathbb{R}^{p_1\times p_2\times p_3}.$$

When $\boldsymbol{Z}$ follows a multivariate normal distribution, by the multivariate version of Stein's Lemma (Liu 1994, Lemma 1), we have the following property regarding $\boldsymbol{\Phi}$.

Proposition 1. If $\boldsymbol{Z} \sim \mathrm{Normal}(\boldsymbol{0}, \boldsymbol{\Sigma}_{\boldsymbol{Z}})$, then $\boldsymbol{\Phi} = \mathrm{E}(\boldsymbol{X} \circ \boldsymbol{Y} \circ \boldsymbol{Z}) \times_3 \boldsymbol{\Sigma}_{\boldsymbol{Z}}^{-1}$, where $\circ$ denotes the outer product, and $\times_3$ denotes the mode-3 product between a tensor and a matrix.

This conceptual extension from univariate to multivariate variables is straightforward. However, we recognize that all $\boldsymbol{X}, \boldsymbol{Y}$ and $\boldsymbol{Z}$ can be high-dimensional in that $p_1, p_2, p_3 > n$, and the dimension of $\boldsymbol{\Phi}$ is $p_1p_2p_3$, which can be ultrahigh-dimensional. Besides, it involves the inversion of a potentially high-dimensional covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{Z}}$, which makes a direct calculation or any operation on $\boldsymbol{\Phi}$ difficult, if not completely infeasible. Finally, the normality assumption can be restrictive, and there may be no simple way to transform a multivariate $\boldsymbol{Z}$ to follow an approximate normal distribution. Next, we develop a dimension reduction model for $\boldsymbol{\Phi}$, which reduces the dimensionality, avoids $\boldsymbol{\Sigma}_{\boldsymbol{Z}}^{-1}$, and improves the interpretability of the result. We also examine the normality assumption carefully, and show that it is *not* absolutely necessary for our generalized liquid association analysis.

2.2. Dimension Reduction Model for Three-Way Association

We next propose a dimension reduction model for the three-way association analysis. Our goal is to seek the linear combinations of $\boldsymbol{X}$ and $\boldsymbol{Y}$ that change the most as $\boldsymbol{Z}$ or its linear combinations change. Specifically, we first postulate that the matrix $\boldsymbol{g}(\boldsymbol{z}) =$

$\mathrm{E}(\boldsymbol{X}\boldsymbol{Y}^\top|\boldsymbol{Z}=\boldsymbol{z}) \in \mathbb{R}^{p_1\times p_2}$ varies within a low-dimensional subspace for all values of $\boldsymbol{z}$, in that

$$\boldsymbol{g}(\boldsymbol{z}) = \mathrm{E}(\boldsymbol{X}\boldsymbol{Y}^\top|\boldsymbol{Z}=\boldsymbol{z}) = \boldsymbol{\Gamma}_1 \boldsymbol{f}_z(\boldsymbol{z})\boldsymbol{\Gamma}_2^\top,$$

for some semi-orthogonal basis matrices $\boldsymbol{\Gamma}_k \in \mathbb{R}^{p_k\times r_k}, k=1,2$, and some latent function $\boldsymbol{f}_z : \mathbb{R}^{p_3} \mapsto \mathbb{R}^{r_1\times r_2}$. This implies that the linear combinations $\boldsymbol{\Gamma}_1^\top\boldsymbol{X}$ and $\boldsymbol{\Gamma}_2^\top\boldsymbol{Y}$ capture all the variations in the first two modes of the generalized liquid association tensor $\boldsymbol{\Phi}$. Next, we further assume that $\boldsymbol{g}(\boldsymbol{z})$ depends on $\boldsymbol{Z}$ only through a few linear combinations $\boldsymbol{\Gamma}_3^\top\boldsymbol{Z}$ of $\boldsymbol{Z}$, for some semi-orthogonal basis matrix $\boldsymbol{\Gamma}_3 \in \mathbb{R}^{p_3\times r_3}$. Putting these two dimension reduction structures together, we obtain our dimension reduction model for the general three-way association analysis:

$$\boldsymbol{g}(\boldsymbol{z}) = \mathrm{E}(\boldsymbol{X}\boldsymbol{Y}^\top|\boldsymbol{Z}=\boldsymbol{z}) = \boldsymbol{\Gamma}_1 \boldsymbol{f}(\boldsymbol{\Gamma}_3^\top\boldsymbol{z})\boldsymbol{\Gamma}_2^\top, \tag{1}$$

for some latent function $\boldsymbol{f} : \mathbb{R}^{r_3} \mapsto \mathbb{R}^{r_1\times r_2}$. This model is to serve as the basis for our subsequent generalized liquid association analysis. Later, we further introduce sparsity to $(\boldsymbol{\Gamma}_1, \boldsymbol{\Gamma}_2, \boldsymbol{\Gamma}_3)$ to improve the interpretability of the linear combinations in model (1).

2.3. Generalized Liquid Association Analysis via Tensor Decomposition

We propose to estimate the linear combination coefficient $\boldsymbol{\Gamma}_k$ in the dimension reduction model (1), or more accurately, the subspace $\mathrm{span}(\boldsymbol{\Gamma}_k)$ spanned by the columns of $\boldsymbol{\Gamma}_k$, $k=1,2,3$, by solving the following optimization problem,

$$\mathrm{minimize}_{\boldsymbol{G}_1,\boldsymbol{G}_2,\boldsymbol{G}_3} \left\| \boldsymbol{\Delta} - \boldsymbol{\Delta} \times_1 \boldsymbol{P}_{\boldsymbol{G}_1} \times_2 \boldsymbol{P}_{\boldsymbol{G}_2} \times_3 \boldsymbol{P}_{\boldsymbol{G}_3} \right\|_F^2, \tag{2}$$

where $\boldsymbol{\Delta} = \mathrm{E}(\boldsymbol{X}\circ\boldsymbol{Y}\circ\boldsymbol{Z}) \in \mathbb{R}^{p_1\times p_2\times p_3}$, $\boldsymbol{G}_k \in \mathbb{R}^{p_k\times r_k}$ is a semi-orthogonal matrix, $\boldsymbol{P}_{\boldsymbol{G}_k} = \boldsymbol{G}_k(\boldsymbol{G}_k^\top\boldsymbol{G}_k)^{-1}\boldsymbol{G}_k^\top$ is the projection onto the subspace $\mathrm{span}(\boldsymbol{G}_k)$, and $\times_k$ is the mode-k product between a tensor and a matrix, $k=1,2,3$. We first note that the optimization in (2) is actually the well-known tensor Tucker decomposition (Kolda and Bader 2009). Let $(\boldsymbol{G}_1', \boldsymbol{G}_2', \boldsymbol{G}_3')$ denote the population minimizer of (2). We next carefully study the connections among $(\boldsymbol{G}_1', \boldsymbol{G}_2', \boldsymbol{G}_3')$, the linear combination coefficient $(\boldsymbol{\Gamma}_1, \boldsymbol{\Gamma}_2, \boldsymbol{\Gamma}_3)$ in model (1), and the GLA measure $\boldsymbol{\Phi}$ in Definition 1.

Toward that end, we introduce an intermediate optimization problem,

$$\mathrm{minimize}_{\boldsymbol{G}_1,\boldsymbol{G}_2,\boldsymbol{G}_3} \left\| \boldsymbol{\Phi} - \boldsymbol{\Phi} \times_1 \boldsymbol{P}_{\boldsymbol{G}_1} \times_2 \boldsymbol{P}_{\boldsymbol{G}_2} \times_3 \boldsymbol{P}_{\boldsymbol{G}_3} \right\|_F^2. \tag{3}$$

Let $(\boldsymbol{G}_1'', \boldsymbol{G}_2'', \boldsymbol{G}_3'')$ denote the population minimizer of (3). Then $(\boldsymbol{G}_1'', \boldsymbol{G}_2'', \boldsymbol{G}_3'')$ is also the solution to the maximization problem,

$$\mathrm{maximize}_{\boldsymbol{G}_1,\boldsymbol{G}_2,\boldsymbol{G}_3} \left\| \boldsymbol{\Phi} \times_1 \boldsymbol{P}_{\boldsymbol{G}_1} \times_2 \boldsymbol{P}_{\boldsymbol{G}_2} \times_3 \boldsymbol{P}_{\boldsymbol{G}_3} \right\|_F^2.$$

In other words, solving (3) helps find the linear combinations of $\boldsymbol{X}$ and $\boldsymbol{Y}$ whose generalized liquid association given some linear combination of $\boldsymbol{Z}$ is maximized. In this sense, it achieves our goal of finding the most contrastive associations of $\boldsymbol{X}$ and $\boldsymbol{Y}$ given $\boldsymbol{Z}$.

The next theorem characterizes the relations among $(\boldsymbol{G}_1'', \boldsymbol{G}_2'', \boldsymbol{G}_3'')$, $(\boldsymbol{G}_1', \boldsymbol{G}_2', \boldsymbol{G}_3')$, and $(\boldsymbol{\Gamma}_1, \boldsymbol{\Gamma}_2, \boldsymbol{\Gamma}_3)$. Basically, it says minimizing (2) and (3) give the same estimates of $\boldsymbol{\Gamma}_1$ and $\boldsymbol{\Gamma}_2$ in model (1), in that they span the same subspaces. Furthermore, if $\boldsymbol{Z}$ is normally distributed, then the estimates of $\boldsymbol{\Gamma}_3$ under the two minimizations differ by a rotation.

Theorem 1. Suppose model (1) holds. Then, (a) $\mathrm{span}(\boldsymbol{\Gamma}_k) = \mathrm{span}(\boldsymbol{G}_k'') = \mathrm{span}(\boldsymbol{G}_k')$, for $k=1,2$; and (b) $\mathrm{span}(\boldsymbol{\Gamma}_3) = \mathrm{span}(\boldsymbol{G}_3'') = \boldsymbol{\Sigma}_{\boldsymbol{Z}}^{-1}\mathrm{span}(\boldsymbol{G}_3')$, if $\boldsymbol{Z}$ is normally distributed.

Theorem 1 justifies that we can achieve our goal of finding the linear combinations of $\boldsymbol{X}$ and $\boldsymbol{Y}$ that are the most contrastive given $\boldsymbol{Z}$ through the optimization problem (2), with two crucial implications. First, (2) only involves the three-way tensor $\boldsymbol{\Delta}$, but does not require the inversion of the potentially high-dimensional matrix $\boldsymbol{\Sigma}_{\boldsymbol{Z}}^{-1}$ as in $\boldsymbol{\Phi}$ in (3). Second, and perhaps more importantly, we do not require the normality of $\boldsymbol{Z}$. This is because, regardless of the distribution of $\boldsymbol{Z}$, the minimizer $(\boldsymbol{G}_1', \boldsymbol{G}_2')$ from (2) is the same as the minimizer $(\boldsymbol{G}_1'', \boldsymbol{G}_2'')$ from (3), and thus, they share the same interpretation. Only if we aim to recover $\boldsymbol{\Gamma}_3$, then we need both $\boldsymbol{\Sigma}_{\boldsymbol{Z}}^{-1}$ and the normality of $\boldsymbol{Z}$. However, we argue that, in our generalized liquid association analysis, our primary goal is to find the linear combinations of $\boldsymbol{X}$ and $\boldsymbol{Y}$ that change the most given $\boldsymbol{Z}$. As such, we are more interested in the estimation of $\boldsymbol{\Gamma}_1$ and $\boldsymbol{\Gamma}_2$, whereas the estimation of $\boldsymbol{\Gamma}_3$ provides additional dimension reduction, but is, relatively speaking, of less interest. Our proposed dimension reduction model (1) essentially serves as a bridge that connects the two optimization problems (2) and (3), which in turn connects the Tucker decomposition formulation in (2) with the generalized liquid association measure $\boldsymbol{\Phi}$ in Definition 1.

Finally, we remark that, our proposal is similar in spirit to an SDR method, the principal Hessian directions (Li 1992). It was also derived based on Stein's Lemma, but was proven useful for finding low-dimensional representations in graphics (Cook 1998), and for detecting interaction terms in regressions (Tang, Fang, and Dong 2020), even without the normality.

3. Sparse Tensor Estimation

Tucker decomposition is usually solved by a tensor SVD type algorithm, for example, a higher-order orthogonal iteration (HOOI) algorithm, which was first proposed by De Lathauwer, De Moor, and Vandewalle (2000), and later studied in statistical models (e.g., Zhang and Xia 2018; Luo et al. 2020). Next, we develop an iterative algorithm to solve (2). We further introduce sparsity in this decomposition to improve the interpretability of the result.

For n iid data observations $\{\boldsymbol{X}_i, \boldsymbol{Y}_i, \boldsymbol{Z}_i, i=1,\ldots,n\}$, without loss of generality, we assume the data is centered, so that $\sum_{i=1}^n \boldsymbol{Z}_i = \boldsymbol{0}$. The centering of $\boldsymbol{X}_i$ and $\boldsymbol{Y}_i$ is not required, but for simplicity, we assume $\sum_{i=1}^n \boldsymbol{X}_i = \sum_{i=1}^n \boldsymbol{Y}_i = \boldsymbol{0}$ as well. Then the sample estimator of $\boldsymbol{\Delta}$ is simply $\tilde{\boldsymbol{\Delta}} = n^{-1}\sum_{i=1}^n \boldsymbol{X}_i \circ \boldsymbol{Y}_i \circ \boldsymbol{Z}_i \in \mathbb{R}^{p_1\times p_2\times p_3}$. Following the dimension reduction model (1) and the optimization problem (2), $\tilde{\boldsymbol{\Delta}}$ admits a Tucker tensor decomposition structure, which can be solved by some version of the higher-order singular value decomposition algorithm. Specifically, we simplify the STAT-SVD algorithm recently proposed by Zhang and Han (2019) for the tensor denoising problem, and tailor it to our generalized liquid association analysis problem to estimate $\boldsymbol{\Gamma}_k$, $k=1,2,3$. It consists of two major components, SVD of a matrix, and hard thresholding to identify important variables. We summarize the estimation procedure in Algorithm 1, then discuss each step in detail. We further

Algorithm 1 Generalized liquid association analysis via sparse tensor decomposition.

Input: The centered data $\{\boldsymbol{X}_i \in \mathbb{R}^{p_1}, \boldsymbol{Y}_i \in \mathbb{R}^{p_2}, \boldsymbol{Z}_i \in \mathbb{R}^{p_3}, i = 1, \ldots, n\}$, the Tucker ranks $r_k \le p_k$, and the sparsity parameters $(\eta_k, \tilde{\eta}_k)$, $k = 1, 2, 3$.

Step 1, initialization: Compute the sample estimate $\tilde{\boldsymbol{\Delta}} = n^{-1} \sum_{i=1}^{n} \boldsymbol{X}_i \circ \boldsymbol{Y}_i \circ \boldsymbol{Z}_i$. Obtain the initial active set $\hat{I}_k^{(0)}$, and the initial basis matrices by,

$$\hat{I}_k^{(0)} = \left\{ j : ||(\tilde{\boldsymbol{\Delta}}_k)_{[j,:]}||_\infty \ge \eta_k \right\},$$

$$\hat{\boldsymbol{\Gamma}}_k^{(0)} = \mathrm{SVD}_{r_k} \left\{ \boldsymbol{D}_{\hat{I}_k^{(0)}} \tilde{\boldsymbol{\Delta}}_k \boldsymbol{D}_{\hat{I}_{-k}^{(0)}} \right\} \in \mathbb{R}^{p_k \times r_k}, \ k = 1, 2, 3.$$

repeat

Step 2a: Update the active set: $\hat{I}_k^{(t)} = \left\{ j : ||(\tilde{\boldsymbol{\Delta}}_k)_{[j,:]} \hat{\boldsymbol{\Gamma}}_{-k}^{(t)}||_2^2 \ge \tilde{\eta}_k \right\}, k = 1, 2, 3.$

Step 2b: Perform SVD: $\hat{\boldsymbol{\Gamma}}_k^{(t)} = \mathrm{SVD}_{r_k} \left\{ \boldsymbol{D}_{\hat{I}_k^{(t)}} \tilde{\boldsymbol{\Delta}}_k \hat{\boldsymbol{\Gamma}}_{-k}^{(t)} \right\} \in \mathbb{R}^{p_k \times r_k}, k = 1, 2, 3.$

until some stopping criterion is met.

Output: The estimated basis matrices $\hat{\boldsymbol{\Gamma}}_k$, $k = 1, 2, 3$, and $\hat{\boldsymbol{\Delta}} = \tilde{\boldsymbol{\Delta}} \times_1 \boldsymbol{P}_{\hat{\boldsymbol{\Gamma}}_1} \times_2 \boldsymbol{P}_{\hat{\boldsymbol{\Gamma}}_2} \times_3 \boldsymbol{P}_{\hat{\boldsymbol{\Gamma}}_3}$.

study the algorithmic convergence of the estimation procedure in Section S2 of the supplementary material. It is also noted that, in our formulation, we allow the number of variables p_k, $k = 1, 2, 3$, to be much larger than the sample size n.

We first introduce some notation. For a vector $\boldsymbol{a} = (a_i) \in \mathbb{R}^p$, define the ℓ_2-norm as $||\boldsymbol{a}||_2 = (\sum_{i=1}^{p} a_i^2)^{1/2}$, and the ℓ_∞-norm as $||\boldsymbol{a}||_\infty = \max_{1 \le i \le p} |a_i|$. For a matrix $\boldsymbol{A} = (a_{ij}) \in \mathbb{R}^{p \times q}$, define the Frobenius norm as $||\boldsymbol{A}||_F = (\sum_{i=1}^{p} \sum_{j=1}^{q} a_{ij}^2)^{1/2}$. For an integer p, let $[p]$ denote the set $\{1, \ldots, p\}$. For the index sets $I \subseteq [p]$, $J \subseteq [q]$, let $\boldsymbol{A}_{[I,J]}$ denote the corresponding $|I| \times |J|$ submatrix, while the whole index set $[p]$ is simplified as ":"; for example, $\boldsymbol{A}_{[[p],J]} = \boldsymbol{A}_{[:,J]}$. Let $\mathrm{SVD}_r(\boldsymbol{A}) \in \mathbb{R}^{p \times r}$ denote the left r singular vectors of $\boldsymbol{A}$, with $r \le q$. Let $\tilde{\boldsymbol{\Delta}}_k$ and $\boldsymbol{\Delta}_k$ denote the mode-k matricization of the tensors $\tilde{\boldsymbol{\Delta}}$ and $\boldsymbol{\Delta}$, $k = 1, 2, 3$. Define $\boldsymbol{\Gamma}_{-1} = \boldsymbol{\Gamma}_2 \otimes \boldsymbol{\Gamma}_3, \boldsymbol{\Gamma}_{-2} = \boldsymbol{\Gamma}_3 \otimes \boldsymbol{\Gamma}_1, \boldsymbol{\Gamma}_{-3} = \boldsymbol{\Gamma}_1 \otimes \boldsymbol{\Gamma}_2$, where $\otimes$ is the Kronecker product. Define the active sets of variables in the generalized liquid association analysis as,

$$I_k = \left\{ j : (\boldsymbol{\Delta}_k)_{[j,:]} \ne \boldsymbol{0}, \ 1 \le j \le p_k \right\} \subseteq [p_k], \quad k = 1, 2, 3. \quad (4)$$

As an example, the jth variable X_j in $\boldsymbol{X}$ corresponds to the jth row of $\boldsymbol{\Gamma}_1 \in \mathbb{R}^{p_1 \times r_1}$, $j = 1, \ldots, p_1$. Therefore, variable selection in $\boldsymbol{X}$ translates to the row-wise sparsity in $\boldsymbol{\Gamma}_1$, and correspondingly, the row-wise sparsity in $\boldsymbol{\Delta}_1 \in \mathbb{R}^{p_1 \times (p_2 p_3)}$. Define the diagonal matrix $\boldsymbol{D}_{I_k} \in \mathbb{R}^{p_k \times p_k}$ that has one on the ith diagonal element if $i \in I_k$ and zero elsewhere. This matrix represents variable selection along each mode, and is used repeatedly in our estimation algorithm. Define $\boldsymbol{D}_{I_{-1}} = \boldsymbol{D}_{I_2} \otimes \boldsymbol{D}_{I_3}$, whereas I_{-1} denotes the pair of subsets I_2 and I_3. Define $\boldsymbol{D}_{I_{-2}}, \boldsymbol{D}_{I_{-3}}, I_{-2}, I_{-3}$ similarly. Also define $\hat{\boldsymbol{\Gamma}}_{-k}, \hat{I}_{-k}$ in a similar fashion.

We start the algorithm by computing the sample estimate $\tilde{\boldsymbol{\Delta}}$, then perform the initial selection of important variables and initial SVD in Step 1 of Algorithm 1. From (4), we see the selection of important variables can be achieved based on $||(\tilde{\boldsymbol{\Delta}}_k)_{[j,:]}||$ for some appropriate norm $||\cdot||$. In the initialization step, we employ the ℓ_∞-norm, and achieve the selection by hard thresholding under the sparsity parameter η_k. The two diagonal matrices $\boldsymbol{D}_{\hat{I}_k^{(0)}}$, $\boldsymbol{D}_{\hat{I}_{-k}^{(0)}}$ operate as the subset selection operator within the SVD operator. Depending on the sparsity parameter η_k, we may keep all the variables in the active set, that is, $\hat{I}_k = [p_k]$.

Next, we iterate the algorithm by repeatedly selecting important variables and performing SVD in Step 2 of Algorithm 1. We continue to do the selection by hard thresholding, but we use a different norm, that is, the ℓ_2-norm rather than the ℓ_∞-norm, and a different sparsity parameter $\tilde{\eta}_k$. This change of the norm is practically useful because of the following consideration. In the initialization step, the column dimension of $\tilde{\boldsymbol{\Delta}}_k$ is $\prod_{k' \ne k} p_{k'}$ and is often very large, and thus, the ℓ_∞-norm is more effective in screening out the zero rows in $\tilde{\boldsymbol{\Delta}}_k$. During the iterations, the active variable set I_k is selected based on $\tilde{\boldsymbol{\Delta}}_k \boldsymbol{\Gamma}_{-k}$, which has a much smaller column dimension $\prod_{k' \ne k} r_{k'}$. As such, the ℓ_2-norm is preferred to being able to pick up potentially weaker signals and to refine the selection from the initialization. Moreover, a different thresholding parameter $\tilde{\eta}_k$ during the iterations gives more flexibility.

We alternate Steps 2a and 2b until some termination criterion is met. That is, we terminate the algorithm if the consecutive estimates are close enough, in that the difference between the squared ℓ_2-norm of the singular values of the two iterations is smaller than 10^{-6}, or if the algorithm reaches the maximum number of iterations, say, 100. In our numerical experiments, we find that the algorithm converges fast, usually within 10–20 iterations. We output the estimated basis matrices $\hat{\boldsymbol{\Gamma}}_k$, $k = 1, 2, 3$, along with the updated estimate $\hat{\boldsymbol{\Delta}} = \tilde{\boldsymbol{\Delta}} \times_1 \boldsymbol{P}_{\hat{\boldsymbol{\Gamma}}_1} \times_2 \boldsymbol{P}_{\hat{\boldsymbol{\Gamma}}_2} \times_3 \boldsymbol{P}_{\hat{\boldsymbol{\Gamma}}_3}$ that follows a Tucker decomposition.

Our algorithm is related to the STAT-SVD algorithm of Zhang and Han (2019), in that we both use hard thresholding SVD iteratively. However, Zhang and Han (2019) targeted a tensor denoising problem involving identically distributed normal errors, and used a double thresholding scheme with a theoretical thresholding value. Their algorithm, after obtaining the variance of the errors, became tuning-free in terms of the thresholding parameter. By contrast, we aim to obtain a low-rank tensor estimator in the context of generalized liquid association analysis. The sample estimator does not have iid entries, and we use a single thresholding scheme with two data-driven tuning parameters. This leads to a more flexible tuning, and consequently an utterly different approach for the asymptotic analysis.

The thresholding values η_k and $\tilde{\eta}_k$ are treated as tuning parameters, and we propose a prediction-based approach for tuning. That is, we first randomly split the data into a training set and a testing set, and obtain the sample estimates $\tilde{\boldsymbol{\Delta}}^{\text{train}}$ and $\tilde{\boldsymbol{\Delta}}^{\text{test}}$ separately. We then apply Algorithm 1 to $\tilde{\boldsymbol{\Delta}}^{\text{train}}$ to obtain $\hat{\boldsymbol{\Gamma}}_k^{\text{train}}(\eta)$, $k = 1, 2, 3$, under a given set of tuning parameters $\eta = (\eta_1, \eta_2, \eta_3, \tilde{\eta}_1, \tilde{\eta}_2, \tilde{\eta}_3)$. We choose η to minimize the discrepancy,

$$L(\eta) = \left\| \tilde{\boldsymbol{\Delta}}^{\text{test}} - \tilde{\boldsymbol{\Delta}}^{\text{test}} \times_1 \boldsymbol{P}_{\hat{\boldsymbol{\Gamma}}_1^{\text{train}}(\eta)} \times_2 \boldsymbol{P}_{\hat{\boldsymbol{\Gamma}}_2^{\text{train}}(\eta)} \times_3 \boldsymbol{P}_{\hat{\boldsymbol{\Gamma}}_3^{\text{train}}(\eta)} \right\|_F. \quad (5)$$

Meanwhile, the ranks (r_1, r_2, r_3) take some pre-specified values. In practice, r_k is often taken as 1 or 2 for exploratory analysis and data visualization. This is similar in spirt as canonical correlation analysis. Actually, rank selection is still an open and active topic in CCA and tensor problems, and we leave a full treatment of rank selection as future research.

4. Theoretical Properties

We establish the theoretical guarantees for the estimated Tucker tensor $\hat{\boldsymbol{\Delta}} = \tilde{\boldsymbol{\Delta}} \times_1 \boldsymbol{P}_{\hat{\boldsymbol{\Gamma}}_1} \times_2 \boldsymbol{P}_{\hat{\boldsymbol{\Gamma}}_2} \times_3 \boldsymbol{P}_{\hat{\boldsymbol{\Gamma}}_3}$, the estimated subspace basis matrices $\hat{\boldsymbol{\Gamma}}_k$, and the estimated active sets $\hat{I}_k^{(t)}$, $k = 1, 2, 3$, from Algorithm 1. In our theoretical analysis, we allow both the tensor dimension $p = \prod_{k=1}^3 p_k$ and the sparsity level $s = \prod_{k=1}^3 s_k$ to diverge with the sample size n, while we fix the tensor rank $r = \prod_{k=1}^3 r_k$. We begin with two mild regularity conditions.

(A1) Suppose $|X_{j_1} Y_{j_2}| \le M$, for some constant $M > 0$, $j_1 = 1, \ldots, p_1, j_2 = 1, \ldots, p_2$, and suppose Z_{j_3} is sub-Gaussian with the parameter $\sigma^2 > 0, j_3 = 1, \ldots, p_3$.

(A2) Suppose $\lambda \ge \max\{C_1\sqrt{s \log p/n}, C_2\}$, for some constants $C_1, C_2 > 0$, where $\lambda \equiv \min\{\lambda_1, \lambda_2, \lambda_3\}$, and λ_k is the smallest nonzero singular value of $\boldsymbol{\Delta}_k$, $k = 1, 2, 3$.

Assumption 4 requires $\boldsymbol{X}$ and $\boldsymbol{Y}$ to be bounded, and $\boldsymbol{Z}$ to be sub-Gaussian, which are necessary to establish the concentration of each element in $\tilde{\boldsymbol{\Delta}}$ to its population counterpart. The sub-Gaussian assumption is weaker than the normality assumption, and is widely used in high-dimensional nonasymptotic analysis (see, e.g., Wainwright 2019). Besides, it assumes each individual Z_k to be sub-Gaussian, which is weaker than assuming the joint distribution $\boldsymbol{Z}$ is sub-Gaussian. The constant σ^2 does not require all Z_{j_3} to have the same variance. If Z_{j_3} is sub-Gaussian with the parameter $\sigma_{j_3}^2$, $j_3 = 1, \ldots, p_3$, then we can set $\sigma^2 = \max_{j_3} \sigma_{j_3}^2$. Assumption 4 ensures that there is a reasonable gap between the zero and nonzero eigenvalues in $\boldsymbol{\Delta}_k$, under which the consistency for the estimator $\hat{\boldsymbol{\Gamma}}_k$ is ensured. This type of assumption on the eigenvalues is frequently used in high-dimensional singular value decomposition (Yu, Wang, and Samworth 2015; Yang, Ma, and Buja 2016; Zhang and Han 2019).

Next, we derive the nonasymptotic error bound and variable selection property of our estimators. Let $\hat{\boldsymbol{\Delta}}$, $\hat{\boldsymbol{\Gamma}}_k$, and $\hat{I}_k$, $k = 1, 2, 3$, denote the estimators and the corresponding active sets returned from Algorithm 1 after $t_{\max}$ iterations, under the theoretical thresholding values $\eta_k = \sqrt{\alpha \log p/n}$, and $\tilde{\eta}_k = \alpha s_{-k} \log p/n$, where $\alpha = 513(M + \sigma)^4$, and M, σ are as defined in Assumption 4. Moreover, since the basis matrix $\hat{\boldsymbol{\Gamma}}_k$ is identifiable only up to an orthogonal rotation, we characterize its bound in terms of the projection matrix $\boldsymbol{P}_{\hat{\boldsymbol{\Gamma}}_k}$.

Theorem 2 (Nonasymptotic properties). Suppose Assumptions 4, 4, and model (1) hold. Then, with probability at least $1 - C\max\left[p^{-\gamma}, p^{-\left\{\sqrt{n(\gamma+1)/(2\log p)}-1\right\}}\right]$, (a) $||\hat{\boldsymbol{\Delta}} - \boldsymbol{\Delta}||_F \le (c_1 + c_2 2^{-t_{\max}})\sqrt{s \log p/n}$; (b) $||\boldsymbol{P}_{\hat{\boldsymbol{\Gamma}}_k} - \boldsymbol{P}_{\boldsymbol{\Gamma}_k}||_F \le (c_3 + c_4 2^{-t_{\max}})\sqrt{s \log p/n}$, $k = 1, 2$, and $||\boldsymbol{P}_{\hat{\boldsymbol{\Gamma}}_3} - \boldsymbol{P}_{\boldsymbol{G}_3'}||_F \le (c_3 + c_4 2^{-t_{\max}})\sqrt{s \log p/n}$; and (c) $\hat{I}_k \subseteq I_k$, $k = 1, 2, 3$, where $\gamma, C, c_1, c_2, c_3, c_4$ are some positive constants.

We make a few remarks. First, statements (a) and (b) establish the nonasymptotic error bound for the Tucker tensor estimator $\hat{\boldsymbol{\Delta}}$, as well as the subspace spanned by the basis matrix $\hat{\boldsymbol{\Gamma}}_k$, $k = 1, 2, 3$. Note that $\hat{\boldsymbol{\Gamma}}_1, \hat{\boldsymbol{\Gamma}}_2$ directly target $\boldsymbol{G}_1', \boldsymbol{G}_2'$ from optimization (2), which, by Theorem 1(a), are the same as $\boldsymbol{\Gamma}_1, \boldsymbol{\Gamma}_2$ in our dimension reduction model (1), as well as $\boldsymbol{G}_1'', \boldsymbol{G}_2''$ from the generalized liquid association measure $\boldsymbol{\Phi}$, in the sense that they span the same subspaces. Meanwhile, $\hat{\boldsymbol{\Gamma}}_3$ targets the population minimizer $\boldsymbol{G}_3'$ from (2), which, by Theorem 1(b), differs from $\boldsymbol{\Gamma}_3$ in model (1) and $\boldsymbol{G}_3''$ from $\boldsymbol{\Phi}$ by a rotation, when $\boldsymbol{Z}$ is normally distributed. However, as we have discussed after Theorem 1, our primary interest is to recover $\boldsymbol{\Gamma}_1, \boldsymbol{\Gamma}_2$, rather than $\boldsymbol{\Gamma}_3$. As such, we do *not* require the normality assumption for Theorem 2. Second, the error bounds in statements (a) and (b) are functions of the maximum number of iterations $t_{\max}$, and they decrease when $t_{\max}$ increases. As such, our estimators are to have improved accuracy with more iterations. Third, statement (c) shows that our method avoids selecting inactive variables with a high probability. This result is similar to that in Zhang and Han (2019), and can be viewed as a weaker version of variable selection consistency when compared to Theorem 3 below. Fourth, we treat $t_{\max}$ as a constant in this section, regardless of the tensor dimension or sample size. This is because the algorithm converges fast, often within 10–20 iterations. On the other hand, we can easily extend the results by allowing $t_{\max}$ to diverge. For instance, parallel to Zhang and Han (2019), we can let $t_{\max}$ diverge at the rate of $o(p)$. Finally, when the sample size n is sufficiently large, that is, when $n \gg \log p$, the statements in Theorem 2 hold with probability at least $1 - Cp^{-\gamma}$. Besides, the constants $c_1, \ldots, c_4$ depend on the constants M and σ in Assumption 4, and their explicit forms are given in Section S5.3 of the supplementary material.

We also briefly comment that, in real applications, the modalities may be low-dimensional or have no sparsity. Accordingly, we can modify Algorithm 1 to a nonsparse version, by setting $\eta_k = 0$ in Step 1, and $\tilde{\eta}_k = 0$ in Step 2, for $k = 1, 2, 3$. We give the corresponding nonasymptotic error bounds in Section S3 of the supplementary material.

Next, we establish the asymptotic consistency of the tensor parameter estimation, subspace estimation and variable selection as n, p, s diverge to infinity. We allow $s \log p = o(n)$, that is, each tensor dimension p_k can diverge faster than the sample size n.

Corollary 1 (Asymptotic consistency). Suppose the conditions in Theorem 2 hold, and $s \log p = o(n)$. Then, as $n, p, s \to \infty$, with probability tending to one, (a) $||\hat{\boldsymbol{\Delta}} - \boldsymbol{\Delta}||_F \to 0$; (b) $||\boldsymbol{P}_{\hat{\boldsymbol{\Gamma}}_k} - \boldsymbol{P}_{\boldsymbol{\Gamma}_k}||_F \to 0$, $k = 1, 2$, and $||\boldsymbol{P}_{\hat{\boldsymbol{\Gamma}}_3} - \boldsymbol{P}_{\boldsymbol{G}_3'}||_F \to 0$; and (c) $\hat{I}_k \subseteq I_k$, $k = 1, 2, 3$.

While Theorem 2(c) shows that our method can exclude the inactive variables from the selection, we show in the next theorem that our method can exactly recover the active variables, with a high probability. We need an additional regularity condition.

(A3) Suppose $\delta_{\min} \geq C_3\sqrt{s\log p/n}$, for some sufficiently large constant C_3, where $\delta_{\min} \equiv \min_{k\in\{1,2,3\}, i\in I_k} ||(\mathbf{\Delta}_k)_{[i,:]}||_2$ denotes the minimal signal strength.

Theorem 3 (Variable selection consistency). Suppose Assumptions 4 to 4, and model (1) hold. Then, with probability at least $1 - C' \max\left[p^{-\gamma}, p^{-\left\{\sqrt{n(\gamma+1)/(2\log p)}-1\right\}}\right]$, we have, $\hat{I}_k = I_k$, for $k = 1, 2, 3$, where γ, C' are some positive constants.

Assumption 4 ensures the signal of the active variables is of a reasonable strength when n, p, s diverge, which in turn leads to the variable selection consistency in Theorem 3. Note that $\delta_{\min}$ is also the minimal Frobenius norm of the nonzero slices in $\mathbf{\Delta}$, that is, the slices of $\mathbf{\Delta}$ corresponding to those variables $i \in I_k$, $k = 1, 2, 3$. We feel this assumption is reasonable. Actually, if we allow s, p to diverge with n at the rate of $s\log p = o(n)$, and suppose the nonzero entries of $\mathbf{\Delta}$ are bounded away from zero, then this assumption is satisfied.

5. Simulation Studies

5.1. Simulation Setup

We carry out the simulations to investigate the empirical performance of the proposed generalized liquid association analysis (GLAA) method. We consider three scenarios. In the first scenario, we fix the dimension of $\boldsymbol{Z}$ at $p_3 = 1$, and increase the dimensions of $\boldsymbol{X}$ and $\boldsymbol{Y}$ as $p_1 = p_2 = \{100, 200, 300, 400, 500\}$. In the second scenario, we fix $p_1 = p_2 = 100$, and increase $p_3 = \{20, 40, 60, 80, 100\}$. In both cases, we fix the sample size at $n = 500$. In the third scenario, we fix $p_1 = 100, p_2 = 25, p_3 = 1$, and increase the sample size $n = \{60, 80, 100, 120, 160\}$. We generate the data in the following way. For $i = 1, \ldots, n$, we first generate $\boldsymbol{Z}_i$ from a normal distribution with mean zero and covariance $\boldsymbol{I}_{p_3}$. We then generate $(\boldsymbol{X}_i, \boldsymbol{Y}_i)$ jointly from a normal distribution with mean zero and covariance,

$$\text{cov}(\boldsymbol{X}, \boldsymbol{Y}|\boldsymbol{Z} = \boldsymbol{Z}_i) = \begin{pmatrix} \mathbf{\Sigma}_{\boldsymbol{X}} & \mathbf{\Gamma}_1 \boldsymbol{f}(\mathbf{\Gamma}_3^\top \boldsymbol{Z}_i)\mathbf{\Gamma}_2^\top \\ \mathbf{\Gamma}_2 \boldsymbol{f}^\top(\mathbf{\Gamma}_3^\top \boldsymbol{Z}_i)\mathbf{\Gamma}_1^\top & \mathbf{\Sigma}_{\boldsymbol{Y}} \end{pmatrix}.$$

To ensure the positive-definiteness of this covariance matrix, we set $\mathbf{\Gamma}_1 = \mathbf{\Sigma}_{\boldsymbol{X}}^{1/2}(\boldsymbol{O}_1^\top, \boldsymbol{0})^\top$ and $\mathbf{\Gamma}_2 = \mathbf{\Sigma}_{\boldsymbol{Y}}^{1/2}(\boldsymbol{O}_2^\top, \boldsymbol{0})^\top$, where $\boldsymbol{O}_1 = \boldsymbol{O}_2 \in \mathbb{R}^{5\times 2}$ with the first column being $(1, 1, 1, 1, 1)/\sqrt{5}$, and the second column being $(0, 0, 0, -1, 1)/\sqrt{2}$. As a result, in this example, for $\boldsymbol{X}$ and $\boldsymbol{Y}$, the ranks are $r_1 = r_2 = 2$, and the sparsity levels are $s_1 = s_2 = 5$. The marginal covariance matrix $\mathbf{\Sigma}_{\boldsymbol{X}}$ is set as a block diagonal matrix, $\mathbf{\Sigma}_{\boldsymbol{X}} = \text{bdiag}(\mathbf{\Sigma}_{\boldsymbol{X},1}, \mathbf{\Sigma}_{\boldsymbol{X},2})$, where $\mathbf{\Sigma}_{\boldsymbol{X},1} \in \mathbb{R}^{s_1\times s_1}$ corresponds to the active variables in $\boldsymbol{X}$ and takes the form of an AR structure such that its (i, j)th entry equals $\sigma_{ij} = 0.3^{|i-j|}, i, j = 1, \ldots, s_1$, and $\mathbf{\Sigma}_{\boldsymbol{X},2} \in \mathbb{R}^{(p_1-s_1)\times(p_1-s_1)}$ is the identity matrix. The marginal covariance matrix $\mathbf{\Sigma}_{\boldsymbol{Y}}$ is constructed in a similar fashion. The matrix $\boldsymbol{f}(\mathbf{\Gamma}_3^\top \boldsymbol{Z}_i)$ is set as $\text{diag}\{f_1(\mathbf{\Gamma}_3^\top \boldsymbol{Z}_i), f_2(\mathbf{\Gamma}_3^\top \boldsymbol{Z}_i)\}$, where $f_1(a) = 0.95\text{sign}(a)$ and $f_2(a) = 0.85\text{sign}(a)$. In the Appendix, we consider additional simulations using $f(a; \rho, \xi) = \rho\{2/(1 + e^{-2\xi a}) - 1\}$, with different parameters $0 < \rho \leq 1$ and $\xi > 0$ that control the magnitude and speed of changes in $\text{cov}(\boldsymbol{X}, \boldsymbol{Y}|\boldsymbol{Z})$. For the first and the third scenarios, $p_3 = 1$ and thus, $\mathbf{\Gamma}_3 = 1$. For the second scenario, where p_3 varies from 20 to 100, we set $\mathbf{\Gamma}_3 = (1, 1, 1, 1, 1, 0, \ldots, 0)$, with $s_3 = 5$ and $r_3 = 1$. When applying the proposed method, we adopt the theoretical forms for the tuning parameters, that is, $\eta_k = \sqrt{\zeta\log p/n}$, and $\tilde{\eta}_k = \zeta s_{-k}\log p/n$, and tune ζ following the approach in (5).

There is no existing method designed to directly address our targeting problem. For the purpose of comparison, we consider three relevant solutions. The first solution we consider is a naive and marginal extension of the univariate liquid association (ULA) method from Li (2002). That is, we construct a tensor estimator $\tilde{\mathbf{\Phi}}$, each entry of which is the sample univariate LA for the triplet $(X_{j_1}, Y_{j_2}, Z_{j_3})$ as defined in Li (2002), $j_1 = 1, \ldots, p_1, j_2 = 1, \ldots, p_2, j_3 = 1, \ldots, p_3$. We then perform the usual SVD to each mode-k matricization of $\tilde{\mathbf{\Phi}}$, denoted by $\tilde{\mathbf{\Phi}}_{(k)}$, under the given rank to obtain the estimates of basis matrices; that is, $\text{SVD}_{r_k}\{\tilde{\mathbf{\Phi}}_{(k)}\}, k = 1, 2, 3$. The second and third solutions we consider are two different versions of canonical correlation analysis, the penalized matrix decomposition (PMD) method of Witten, Tibshirani, and Hastie (2009), and the sparse canonical correlation analysis (SCCA) method of Mai and Zhang (2019). We have chosen these two versions due to their computational simplicity and superior empirical performance. We note that CCA is not designed to incorporate the third set of variables $\boldsymbol{Z}$. We thus, simply take the first r_1 and r_2 directions of $\boldsymbol{X}$ and $\boldsymbol{Y}$ from CCA as the estimated basis matrices corresponding to $\mathbf{\Gamma}_1$ and $\mathbf{\Gamma}_2$. We evaluate the performance of each method in terms of the variable selection accuracy and the subspace estimation accuracy.

For variable selection, we record the true positive rate (TPR) and false positive rate (FPR) for each mode. Recall from (4), the active set of variables is I_k, which is also the index set of nonzero rows in $\mathbf{\Gamma}_k$. Let $\hat{I}_k$ denote the estimated active set corresponding to $\hat{\mathbf{\Gamma}}_k$, then TPR-$k = |I_k \cap \hat{I}_k|/s_k$ and FPR-$k = |I_k^c \cap \hat{I}_k|/(p_k - s_k)$, $k = 1, 2, 3$. For GLAA, PMD and SCCA, we estimate the active set as $\hat{I}_k = \{i : \text{there exist nonzero elements in the } i\text{th row of } \hat{\mathbf{\Gamma}}_k\}$. For ULA, it does not perform any variable selection. For the purpose of comparison, we simply calculate the ℓ_2-norm of each row for the kth matricization $\tilde{\mathbf{\Phi}}_{(k)}$, arrange the row indices in a descending order by the ℓ_2-norms, then select the first s_k rows for each mode, $k = 1, 2, 3$. Of course, the information about s_k is generally unknown in practice, and this solution utilizes such knowledge. Even so, as we show later, ULA is still far less effective compared to the proposed GLAA method.

For subspace estimation, we compute the average distance between the true and the estimated subspaces, $D = \sum_{k=1}^{\tilde{k}} D(\mathbf{\Gamma}_k, \hat{\mathbf{\Gamma}}_k)/\tilde{k}$, where $D(\mathbf{\Gamma}_k, \hat{\mathbf{\Gamma}}_k) = \left\|\boldsymbol{P}_{\mathbf{\Gamma}_k} - \boldsymbol{P}_{\hat{\mathbf{\Gamma}}_k}\right\|_F/\sqrt{2r_k}$. Note that, since $\boldsymbol{Z}$ is normal with an identity matrix in this example, $\hat{\mathbf{\Gamma}}_3$ is estimating $\text{span}(\mathbf{\Gamma}_3) = \text{span}(\boldsymbol{G}_3') = \text{span}(\boldsymbol{G}_3'')$. For PMD and SCCA, this distance measure is averaged over the first two modes X, Y, so $\tilde{k} = 2$. For GLAA and ULA, it is averaged over the first two modes in the first and the third scenarios, then $\tilde{k} = 2$, and it is averaged over all three modes of X, Y, and Z in the second scenario, so $\tilde{k} = 3$. By definition, this distance measure is between 0 and 1, where 0 indicates a perfect recovery.

5.2. Simulation Results

Tables 1–3 summarize the simulation results over 100 replications for the three scenarios.

Table 1. Simulation results for Scenario 1 where $p_1 = p_2$ varies.

p_1, p_2	Method	TPR-1	FPR-1	TPR-2	FPR-2	D
100	GLAA	1.000 (0.000)	0.000 (0.000)	1.000 (0.000)	0.000 (0.000)	0.095 (0.002)
	ULA	0.998 (0.002)	0.000 (0.000)	0.996 (0.003)	0.000 (0.000)	0.776 (0.002)
	PMD	0.804 (0.028)	0.735 (0.029)	0.802 (0.029)	0.731 (0.028)	0.971 (0.002)
	SCCA	0.584 (0.030)	0.626 (0.027)	0.634 (0.030)	0.629 (0.027)	0.989 (0.001)
200	GLAA	1.000 (0.000)	0.000 (0.000)	1.000 (0.000)	0.000 (0.000)	0.100 (0.003)
	ULA	0.928 (0.010)	0.002 (0.000)	0.944 (0.009)	0.001 (0.000)	0.873 (0.002)
	PMD	0.766 (0.033)	0.731 (0.029)	0.762 (0.033)	0.727 (0.029)	0.989 (0.001)
	SCCA	0.596 (0.031)	0.611 (0.027)	0.626 (0.035)	0.611 (0.027)	0.993 (0.001)
300	GLAA	1.000 (0.000)	0.000 (0.000)	1.000 (0.000)	0.000 (0.000)	0.100 (0.003)
	ULA	0.808 (0.016)	0.003 (0.000)	0.806 (0.016)	0.003 (0.000)	0.945 (0.003)
	PMD	0.718 (0.032)	0.693 (0.030)	0.730 (0.034)	0.696 (0.029)	0.992 (0.001)
	SCCA	0.670 (0.028)	0.651 (0.019)	0.624 (0.028)	0.651 (0.018)	0.997 (0.000)
400	GLAA	0.932 (0.024)	0.008 (0.004)	0.930 (0.025)	0.008 (0.004)	0.186 (0.025)
	ULA	0.652 (0.018)	0.004 (0.000)	0.678 (0.019)	0.004 (0.000)	0.980 (0.002)
	PMD	0.798 (0.029)	0.766 (0.026)	0.800 (0.029)	0.762 (0.026)	0.994 (0.001)
	SCCA	0.594 (0.022)	0.601 (0.010)	0.590 (0.024)	0.602 (0.010)	0.997 (0.000)
500	GLAA	0.848 (0.032)	0.041 (0.010)	0.848 (0.033)	0.040 (0.009)	0.304 (0.037)
	ULA	0.526 (0.021)	0.005 (0.000)	0.526 (0.020)	0.005 (0.000)	0.989 (0.001)
	PMD	0.788 (0.031)	0.727 (0.027)	0.794 (0.030)	0.729 (0.028)	0.995 (0.001)
	SCCA	0.532 (0.024)	0.518 (0.008)	0.532 (0.024)	0.516 (0.008)	0.997 (0.000)

The reported are the average TPR and FPR for the variable selection accuracy, and D for the subspace estimation accuracy, with the standard errors in the parenthesis. The results are over 100 replications.

Table 1 reports the accuracy of variable selection and subspace estimation for Scenario 1 when the dimension $p_1 = p_2$ of $\boldsymbol{X}$ and $\boldsymbol{Y}$ increases. It is clearly seen that GLAA dominates all the competing solutions. For ULA, even with the oracle knowledge of the true sparsity level, the naive variable selection of ULA still performs worse, since it only utilizes the marginal information of each mode. Besides, the estimated subspace is distant away from the true subspace. Moreover, as p_1, p_2 increase, the performance of ULA degrades fast, while GLAA remains competitive. For PMD and SCCA, both suffer large false positive rates in selection, while the estimated subspaces are almost orthogonal to the true subspace with D being almost one. This is because, by design, neither method takes into account the conditioning variable $\boldsymbol{Z}$ when studying the association between $\boldsymbol{X}$ and $\boldsymbol{Y}$.

Table 2 reports the results for Scenario 2 when the dimension p_3 of $\boldsymbol{Z}$ increases. In this case, our goal is to estimate $\boldsymbol{\Gamma}_1, \boldsymbol{\Gamma}_2$ accurately, meanwhile select the variables in $\boldsymbol{X}$ and $\boldsymbol{Y}$ accurately. It is seen that GLAA outperforms all other methods considerably. Besides, it shows a competitive performance of GLAA even with a relatively large dimension of $\boldsymbol{Z}$. This complements our real data example where the dimension of $\boldsymbol{Z}$ is one.

Table 3 reports the results for Scenario 3 when the sample size n increases. Here we examine n that is comparable to the sample size in our multimodal PET example. It is seen that GLAA performs the best, even under a relatively small n. Moreover, the performances of all methods improve as n increases. However, ULA suffers a poor subspace estimation, while both PMD and SCCA continue to suffer both high false positive rates and poor subspace estimation accuracy, even for a relatively large n.

6. Multimodal PET Analysis

6.1. Study and Data Description

We revisit the multimodal PET study introduced in Section 1.1. It is part of the ongoing Berkeley Aging Cohort Study that targets Alzheimer's disease (AD) as well as normal aging. AD is an irreversible neurodegenerative disorder and the leading form of dementia. It is characterized by progressive impairment of cognitive capabilities, then loss of bodily functions, and ultimately death. AD currently affects more than 10% of adults aged 65 or older, and the prevalence is continuously growing. It has now become an international imperative to understand, diagnose, and treat this disorder (Alzheimer's Association 2020).

The data consist of $n = 81$ elderly subjects, with the average age 77.5 years, and the standard deviation 6.2 years. For each subject, three types of neuroimages were acquired, including a Pittsburgh Compound B (PiB) PET scan that measures amyloid-beta protein, an AV-1451 PET scan that measures tau protein, and a 1.5T structural MRI scan for coregistration. MRI and PET images have all been preprocessed, and PET images were both coregistered to each participant's MRI image. Moreover, a mask representing voxels likely to accumulate cortical amyloid and tau pathology was created (Lockhart et al. 2017). Then a set of MNI-space regions of interest were created, and the amount of amyloid-beta and tau deposition was summarized for each region. This results in $p_1 = 60$ regions for amyloid-beta PET, and $p_2 = 26$ regions for tau-PET. We note that brain region parcellation is particularly useful to facilitate the interpretation, and has been frequently employed in brain imaging analysis (Fornito, Zalesky, and Breakspear 2013; Kang et al. 2016).

6.2. Analyses and Results

One of the primary goals of this study is to identify brain regions where the association of amyloid-beta and tau changes the most as age varies, and to further understand this association change. This would offer useful insight about how these two AD pathological proteins interact in the aging brains, which in turn may enable more accurate prediction of individual subjects demonstrating in vivo neuropathology, and allow better design and subject recruitment of clinical trials to potentially slow the

Table 2. Simulation results for Scenario 2 where p_3 varies.

p_3	Method	TPR-1	FPR-1	TPR-2	FPR-2	D
20	GLAA	1.000 (0.000)	0.000 (0.000)	1.000 (0.000)	0.000 (0.000)	0.132 (0.004)
	ULA	0.642 (0.019)	0.019 (0.001)	0.638 (0.019)	0.019 (0.001)	0.872 (0.003)
	PMD	0.774 (0.031)	0.733 (0.029)	0.776 (0.031)	0.732 (0.029)	0.978 (0.002)
	SCCA	0.518 (0.034)	0.534 (0.029)	0.536 (0.032)	0.532 (0.030)	0.989 (0.001)
40	GLAA	1.000 (0.000)	0.000 (0.000)	1.000 (0.000)	0.000 (0.000)	0.131 (0.004)
	ULA	0.488 (0.020)	0.027 (0.001)	0.498 (0.020)	0.026 (0.001)	0.888 (0.002)
	PMD	0.774 (0.031)	0.695 (0.030)	0.772 (0.032)	0.695 (0.029)	0.973 (0.002)
	SCCA	0.590 (0.034)	0.596 (0.031)	0.624 (0.033)	0.589 (0.031)	0.987 (0.001)
60	GLAA	1.000 (0.000)	0.000 (0.000)	1.000 (0.000)	0.000 (0.000)	0.136 (0.004)
	ULA	0.384 (0.020)	0.032 (0.001)	0.404 (0.020)	0.031 (0.001)	0.898 (0.002)
	PMD	0.748 (0.030)	0.694 (0.030)	0.754 (0.033)	0.701 (0.030)	0.973 (0.002)
	SCCA	0.564 (0.031)	0.537 (0.029)	0.572 (0.031)	0.541 (0.029)	0.985 (0.002)
80	GLAA	0.998 (0.002)	0.000 (0.000)	0.998 (0.002)	0.000 (0.000)	0.145 (0.005)
	ULA	0.378 (0.021)	0.033 (0.001)	0.368 (0.019)	0.033 (0.001)	0.903 (0.001)
	PMD	0.706 (0.033)	0.689 (0.031)	0.760 (0.032)	0.688 (0.031)	0.975 (0.002)
	SCCA	0.624 (0.031)	0.613 (0.027)	0.592 (0.033)	0.609 (0.027)	0.987 (0.002)
100	GLAA	0.996 (0.003)	0.005 (0.005)	0.996 (0.003)	0.005 (0.005)	0.158 (0.010)
	ULA	0.332 (0.020)	0.035 (0.001)	0.360 (0.018)	0.034 (0.001)	0.905 (0.001)
	PMD	0.712 (0.034)	0.633 (0.033)	0.662 (0.038)	0.633 (0.033)	0.974 (0.002)
	SCCA	0.560 (0.033)	0.565 (0.030)	0.586 (0.032)	0.567 (0.030)	0.989 (0.001)

Note: The rest is the same as Table 1.

Table 3. Simulation results for Scenario 3 where n varies.

n	Method	TPR-1	FPR-1	TPR-2	FPR-2	D
60	GLAA	0.606 (0.025)	0.048 (0.008)	0.664 (0.027)	0.154 (0.021)	0.743 (0.015)
	ULA	0.410 (0.022)	0.031 (0.001)	0.362 (0.018)	0.160 (0.004)	0.920 (0.004)
	PMD	0.498 (0.040)	0.486 (0.036)	0.552 (0.039)	0.466 (0.036)	0.957 (0.004)
	SCCA	0.604 (0.023)	0.590 (0.012)	0.688 (0.023)	0.643 (0.017)	0.968 (0.002)
80	GLAA	0.638 (0.027)	0.018 (0.002)	0.738 (0.025)	0.137 (0.027)	0.662 (0.021)
	ULA	0.550 (0.021)	0.024 (0.001)	0.412 (0.018)	0.147 (0.005)	0.895 (0.004)
	PMD	0.494 (0.038)	0.441 (0.034)	0.502 (0.038)	0.429 (0.034)	0.953 (0.004)
	SCCA	0.672 (0.026)	0.665 (0.016)	0.716 (0.025)	0.718 (0.019)	0.966 (0.003)
100	GLAA	0.820 (0.022)	0.013 (0.004)	0.848 (0.020)	0.060 (0.016)	0.483 (0.024)
	ULA	0.686 (0.019)	0.017 (0.001)	0.534 (0.021)	0.117 (0.005)	0.870 (0.004)
	PMD	0.510 (0.035)	0.476 (0.032)	0.560 (0.036)	0.442 (0.033)	0.947 (0.004)
	SCCA	0.702 (0.027)	0.677 (0.021)	0.732 (0.025)	0.714 (0.024)	0.961 (0.003)
120	GLAA	0.896 (0.018)	0.004 (0.001)	0.914 (0.018)	0.010 (0.003)	0.350 (0.021)
	ULA	0.802 (0.017)	0.010 (0.001)	0.622 (0.018)	0.094 (0.004)	0.850 (0.003)
	PMD	0.576 (0.035)	0.564 (0.032)	0.662 (0.035)	0.530 (0.032)	0.941 (0.004)
	SCCA	0.640 (0.030)	0.657 (0.024)	0.708 (0.028)	0.705 (0.022)	0.970 (0.003)
160	GLAA	0.990 (0.004)	0.001 (0.000)	0.984 (0.005)	0.002 (0.001)	0.209 (0.012)
	ULA	0.920 (0.011)	0.004 (0.001)	0.732 (0.016)	0.067 (0.004)	0.812 (0.003)
	PMD	0.610 (0.037)	0.570 (0.034)	0.652 (0.038)	0.544 (0.035)	0.944 (0.004)
	SCCA	0.626 (0.031)	0.651 (0.024)	0.698 (0.029)	0.691 (0.024)	0.969 (0.003)

Note: The rest is the same as Table 1.

spread of AD. For instance, clinical trials that aim at testing anti-amyloid-beta or anti-tau agents would need to know not only that participants have amyloid-beta and tau in the brain, but also how the relative levels of each pathological protein are spatially associated with each other given their ages. We cast this problem in the framework of liquid association analysis. Let $\boldsymbol{X} \in \mathbb{R}^{60}$, $\boldsymbol{Y} \in \mathbb{R}^{26}$ denote the amyloid-beta accumulation and tau accumulation in various brain regions, respectively, and $Z \in \mathbb{R}$ denote the subject's age. We first log-transform each variable in $\boldsymbol{X}$ and $\boldsymbol{Y}$, and standardize $\boldsymbol{X}$, $\boldsymbol{Y}$ and Z marginally. We then apply the proposed generalized liquid association analysis (GLAA) method to this data. We choose the thresholding parameters η_1 and η_2 for the initialization step, so that about half of the variables in $\boldsymbol{X}$ and in $\boldsymbol{Y}$ are kept for subsequent iterations. We then tune the thresholding parameters $\tilde{\eta}_1$ and $\tilde{\eta}_2$ used in iterative sparse SVD by cross-validation over a grid of candidate values. We choose the ranks r_1, r_2, that is, the numbers of linear combinations for $\boldsymbol{X}$ and $\boldsymbol{Y}$, to be one, which is most common in canonical correlation analysis.

After obtaining the two estimated linear combinations $\hat{\boldsymbol{\Gamma}}_1^\top \boldsymbol{X}$ and $\hat{\boldsymbol{\Gamma}}_2^\top \boldsymbol{Y}$, we plot them as the value of Z changes. We divide the interval of Z into six equal-sized intervals with overlaps, then draw the scatterplot of $\hat{\boldsymbol{\Gamma}}_2^\top \boldsymbol{Y}$ versus $\hat{\boldsymbol{\Gamma}}_1^\top \boldsymbol{X}$ within each interval. We also add a fitted linear regression line in each panel to reflect the correlation between $\hat{\boldsymbol{\Gamma}}_1^\top \boldsymbol{X}$ and $\hat{\boldsymbol{\Gamma}}_2^\top \boldsymbol{Y}$. Figure 1 shows the trellis plots, where the stripe at the top of each panel represents the range of Z it covers. It is interesting to see from the GLAA estimation, the correlation between $\boldsymbol{\Gamma}_1^\top \boldsymbol{X}$ and $\boldsymbol{\Gamma}_2^\top \boldsymbol{Y}$ changes from negative to positive gradually, as the age variable Z increases. This may be due to different deposition patterns of amyloid-beta and tau. In particular, amyloid-beta plaques

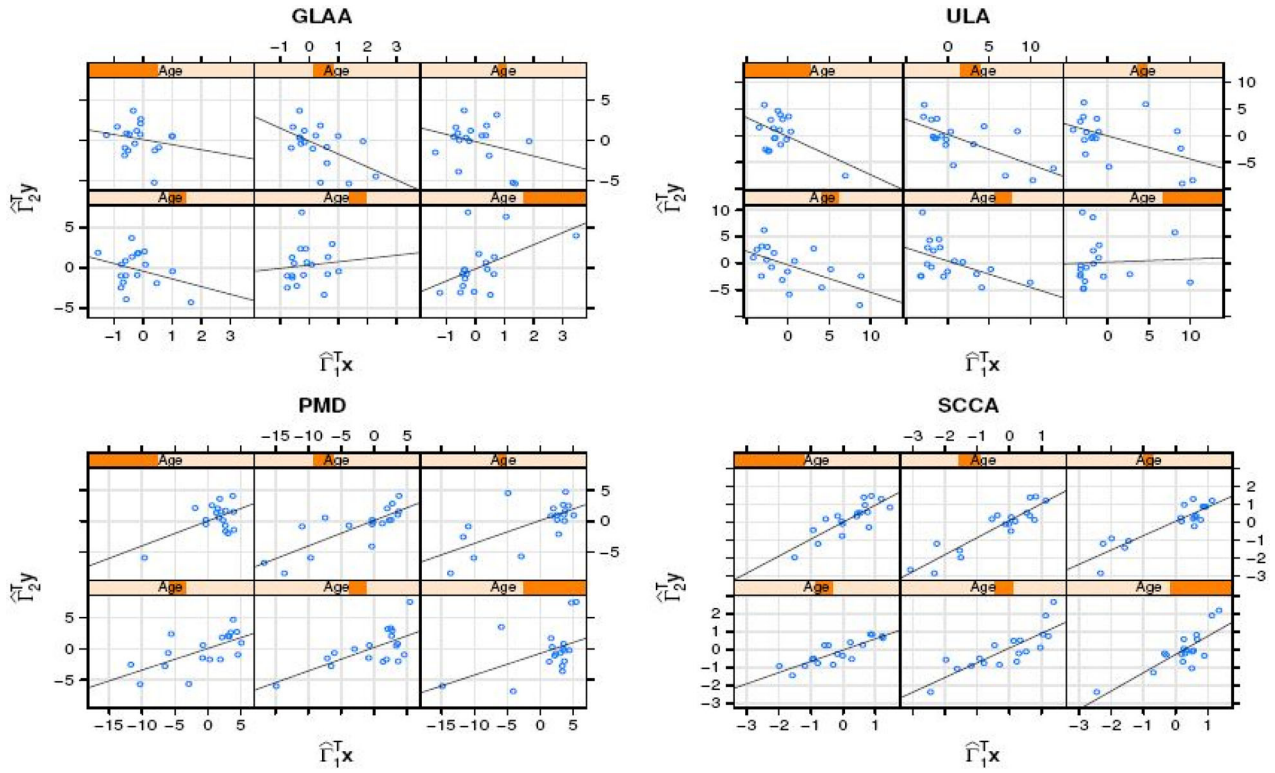

Figure 1. Trellis plots of the estimated linear combinations $\hat{\mathbf{\Gamma}}_2^\top \boldsymbol{Y}$ versus $\hat{\mathbf{\Gamma}}_1^\top \boldsymbol{X}$ as Z varies. Each panel represents an interval of Z, with a linear line added. The methods include: generalized liquid association analysis (GLAA), univariate liquid association (ULA), penalized matrix decomposition (PMD), and sparse canonical correlation analysis (SCCA).

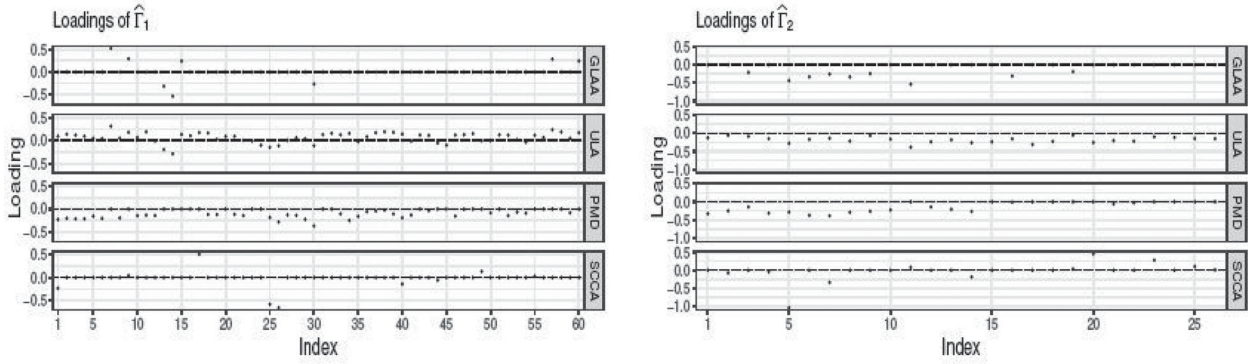

Figure 2. Estimated loadings in $\hat{\mathbf{\Gamma}}_1$ and $\hat{\mathbf{\Gamma}}_2$. The number of nonzero loading entries estimated by GLAA, ULA, PMD, SCCA are 8, 60, 37, 9 for $\hat{\mathbf{\Gamma}}_1$, and 9, 26, 16, 11 for $\hat{\mathbf{\Gamma}}_2$.

are detectable in the brain many years before dementia onset, while tau neurofibrillary tangles aggregate specifically in the medial temporal lobes in normal aging. The spread of tau out of medial temporal lobes and into the surrounding isocortex at elder age coincides with cognitive impairment, and the process is hypothesized to be potentiated or accelerated by the presence of amyloid-beta (He et al. 2018; Vogel et al. 2020). The change from a negative association in early years to a positive association in later years between amyloid-beta and tau found by our GLAA method may offer some support to this hypothesis. As a comparison, no clear changing pattern is observed from the other three estimation methods.

Next, we examine more closely the brain regions identified by GLAA that demonstrate dynamic association patterns. Figure 2 plots the loadings of the estimated $\hat{\mathbf{\Gamma}}_1$ and $\hat{\mathbf{\Gamma}}_2$, where the indices of nonzero loadings correspond to the selected regions. The number of nonzero loading entries estimated by GLAA, ULA, PMD, SCCA are $8, 60, 37, 9$ for $\hat{\mathbf{\Gamma}}_1$, and $9, 26, 16, 11$ for $\hat{\mathbf{\Gamma}}_2$, respectively. Note that, the ULA method does not do variable selection, and for the real data, no information on the true sparsity level is known, so its estimated loadings are nonsparse. Moreover, the PMD method yields a large number of nonzero estimates, making the interpretation difficult. The SCCA method selects about the same number of nonzero regions as GLAA, but the selected regions are less meaningful.

Table 4 reports the identified brain regions by GLAA for amyloid-beta and tau, respectively, while Figure 3 visualizes those regions on a template brain using BrainNet Viewer (Xia, Wang, and He 2013). Many of these regions are known to be closely related to AD, and the dynamic associations between amyloid-beta and tau of those regions reveal interesting and new insights. Particularly, for both amyloid-beta and tau, the

Table 4. Identified brain regions for amyloid-beta and tau by GLAA.

Modality	Identified regions				
amyloid-beta	Entorhinal R	Entorhinal L	Hippocampus R	Hippocampus L	Amygdala R
	Orbitofrontal L	Posterior Cingulate L	Middle Frontal R		
tau	Entorhinal R	Entorhinal L	Hippocampus R	Parahippocampal R	Fusiform L
	Middle Temporal R	Middle Temporal L	Insula L	Rostral Anterior Cingulate R	

Note: Regions in the left hemisphere are denoted by "L", and regions in the right hemisphere are denoted by "R"

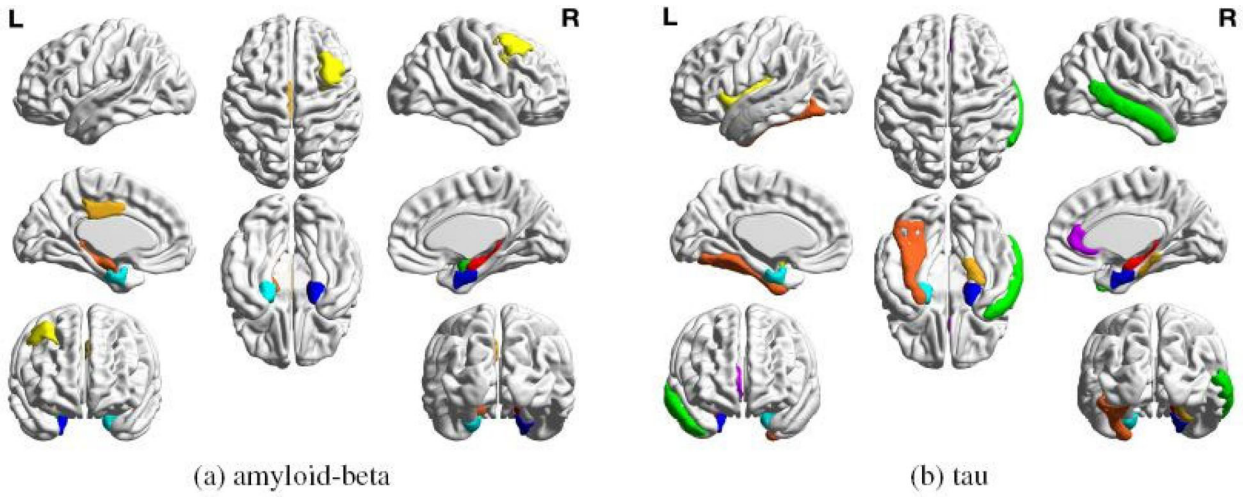

Figure 3. Identified brain regions for amyloid-beta and tau by GLAA.

identified regions include hippocampus and entorhinal cortex. Hippocampus is a major component functionally involved in response inhibition, episodic memory, and spatial cognition. It is one of the first brain regions to suffer damage from AD (Jack et al. 2011). Entorhinal cortex is a brain region that functions as a hub in a widespread network for memory and navigation. Entorhinal cortex and hippocampus together play an important role in memories. Atrophy in entorhinal cortex has been consistently reported in AD (Pini et al. 2016). Moreover, animal models have suggested that neurofibrillary tangles of tau first appear in entorhinal cortex, then spread to hippocampus (Cho et al. 2016). For amyloid-beta, other identified regions include amygdala, orbitofrontal cortex, posterior cingulate cortex and areas of middle frontal cortices. Amygdala is responsible for memory processing, decision-making and emotional responses. Amygdala atrophy is found prominent in early AD (Poulin et al. 2011). Orbitofrontal cortex is involved in decision-making, while posterior cingulate cortex is one of the most metabolically active brain regions, and is linked to emotion and memory. Atrophy of both regions and middle frontal cortices have been found associated with AD (Pini et al. 2016). For tau, other identified regions include parahippocampal gyrus, middle temporal gyrus, fusiform, insula, and rostral anterior cingulate cortex. Parahippocampal gyrus is central for memory encoding and retrieval. Atrophy in parahippocampal gyrus has been identified as an early biomarker of AD (Echavarri et al. 2011). Middle temporal gyrus is connected with recognition of known faces and accessing word meaning. Fusiform is linked with various neural pathways related to recognition. Insula is involved in consciousness and emotion. Rostral anterior cingulate cortex is involved in attention allocation, decision-making and emotion. There have been numerous evidences suggesting associations between these regions and AD (Convit et al. 2000; Pini et al. 2016).

In summary, GLAA identifies interesting dynamic association patterns among a number of important brain regions between amyloid-beta and tau as age increases. Moreover, GLAA provides a useful dimension reduction tool to help visualize such patterns.

7. Discussion

In this article, we have proposed generalized liquid association analysis, which offers a new angle to study three-way associations among random variables, and is particularly useful for multimodal integrative data analysis. We have illustrated with a multimodal neuroimaging study of Alzheimer's disease in detail. Meanwhile, the proposal is potentially applicable to other multimodal data problems, e.g., to understand gene co-expressions given single nucleotide polymorphisms or under varying physiological states (Chen, Xie, and Li 2011; Yu 2018), or to understand interactions between gene expressions and microRNA levels or comparative genomic hybridizations given cancer states and demographics (Cai, Cai, and Zhang 2016; Mai and Zhang 2019). Next, we discuss some potential extensions.

First, we begin with the situation when there is a univariate and categorical Z, whereas the analysis so far has primarily concentrated on the case when each variable in $\boldsymbol{Z}$ is continous. In general, it remains an open question on how to define liquid association for a categorical variable, since the function $g(z)$ is no longer differentiable for a categorical Z. For a binary $Z \in \{0, 1\}$, we propose to replace the derivative of the conditional mean function with the absolute change in the conditional means across the two groups, that is, $\mathrm{LA}(X, Y|Z) =$

$|\mathrm{E}(XY|Z=1)-\mathrm{E}(XY|Z=0)|$, where the absolute value is used because the class labels are interchangeable. This naturally fits the original interpretation of LA. Similarly, for a categorical or ordinal $Z \in \{1,\ldots,K\}$, we can use the weighted sum of pairwise absolute mean difference between the pairs of groups. Accordingly, the liquid association of $\boldsymbol{X}$ and $\boldsymbol{Y}$ given Z is defined as a $p_1 \times p_2$ matrix.

Next, for a multivariate mixed type $\boldsymbol{Z}$, we first organize $\boldsymbol{Z} = (\boldsymbol{Z}_1, \boldsymbol{Z}_2)^\top$ to separate the continuous variables, $\boldsymbol{Z}_1 = (Z_1,\ldots,Z_q)^\top \in \mathbb{R}^q$, from the categorical variables, $\boldsymbol{Z}_2 = (Z_{q+1},\ldots,Z_{p_3})^\top \in \mathbb{R}^{p_3-q}$. Directly imposing a low-dimensional structure on entire $\boldsymbol{Z}$ would lead to difficulty in interpretation. Alternatively, we propose a dimension reduction approach, by recognizing the reduction on $\boldsymbol{Z}$ in model (1) is indeed a sufficient dimension reduction model. Specifically, when $\boldsymbol{Z}$ is continuous, by model (1), we have $\boldsymbol{g}(\boldsymbol{Z}) = \boldsymbol{g}(\boldsymbol{P}_{\mathcal{S}}\boldsymbol{Z})$, where $\mathcal{S} = \mathrm{span}(\boldsymbol{\Gamma}_3)$. Therefore, $\boldsymbol{g}(\boldsymbol{Z}) \perp\!\!\!\perp \boldsymbol{Z}|\boldsymbol{P}_{\mathcal{S}}\boldsymbol{Z}$. This leads to a sufficient dimension reduction model of $\boldsymbol{Z}$ for the conditional mean function $g(\boldsymbol{Z}) = \mathrm{E}(\boldsymbol{X}\boldsymbol{Y}^\top|\boldsymbol{Z})$, in the sense that all the mean information of the regression of the matrix response $\boldsymbol{X}\boldsymbol{Y}^\top$ given the predictor vector $\boldsymbol{Z}$ is fully captured by the linear combinations $\boldsymbol{\Gamma}_3^\top\boldsymbol{Z}$. As such, model (1) can be viewed as a generalization of the notion of sufficient mean reduction (Cook and Li 2002). Now for the mixed type $\boldsymbol{Z} = (\boldsymbol{Z}_1^\top, \boldsymbol{Z}_2^\top)^\top$, we adopt the idea of partial dimension reduction (Chiaromonte, Cook, and Li 2002), or groupwise dimension reduction (Li, Li, and Zhu 2010), and estimate the subspace $\mathcal{S} \subseteq \mathbb{R}^q$ such that $\boldsymbol{g}(\boldsymbol{Z}) \perp\!\!\!\perp \boldsymbol{Z}|(\boldsymbol{P}_{\mathcal{S}}\boldsymbol{Z}_1, \boldsymbol{Z}_2)$.

We have so far focused on studying the associations between two sets of random variables conditioning on another set. It is possible to generalize to the associations of more than two sets of variables conditioning on another set. It involves a higher-order tensor than an order-3 tensor. Nevertheless, both the estimation algorithm and the theory can be extended to a general order tensor in a relatively straightforward fashion. Moreover, from the simulation results in Table 2, we observe that, the higher the dimension of $\boldsymbol{Z}$, the more challenging the problem becomes. We believe it is possible to employ an alternative modeling strategy such as Chen, Xie, and Li (2011) when the dimension of $\boldsymbol{Z}$ is ultrahigh. This is also true when the scientific interest is to select important variables in $\boldsymbol{Z}$, while our current interest concentrates on the selection of variables in $\boldsymbol{X}$ and $\boldsymbol{Y}$, but not in $\boldsymbol{Z}$. We leave the pursuit of these lines of research as our future work.

Supplementary Materials

The supplementary material contains additional simulations, new theoretical results, and technical proofs.

Acknowledgments

The authors thank the Editor, the Associate Editor, and three referees for their constructive comments and suggestions. All authors contributed equally and are listed in alphabetical order.

Funding

Li's research was supported by NSF grant CIF-2102227, and NIH grants R01AG061303, R01AG062542, and R01AG034570. Zeng's research was supported by NSF grant CCF-1908969. Zhang's research was supported by NSF grant DMS-2053697, and NIH grant DE-030509.

ORCID

Jing Zeng http://orcid.org/0000-0002-4886-5682

References

Abid, A., Zhang, M. J., Bagaria, V. K., and Zou, J. (2018), "Exploring Patterns Enriched in a Dataset with Contrastive Principal Component Analysis," *Nature Communications*, 9, 1–7. [1985]

Alzheimer's Association. (2020), "2020 Alzheimer's Disease Facts and Figures," *Alzheimer's & Dementia*, 16, 391–460. [1991]

Bi, X., Qu, A., and Shen, X. (2018), "Multilayer Tensor Factorization with Applications to Recommender Systems," *The Annals of Statistics*, 46, 3308–3333. [1985]

Bi, X., Tang, X., Yuan, Y., Zhang, Y., and Qu, A. (2021), "Tensor in Statistics," *Annual Review of Statistics and Its Application*, 8, 345–368. [1985]

Braak, H., and Braak, E. (1991), "Neuropathological Stageing of Alzheimer-Related Changes," *Acta Neuropathologica*, 82, 239–259. [1984]

Cai, T., Cai, T. T., and Zhang, A. (2016), "Structured Matrix Completion with Applications to Genomic Data Integration," *Journal of the American Statistical Association*, 111, 621–633. [1984,1994]

Chen, J., Xie, J., and Li, H. (2011), "A Penalized Likelihood Approach for Bivariate Conditional Normal Models for Dynamic Co-expression Analysis," *Biometrics*, 67, 299–308. [1985,1994,1995]

Chiaromonte, F., Cook, R. D., and Li, B. (2002), "Sufficient Dimensions Reduction in Regressions with Categorical Predictors," *The Annals of Statistics*, 30, 475–497. [1985,1995]

Cho, H., Choi, J. Y., Hwang, M. S., Kim, Y. J., Lee, H. M., Lee, H. S., Lee, J. H., Ryu, Y. H., Lee, M. S., and Lyoo, C. H. (2016), "In Vivo Cortical Spreading Pattern of Tau and Amyloid in the Alzheimer Disease Spectrum," *Annals of Neurology*, 80, 247–258. [1994]

Convit, A., de Asis, J., de Leon, M. J., Tarshish, C. Y., De Santi, S., and Rusinek, H. (2000), "Atrophy of the Medial Occipitotemporal, Inferior, and Middle Temporal Gyri in Non-demented Elderly Predict Decline to Alzheimer's Disease," *Neurobiology of Aging*, 21, 19–26. [1994]

Cook, R. D. (1998), "Principal Hessian Directions Revisited," *Journal of the American Statistical Association*, 93, 84–94. [1985,1987]

——— (2007), "Fisher Lecture: Dimension Reduction in Regression," *Statistical Science*, 22, 1–26. [1985]

Cook, R. D., and Li, B. (2002), "Dimension Reduction for Conditional Mean in Regression," *The Annals of Statistics*, 30, 455–474. [1995]

De Lathauwer, L., De Moor, B., and Vandewalle, J. (2000), "On the Best Rank-1 and Rank-(r1, r2,..., rn) Approximation of Higher-Order Tensors," *SIAM Journal on Matrix Analysis and Applications*, 21, 1324–1342. [1987]

Echavarri, C., Aalten, P., Uylings, H., Jacobs, H., Visser, P., Gronenschild, E., Verhey, F., and Burgmans, S. (2011), "Atrophy in the Parahippocampal Gyrus as an Early Biomarker of Alzheimer's Disease," *Brain Structure & Function*, 215, 265–271. [1994]

Fornito, A., Zalesky, A., and Breakspear, M. (2013), "Graph Analysis of the Human Connectome: Promise, Progress, and Pitfalls," *NeuroImage*, 80, 426–444. [1991]

Gao, C., Ma, Z., Ren, Z., and Zhou, H. H. (2015), "Minimax Estimation in Sparse Canonical Correlation Analysis," *The Annals of Statistics*, 43, 2168–2197. [1985]

Hao, B., Zhang, A., and Cheng, G. (2020), "Sparse and Low-Rank Tensor Estimation via Cubic Sketchings," *IEEE Transactions on Information Theory*, 66, 5927–5964. [1985]

He, Z., Guo, J. L., McBride, J. D., Narasimhan, S., Kim, H., Changolkar, L., Zhang, B., Gathagan, R. J., Yue, C., Dengler, C., Stieber, A., Nitla, M., Coulter, D. A., Abel, T., Brunden, K. R., Trojanowski, J. Q., and Lee, V. M.-Y. (2018), "Amyloid-Beta Plaques Enhance Alzheimer's Brain Tau-Seeded Pathologies by Facilitating Neuritic Plaque Tau Aggregation," *Nature Medicine*, 24, 29–38. [1993]

Ho, Y.-Y., Parmigiani, G., Louis, T. A., and Cope, L. M. (2011), "Modeling Liquid Association," *Biometrics*, 67, 133–141. [1985]

Jack, C. R., Barkhof, F., Bernstein, M. A., Cantillon, M., Cole, P. E., DeCarli, C., Dubois, B., Duchesne, S., Fox, N. C., Frisoni, G. B., Hampel, H., Hill, D. L. G., Johnson, K., Mangin, J.-F., Scheltens, P., Schwarz, A. J., Sperling, R., Suhy, J., Thompson, P. M., Weiner, M., and Foster, N. L. (2011), "Steps to Standardization and Validation of Hippocampal Volumetry as a Biomarker in Clinical Trials and Diagnostic Criterion for Alzheimer's Disease," *Alzheimer's & Dementia*, 7, 474–485.e4. [1994]

Kang, J., Bowman, F. D., Mayberg, H., and Liu, H. (2016). "A Depression Network of Functionally Connected Regions Discovered via Multi-Attribute Canonical Correlation Graphs," *NeuroImage*, 141, 431–441. [1991]

Kolda, T. G., and Bader, B. W. (2009), "Tensor Decompositions and Applications," *SIAM Review*, 51, 455–500. [1985,1987]

Li, B. (2018), *Sufficient Dimension Reduction: Methods and Applications with R*. Chapman & Hall/CRC Monographs on Statistics and Applied Probability. Boca Raton, FL: CRC Press. [1985]

Li, G., and Gaynanova, I. (2018), "A General Framework for Association Analysis of Heterogeneous Data," *The Annals of Applied Statistics*, 12, 1700–1726. [1985]

Li, G., and Jung, S. (2017), "Incorporating Covariates into Integrated Factor Analysis of Multi-View Data," *Biometrics*, 73, 1433–1442. [1985]

Li, K.-C. (1992), "On Principal Hessian Directions for Data Visualization and Dimension Reduction: Another Application of Stein's Lemma," *Journal of the American Statistical Association*, 87, 1025–1039. [1985,1987]

——— (2002), "Genome-Wide Coexpression Dynamics: Theory and Application," *Proceedings of the National Academy of Sciences*, 99, 16875–16880. [1985,1986,1990]

Li, K.-C., Liu, C.-T., Sun, W., Yuan, S., and Yu, T. (2004), "A System for Enhancing Genome-Wide Coexpression Dynamics Study," *Proceedings of the National Academy of Sciences*, 101, 15561–15566. [1985,1986]

Li, L., Li, B., and Zhu, L.-X. (2010), "Groupwise Dimension Reduction," *Journal of the American Statistical Association*, 105, 1188–1201. [1985,1995]

Liu, J. S. (1994), "Siegel's Formula via Stein's Identities," *Statistics & Probability Letters*, 21, 247–251. [1986]

Lock, E. F., Hoadley, K. A., Marron, J. S., and Nobel, A. B. (2013), "Joint and Individual Variation Explained (JIVE) for Integrated Analysis of Multiple Data Types," *The Annals of Applied Statistics*, 7, 523–542. [1985]

Lockhart, S. N., Schöll, M., Baker, S. L., Ayakta, N., Swinnerton, K. N., Bell, R. K., Mellinger, T. J., Shah, V. D., O'Neil, J. P., Janabi, M., and Jagust, W. J. (2017), "Amyloid and Tau PET Demonstrate Region-Specific Associations in Normal Older People," *NeuroImage*, 150, 191–199. [1991]

Luo, Y., Raskutti, G., Yuan, M., and Zhang, A. R. (2020), "A Sharp Blockwise Tensor Perturbation Bound for Orthogonal Iteration, *arXiv preprint arXiv:2008.02437*. [1987]

Mai, Q., and Zhang, X. (2019), "An Iterative Penalized Least Squares Approach to Sparse Canonical Correlation Analysis," *Biometrics*, 75, 734–744. [1984,1985,1990,1994]

Nathoo, F. S., Kong, L., and Zhu, H. (2017), "Inference on High-Dimensional Differential Correlation Matrix," *arXiv preprint arXiv:1707.07332*. [1984]

Pini, L., Pievani, M., Bocchetta, M., Altomare, D., Bosco, P., Cavedo, E., Galluzzi, S., Marizzoni, M., and Frisoni, G. B. (2016), "Brain Atrophy in Alzheimer's Disease and Aging," *Ageing Research Reviews*, 30, 25–48. [1994]

Poulin, S. P., Dautoff, R., Morris, J. C., Barrett, L. F., and Dickerson, B. C. (2011), "Amygdala Atrophy is Prominent in Early Alzheimer's Disease and Relates to Symptom Severity," *Psychiatry Research: Neuroimaging*, 194, 7–13. [1994]

Richardson, S., Tseng, G. C., and Sun, W. (2016), "Statistical Methods in Integrative Genomics," *Annual Review of Statistics and Its Application*, 3, 181–209. [1984]

Shu, H., Wang, X., and Zhu, H. (2019), "D-CCA: A Decomposition-based Canonical Correlation Analysis for High-Dimensional Datasets," *Journal of the American Statistical Association*, 115, 292–306. [1985]

Stein, C. M. (1981), "Estimation of the Mean of a Multivariate Normal Distribution," *The Annals of Statistics*, 9, 1135–1151. [1986]

Sun, W., Lu, J., Liu, H., and Cheng, G. (2017), "Provable Sparse Tensor Decomposition," *Journal of the Royal Statistical Society*, Series B, 79, 899–916. [1985]

Tang, C. Y., Fang, E. X., and Dong, Y. (2020), "High-Dimensional Interactions Detection with Sparse Principal Hessian Matrix," *Journal of Machine Learning Research*, 21, 1–25. [1987]

Tang, X., Bi, X., and Qu, A. (2019), "Individualized Multilayer Tensor Learning with an Application in Imaging Analysis," *Journal of the American Statistical Association*, 115, 836–851. [1985]

Uludag, K., and Roebroeck, A. (2014), "General Overview on the Merits of Multimodal Neuroimaging Data Fusion," *NeuroImage*, 102, 3–10. [1984]

Vogel, J. W., Iturria-Medina, Y., and et al. (2020), "Spread of Pathological Tau Proteins Through Communicating Neurons in Human Alzheimer's Disease," *Nature Communications*, 11, 1–15. [1993]

Wainwright, M. J. (2019), *High-Dimensional Statistics: A Non-asymptotic Viewpoint* (Vol. 48), Cambridge, UK: Cambridge University Press. [1989]

Witten, D. M., Tibshirani, R., and Hastie, T. (2009), "A Penalized Matrix Decomposition, with Applications to Sparse Principal Components and Canonical Correlation Analysis," *Biostatistics*, 10, 515–534. [1985,1990]

Xia, M., Wang, J., and He, Y. (2013), "Brainnet Viewer: A Network Visualization Tool for Human Brain Connectomics," *PLOS ONE*, 8, 1–15. [1993]

Xia, Y., Li, L., Lockhart, S. N., and Jagust, W. J. (2019), "Simultaneous Covariance Inference for Multimodal Integrative Analysis," *Journal of the American Statistical Association*, 115, 1279–1291. [1985]

Yang, D., Ma, Z., and Buja, A. (2016), "Rate Optimal Denoising of Simultaneously Sparse and Low Rank Matrices," *The Journal of Machine Learning Research*, 17, 3163–3189. [1986,1989]

Yu, T. (2018), "A New Dynamic Correlation Algorithm Reveals Novel Functional Aspects in Single Cell and Bulk RNA-seq Data," *PLOS Computational Biology*, 14, 1–22. [1985,1994]

Yu, Y., Wang, T., and Samworth, R. J. (2015), "A Useful Variant of the Davis–Kahan Theorem for Statisticians," *Biometrika*, 102, 315–323. [1989]

Zhang, A., and Han, R. (2019), "Optimal Sparse Singular Value Decomposition for High-Dimensional High-Order Data," *Journal of the American Statistical Association*, 114, 1078–1725. [1985,1986,1987,1988,1989]

Zhang, A., and Xia, D. (2018), "Tensor SVD: Statistical and Computational Limits," *IEEE Transactions on Information Theory*, 64, 7311–7338. [1987]

Zhou, H., Li, L., and Zhu, H. (2013), "Tensor Regression with Applications in Neuroimaging Data Analysis," *Journal of the American Statistical Association*, 108, 540–552. [1985]