Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

Group linear non-Gaussian component analysis with applications to neuroimaging

Yuxuan Zhao [a], David S. Matteson [a], Stewart H. Mostofsky [b,c,d],
Mary Beth Nebel [b,c], Benjamin B. Risk [e,*,1]

[a] *Department of Statistics and Data Science, Cornell University, United States of America*
[b] *Center for Neurodevelopmental and Imaging Research, Kennedy Krieger Institute, United States of America*
[c] *Department of Neurology, Johns Hopkins University School of Medicine, United States of America*
[d] *Department of Psychiatry and Behavioral Science, Johns Hopkins University School of Medicine, United States of America*
[e] *Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University, United States of America*

ARTICLE INFO

ABSTRACT

Independent component analysis (ICA) is an unsupervised learning method popular in functional magnetic resonance imaging (fMRI). Group ICA has been used to search for biomarkers in neurological disorders including autism spectrum disorder and dementia. However, current methods use a principal component analysis (PCA) step that may remove low-variance features. Linear non-Gaussian component analysis (LNGCA) enables simultaneous dimension reduction and feature estimation including low-variance features in single-subject fMRI. A group LNGCA model is proposed to extract group components shared by more than one subject. Unlike group ICA methods, this novel approach also estimates individual (subject-specific) components orthogonal to the group components. To determine the total number of components in each subject, a parametric resampling test is proposed that samples spatially correlated Gaussian noise to match the spatial dependence observed in data. In simulations, estimated group components achieve higher accuracy compared to group ICA. The method is applied to a resting-state fMRI study on autism spectrum disorder in 342 children (252 typically developing, 90 with autism), where the group signals include resting-state networks. The discovered group components appear to exhibit different levels of temporal engagement in autism versus typically developing children, as revealed using group LNGCA. This novel approach to matrix decomposition is a promising direction for feature detection in neuroimaging.

1. Introduction

Independent component analysis (ICA) is a popular unsupervised learning method to identify brain networks in functional magnetic resonance imaging (fMRI) studies (Beckmann, 2012). In fMRI experiments, the observed fMRI data represent a combination of neural activity and nuisance-related variation across many different source signals. Assuming that latent signals are statistically independent and have non-Gaussian distributions, ICA linearly decomposes the observed fMRI data into independent spatial maps and corresponding time courses. The maps related to neural activity are commonly called

---

* Corresponding author.
  E-mail address: brisk@emory.edu (B.B. Risk).
[1] 1518 Clifton Rd NE, Atlanta, GA 30312.

resting-state "networks." Previous studies have used group ICA to examine differential levels of intrinsic engagement between neurotypical and atypical individuals, for example, in schizophrenia (Calhoun et al., 2009; Du et al., 2017, 2019). Group ICA of fMRI relies upon a PCA step that may discard low variance features. In ICA, a low variance feature is a spatial component that has a time course with small variance. Low variance features contained in the "noise" subspace in PCA may be related to neural activity, and in particular may exhibit differential levels of engagement in neurological disorders. In contrast to ICA with PCA, linear non-Gaussian component analysis (LNGCA) can extract low-variance features and has been successfully applied to single-subject fMRI data (Risk et al., 2019). LNGCA was recently extended to discover novel non-Gaussian features shared by multiple types of neuroimaging data (Risk and Gaynanova, 2021). Functionally related brain regions tend to correspond to sparse images, and these vectorized images are highly non-Gaussian. Hence, estimating the signal subspace by maximizing non-Gaussianity can offer new insights. However, most fMRI studies involve data from multiple subjects, and it is unclear how to extend LNGCA to such a setting.

Extending non-Gaussian matrix decomposition methods to group analyses has been challenging because considerable between-subject variability exists in both the spatial configuration and the temporal engagement of the estimated functional brain networks. Some approaches to ICA-based group inference model between-subject variability in the temporal domain while assuming spatial signals are equal across all subjects (Calhoun et al., 2001; Guo, 2011; Eloyan et al., 2013). Models allowing subject-specific spatial deviations from the group components have been proposed (Beckmann et al., 2009; Guo and Tang, 2013; Mejia et al., 2019; Du and Fan, 2013), but in practice, all of these options are preceded by a dimension reduction step using PCA to alleviate computational demands. Hereafter, when we say group ICA, we are referring to this process of performing PCA prior to estimating independent components for each subject.

PCA processing prior to ICA (PCA+ICA) can be problematic because principal components are ranked in terms of variance explained during dimension reduction. Hence, PCA+ICA may discard important low-variance spatial signals. As an alternative, LNGCA can simultaneously perform dimension reduction and extract latent signals by non-Gaussianity (Risk et al., 2019). Because components are ranked in terms of how non-Gaussian they are instead of by how much variance they explain during dimension reduction, LNGCA is able to recover low variance non-Gaussian signals which would otherwise be discarded by PCA+ICA.

We propose group linear non-Gaussian component analysis (group LNGCA) to extract spatial signals from fMRI data that are common across subjects, as well as signals unique to each subject. The proposed model is an extension of LNGCA to multi-subject data, comprising two stages where the first stage applies LNGCA to each subject to extract the top non-Gaussian signals, and the second stage decomposes each subject's non-Gaussian signals into group signals and individual signals. Compared with current group ICA methods, the proposed group LNGCA model has the following advantages. (1) The proposed model can recover low-variance group signals that may contain important biological structure by using subject-level LNGCA for dimension reduction instead of PCA. (2) The proposed model allows subject-specific spatial deviations by decomposing non-Gaussian components into group and individual signals at the second stage. It is hypothesized some of the individual components will correspond to artifacts that are unique to each subject, such as motion artifacts that have subject-specific spatial features. These individual components can also capture subject-specific deviations from group components. We also propose a new approach to estimate the number of non-Gaussian signals in each subject that accounts for the spatial correlation existing in fMRI data. Existing approaches to dimension estimation either assume a Gaussian model (Li et al., 2007; Beckmann and Smith, 2004; Minka, 2001) or ignore the spatial correlation inherent to fMRI data (Nordhausen et al., 2017a; Jin et al., 2017). Through simulation studies, we show that ignoring spatial correlation led to overestimates of the number of signals, while the proposed method achieved more accurate estimation.

In Section 2, we introduce the proposed group LNGCA model, its estimation mechanism and the proposed test of the number of non-Gaussian signals. In Section 3, we use simulations to assess model performance in terms of group components extraction and non-Gaussian signal subspace dimension estimation. In Section 4, we use group LNGCA to estimate components from resting-state fMRI data collected from 342 school age children, including 90 with autism and 252 typically developing, and we compare our results with those obtained from group ICA estimated using GIFT software.

2. Methods

2.1. Group non-Gaussian component analysis model

We propose a group LNGCA model for multi-subject fMRI data, which decomposes non-Gaussian (NG) signals into group signals and individual signals for each subject. Let $i = 1, \ldots, k$ index subjects, $t = 1, \ldots, T$ index time points, and $v = 1, \ldots, V$ index voxels (volumetric pixels). Let $\mathbf{x}_i(v) \in \mathbb{R}^T$ be a data vector of observed fMRI data from subject i at voxel v.

2.1.1. Subject-level LNGCA model

To extract group signals, we first decompose observations into an orthogonal NG subspace and a Gaussian subspace for each subject. Specifically, the LNGCA model decomposes observation $\mathbf{x}_i(v)$ as

$$\mathbf{x}_i(v) = \mathbf{M}_i^{\mathbf{s}} \mathbf{s}_i(v) + \mathbf{M}_i^{\mathbf{n}} \mathbf{n}_i(v), \text{ for } v = 1, \ldots, V, \tag{1}$$

where $\mathbf{s}_i(v) \in \mathbb{R}^{q_i}$ is a vector of mutually independent NG signals with $1 \leq q_i \leq T$, and $\mathbf{n}_i(v) \in \mathbb{R}^{T-q_i}$ is a Gaussian noise vector. The number of NG signals q_i may vary across i. Mixing matrices $\mathbf{M}_i^{\mathbf{s}} \in \mathbb{R}^{T \times q_i}$, $\mathbf{M}_i^{\mathbf{n}} \in \mathbb{R}^{T \times (T-q_i)}$ satisfy that $[\mathbf{M}_i^{\mathbf{s}}, \mathbf{M}_i^{\mathbf{n}}]$

has full rank for any i. Here, $\{\mathbf{x}_i(v)\}_{v=1,\ldots,V}$ are observed while $\{\mathbf{s}_i(v)\}_{v=1,\ldots,V}$ and $\{\mathbf{n}_i(v)\}_{v=1,\ldots,V}$ are latent. We assume $\mathrm{E}\,\mathbf{s}_i(v) = \mathbf{0}$ with $\mathrm{E}\,\mathbf{s}_i(v)\mathbf{s}_i'(v) = \mathbf{I}$. Additionally, assume $\mathrm{E}\,\mathbf{n}_i(v) = 0$ such that $\mathrm{E}\,\mathbf{x}_i(v) = 0$ and $\mathbf{n}_i(v)$ has unit variance for identifiability. In practice, data are centered by their sample mean followed by pre-whitening, which enforces orthogonality of NG components, as described in Section 2.2. We call $\mathbf{s}_i(v)$ NG signals and $\mathbf{n}_i(v)$ Gaussian noise.

PCA+ICA projects observations into a smaller dimensional subspace spanned by the top principal components and extract top NG signals from that subspace. In contrast, subject-level LNGCA directly projects observations into a smaller subspace spanned by the top NG signals, which are ordered by a measure of non-Gaussianity. Thus LNGCA can capture NG signals that have small variance that may be discarded in the PCA step in PCA+ICA. Notice the key difference between LNGCA and PCA+ICA is how they search for a low rank space, and they are equivalent if there is no dimension reduction.

2.1.2. Group-level LNGCA model

Assume there is at least one group signal. We further decompose $\mathbf{s}_i(v)$ into orthogonal group signals $\mathbf{s}_g(v)$ and individual signals $\mathbf{s}_{I,i}(v)$,

$$\mathbf{M}_i^{\mathbf{s}}\mathbf{s}_i(v) = \mathbf{M}_i^{g}\mathbf{s}_g(v) + \mathbf{M}_i^{I}\mathbf{s}_{I,i}(v), \text{ for } v = 1, \ldots, V, \tag{2}$$

where $\mathbf{s}_g(v) \in \mathbb{R}^{q_g}$ is a vector of mutually independent group signals with $1 \leq q_g \leq \min_i q_i$, $\mathbf{s}_{I,i}(v) \in \mathbb{R}^{q_{I,i}}$ is a vector of mutually independent individual signals with $q_{I,i} = q_i - q_g$. Group signals $\mathbf{s}_g(v)$ are shared across all subjects. Combining (1) and (2), we have the complete decomposition, which we call the group linear non-Gaussian component analysis model (group LNGCA):

$$\mathbf{x}_i(v) = \mathbf{M}_i^{g}\mathbf{s}_g(v) + \mathbf{M}_i^{I}\mathbf{s}_{I,i}(v) + \mathbf{M}_i^{\mathbf{n}}\mathbf{n}_i(v). \tag{3}$$

In subject-level LNGCA, the NG signals and the mixing matrix in (1), i.e. $\mathbf{s}_i(v) = [\mathbf{s}_g(v)', \mathbf{s}_{I,i}(v)']'$ and $\mathbf{M}_i^{\mathbf{s}} = [\mathbf{M}_i^{g}, \mathbf{M}_i^{I}]$, are identifiable up to sign and permutation (Risk et al., 2019). Consider the matrix of subject-level components for the ith subject: $\mathbf{S}_i \in \mathbb{R}^{q_i \times V}$, with rows $[s_{ij}(1), \ldots, s_{ij}(V)]$ for $j = 1, \ldots, q_i$. Then the group components are matched rows from different subjects, and individual components are rows that are not equal.

Risk et al. (2019) define an estimator and conditions for consistency in single-subject LNGCA when $V \to \infty$. It follows immediately that in the group LNGCA model, the estimate $\hat{\mathbf{M}}_i^{\mathbf{s}}$ from a single-subject analysis is a consistent estimator of $\mathbf{M}_i^{\mathbf{s}}$. As described in Section 2.2, we will use the logistic non-linearity, which is consistent for most super-Gaussian distributions. The family of consistent distributions is defined by Assumption 5 in Risk et al. (2019) with additional details for the logistic density in their Web Supplement B. A key observation is that $\mathbf{M}_i^{\mathbf{s}}\mathbf{s}_i(v)$ can have smaller variance than $\mathbf{M}_i^{\mathbf{n}}\mathbf{n}_i(v)$, such that in a sample $v = 1, \ldots, V$, extracting the "signal" subspace with the PCA step in group ICA will discard the non-Gaussian components. In our framework, this means that some brain components may be missed by group ICA.

In practice, we will use a singular value decomposition of the concatenated subject-level NG components (subspaces of $\mathbb{R}^V$) to find the group subspace, where the size of the singular values characterizes the extent to which a direction of the non-Gaussian subspace is present across subjects; this is discussed in the next section. In applications, if a group component is prominent in one subgroup but reduced in another subgroup, then this can be captured with a group component in which the variances of the subject-specific time courses (columns of $\mathbf{M}_i^{g}$) reflect differing roles, as explored in Section 4. We expect that subject-specific components will arise from components that are unique to a subject, such as certain types of motion artifacts. Additionally, subject-specific components may arise from deviations in spatial patterns of common resting-state networks that are unique to a subject. For example, in fMRI, different subjects may have similar, but not identical, default mode networks (DMNs) (Braga and Buckner, 2017).

2.2. Estimation

In the population formulation of group LNGCA, the decomposition for each subject following (3) is a re-labeling of the subject components in (1) into group and individual NG components. In a sample of observations, the subject-specific LNGCAs will result in noisy estimates of the population signals, and we propose a scalable algorithm to extract the group components. We summarize our estimation procedure in Algorithm 1.

In Step 1, we estimate NG signals for each subject. We first center and scale the rows of each subject's $T \times V$ data matrix, take the right singular vectors, and then multiply by $\sqrt{V-1}$. The resulting data matrix, denoted $\mathbf{X}_i$, satisfies $\mathbf{X}_i\mathbf{1} = \mathbf{0}$ and $\frac{1}{V-1}\mathbf{X}_i\mathbf{X}_i' = \mathbf{I}$. This pre-whitening is common in ICA (Beckmann and Smith, 2004). Then the initial subject-level NGs are estimated by finding a semiorthogonal matrix, called the unmixing matrix, that achieves dimension reduction by maximizing non-Gaussianity. We use the logistic non-linearity as our measure of non-Gaussianity with the fastICA algorithm (Hyvarinen, 1999) modified for LNGCA because the logistic non-linearity is used in Infomax and performs well in fMRI (Correa et al., 2007; Plis et al., 2014). We find this algorithm works well for the super-Gaussian distributions found in fMRI (see simulations in Section 3), and it is computationally tractable and less sensitive to initializations. Our approach can be adapted to more flexible (but typically computationally costly) non-linearities that also allow the estimation of sub-Gaussian densities, for example, ProDenICA (Hastie and Tibshirani, 2003), if applied to other applications with sub-Gaussian distributions.

Algorithm 1 Group LNGCA Estimation Algorithm.

Input: pre-whitened $\mathbf{X}_i \in \mathbb{R}^{T \times V}$ with $\mathbf{X}_i \mathbf{1} = \mathbf{0}$ and $\frac{1}{V}\mathbf{X}_i \mathbf{X}_i' = \mathbf{I}$, NG subspace dimension q_i for each subject i; group NG subspace dimension q_g.
Output: Group signals $\hat{\mathbf{S}}_g \in \mathbb{R}^{q_g \times V}$; individual signals $\hat{\mathbf{S}}_{I,i} \in \mathbb{R}^{(q_i - q_g) \times V}$ for each subject i.

1. Perform the subject-level LNGCA for each i:

$$\mathbf{X}_i = \hat{\mathbf{M}}_i^{\mathbf{s}} \hat{\mathbf{S}}_i + \hat{\mathbf{E}}_i$$

where $\hat{\mathbf{S}}_i \in \mathbb{R}^{q_i \times V}$.
2. Concatenate $\hat{\mathbf{S}}_i$: $\hat{\mathbf{S}}_{1:k} = [\hat{\mathbf{S}}_1', \dots, \hat{\mathbf{S}}_k']' \in \mathbb{R}^{(\sum_i q_i) \times V}$. Conduct an SVD of $\hat{\mathbf{S}}_{1:k}$. Retain the first q_g right singular vectors of $\hat{\mathbf{S}}_{1:k}$, denoted as $\hat{\mathbf{Y}}_g \in \mathbb{R}^{q_g \times V}$.
3. Perform noise-free ICA on the group signals subspace:

$$\hat{\mathbf{Y}}_g = \hat{\mathbf{M}} \hat{\mathbf{S}}_g$$

where $\hat{\mathbf{S}}_g \in \mathbb{R}^{q_g \times V}$ are the estimated q_g group signals satisfying $\frac{1}{V}\hat{\mathbf{S}}_g \hat{\mathbf{S}}_g' = \mathbf{I}$.
4. Subtract group subspace from each subject's non-Gaussian subspace and apply LNGCA:

$$\hat{\mathbf{M}}_i \hat{\mathbf{S}}_i (\mathbf{I} - \frac{1}{V}\hat{\mathbf{S}}_g' \hat{\mathbf{S}}_g) = \hat{\mathbf{A}}_i,$$

$$\hat{\mathbf{A}}_i = \hat{\mathbf{M}}_i^I \hat{\mathbf{S}}_{I,i}$$

where $\hat{\mathbf{S}}_{I,i} \in \mathbb{R}^{(q_i - q_g) \times V}$ is the estimated $(q_i - q_g)$ individual signals for subject i.

In Step 2, we construct the subspace spanned by all NG signals across subjects by concatenating the subject-level NGs. To estimate the group non-Gaussian subspace, we conduct an SVD and extract the first q_g singular vectors, which is equivalent to performing PCA on the concatenated subject NGs. Insight into this step can be gained by viewing the SVD as a principal angle analysis. For ease of notation in what follows, rescale the NG components to have norm equal to 1 prior to concatenation. Assume observations $v = 1, \dots, V$ from k subjects, and let $\mathbf{S}_{i,d}^g \in \mathbb{R}^{1 \times V}$ be the dth group component in the ith subject. In the concatenated NG components, this group component has multiplicity k. Then the size of the singular value from a group component shared by all subjects is equal to $\sqrt{k}$. In practice, this group subspace will be an average of the information shared across subspaces. For two subjects, the principal angles between their subspaces are equal to $\arccos(\sigma_d^2 - 1)$, $d = 1, \dots, (q_1 + q_2)$, where σ_d is the dth singular value and q_i is the number of NG components in the ith subject. In the case of $q_1 = q_2 = 1$, the first right singular vector is the average of the NG components, and $\sigma_1^2 - 1$ is the correlation (for mean centered data). The use of the SVD is similar to the approach in Feng et al. (2018) in which joint structure is extracted from multiple datasets, but here we apply the procedure to the concatenated NG signals from multiple subjects to obtain a subspace that is shared across multiple subjects.

We discuss how to estimate $\{q_i\}_{i=1,\dots,k}$ in Section 3.2. Determining q_G is beyond the scope of this work but discussed in Section 5. Notice the choice of q_G is more of a practical consideration for large noisy fMRI data, since those data usually do not admit a clear gap between group and individual components. With small q_G, only a small portion of NG signals can be estimated, but fast and accurate. With large q_G, the returned components are likely to contain all NG signals, although with slow computation, and visual inspection is required to exclude noise components. We show such phenomena through simulation study.

In Step 3, we search for the orthogonal transformation of the group subspace that maximizes non-Gaussianity using noise-free ICA. This produces an estimate of the group components.

In Step 4, we estimate the subject-specific signal subspace as the orthogonal complement of the group signal subspace in the estimated NG subspace for that subject. Then we apply LNGCA to the subject-specific subspace to extract subject-specific NG signals.

2.3. Comparison with group ICA model and algorithm

A major difference between group ICA and group LNGCA is that LNGCA is conducted for subject dimension reduction rather than PCA. In simulations, we examine how this impacts the estimation under different signal variance regimes. Although both algorithms use PCA for the group stage dimension reduction, it is performed on distinct subspaces. For group LNGCA, as long as the NG signals are extracted in many subjects, they will be the top principal components in the concatenated NG components. For group ICA, there is no such guarantee. An NG signal is included in the top principal components only when it has relatively large variance and thus is not largely discarded in the subject PCA step. Although one could keep a very large ratio of variance in the subject PCA step to avoid the issue, that would result in challenges to accurately conduct the group stage PCA (usually through iterative methods) due to much larger data size compared to group LNGCA. When there is no dimension reduction on the subject data, the group LNGCA algorithm is equivalent to group ICA algorithm as in Calhoun et al. (2001).

In addition to the difference in how the group subspace is defined, group LNGCA estimates individual signals, which is useful for examining non-Gaussian signal that is not captured in the group component. In contrast, group ICA discards all information not captured by the group components.

2.4. Test the dimension of the non-Gaussian signals subspace

To estimate the NG subspace dimension q_i for each subject, recently proposed methods (Nordhausen et al., 2017a; Jin et al., 2017) sequentially test the dimension of the NG subspace:

$$H_0^k : \text{There are at most } k \text{ NG signals.} \quad \text{versus} \quad H_A^k : \text{There are at least } k+1 \text{ NG signals.}$$

for $0 \le k \le T-1$. Suppose the true dimension is k_0. With increasing sample size, we expect a good test to perform as: (1) for $k < k_0$, the power of the test for H_0^k approaches one; (2) for $k = k_0$, the size of the test for $H_0^{k_0}$ approaches prespecified α; and (3) for $k > k_0$, the rejection probability for H_0^k tends to be smaller than α.

Denote the p-value associated with H_0^k by p_k. The estimate is $\hat{k} = \{k \mid p_k \le \alpha, \ p_{\hat{k}-1} > \alpha\}$. It is very expensive to test H_0^k for all possible k. However, using a binary search in k, we can expect no more than $\lceil \log_2 T \rceil$ tests for dimension T (Jin et al., 2017). The estimated $\hat{k}$ then relies on multiple tests, which may be problematic for large T. One may consider adjusting the obtained p-values using Bonferroni correction or FDR control. However, the sequential tests result in a set of nested hypotheses that are highly dependent in that: for any $k_1 < k_2$, when $H_0^{k_1}$ holds, $H_0^{k_2}$ must hold. Adjusting p-values without considering this special dependence structure may largely decrease the power of sequential tests and damage the estimated $\hat{k}$. Thus we use the original p-values and show through simulations in Sec 3.2 that we obtain accurate estimation of k.

2.4.1. Motivation

Suppose $D(\cdot)$ is some non-Gaussianity measure for a signal, e.g., skewness, excess kurtosis, or the Jarque-Bera statistic. For a matrix $\mathbf{Y} \in \mathbb{R}^{T \times V}$, sort $\mathbf{Y} = [\mathbf{Y}'_{(1)}, \cdots, \mathbf{Y}'_{(T)}]'$ such that $D(\mathbf{Y}_{(1)}) > \cdots > D(\mathbf{Y}_{(T)})$, where $\mathbf{Y}_{(t)} \in \mathbb{R}^V$ is a row vector of $\mathbf{Y}$, for $t = 1, \ldots, T$. In this section, we drop the subject index, but in practice, these tests are applied separately to each subject. We will use this sorting (rank) notation for other data matrices below. Motivated by Jin et al. (2017), we assume $\mathbf{Y}_{(j_1)}$ is more non-Gaussian than $\mathbf{Y}_{(j_2)}$ in terms of $D(\cdot)$ for $j_1 < j_2$. Consequently, under H_0^k, $[\mathbf{Y}'_{(1)}, \ldots, \mathbf{Y}'_{(k)}]'$ are NG signals and $[\mathbf{Y}'_{(k+1)}, \ldots, \mathbf{Y}'_{(T)}]'$ are Gaussian noise. Jin et al. (2017) proposed an algorithm to test H_0^k based on a max-min estimator which maximizes the non-Gaussianity of k components while minimizing the non-Gaussianity of the $(T-k)$ Gaussian components. The key idea is to extract the NG signals and Gaussian noise from the original data, then generate new data samples mixed from the extracted NG signals and new sampled Gaussian noise. However, for every generated new sample, the algorithm conducts a full rank ICA-like algorithm on a matrix of size $V \times T$, which is very expensive. To overcome such computational burden, we choose to only sample the Gaussian noise rather than the whole data set. The computation time of our procedure is over 100 times faster than the approach in Jin et al. (2017).

2.4.2. Resampling test for dimension estimation

We propose an algorithm based on the sample distribution of the maximum component-wise non-Gaussianity, i.e. $D(\mathbf{G}_{(1)})$, when ICA is applied to $T-k$ Gaussian components $\mathbf{G}$. We state our method in Algorithm 2.

Algorithm 2 Our NG subspace dimension test H_0^k.

Input: pre-whitened $\mathbf{X} \in \mathbb{R}^{T \times V}$ with $\mathbf{X1} = 0$ and $\frac{1}{V}\mathbf{XX}' = \mathbf{I}$, number of samples B.
Output: Empirical p-value $\hat{p}$.

1. Estimate $\hat{\mathbf{S}} = \hat{\mathbf{W}}\mathbf{X}$ with $\hat{\mathbf{S}} \in \mathbb{R}^{T \times V}, \hat{\mathbf{W}} \in \mathcal{O}^{T \times T}$.
2. Sort $\hat{\mathbf{S}} = [\hat{\mathbf{S}}'_{(1)}, \cdots, \hat{\mathbf{S}}'_{(k+1)}, \ldots, \hat{\mathbf{S}}'_{(T)}]'$ to obtain $D(\hat{\mathbf{S}}_{(k+1)})$.
3. Repeat for $b = 1, \cdots, B$:
   (a) Generate $(T-k)$ independent Gaussian components $\mathbf{G}^{(b)} \in \mathbb{R}^{(T-k) \times V}$.
   (b) Estimate $\hat{\mathbf{S}}_G^{(b)} = \hat{\mathbf{W}}_G^{(b)}\mathbf{G}^{(b)}$ with $\hat{\mathbf{S}}_G^{(b)} \in \mathbb{R}^{(T-k) \times V}, \hat{\mathbf{W}}_G^{(b)} \in \mathcal{O}^{(T-k) \times (T-k)}$.
   (c) Sort the rows of $\hat{\mathbf{S}}_G^{(b)}$ to obtain $D(\hat{\mathbf{S}}_{G,(1)}^{(b)})$.
4. Calculate the empirical p-value: $\hat{p} = \frac{\#\{D(\hat{\mathbf{S}}_{(k+1)}) < D(\hat{\mathbf{S}}_{G,(1)}^{(b)})\}}{B}$

Now in each repetition, the ICA is implemented on a matrix with only $T-k$ components at Algorithm 2 step 3(b), to extract the component $\hat{\mathbf{S}}_{G,(1)}^{(b)}$ with highest $D(\cdot)$ among orthogonally transformed $(T-k)$-variate Gaussian components. We assume that $\hat{\mathbf{S}}_{(k+1)}$ shares the same distribution as $\hat{\mathbf{S}}_{G,(1)}^{(b)}$. As a result, $D(\hat{\mathbf{S}}_{(k+1)})$ and $D(\hat{\mathbf{S}}_{G,(1)}^{(b)})$ have the same distribution.

To implement our proposed algorithm, we use FOBI (Cardoso, 1989) with $D(\cdot)$ equal to kurtosis. Unlike other methods, FOBI has a closed-form solution, which makes it computationally scalable and avoids the need for multiple restarts. Although in theory other more computationally intensive methods could be used, our goal is to develop a method that can be applied

to real fMRI data. Previous approaches without spatial correlation have found FOBI works well for dimensionality tests (Nordhausen et al., 2017a). Note the goal here is to estimate the dimensions, rather than accurate components.

2.4.3. Spatially correlated Gaussian noise

For fMRI data, the pre-processing steps may apply Gaussian smoothing, and interpolation steps also introduce smoothness (Chen and Calhoun, 2018; Coalson et al., 2018). For example, suppose $\mathbf{G} \in \mathbb{R}^{(T-k) \times V}$ has each row generated from a Gaussian random field. Then unmixing these components can result in disk-like features in spurious NG components (Risk et al., 2019). As we will show, methods assuming independent entries in Gaussian noise can overestimate k_0, and consequently too large k_0 cannot help achieve dimension reduction. We can account for the spatial autocorrelation by adjusting Algorithm 2 step 3(a).

When there is no spatial correlation, we generate row vectors $\mathbf{G}_l^{(b)} \overset{i.i.d.}{\sim} \mathcal{N}(0, I_V)$ for $l = 1, \ldots, T - k$. When there is spatial correlation, we can modify Step 3(a) in Algorithm 2. We first specify or estimate the spatial correlation matrix $\Sigma \in \mathbb{R}^{V \times V}$, then generate row vectors $\mathbf{G}_l^{(b)} \overset{i.i.d.}{\sim} \mathcal{N}(0, \Sigma)$ for $l = 1, \ldots, T - k$. In practice, we can generate random fields by applying smoothers to iid Gaussian data, as implemented in neuRosim (Welvaert et al., 2011). Such modification ensures a component is only classified as NG signal when it is more non-Gaussian than the spatially correlated Gaussian noise.

3. Simulations: spatio-temporal signals

We evaluate the performance of group LNGCA versus group ICA (Calhoun et al., 2001) in a simulation study with twenty subjects. We also compare the performance of our NG subspace dimension test with those proposed in Nordhausen et al. (2017a). All algorithms are repeated with 30 random initializations. Our code is available at https://github.com/yuxuanzhao2295/Group_LNGCA_for_Neuroimaging.

3.1. Data generation

For each of the twenty subjects, we used 3 group signals, 22 individual signals from gamma random fields, and 25 Gaussian noise components from Gaussian random fields. Example components are depicted in Fig. 1. The group signals have active pixels in the shape of a "1", "2 2", or "3 3 3" with values between 0.5 and 1 and inactive pixels as iid normal with mean 0 and variance 0.001. These generate mean logistic non-linearities (log logistic density with scale parameter equal to $\sqrt{3}/\pi$) equal to -1.09, -1.08, and -1.11. For reference, Gaussian data have mean logistic non-linearity equal to -1.43, and in our real data application, the 59 group components ranged from -1.36 to -0.92. Random fields (RF) were simulated using the R package neuRosim (Welvaert et al., 2011). For both gamma and Gaussian random fields, the full-width at half maximum of the Gaussian kernel, which controls spatial correlation, was set to 9. For the marginal distributions of the individual signals, we selected gamma parameters (shape parameter 0.02, rate parameter 10^{-4}) that resulted in individual signals with mean logistic non-linearity ranging from -1.17 to -0.87. The corresponding range found in our real data application is from -1.39 to -0.82. We do not strictly enforce orthogonality of components, but the areas of activation in the group components are disjoint and the random individual components are independent, such that the model assumptions are approximately met. For the columns of the mixing matrix, we simulate AR(1) processes per subject with $\phi = 0.37$, estimated from our real data application (detailed in the Web Appendix Section S.1.1. The scaling of the columns was chosen to result in variance proportions detailed below.

To control signal and noise strength, we define the subspace variance ratio (SVAR) among the group NG subspace, individual NG subspace, and Gaussian noise subspace. Fixing the subscript i, let $\lambda_1, \ldots, \lambda_{q_g}$ be the nonzero eigenvalues from the eigenvalue decomposition (EVD) of the covariance matrix of $\mathbf{M}_i^g \mathbf{S}_g$; $\mu_1, \ldots, \mu_{q_i - q_g}$ be the nonzero eigenvalues from the EVD of the covariance matrix of $\mathbf{M}_i^I \mathbf{S}_{I,i}$; and $\nu_1, \ldots, \nu_{T-q_i}$ be the nonzero eigenvalues from the EVD of the covariance matrix of $\mathbf{M}_i^n \mathbf{N}_i$. The SVAR is defined as $\sum_{l=1}^{q_g} \lambda_l : \sum_{l=1}^{q_i - q_g} \mu_l : \sum_{l=1}^{T-q_i} \nu_l$. We can also normalize these values by the total variance, and then this is equivalent to the proportion of variance from the group signals, individual signals, and Gaussian noise. Here, we focus on the variance ratio of the group NG subspace, so we designed three simulation settings corresponding to a high SVAR (large $\sum_{l=1}^{q_g} \lambda_l$), a medium SVAR, and a low SVAR in the group NG subspace. In the discussion that follows, we refer to the variance of a component as the sum of squares of the corresponding column of the mixing matrix.

First, we describe how we allocated the variance in a given subspace to the components in that subspace. For all settings, we assumed the SVAR was equal for all 20 subjects. Then in the group subspace, we defined a low variance component, a medium variance component, and a higher variance component using the 0.1, 0.5, and 0.9 quantiles of the variance of the group signals estimated in the real data application (15.4% : 29.8% : 54.8%, based on the 59 group components estimated from 342 subjects). This is designed to examine whether signals with lower variance may be discarded by group ICA. In the individual NG subspace and the Gaussian subspace, we assumed all components had equal variance.

Next, we describe how we allocated the total variance to each subspace. To determine the high SVAR scenario, we calculated the proportion of variance in each of these subspaces for each subject in the real data application, wherein we estimated 59 group signals and 85 total NG signals with either 128 or 156 time points, see Section 4. We then calculated the median across subjects and re-scaled the proportions to sum to one ($\sum_{l=1}^{q_g} \lambda_l : \sum_{l=1}^{q_i - q_g} \mu_l : \sum_{l=1}^{T-q_i} \nu_l = 33.5\% : 29.9\% : 36.6\%$).
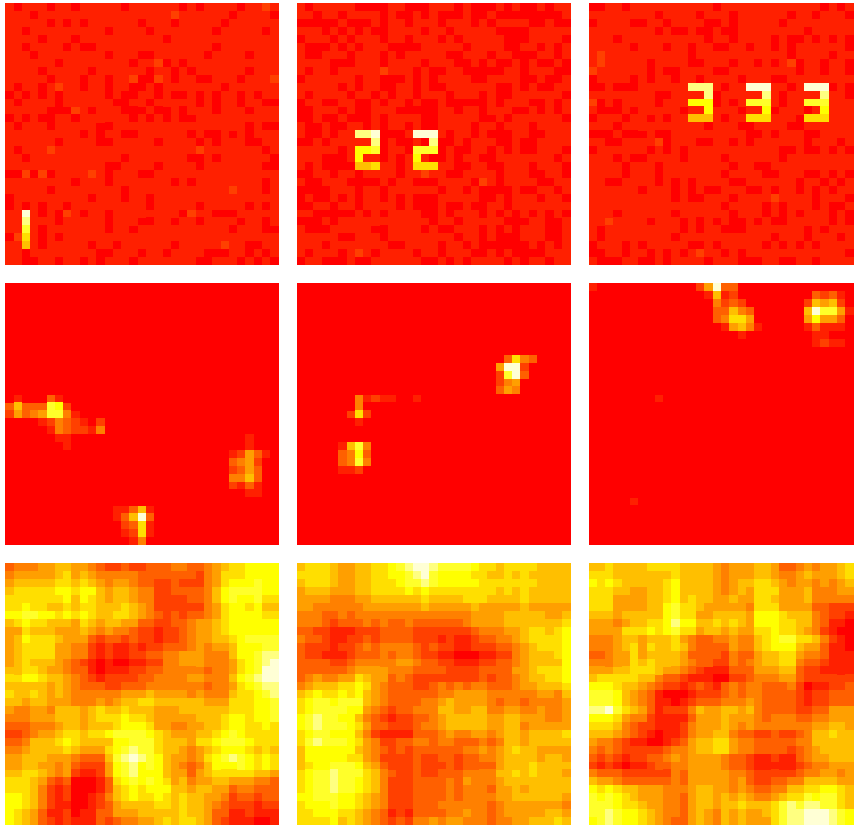
Fig. 1. Simulated non-Gaussian and Gaussian components. First row depicts 3 group NG signals. Second row depicts 3 (of 22) individual NG signals. Last row depicts 3 (of 25) Gaussian noise components. Each component is a 33×33 image corresponding to $V = 1089$.

Note this distributes the variance of the NG subspace formed from 59 signals to 3 signals, and thus these signals have a proportionately very large variance. For the low SVAR scenario, we set the variance of each of the group signals equal to the 0.1, 0.5, and 0.9 quantiles of the variances of the 59 group signals. This results in a much lower variance relative to the high SVAR scenario: $\sum_{l=1}^{q_g} \lambda_l : \sum_{l=1}^{q_i - q_g} \mu_l : \sum_{l=1}^{T-q_i} \nu_l = 1.7\% : 46\% : 52.3\%$. For the medium SVAR scenario, we set $\sum_{l=1}^{q_g} \lambda_l$ to be half of the sum of that in high and low SVAR scenario. Thus we have medium SVAR: $\sum_{l=1}^{q_g} \lambda_l : \sum_{l=1}^{q_i - q_g} \mu_l : \sum_{l=1}^{T-q_i} \nu_l = 17.6\% : 38.6\% : 43.8\%$.

3.2. Non-Gaussian subspace dimension estimation

We implemented our test in Algorithm 2 with spatially correlated Gaussian noise, denoted as `FOBI-GRF`, and compare its performance with tests `FOBIasymp` and `FOBIboot` from the R package `ICtest` (Nordhausen et al., 2017b). For each test, we apply a binary search to estimate the NG subspace dimension. For each SVAR setting, one experiment contains 20 subjects and the experiment is repeated 40 times, thus 800 subjects in total. The true NG subspace dimension is 25 for all settings. We report the results under the medium SVAR setting in Fig. 2. The results under the high and low SVAR settings are very similar and appear in Web Appendix Fig. S.2.

`FOBIasymp` and `FOBIboot` do not consider the spatial correlation, and they clearly overestimate the dimension. `FO-BIboot` is a parametric bootstrap method that simulates Gaussian noise with iid entries. We see `FOBIboot` estimates almost all components as NG when the assumption is violated. Our test `FOBI-GRF` performs the best in terms of accuracy and stability. Sometimes our test underestimates the number of NG components, but as we will show in the next section, group LNGCA is still able to extract all group NG components. The underestimation can occur due to the event of "unmixed" Gaussian RF components having higher non-Gaussianity than gamma RF components. This is similar to the gap between signal and noise eigenvalues in principal component analysis (Zhu and Ghodsi, 2006). In our context, a smaller gap between the NG signals and transformations of the spatially correlated Gaussian signals can result in a decrease in the accuracy of our testing procedure. In applications, the individual components tend to correspond to structured artifacts or individual resting-state networks, which may be more distinct from Gaussian RFs than the gamma RFs used in our simulations. The choice of gamma RFs here is to allow computational scalability, since it is difficult to generate 22*20 individual NG components.
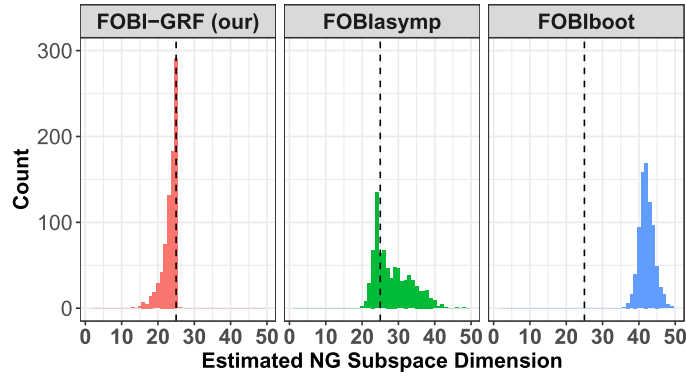
Fig. 2. Estimated non-Gaussian subspace dimension across 800 subjects under medium SVAR setting. The significance level $\alpha = 0.05$. Dashed line indicates true dimension 25. The most frequently selected dimension using our test, FOBI-GRF, corresponds to the true dimension. Although the test underestimated the dimensions in many simulations, this was due to possibly missing individual components, while it always retained the group components, as described in Section 3.3.

We comment here the NG subspace dimension is a challenging problem, and in particular is more difficult for higher dimensions. Suppose we have a fixed number of NG signals and linearly mix them with Gaussian components. As the number of Gaussian components increase, there will also be more components displaying high non-Gaussianity among extracted independent components. Bickel et al. (2018) showed the projection of multivariate Gaussian distribution can approximate any non-Gaussian distribution for large T, in particular, for $T > V$. Here, it appears that with low frequency some components estimated from Gaussian RFs can be more non-Gaussian than the NGs. Thus when estimating the NG subspace dimension k, one could consider using a higher value of α with increased dimension T, as the NG component can be hidden between spurious components. In practice, a visual inspection may suffice to identify the spurious, disc-like components that arise from maximizing non-Gaussianity of Gaussian RFs.

3.3. Group and individual non-Gaussian components extraction

Using the estimated subject NG subspace dimension from our test as reported in Sec 3.2, we implement group LNGCA as in Algorithm 1 to estimate the group signals. For group ICA, we conducted the subject-level PCA with the number of signals selected to retain at least 82% of the variance (selected to be consistent with the real data application, detailed in Section 4) for each subject. For both group LNGCA and group ICA, we first examine their performance with correctly specified group NG subspace dimension $q_g = 3$. After extracting group signals, we matched them to true signals using a modification of the Hungarian algorithm (Risk et al., 2014), and then calculated the correlation between matched signals, for each method respectively. For group LNGCA, we also calculate the correlation for 22 individual components. To be consistent when computing the correlation for individual components, we implement Step 4 of Algorithm 2 with the number of individual components equal to 22 for all subjects. The simulation is repeated 40 times for each setting. The computation time for group LNGCA for each repetition is 3-4 minutes on a laptop with an Intel i5 3.1 GHz processor with 8 GB RAM. We report the results, the correlation between matched components, in Fig. 3, and the estimated signals from the repetition associated with the median matching error (with true signal) in each setting are depicted in Fig. 4.

When SVAR is high, both methods can recover three group signals with high accuracy. However, accuracy across methods diverges as SVAR decreases. Group ICA failed to recover the low variance group signal in the medium SVAR setting and failed to extract any group signals under the low SVAR setting, while group LNGCA was highly accurate under all three SVAR settings. Note that these group LNGCA estimates used the estimated subject-level subspaces from Section 3.2 with the estimated $\hat{q}_i$, which were often less than the true $q_i = 25$ (Fig. 2). Even in these cases, the group components are accurately estimated. Although we do not examine the case where we deliberately mis-specify $\hat{q}_i > q_i$, we expect similar results in that setting, since the subject-level subspace would contain the group signal. Also notice group ICA is better able to estimate group signals with larger variance, which is consistent with the fact it reduces the data dimension by signals' variance and may lose low variance signals.

Overall, group LNGCA extracts most individual components with high accuracy (right column Fig. 3). Group LNGCA appears to occasionally miss one or two individual components in a subject, represented by the outliers in the boxplot. Since the individual components are randomly generated from gamma random fields, some realizations can be closer to Gaussian than others. At the same time, some linear combinations of the Gaussian random fields can result in non-trivial non-Gaussianity. Then these individual components can be missed by group LNGCA, which appear as outliers in Fig. 3. However, this is uncommon ($< 0.2\%$ of the 22*20*40 individual components' correlations were less than 0.5).
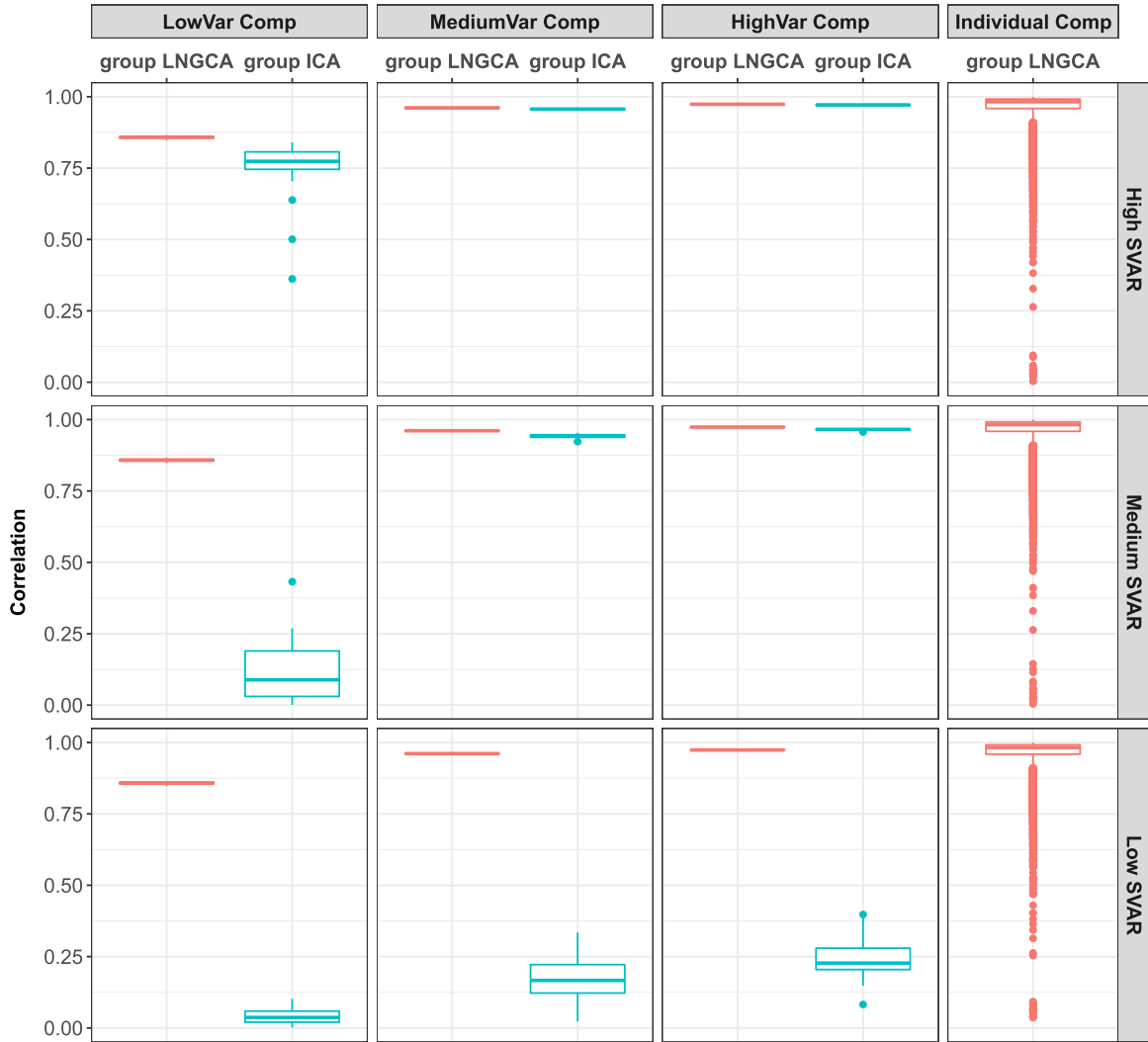
Fig. 3. Correlation between estimated components and their corresponding true components. The plot for individual components is over all 22 individual components in 20 subjects across all 40 repetitions (17,600 in total). The correlation on three group components under all settings for group LNGCA concentrates at a high correlation value with vanishing variance. The percentage of extracted individual components with low correlation ($< .5$) is below 0.2% for all SVAR settings.

3.4. Additional simulations evaluating robustness

In Web Appendix Section S.2.1, we examine robustness to mis-specifying the number of group components. When $q_g = 4$, all three group signals can still be recovered with high accuracy (Figs. S.3 and S.4). When $q_g = 2$, two of the three group signals can be recovered with high accuracy (Figs. S.5 and S.6).

In Web Appendix Section S.2.2, we examine accuracy when we vary the number of time points to $T = 30$ and $T = 70$, by changing the number of Gaussian random field components. This is similar to changing scan duration in fMRI scans. Group LNGCA is accurate in both settings (Fig. S.7 and Fig. S.8).

In Web Appendix Section S.2.3, we create subject-specific deviations from the group components in two subjects, as detailed in Fig. S.9. Group LNGCA is able to recover both the group signal, which corresponds to the group feature shared across all subjects, and also the subject deviations, which are captured as individual components.

4. Resting-state fMRI data example

We applied group LNGCA and group ICA to resting-state fMRI data from 342 school-age children, ages 8 through 12 years, recruited at the Kennedy Krieger Institute (PI: S. Mostofsky) including 90 with autism spectrum disorder (ASD) and 252 typically developing (TD) controls. Resting-state fMRI was acquired during either a 5 min 20 s- or 6 min 30 s-long scan on a 3.0 T Philips scanner using a single-shot, partially parallel, gradient-recalled echo planar sequence (repetition time
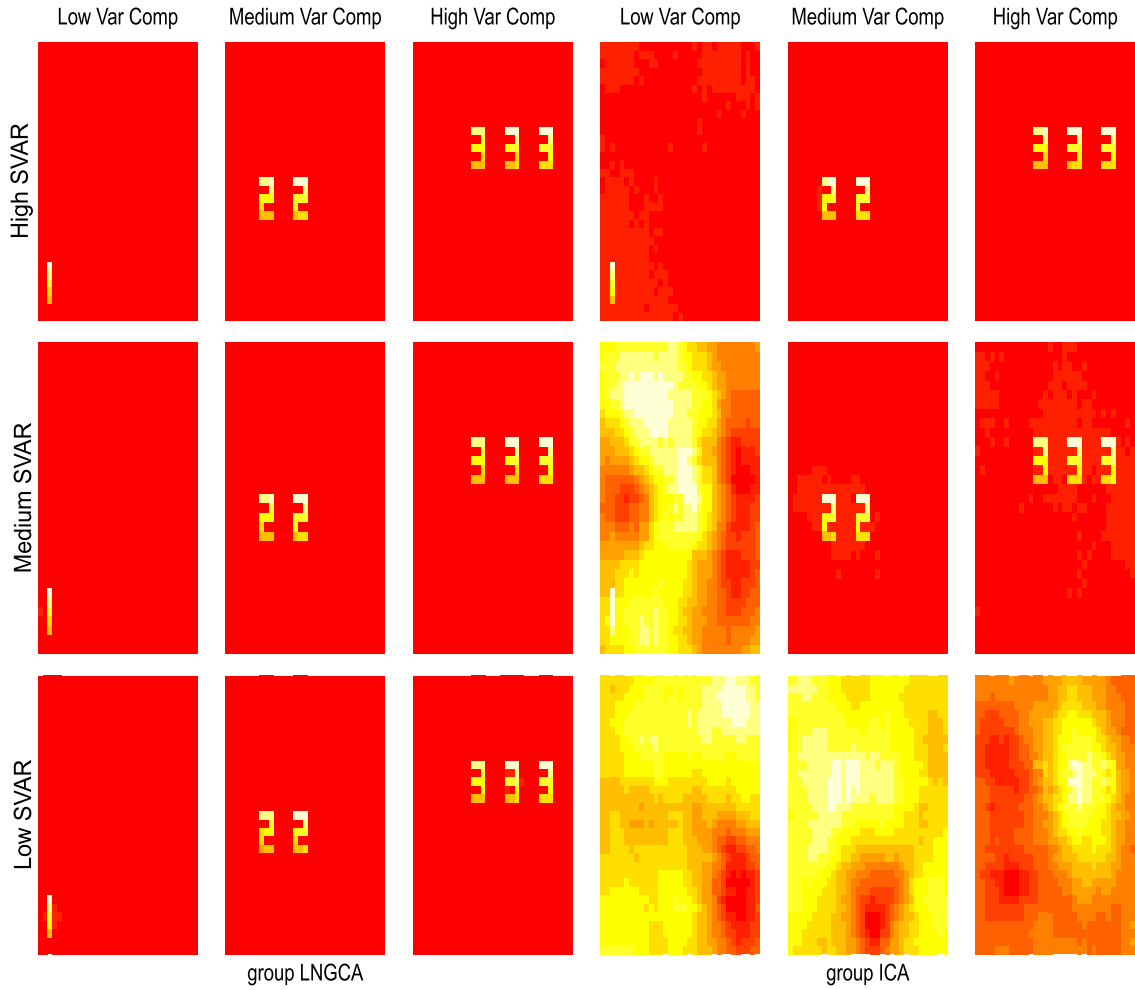
Fig. 4. The estimated group components from a representative simulation (median matching error) when $q_G = 3$. Left three columns display results from group LNGCA, while right three columns display results from group ICA. For each method, the allocated variance increases from left to right among three signals. Three rows represent high, medium and low SVAR settings respectively, from top to bottom.

2500 ms, echo time 30 ms, flip angle = 75°, SENSE factor of 2, 3-mm axial slices with no slice gap, in-plane resolution of 3.05×3.15 mm resulting in $84 \times 81 \times 47$ voxels). Data were registered to MNI space and smoothed using a 6-mm FWHM Gaussian filter. Each participant's preprocessed data were mean-centered and variance normalized on a voxel-wise basis. Additional details on preprocessing and motion exclusion criteria are provided in the Web Appendix Section S.3.1.

4.1. Subject-level and group-level dimension reduction

We applied our test in Algorithm 2 to six participants. Gaussian RFs were generated using the estimated FWHM of the data from 3dFWHMx in AFNI. The estimated dimension was 65, 38, 89, 47, 73 and 83, corresponding to participants with 128 time points for first three participants and 156 for last three participants (detailed in the Web Appendix Section S.3.2). Thus the number of components varied by participant, but in general indicated the NG signal was contained in a lower-dimensional subspace. The test is applied to only six participants because it is computationally intensive on fMRI data (approximately two hours per subject, Web Appendix Table S.1), and we only need to get a ballpark number here.

For the subject-level PCA in group ICA, we performed dimension estimation on each participant's preprocessed data using an information theoretic approach implemented within GIFT (Li et al., 2007), which calculates minimum description length (MDL) assuming a Gaussian model. The maximum of the estimated number of PCs from the 342 datasets was 59. Following recommendations from Erhardt et al. (2011), we conservatively chose to use 85 components for all subjects. This kept at least 82% of the variance in each subject's data. We note that this approach retains more variance than suggested using GIFT's dimensionality tests, and in this respect will make our results with group LNGCA more similar to group ICA. For group LNGCA, we also chose 85 NGs for all participants, to make fair comparison between two methods. Notice 85 is also near the maximum of our estimated NG dimension from the six participants.
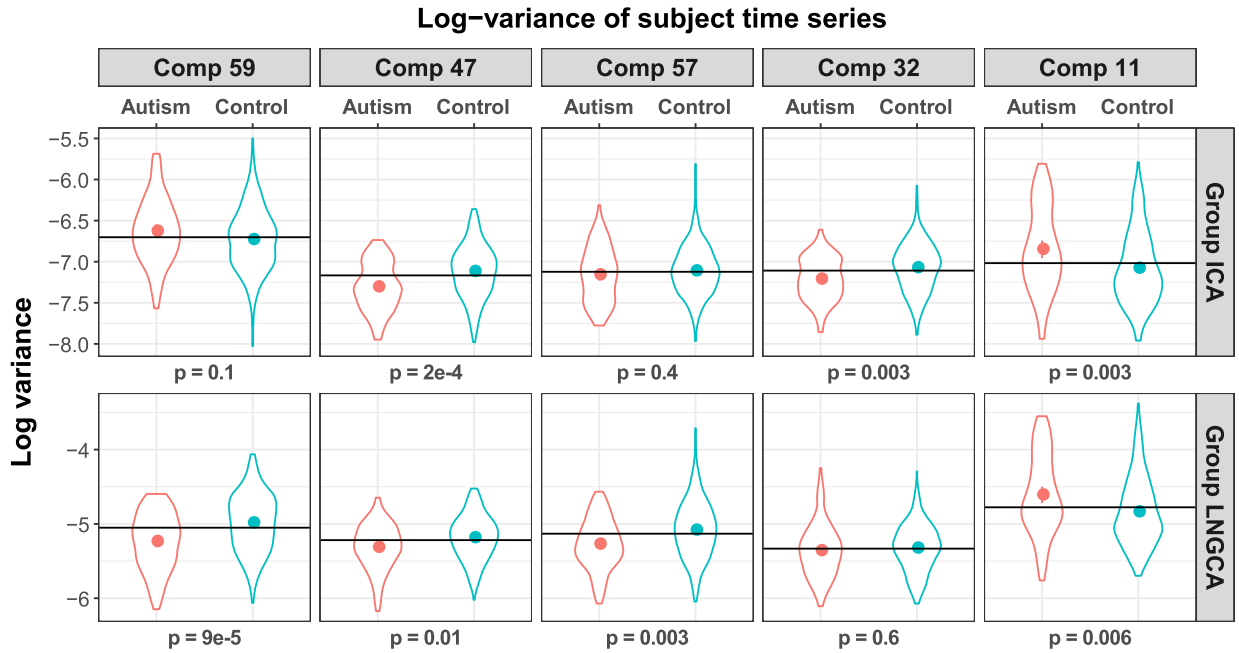
Fig. 5. Five components with significantly different log variance between ASD and TD children. The corresponding spatial maps are shown in Fig. 6. 1) Group LNGCA component 59 (p = 9e−5, all p-values FDR corrected), its matched group ICA component 59 is not significant (p = 0.1); 2) group ICA component 47 (p = 2e−4), its matched group LNGCA component 47 is also significant (p = 0.01); 3) group LNGCA component 57 (p = 0.003), its matched group ICA component 57 is not significant (p = 0.4); 4) group ICA component 32 (p = 0.003), its matched group LNGCA component is not significant (p = 0.6); 5) group ICA component 11 (p = 0.003), its matched group LNGCA component is also significant (p = 0.006). Horizontal line marks the mean across all children.

For the group-level PCA, we retain 59 components for both group LNGCA and GIFT to make fair comparison. As mentioned above, 59 is the maximum estimated number of PCs across all subjects.

4.2. Group ICA and group LNGCA

For group ICA, participant-specific PCs were temporally concatenated and a second PCA was performed to extract group-level PCs. Noise-free ICA was repeated on the group-level PCs 100 times using the Infomax algorithm (Bell and Sejnowski, 1995) and the ICASSO toolbox (Himberg et al., 2004) with randomized initial conditions in GIFT. Group components from GIFT were labeled using ICs from (Allen et al., 2011).

For group LNGCA, we used 40 randomized initializations for each subject-level LNGCA and 100 in the group-level ICA. We used the logistic (Infomax) non-linearity. The main computational expense is the subject-level LNGCA (Step 1 in Algorithm 1). We run the 40 random starts for each subject in parallel using 20 cores on a high-performance computing cluster, which takes 35-40 minutes for each subject and is not memory intensive for our datasets ($T = 128$ or 156). We performed the SVD (Step 2 in Algorithm 1) using power iterations on a laptop with 8 Gb of RAM. We matched the components in GIFT and group LNGCA using the Hungarian algorithm. The majority of components were very similar, with correlations greater than 0.8; twenty components had correlations less than 0.8, and of those twenty, five had correlations less than 0.5.

The variance of subjects' time series corresponding to extracted group components may reveal different levels of intrinsic activity between children with autism and healthy controls. For a given subject, it represents the energy that a component exhibits. If a component is orthogonal to the subjects' data, then the variance will be zero. If a component is more intrinsically active in children with ASD compared to TD children, then we expect the variance of that component to be higher on average in the former. For each estimated group component (59 from group LNGCA and 59 from GIFT), we applied t-tests to the log variance of the corresponding subject time series between two groups after controlling the age and sex. After adjusting the p-values for 118 hypothesis tests using FDR control, we found six tests were significant at alpha=0.01: ordered by their p-values, component 59 in group LNGCA (p = 9e−5, all p-values FDR corrected), component 47 in group ICA (p = 2e−4), component 57 in group LNGCA (p = 3e−3), component 32 in group ICA (p = 3e−03), component 11 in group ICA (p = 3e−03), and component 11 in group LNGCA (p = 6e−03). We report the log-variance of subjects' time series of these components matched across methods in Fig. 5, which depicts means +/- 2SE and violin plots of the densities for each group.

In Fig. 6, we see that component 59 differs greatly between the two methods, component 47 is highly similar, components 57 and 32 have some differences between methods, and component 11 appears similar. Component 59 is particularly poorly matched across methods (cross method correlation = .26); the group ICA version of this component was labeled as
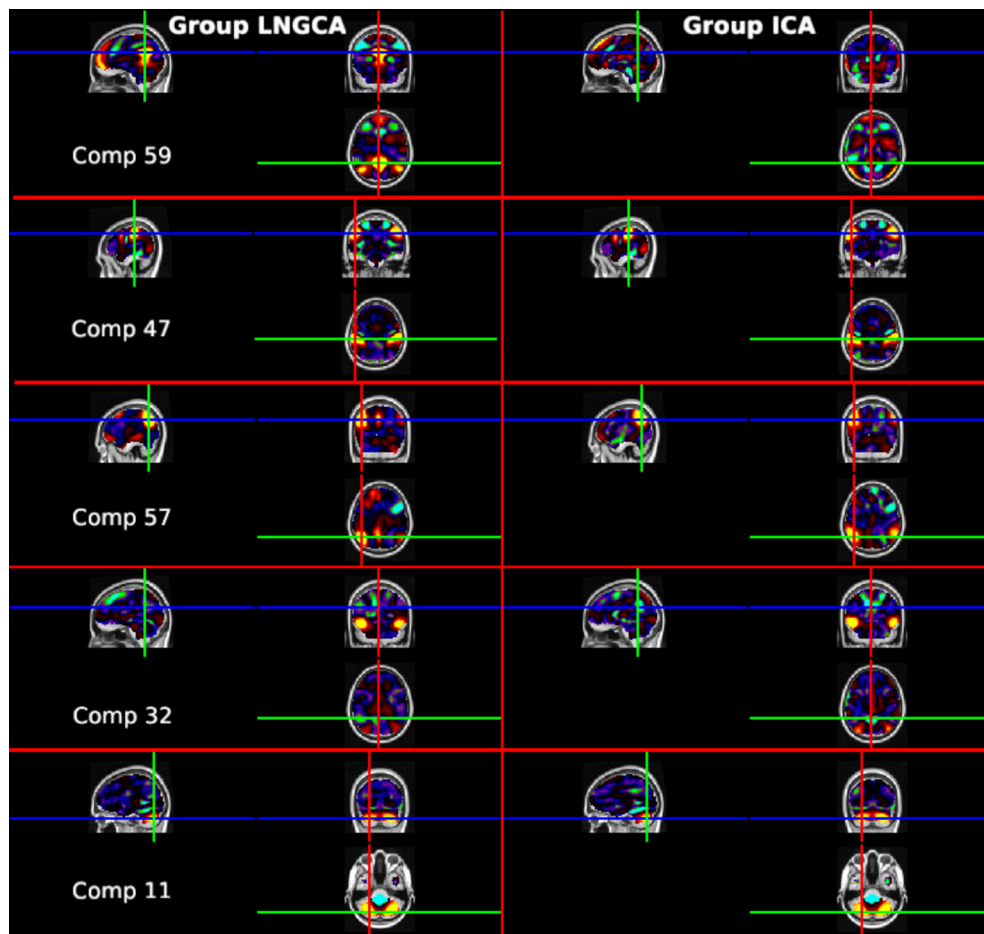
Fig. 6. Comparison of group LNGCA (left) and matched group ICA components (right) associated with the five most significant ASD versus TD differences in log variance. References to significance and the direction of effects are from Fig. 5. Color bars are based on the second and ninety-eighth percentiles for the positive and negative values of each component. Top to bottom: Component 59: Group LNGCA characterizes the default mode network while the matched group ICA component may be related to motion (high activation near the edges), and significantly lower engagement in ASD than TD in log variance is only observed for group LNGCA (Fig. 5). Component 47: For both methods, this component is the frontoparietal network and is significantly less engaged in ASD vs TD children. Component 57: Contains portions of the default mode network for both methods but is more left-lateralized for group LNGCA, and we only observe significantly lower ASD engagement versus TD for group LNGCA. Component 32: Large weights in the inferior temporal gyrus from the task positive network in both methods, and the group ICA component also contains large negative values in the posterior cingulate cortex in the default mode network, including portions that were allocated to component 59 by group LNGCA. Significantly lower engagement in ASD is observed in group ICA but not group LNGCA. Component 11: Includes portions of the cerebellum and has significantly greater engagement in ASD in both methods. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

artifact, while the group LNGCA version is more easily identified as part of the DMN. The group LNGCA result suggests that component 59 is less intrinsically active in the ASD group than the TD group (Fig. 5 bottom right). Extant ASD research (Padmanabhan et al., 2017) has implicated disruptions in intrinsic activity within the DMN, which purportedly contributes to ASD-associated difficulties integrating information about the self in the context of others. The group difference we observed was consistent across methods for component 47. The spatial representations of this component were very similar across group ICA and group LNGCA (correlation=0.91, Fig. 6 middle panel). Component 47, which was labeled as belonging to the frontoparietal network, was less engaged in the ASD group than in the TD group using estimates from both group ICA and group LNGCA (Fig. 5 left). This finding is also consistent with existing literature (Yerys et al., 2019) implicating reduced connectivity within the frontoparietal network in children with ASD. The spatial representation of component 57 is moderately similar between the two methods (cross method correlation = .62) and also appears to be driven by regions of the DMN, but the group LNGCA version is slightly more left-lateralized. Consistent with the group LNGCA result for component 59, the log variances of the subject-specific time courses for component 57 suggest it is less intrinsically engaged in the ASD group compared to the typically developing group, as shown in Fig. 5. Component 32 contains portions of the task positive network in the inferior temporal gyrus, and additionally, the group ICA component contains prominent negative weights in the posterior cingulate cortex of the default mode network. The log variance in component 32 was significantly lower in group ICA but not group LNGCA. Arguably, the large weights in the posterior cingulate cortex highlighted in group LNGCA component 59 conform better with a distinct default mode network, rather than the combined task positive and default

mode network in group ICA component 32. Finally, the cerebellum had large weights in component 11, and was significant in both methods. We also examined subject-specific (individual) components (Web Appendix Fig. S.10). The subject specific components include motion-related artifacts characterized by activation near the edge of the brain (first and third rows), as well as other scanner artifacts. Recall subject specific components are not part of the group ICA model.

5. Discussion

We propose a method to extract group non-Gaussian components from hundreds of subjects. We demonstrate in simulations that our method can extract low variance features that are discarded using group ICA. Our method involves a first-stage LNGCA for each subject. In this stage, we present a novel test of the number of non-Gaussian components in the presence of spatially correlated noise that improves upon methods assuming uncorrelated noise, which dramatically overestimate the dimensionality. We apply group LNGCA to an rsfMRI study and discover components that exhibit different levels of activity in children with ASD as compared with typically developing children. This provides an example of how information extracted using non-Gaussianity in group LNGCA may be relevant to neurological disorders. Group LNGCA component 59 explained a median of 1% (range: 0-2%) of the variance in each participant. Despite this small variance, it had the smallest p-value in Fig. 5 and was not extracted by group ICA. In general, methods emphasizing non-Gaussianity in dimension reduction can be useful in neuroimaging studies.

Group LNGCA divides each subject's components into group, individual, and Gaussian noise components. Current group ICA methods generally calculate the subject time courses using the group components, including back projection and dual regression, and these time courses can then be used to re-calculate subject versions of the group components (e.g., Erhardt et al., 2011). Subject-specific deviations of group components are modeled using random effects in hierarchical ICA (Guo and Tang, 2013). Unlike these other approaches, the subject-level (individual) components are orthogonal to the group components in group LNGCA. In our data example, the individual components include subject-level artifacts, and the subject-level LNGCA step appears to separate artifacts from neuronally related components. In theory, our algorithm can allow subject-specific deviations to be captured by individual components. We illustrated this in Web Appendix Fig. S.9. This is related to partially shared structure, for example, as modeled in Gaynanova and Li (2019). As a consequence, q_i (the rank of the non-Gaussian subspace for the ith subject) may not be equal to q_g (the number of group components) plus the rank of the subject's individual subspace. In our Algorithm, Step 2 averages "noisy" estimates of the group components. We suggest this will tend to average out subject differences and allocate those differences to individual components. Our data example involved large voxel sizes and spatial smoothing, which may reduce subject-specific deviations. Moreover, the scan times were relatively short (5.33-6.5 minutes), and longer scans may be necessary to uncover individual differences in rsfMRI studies (Gordon et al., 2017). Future research should examine whether subject-specific deviations are captured in individual components in higher spatial resolution and longer rsfMRI scans.

Lower dimensional PCA in PCA+ICA can mistakenly aggregate features, as examined in Risk et al. (2019), although here we have compared LNGCA to PCA+ICA with relatively high dimensional PCA. In some respects, our formulation is more flexible since it can detect components unique to one subject. Our framework is related to Joint and Individual Variation Explained (Lock et al., 2013) and Common Orthogonal Basis Extraction (COBE) (Zhou et al., 2015). Individual subspaces in rsfMRI improved the prediction of behavior (Kashyap et al., 2019). Ideally, one would simultaneously estimate all individual and group signals, but this is computationally infeasible with LNGCA. We subtract the projected group components treating the estimated non-Gaussian subject subspace as fixed, and then conduct LNGCA to estimate individual components, as in Algorithm 1. Alternatively, one could modify Step 4 and formulate an objective function treating the group components as fixed while estimating the non-Gaussian subspace from the full data, which would result in a new estimate of the individual subspace that may improve estimation of individual components. These are promising directions for future research in group LNGCA.

In GIFT, there is arguably no clear guidance how to select the number of PCs for both the subject and group level. For the subject-level, it is often selected as many as the computation can afford, but not smaller than the estimated number from the MDL-based method, which assumes Gaussianity. In group LNGCA, our proposed test can be used to determine the number of components to be kept for the subject level. Determining the number of group components is unresolved in the group ICA literature and is beyond the scope of this work. Methods from JIVE could be examined to determine whether subject-level non-Gaussian components are noisy estimates of group components (Lock et al., 2013; Feng et al., 2018).

In our study, we use the variance of the subject-specific time course for each group component to summarize the participant's brain activity. Future research should examine temporal dynamics, which may provide additional insights into the autistic mind. Tang et al. (2021) propose a Bayesian latent factor model in which the presence or absence of subject-specific spatial loadings (what we call components) are governed by population-level probabilities, although applications to high spatial resolutions may be challenging. Their model allows the variances of the latent factors to change across time similar to volatility modeling of financial markets. As an alternative to working with the time series of locations or components, models for dynamic connectivity may use time varying correlation matrices between predefined regions of interest calculated for each time point using sliding windows (Ting et al., 2021b) or time varying adjacency matrices calculated by thresholding these correlations (Ting et al., 2021a). One possible advantage of group LNGCA is that it allows a study-specific estimation of voxels that co-activate, which are described by the non-Gaussian components. This can be especially useful in studies with children, such as ours, since regions of interest defined using conventional atlases are

based on adult brains. The time courses of the group LNGCA components could then be utilized to generate the connectivity matrices and input into the aforementioned models of dynamic connectivity.

Lastly, our application to resting-state fMRI may be a conservative illustration of the difference between group ICA and group LNGCA. We chose 85 PCs and 85 NGs for group ICA and group LNGCA, respectively, which is higher than suggested by GIFT's criterion, and hence is higher than would be used in many neuroimaging studies. Even with this conservative approach, we find important differences in resting-state networks. Most notably, the component 59 showed significantly greater activity in the ASD versus typically developing group in group LNGCA, but these differences were not detected using the matched component from group ICA. We hypothesize that greater dimension reduction in the subject-level analyses would lead to larger differences in the estimated resting-state networks.

Acknowledgements

Appendix A. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.csda.2022.107454.

References

Allen, E.A., Erhardt, E.B., Damaraju, E., Gruner, W., Segall, J.M., Silva, R.F., Havlicek, M., Rachakonda, S., Fries, J., Kalyanam, R., et al., 2011. A baseline for the multivariate comparison of resting-state networks. Front. Syst. Neurosci. 5 (2).

Beckmann, C.F., 2012. Modelling with independent components. NeuroImage 62 (2), 891–901.

Beckmann, C.F., Mackay, C.E., Filippini, N., Smith, S.M., 2009. Group comparison of resting-state fmri data using multi-subject ica and dual regression. NeuroImage 47 (Suppl 1), S148.

Beckmann, C.F., Smith, S.M., 2004. Probabilistic independent component analysis for functional magnetic resonance imaging. IEEE Trans. Med. Imaging 23 (2), 137–152.

Bell, A.J., Sejnowski, T.J., 1995. An information-maximization approach to blind separation and blind deconvolution. Neural Comput. 7 (6), 1129–1159.

Bickel, P.J., Kur, G., Nadler, B., 2018. Projection pursuit in high dimensions. Proc. Natl. Acad. Sci. 115 (37), 9151–9156.

Braga, R.M., Buckner, R.L., 2017. Parallel interdigitated distributed networks within the individual estimated by intrinsic functional connectivity. Neuron 95 (2), 457–471.

Calhoun, V.D., Adali, T., Pearlson, G.D., Pekar, J., 2001. A method for making group inferences from functional mri data using independent component analysis. Hum. Brain Mapp. 14 (3), 140–151.

Calhoun, V.D., Eichele, T., Pearlson, G., 2009. Functional brain networks in schizophrenia: a review. Front. Human Neurosci. 3 (AUG), 17.

Cardoso, J.-F., 1989. Source separation using higher order moments. In: International Conference on Acoustics, Speech, and Signal Processing. IEEE, pp. 2109–2112.

Chen, Z., Calhoun, V., 2018. Effect of spatial smoothing on task fMRI ICA and functional connectivity. Front. Neurosci. 12 (FEB), 15.

Coalson, T.S., Van Essen, D.C., Glasser, M.F., 2018. The impact of traditional neuroimaging methods on the spatial localization of cortical areas. Proc. Natl. Acad. Sci. USA 115 (27), E6356–E6365.

Correa, N., Adalı, T., Calhoun, V.D., 2007. Performance of blind source separation algorithms for fmri analysis using a group ica method. Magn. Reson. Imaging 25 (5), 684–694.

Du, Y., Fan, Y., 2013. Group information guided ica for fmri data analysis. NeuroImage 69, 157–197.

Du, Y., Fu, Z., Sui, J., Gao, S., Xing, Y., Lin, D., Salman, M., Rahaman, M.A., Abrol, A., Chen, J., et al., 2019. Neuromark: a fully automated ica method to identify effective fmri markers of brain disorders. medRxiv, p. 19008631.

Du, Y., Pearlson, G.D., Lin, D., Sui, J., Chen, J., Salman, M., Tamminga, C.A., Ivleva, E.I., Sweeney, J.A., Keshavan, M.S., et al., 2017. Identifying dynamic functional connectivity biomarkers using gig-ica: application to schizophrenia, schizoaffective disorder, and psychotic bipolar disorder. Hum. Brain Mapp. 38 (5), 2683–2708.

Eloyan, A., Crainiceanu, C.M., Caffo, B.S., 2013. Likelihood-based population independent component analysis. Biostatistics 14 (3), 514–527.

Erhardt, E.B., Rachakonda, S., Bedrick, E.J., Allen, E.A., Adali, T., Calhoun, V.D., 2011. Comparison of multi-subject ica methods for analysis of fmri data. Hum. Brain Mapp. 32 (12), 2075–2095.

Feng, Q., Jiang, M., Hannig, J., Marron, J., 2018. Angle-based joint and individual variation explained. J. Multivar. Anal. 166, 241–265.

Gaynanova, I., Li, G., 2019. Structural learning and integrative decomposition of multi-view data. Biometrics 75 (4), 1121–1132.

Gordon, E.M., Laumann, T.O., Gilmore, A.W., Newbold, D.J., Greene, D.J., Berg, J.J., Ortega, M., Hoyt-Drazen, C., Gratton, C., Sun, H., Hampton, J.M., Coalson, R.S., Nguyen, A.L., McDermott, K.B., Shimony, J.S., Snyder, A.Z., Schlaggar, B.L., Petersen, S.E., Nelson, S.M., Dosenbach, N.U., 2017. Precision functional mapping of individual human brains. Neuron 95 (4), 791–807.

Guo, Y., 2011. A general probabilistic model for group independent component analysis and its estimation methods. Biometrics 67 (4), 1532–1542.

Guo, Y., Tang, L., 2013. A hierarchical model for probabilistic independent component analysis of multi-subject fmri studies. Biometrics 69 (4), 970–981.

Hastie, T., Tibshirani, R., 2003. Independent components analysis through product density estimation. Adv. Neural Inf. Process. Syst. 15, 649–656.

Himberg, J., Hyvärinen, A., Esposito, F., 2004. Validating the independent components of neuroimaging time series via clustering and visualization. NeuroImage 22 (3), 1214–1222.

Hyvarinen, A., 1999. Fast and robust fixed-point algorithms for independent component analysis. IEEE Trans. Neural Netw. 10 (3), 626–634.

Jin, Z., Risk, B.B., Matteson, D.S., 2017. Optimization and testing in linear non-Gaussian component analysis. Stat. Anal. Data Min. ASA Data Sci. J.

Kashyap, R., Kong, R., Bhattacharjee, S., Li, J., Zhou, J., Yeo, B.T., 2019. Individual-specific fmri-subspaces improve functional connectivity prediction of behavior. NeuroImage 189, 804–812.

Li, Y.-O., Adalı, T., Calhoun, V.D., 2007. Estimating the number of independent components for functional magnetic resonance imaging data. Hum. Brain Mapp. 28 (11), 1251–1266.

Lock, E.F., Hoadley, K.A., Marron, J.S., Nobel, A.B., 2013. Joint and individual variation explained (jive) for integrated analysis of multiple data types. Ann. Appl. Stat. 7 (1), 523.

Mejia, A.F., Nebel, M.B., Wang, Y., Caffo, B.S., Guo, Y., 2019. Template independent component analysis: targeted and reliable estimation of subject-level brain networks using big data population priors. arXiv preprint. arXiv:1906.07294.

Minka, T.P., 2001. Automatic choice of dimensionality for pca. In: Advances in Neural Information Processing Systems, pp. 598–604.

Nordhausen, K., Oja, H., Tyler, D.E., Virta, J., 2017a. Asymptotic and bootstrap tests for the dimension of the non-Gaussian subspace. IEEE Signal Process. Lett. 24 (6), 887–891.

Nordhausen, K., Oja, H., Tyler, D.E., Virta, J., 2017b. Ictest: estimating and testing the number of interesting components in linear dimension reduction. URL: https://CRAN. R-project.org/package= ICtest. R package version 0.3.

Padmanabhan, A., Lynch, C.J., Schaer, M., Menon, V., 2017. The default mode network in autism. Biological Psychiatry: Cogn. Neurosci. Neuroimaging 2 (6), 476–486.

Plis, S.M., Hjelm, D.R., Salakhutdinov, R., Allen, E.A., Bockholt, H.J., Long, J.D., Johnson, H.J., Paulsen, J.S., Turner, J.A., Calhoun, V.D., 2014. Deep learning for neuroimaging: a validation study. Front. Neurosci. 8, 229.

Risk, B., Gaynanova, I., 2021. Simultaneous non-Gaussian component analysis (SING) for data integration in neuroimaging. Ann. Appl. Stat. 15 (3), 1431–1454.

Risk, B.B., Matteson, D.S., Ruppert, D., 2019. Linear non-Gaussian component analysis via maximum likelihood. J. Am. Stat. Assoc. 114 (525), 332–343.

Risk, B.B., Matteson, D.S., Ruppert, D., Eloyan, A., Caffo, B.S., 2014. An evaluation of independent component analyses with an application to resting-state fmri. Biometrics 70 (1), 224–236.

Tang, M., Ting, C.-M., Ombao, H., 2021. BICNet: A Bayesian approach for estimating task effects on intrinsic connectivity networks in fMRI data. arXiv.

Ting, C.M., Samdin, S.B., Tang, M., Ombao, H., 2021a. Detecting dynamic community structure in functional brain networks across individuals: a multilayer approach. IEEE Trans. Med. Imaging 40 (2), 468–480.

Ting, C.-M., Skipper, J.I., Noman, F., Small, S.L., Ombao, H., 2021b. Separating stimulus-induced and background components of dynamic functional connectivity in naturalistic fMRI. IEEE Trans. Med. Imaging.

Welvaert, M., Durnez, J., Moerkerke, B., Verdoolaege, G., Rosseel, Y., 2011. neurosim: an r package for generating fmri data. J. Stat. Softw. 44 (10), 1–18.

Yerys, B.E., Tunç, B., Satterthwaite, T.D., Antezana, L., Mosner, M.G., Bertollo, J.R., Guy, L., Schultz, R.T., Herrington, J.D., 2019. Functional connectivity of frontoparietal and salience/ventral attention networks have independent associations with co-occurring attention-deficit/hyperactivity disorder symptoms in children with autism. Biological Psychiatry: Cogn. Neurosci. Neuroimaging 4 (4), 343–351.

Zhou, G., Cichocki, A., Zhang, Y., Mandic, D.P., 2015. Group component analysis for multiblock data: common and individual feature extraction. IEEE Trans. Neural Netw. Learn. Syst. 27 (11), 2426–2439.

Zhu, M., Ghodsi, A., 2006. Automatic dimensionality selection from the scree plot via the use of profile likelihood. Comput. Stat. Data Anal. 51 (2), 918–930.