Likelihood Component Analysis

author here

September 29, 2015

Abstract

Independent component analysis (ICA) is popular in many applications, including cognitive neuroscience and signal processing. Due to computational constraints, principal component analysis is used for dimension reduction prior to ICA (PCA+ICA), which could remove important information. The problem is that interesting independent components (ICs) could be mixed in several principal components that are discarded and then these ICs cannot be recovered. To address this issue, we propose likelihood component analysis (LCA), a novel methodology in which dimension reduction and latent variable estimation is achieved simultaneously by maximizing a likelihood with Gaussian and non-Gaussian components. We present a parametric LCA model using the logistic density and a semi-parametric LCA model using tilted Gaussians with cubic B-splines. We implement an algorithm scalable to datasets common in applications (e.g., hundreds of thousands of observations across hundreds of variables with dozens of latent components). In simulations, our methods recover latent components that are discarded by PCA+ICA methods. We apply our method to dependent multivariate data and demonstrate that LCA is a useful data visualization and dimension reduction tool that reveals features not apparent from PCA or PCA+ICA. We also apply our method to an experiment from the Human Connectome Project with state-of-the-art temporal and spatial resolution and identify an artifact using LCA that was missed by PCA+ICA. We present theoretical results on identifiability of the LCA
model and consistency of our estimator. Proofs are in a web supplement.

Keywords: Functional Magnetic Resonance Imaging, Independent Component Analysis, Neuroimaging, Non-Gaussian Component Analysis, Principal Component Analysis, Projection Pursuit

1 Introduction

The basic independent component analysis (ICA) model is \( X = MS \) where \( X \) is an observed vector, \( S \) is a latent vector of independent random variables, and \( M \) is a square matrix called the mixing matrix. It is assumed that we have a sample \( X_1, \ldots, X_V \) with corresponding latent \( S_1, \ldots, S_V \). The goal is to estimate \( M \) and to recover \( S_1, \ldots, S_V \). Except for Matteson and Tsay (2013), ICA methodology does not directly attempt to find components that are independent but rather components that are as non-Gaussian as possible. The principle here is that any sum of ICs will be closer to Gaussian distributed than the ICs themselves. Thus, the \( S_v \) are correctly recovered if they maximize some measure of non-Gaussianity.

Transformations that maximize non-Gaussianity play a prominent role in many applications including separating audio recordings in signal processing (Bell and Sejnowski, 1995), denoising in image processing (Hyvärinen et al., 1999), face recognition in computer learning (Bartlett et al., 2002), artifact removal in electrophysiology data (Delorme et al., 2007), and estimating brain networks in cognitive neuroscience (Beckmann, 2012). We propose a novel approach for modeling non-Gaussian signals and Gaussian noise that we call Likelihood Component Analysis (LCA). Consider a sample \((x_v, s_v, \nu_v), v = 1, \ldots, V\), from the LCA model:

\[
X = M_S S + M_N N
\]

where \( X \in \mathbb{R}^T; S \in \mathbb{R}^Q \) is a vector of mutually independent non-Gaussian random variables with \( 1 \leq Q \leq T; M_S \in \mathbb{R}^{T \times Q}; M_N \in \mathbb{R}^{T \times (T-Q)}; M = [M_S, M_N] \) (the concatenation of
\( \mathbf{M}_S \) and \( \mathbf{M}_N \) is full rank; and \( \mathbf{N} \) is \((T - Q)\)-variate normal. One observes \( \mathbf{x}_1, \ldots, \mathbf{x}_V \) and \( s_1, \ldots, s_V \) and \( v_1, \ldots, v_V \) are latent. We assume \( \mathbf{E} \mathbf{S} = \mathbf{0} \) and \( \mathbf{E} \mathbf{N} = \mathbf{0} \) in (1) (in practice, data are centered by their sample mean). Our goal is to estimate \( \mathbf{M}_S \) and the realizations \( s_1, \ldots, s_V \) of \( \mathbf{S} \), which we call likelihood components (LCs). By simultaneously performing dimension reduction and latent variable estimation, we will demonstrate through simulations and two real datasets that estimation of the proposed model allows the discovery of non-Gaussian signals discarded by other popular methods.

Classic independent component analysis (ICA) and principal component analysis followed by ICA (hereafter, PCA+ICA) are among the most commonly used models for extracting non-Gaussian signals. Unlike (1), the classic ICA model assumes the number of components is equal to the dimension of the observations such that \( \mathbf{M}_S \) is square and \( \mathbf{N} = \mathbf{0} \) (Hyvärinen and Oja, 2000). In practice, PCA is applied to the observations \( \mathbf{x}_1, \ldots, \mathbf{x}_V \) prior to classic ICA to meet the assumption of square mixing and to reduce computational costs (Hyvärinen et al., 2001). We will demonstrate that removing the smallest principal components (PCs) can discard the relevant signal (see also Green et al. 2002). Non-Gaussian signals are often discarded by PCA+ICA when they are associated with small variance. When the motivating scientific problem has a low signal-to-noise ratio, we will demonstrate that LCA is particularly well-suited to recovering the non-Gaussian signals.

PCA+ICA is commonly used to identify brain networks and artifacts in neuroimagery (Beckmann, 2012). In fMRI, the blood oxygen level dependence (BOLD) signal is measured across time at thousands of voxels (three-dimensional analogue of a pixel) across the brain. ICA of fMRI requires dimension reduction via PCA prior to the application of ICA. It is believed that ICA can ‘unmix’ the BOLD signal to reveal the underlying functional architecture of the brain. The existence and importance of these networks has been corroborated by other neuroimaging modalities and by the application of other statistical methods (Sporns, 2011). Additionally, ICA is commonly used for artifact removal in electroencephalography (EEG) and fMRI. Independent components (ICs) are identified that correspond to physiolog-
ical noise and/or motion, and accounting for these artifacts can improve subsequent analyses (Griffanti et al., 2014; Delorme et al., 2007). Even though the results from the two-stage PCA+ICA approach have been useful in the applied sciences, as we show in an example, a single analysis that uses non-Gaussianity for both dimension reduction and extracting LCs can provide novel insight.

As an alternative to classic ICA, the noisy-ICA model posits that the number of noise components is equal to the dimension of the data and typically assumes isotropic noise: $\mathbf{M}_N \mathbf{N} \sim N(0, \sigma^2 \mathbf{I}_T)$, where $\mathbf{I}_T$ is the $T \times T$ identity matrix. Beckmann and Smith (2004) propose a variant of PCA+ICA as an approximation to the noisy-ICA model, where they estimate the number of ICs and achieve dimension reduction using probabilistic PCA (Tipping and Bishop, 1999). Alternatively, independent factor analysis (IFA) could be used for simultaneous dimension reduction and latent variable estimation wherein the ICs are modeled as Gaussian mixtures (Attias, 1999). Allassonniere and Younes (2012) developed stochastic EM algorithms to estimate the IFA model and proposed a number of plausible parametric methods. Nonetheless, it is difficult to apply IFA to moderately sized datasets. Letting $m$ denote the number of elements in each Gaussian mixture and $Q$ the number of non-Gaussian components, an $m^Q$-dimensional integral must be approximated at each iteration, which quickly becomes computationally intractable (Allassonniere and Younes, 2012). Guo and Tang (2013) developed a multi-subject IFA model, although their application to big data utilizes PCA. Amato et al. (2010) develop non-parametric density estimators of the component densities in the noisy-ICA model but assume $\mathbf{M}_S$ is semi-orthogonal (has orthogonal columns), which is not realistic for our application.

There are a number of other methods that explore structure in multivariate data using non-Gaussianity. Non-Gaussian measures of information such as kurtosis were first explored in projection pursuit algorithms (Huber, 1985), which sequentially extract “interesting” directions of information using a fixed projection pursuit index. Miettinen et al. (2014) developed the deflationary FastICA algorithm to adaptively select a parametric projection
pursuit index for each non-Gaussian direction; however, their method assumes $Q = T$. Non-Gaussian component analysis (NGCA) is a variant of independent subspace analysis (Theis, 2006) which represents data using a Gaussian subspace and an independent non-Gaussian subspace, and the non-Gaussian subspace is estimated using multiple projection pursuit indices or radial basis functions (Kawanabe et al., 2007). However, NGCA does not model independent components, and thus does not lend itself to identifying brain networks and/or artifacts. The ICA model for the case when the number of components is less than the dimension of $X$ is sometimes called under-complete ICA (Amari, 1999) or over-determined ICA, and can be estimated using a natural gradient descent algorithm. Overall, the LCA model is unique in that it specifies a generative model for the non-Gaussian signal while also defining a subspace containing Gaussian noise.

In this paper, we present a method for simultaneous dimension reduction and latent variable extraction that uncovers features that are not detected using current models. In Section 2, we define conditions for the identifiability of the LCA model in (1) and propose parametric and semi-parametric estimators. In Section 3, we investigate simulations when the observations of the latent variables are independently and identically distributed. In Section 4, we examine model robustness by applying our method to temporally and spatially structured simulated data. In Section 5, we use LCA for data visualization and dimension reduction in multivariate data. In Section 6, we estimate brain networks and artifacts from high-resolution fMRI data from the Human Connectome Project. In Section 7, we present our conclusions and discuss avenues for future research. Code implementing our method and proofs of the theorems are in the Web Supplement.
2 Methodology

2.1 Model identifiability

The identifiability of the LCA model can be established using the theorem on the uniqueness of decomposition of the “linear structure model” described in Kagan et al. (1973). Throughout this section we will assume all random variables are mean zero. Define the equivalence relation $B \cong C$ if $B$ equals $C$ up to scaling and permutation of columns. The following theorem can be established using Theorem 10.3.9 in Kagan et al. (1973).

**Theorem 1.** Let $X = M_S S + M_N N$, where $M_S \in \mathbb{R}^{T \times Q}$, the elements of $S$ are mutually independent non-Gaussian components, $M_N \in \mathbb{R}^{T \times (T-Q)}$, $N$ is $(T-Q)$-variate normal, and $[M_S, M_N]$ is full rank. Then for any other representation $X = M'_S S' + E'$ where $S' \in \mathbb{R}^Q$ are independent non-Gaussian components and $E'$ is multivariate normal, we have: $M'_S \cong M_S$; $S' \distr S$ up to scaling and permutations; and $E' \distr M_N N$.

All proofs are in the Web Supplement.

From Theorem 1, the signal has a unique decomposition (on the equivalence class of scalings and permutations) into a fixed matrix and independent components. Note that the noise does not have a unique decomposition (e.g., if $N$ comprises independent normals with equal variance, then $M_N O'$ and $O'N$ for orthogonal $O$ is another decomposition with independent components). Let $S = [S_1, \ldots, S_Q]'$. We state the assumptions of the LCA model below.

**Assumption 1.** Assume that model in (1) holds with

(i) $S_1, \ldots, S_Q$ mutually independent, non-Gaussian random variables with $E S = 0$ and $E S S^T = I_Q$.

(ii) $\text{rank}([M_S, M_N]) = T$.

(iii) $N$ is $(T-Q)$-variate (non-degenerate) normal with $E N = 0$. 

6
Without loss of generality, we will assume that $N$ is standard multivariate normal. Let $f_1, \ldots, f_Q$ be the true densities of the LCs (the signal components). For the purposes of this paper, we will also assume $f_1, \ldots, f_Q$ are absolutely continuous, although identifiability holds more generally. Denote the eigenvalue decomposition (EVD) of the covariance matrix of $X$ by $\Sigma = U\Lambda U'$. Let $L = U\Lambda^{-1/2}U'$ be a whitening matrix (the covariance matrix of $L^{-1}X$ is $I_T$), and define $W = M^{-1}L^{-1}$ where $M = [M_S, M_N]$. Note that $W \in O_{T \times T}$, where $O_{T \times T}$ is the class of $T \times T$ orthogonal matrices. Let $w'_q$ denote the $q$th row of $W$, and let $W_S$ denote the first $q$ rows. Let $\phi(x)$ denote the standard normal density. Noting that $|\det W| = 1$, we have

$$f_X(x|W, L) = \det(L) \prod_{q=1}^{Q} f_q(w'_q Lx) \prod_{k=1}^{T-Q} \phi(w'_{k+Q} Lx).$$

(2)

We can now establish the identifiability of the LCA model.

**Corollary 1.** Suppose the linear structure model in (1) with density defined in (2) and suppose that Assumption 1 holds. Then the densities $f_1, \ldots, f_Q$ are identifiable, and the vectors $w_q$ for $q = 1, \ldots, Q$ are identifiable up to sign (and exact if the density is asymmetric). Note that the ordering of $f_q$ and $w_q$ is not identifiable, nor are the rows $w_{k+Q}$ for $k = 1, \ldots, T - Q$.

All proofs are in the Web Supplement.

### 2.2 The general LCA estimator

Now let $x_1, \ldots, x_V$ be an iid sample of $X$. Since $\mathbf{E}X = \mathbf{0}$, we will demean the data so that $\sum_{v=1}^{V} x_v = 0$, and assume such hereafter. Let $\mathbb{R}_+^{T \times T}$ denote the set of $T \times T$ positive definite matrices. Assume $V \geq T$. Let $\tilde{\Sigma}$ be the sample covariance matrix of $x_1, \ldots, x_V$. Consider its eigenvalue decomposition, $\tilde{\Sigma} = \tilde{U}\Lambda\tilde{U}'$. Then define $\hat{L} = \tilde{U}\Lambda^{-1/2}\tilde{U}'$. Note that
We have

\[ \sum_{v=1}^{V} o_q' \hat{L} x_v = 0. \]

Let \( \mathcal{O}_{Q \times T} \) be the class of \( Q \times T \) semi-orthogonal matrices. Define the estimator

\[ \hat{W}_s^{Tr} = \arg \max_{O_s \in \mathcal{O}_{Q \times T}} \sum_{v=1}^{V} \sum_{q=1}^{Q} \log f_q \left( o_q' \hat{L} x_v \right). \] (4)

The superscript “Tr” indicates that \( \hat{W}_s^{Tr} \) is computed when \( Q \) and the true component densities are known so \( \hat{W}_s^{Tr} \) is an oracle estimator that cannot be used in practice. In our Logis-LCA estimator, \( f_1, \ldots, f_Q \) are replaced by the logistic density, while in Spline-LCA we alternate between estimating \( f_1, \ldots, f_Q \) by tilted Gaussians with \( \hat{W}_s \) fixed and computing \( \hat{W}_s \) by (4) with \( f_1, \ldots, f_Q \) fixed at estimates.

Observe that the problem of estimating \( W_s \) is equivalent to the problem of estimating the LCs because \( \hat{s}_v = \hat{W}_s \hat{L} x_v \) for all \( v \). Thus we would like a consistent estimator of \( W_s \). Towards this, we have the following proposition.

**Proposition 1.** Consider a random vector \( Y \in \mathbb{R}^T \) with density \( f_Y \) such that \( E_Y = 0 \) and \( E YY' = I_T \). Then for any \( o \) and \( w \) such that \( o'o = w'w = 1 \), we have

\[ E \log \phi(o'Y) = E \log \phi(w'Y). \]

Next, we show the consistency of the oracle estimate of \( W_s \).

**Theorem 2.** Suppose \( X \) follows the LCA model in (1) with Assumption 1 holding and assume the non-Gaussian components have bounded absolutely continuous densities (satisfied by the classes considered below). Additionally assume \( E X = 0 \) and \( E XX' = I \) (here, \( W_s \) comprises the first \( Q \) rows of \( M^{-1} \)). Given an iid sample \( x_1, \ldots, x_V \), \( \hat{W}_s^{Tr} \rightarrow W_s \) almost surely on the equivalence class of signed permutations.
Since $\hat{W}_S$ is not invertible, we must also define an estimator of $M_S$:

$$\hat{M}_S = \arg\min_{B \in \mathbb{R}^{T \times Q}} \sum_{v=1}^{V} ||x_v - Bs_v||_2^2.$$  (5)

Although we assume iid observations in the construction of (4), the LCA model is capable of recovering many forms of dependent data, as is also the case in ICA. This will be demonstrated in simulations.

There is a natural ordering of the LCs when the component densities are not equal. Additionally, if the component densities are skewed, we can assume positive skewness and then the LCA model is fully identifiable (as in ICA, Eloyan and Ghosh 2013). We define a canonical ordering for a sample $x_1, \ldots, x_V$:

$$\sum_{v=1}^{V} f_1(w_1' Lx_v) > \sum_{v=1}^{V} f_2(w_2' Lx_v) > \cdots > f_Q(w_Q' Lx_v).$$

with

$$\sum_{v=1}^{V} (w_q' Lx_v)^3 > 0$$

for $q = 1, \ldots, Q$. In practice, we force the sample third moment to be positive and order components by their likelihoods.

### 2.3 A sign- and permutation-invariant measure for non-square matrices

To assess the accuracy of our estimates and/or compare multiple estimates, we need a discrepancy measure that is invariant on the equivalence class of signed permutation matrices, and we would like a measure that can apply to matrices of differing dimensions when the estimated number of components may not equal $Q$. We cannot use the Amari or the minimum distance (Ilmonen et al., 2010) measures because $M_S$ is non-square. We propose a novel
measure of dissimilarity that uses the Hungarian algorithm to match rows of the unmixing matrix as in Risk et al. (2014) but applies to non-square unmixing. We also generalize the measure to apply to matrices that may have a different number of columns, in which case the measure only compares matching columns. In addition to diagnosing accuracy, this measure will also be used to assess convergence in our algorithms.

Consider \( \mathbf{M}_1 \in \mathbb{R}^{T \times Q} \) and \( \mathbf{M}_2 \in \mathbb{R}^{T \times R} \) with \( Q \leq R \). Let \( \mathcal{P}_\pm \) be the class of \( R \times Q \) signed permutation matrices, so that postmultiplication of \( \mathbf{M}_2 \) by \( \mathbf{P}_\pm \in \mathcal{P}_\pm \) results in a subset of \( Q \) (permuted) columns of \( \mathbf{M}_2 \) for \( Q < R \). Define the permutation-invariant mean-squared error:

\[
PMSE(\mathbf{M}_1, \mathbf{M}_2) = \arg\min_{\mathbf{P}_\pm \in \mathcal{P}_\pm} ||\mathbf{M}_1 - \mathbf{M}_2 \mathbf{P}_\pm||_F^2, \tag{6}
\]

where \( || \cdot ||_F \) is the Frobenius norm and the optimal \( \mathbf{P}_\pm \) is found using the Hungarian algorithm. In practice, we also standardize the columns of \( \mathbf{M}_1 \) and \( \mathbf{M}_2 \) to have unit norm, and thus the measure is scale invariant. Another advantage of this measure is that it can be used to compare independent components directly. If \( \mathbf{S}_1 \) is a \( V \times Q \) matrix in which each row is a realization of the LC in \( \mathbb{R}^Q \), and if \( \mathbf{S}_2 \in \mathbb{R}^{V \times R} \), then we define their discrepancy as \( PMSE(\mathbf{S}_1, \mathbf{S}_2) \).

### 2.4 A parametric model: Logis-LCA

First we present a parametric method called Logis-LCA in which the densities of the LCs are assumed to be logistic. The logistic density is used in the Infomax algorithm, where it appears to work well for unmixing audio signals (Bell and Sejnowski, 1995) and brain networks (Correa et al., 2007). Under the constraint of zero mean and unit variance, the logistic density has the form

\[
f_\theta (x) = \frac{\exp \left(-x/\sqrt{\frac{3\pi}{\pi}}\right)}{\sqrt{\frac{3\pi}{\pi}} \left\{1 + \exp(-x/\sqrt{\frac{3\pi}{\pi}})\right\}^2}. \tag{7}
\]
We define our estimator for some $Q^* \leq T$ such that $Q^*$ may or may not equal the true number of LCs, $Q$. Applying (7) to (4) and the centered data $x_1, \ldots, x_V$, the Logis-LCA estimator of $W_S$ can be written as

$$
\hat{W}_S^L = \arg\max_{O_S \in O_{Q \times T}} - \sum_{v=1}^V \sum_{q=1}^{Q^*} \log \left\{ 1 + \exp \left( -o'_q \hat{L}x_v \frac{\pi}{\sqrt{3}} \right) \right\}.
$$

We maximize (8) using a modification of the symmetric fixed-point ICA algorithm (Hyvarinen, 1999). ICA implementations require the estimator to be a square matrix. We orthogonalize intermediate estimates of the rows of $W_S^L$ by calculating the SVD and setting the singular values equal to one. Additionally, we assess convergence using (6). See the Web Supplement for details.

2.5 A semi-parametric model: Spline-LCA

In this section, we use the flexible family of tilted Gaussian densities to model the LCs. The proposed model is equivalent to ProDenICA (Hastie and Tibshirani, 2003) when $Q = T$. For $Q < T$, it can be shown that the likelihood extends the semiparametric likelihood in Blanchard et al. (2006) to include an independence model for the LCs (see Web Supplement). The independence assumption is necessary for physically and biologically useful interpretations.

Suppose the LCs have tilted Gaussian distributions of the form $\phi(x)e^{g(x)}$, where $g(x)$ is some twice-differentiable function. Define the log-likelihood for some $O \in O_{T \times T}$:

$$
\ell(O, g_1, \ldots, g_{Q^*} | \hat{L}, Q^*, x_1, \ldots, x_V) = \sum_{v=1}^V \left[ \sum_{q=1}^{Q^*} \left\{ \log \phi(o'_q \hat{L}x_v) + g_q(o'_q \hat{L}x_v) \right\} + \sum_{k=1}^{T-Q^*} \log \phi(o'_{k+Q^*} \hat{L}x_v) \right].
$$
This log-likelihood does not have an upper bound, so we define a penalized log-likelihood:

$$
\ell(O, g_1, \ldots, g_{Q^*} \mid \hat{L}, Q^*, x_1, \ldots, x_V) = -\sum_{q=1}^{Q^*} \lambda_q \int \{g_q''(x)\}^2 dx - \int \phi(x)e^{g_q(x)} dx
$$

(9)

$$
+ \frac{1}{V} \sum_{v=1}^{V} \sum_{q=1}^{Q^*} g_q(o_q'\hat{L}x_v) - \frac{T}{2} (\log 2\pi + 1),
$$

where we have used (3) to simplify the Gaussian components.

Now consider the problem of estimating $g(x)$ for fixed $O$.

**Proposition 2.** Let $G$ be the class of all cubic splines $g : \mathbb{R} \rightarrow \mathbb{R}$. Consider the argmax of (9) for $g_q \in G$. Then (i) $\int \phi(x)e^{g_q(x)} dx = 1$ and (ii) $\int x\phi(x)e^{g_q(x)} dx = 0$ for each $q$.

We adapt the ProDenICA algorithm of Hastie and Tibshirani (2003) to LCA in which we alternate between estimating $W_S$ for fixed $\hat{f}_1, \ldots, \hat{f}_{Q^*}$ via the fixed point algorithm and estimating $f_1, \ldots, f_{Q^*}$ for fixed $\hat{W}_S$ using the “Poisson trick”. Our account largely follows the description in Hastie et al. (2009) but for semi-orthogonal (rather than orthogonal) matrices.

Suppose $W_S$ is given and define $s_{vq} = w_q'z_v$ where $z_v = \hat{L}x_v$. Let $x_1^*, \ldots, x_{L+1}^*$ define a discretization, $[x_1^*, x_2^*), [x_2^*, x_3^*), \ldots, [x_L^*, x_{L+1}^*)$, of the support of the tilt function of the non-Gaussian densities such that $\Delta = x_\ell^* - x_{\ell-1}^*$ for all $\ell = 2, \ldots, L + 1$. It suffices to take $x_1^* = \min(s_{11}, \ldots, s_{nd}) - 0.1\hat{\sigma}_z$ and $x_{L+1}^* = \max(s_{11}, \ldots, s_{nd}) + 0.1\hat{\sigma}_z$, where $\hat{\sigma}_z$ denotes the sample standard deviation, which here is equal to one. Next, let $x_\ell = \frac{1}{2}(x_\ell^* + x_{\ell+1}^*)$. For each $q \in \{1, \ldots, Q^*\}$ and $\ell \in \{1, \ldots, L\}$, define

$$
y_{\ell q} = \sum_{v=1}^{V} \mathbb{I}\{s_{vq} \in [x_\ell^*, x_{\ell+1}^*)\}.
$$

We approximate (9) by discretizing the first integral and estimating the sum over $V$ as a weighted sum over $L$. Restricting our attention to a single $q$, we have

$$
-\lambda_q \int \{g_q''(x)\}^2 dx + \sum_{\ell=1}^{L} \left[ \frac{y_{\ell q}}{V} \left\{ g_q(x_\ell) + \log \phi(x_\ell) \right\} - \Delta \phi(x_\ell)e^{g_q(x_\ell)} \right].
$$
Dividing by $\Delta$, we have

$$-\beta_q \int \left( \frac{d^2 g_q}{dx^2} \right)^2 dx + \sum_{\ell=1}^L \left[ \frac{y_{q\ell}}{V\Delta} \left( g_q(x_\ell) + \log \phi(x_\ell) \right) - \phi(x_\ell) e^{g_q(x_\ell)} \right]$$

for some penalty $\beta_q$. This is proportional to a Poisson generalized additive model (GAM), where $\frac{y_{q\ell}}{V\Delta}$ is the response and the expected response is equal to $\phi(x_\ell) e^{g_q(x_\ell)}$. This can be fit using the `gam` package in R (Hastie, 2013) where $\beta_q$ is chosen to result in a user-specified number of (approximate) degrees of freedom. We find that $df = 8$ and $L = 100$ produce fast and accurate density estimates in simulations for a variety of densities when the sample size is equal to 1,000. This method also easily scales to tens of thousands of observations.

The algorithm to estimate both $W_S$ and $f_1, \ldots, f_Q$ is summarized below. Note that step 3 requires the first and second derivatives of the log densities of the LCs, which makes the use of B-splines convenient.

**Algorithm 1:** The Spline-LCA algorithm.

**Inputs**: The whitened $V \times T$ data matrix $Z$; initial $W_S^0$; tolerance $\epsilon$.

**Result**: Estimates of the latent components, $\hat{S}$, and their densities, $\hat{f}_1, \ldots, \hat{f}_Q$.

1. Let $n = 0$ and define $\mathbf{S}^{(n)} = ZW_S^{(n)}$.
2. Estimate $f_q^{(n+1)}$ for $q = 1, \ldots, Q$.
3. Using (4), update $W_S^{(n+1)}$ given $f_1^{(n+1)}, \ldots, f_Q^{(n+1)}$ and $\mathbf{S}^{(n)}$ with one-step of the fixed-point algorithm (see Web Supplement).
4. Let $\mathbf{S}^{(n+1)} = ZW_S^{(n+1)}$.
5. If $PMSE(W_S^{(n+1)}, W_S^{(n)}) < \epsilon$, stop, else increment $n$ and repeat (2)-(4).
3 Simulations: Distributional and Noise-rank Assumptions

In this section, we simulate the LCA model [given by (1) with \( \mathbf{M}_S \in \mathbb{R}^{T \times Q} \)] and the noisy-ICA model [again given by (1) with \( \mathbf{M}_S \in \mathbb{R}^{T \times Q} \) but now with \( \mathbf{M}_N \mathbf{N} \sim N(0, \sigma^2 \mathbf{I}_T) \)] under a variety of source distributions in which the components are iid as well as a scenario in which the sources are sparse images. We compare (i) deflationary FastICA with the ‘log cosh’ nonlinearity (D-FastICA), where the deflation option estimates components one-by-one such that the algorithm is considered a projection pursuit method (Hyvärinen and Oja, 2000); (ii) two-class IFA with isotropic noise (IFA); (iii) PCA followed by Infomax (PCA+Infomax); (iv) PCA followed by ProDenICA (PCA+ProDenICA) (v) Logis-LCA; and (vi) Spline-LCA. We evaluate the robustness of these methods with respect to assumptions on the rank of the noise components, distribution of the latent components, and the signal-to-noise ratio (SNR). We define the SNR as the ratio of the total variance from the mixed non-Gaussian components to the total variance from the noise components. Formally, consider the non-zero eigenvalues \( \lambda_1, \ldots, \lambda_Q \) from the covariance matrix of \( \mathbf{M}_S \mathbf{S} \). For LCA, let \( \lambda_{e_1}, \ldots, \lambda_{e_{T-Q}} \) denote the eigenvalues from the EVD of the covariance matrix of \( \mathbf{M}_N \mathbf{N} \). Then,

\[
SNR = \frac{\sum_{q=1}^Q \lambda_q}{\sum_{k=1}^{T-Q} \lambda_{e_k}}.
\]

(11)

For the noisy-ICA model, we have \( T \) non-zero eigenvalues in the denominator summand.

We fit D-FastICA using the ‘deflation’ option in the fastICA function with \( T \) components from the fastICA R package (Marchini et al., 2010) and select the first \( Q \) components. We fit PCA+Infomax using our own implementation of the Infomax algorithm. We fit PCA+ProDenICA using the ProDenICA function from the R package of that name (Hastie and Tibshirani, 2010). Note that these methods can provide an estimate of \( \mathbf{S} \) but not the mixing matrix, which we estimated using (5). We fit the IFA model with two-component
mixtures of normals using our own implementation in which $M_S$ was estimated by maximizing the log likelihood using a numerical optimizer, and the ICs were estimated by their conditional means (see equation (81) in Attias 1999).

### 3.1 Simulation Design

Data were generated with $T = 5$ and $Q = 2$ according to a $2^2 \times 6$ full factorial design. The three factors were

i) **The model**: the levels were (a) the LCA model with rank-$(T - Q)$ noise and (b) the noisy-ICA model with rank-$T$ noise. In both models the signal was $M_S S$ where $M_S$ is $T \times Q$ with $Q \leq T$.

ii) **The signal to noise (SNR) ratio**: the levels were (a) high where the ratio of the variance from the signal components to the variance from the noise components was 5:1 and (b) low where that ratio was 1:5.

iii) **Signal distribution**: the levels were (a) logistic, (b) t, (c) Gumbel, (d) sub-Gaussian mixture of normals, (e) super-Gaussian mixture of normals, (f) with values determined by a sparse image, as described below. The two signal components were each iid and had the same distributions in cases (a)–(e) but different distributions in the sparse signal case.

Since we generated $Q = 2$ signal components for all simulations, there were $T - Q = 3$ and $T = 5$ noise components for the LCA model and noisy-ICA model, respectively. Observations in the noise components were iid isotropic normal except for the sparse image scenario, in which we used the R-package neuRosim (Welvaert et al., 2011) to generate three-dimensional Gaussian random fields with full width at half maximum (FWHM) equal to 6 for each noise component.

The signal components had scale parameter equal to $\sqrt{3}/\pi$ for the logistic, 5 degrees of freedom for the t, and scale parameter equal to $\sqrt{6}/\pi$ for the Gumbel. For the super-
Gaussian mixture of normals, we simulated a two-class model with the first centered at 0 with variance $2/3$ with probability 0.95 and the second centered at 5 with unit variance (excess kurtosis $\approx 9$), which is motivated by a brain network with 5% of voxels activated. For the sub-Gaussian mixture of normals, we used the two-class model with the first centered at $-1.7$ with unit variance and probability 0.75 and the second centered at 1.7 with unit variance and probability equal to 0.25, which is equivalent to distribution ‘l’ from Hastie and Tibshirani (2003) (excess kurtosis $\approx -0.3$). For the sparse image, we used neuRosim to generate a three-dimensional image in which all voxels were iid normal with variance equal to 0.0001 except, in the first component, a sphere of radius two in which the center was located at (5, 5, 5) with voxel-value equal to one and the exponential decay rate set to 0.5. The second sparse image component was similar except the feature was a cube centered at (7, 7, 7) with both radius and exponential decay rate equal to one.

We conducted 112 simulations with $V = 1,000$ observations in which $\mathbf{M}_S$ and $\mathbf{M}_N$ were randomly generated to have condition number between one and ten for each combination of factors. Since neither the set of orthogonal matrices (PCA+ICA methods) nor semi-orthogonal matrices (LCA methods) is convex, we approximated the argmax by initializing D-FastICA, PCA+Infomax, Logis-LCA, PCA+ProDenICA, and Spline-LCA from twenty random matrices and selecting the estimate associated with the largest objective function value. For Logis-LCA and Spline-LCA, ten of these twenty initializations were from random matrices constrained to the principal subspace. Let $\hat{\mathbf{U}}_{1:Q}$ denote the first $Q$ rows from $\hat{\mathbf{U}}$ in the decomposition $\hat{\Sigma} = \hat{\mathbf{U}} \hat{\Lambda} \hat{\mathbf{U}}'$. Then constraining the initial matrix, $\mathbf{W}_S^0$, to the principal subspace is equivalent to $\mathbf{W}_S^0 = \hat{\mathbf{U}}_{1:Q} \mathbf{O}$ with $\mathbf{O} \in \mathbf{O}_{Q \times Q}$. For IFA, one must specify initial values for the unmixing matrix, the variance of the isotropic noise, and the parameters of the Gaussian mixtures, and here we had four strategies to find the argmax. See Section C of the Web Supplement for additional details.
Figure 1: Boxplots of $PMSE$ for estimated columns of $S$ where the rank of the noise was $T-Q$ (LCA Model) or $T$ (Noisy-ICA Model) in high SNR (‘HI') and low SNR (‘LO') scenarios for various latent distributions. ‘DF’ = D-FastICA; ‘IFA’ = independent factor analysis; ‘PI’ = PCA+Infomax; ‘LL’ = Logis-LCA; ‘PP’ = PCA+ProDenICA; ‘SL’ = Spline-LCA.

### 3.2 Results

When the LCA model was true and there was a high SNR, all methods generally produced accurate estimates of $S$ for the logistic, t, Gumbel, super-Gaussian mixture of normals, and sparse images, but only Spline-LCA was accurate for the sub-Gaussian mixture of normals, and the performance of IFA was more variable than other methods for all distributions (Figure 1). In these simulations, boxplots examining the accuracy of $\hat{M}_S$ showed patterns similar to those found in Figure 1 and consequently are not presented.

When the LCA model was true and there was a low SNR, IFA, PCA+Infomax, and PCA+ProDenICA failed to recover the LCs for all distributions, while D-FastICA and Logis-LCA recovered all distributions except for the sub-Gaussian mixture of normals. Spline-LCA was the method most robust to distributional assumptions and was the only method that recovered the sub-Gaussian mixture.

However, when the noisy-ICA model was true and there was a high SNR, all methods
generally produced accurate estimates for the logistic, t, Gumbel, super-Gaussian, and sparse image. IFA and Spline-LCA were the only methods that recovered ICs with sub-Gaussian distributions. When the noisy-ICA model was true and there was a low SNR, all methods performed poorly, although IFA, PCA+Infomax, and PCA+ProDenICA outperformed LCA algorithms for some distributions.

Overall, LCA methods were robust to the SNR for rank-$(T - Q)$ noise, and performed well in the high SNR scenario for rank-$T$ noise. Additionally, Spline-LCA was most robust to distributional assumptions. In contrast, IFA, PCA+Infomax, and PCA+ProDenICA performed poorly in the low SNR scenario for both the rank-$(T - Q)$ and rank-$T$ noise.

4 Simulations: Spatio-temporal Networks

Next, we examine the ability of D-FastICA, PCA+Infomax, Logis-LCA, and Spline-LCA to recover simulated networks whose loadings vary deterministically with time in the presence of spatially and temporally correlated noise, so that the simulations resemble the structure found in task-based fMRI. We also examine the effect of using $Q^* \neq Q$ on network recovery. In this way, we assess whether the LCA algorithm can recover brain networks and their temporal loadings from spatiotemporal neuroimagery. We did not include IFA in these simulations because it was difficult to estimate the mixing matrix when $T$ was relatively large (e.g., $T = 50$). Additionally, IFA, PCA+Infomax, and PCA+ProDenICA produced similar results for most distributions in the previous simulations, and PCA+Infomax and PCA+ProDenICA were more accurate than IFA in the high SNR scenario for the LCA model. Hence, our previous simulations suggest there would be little insight gained from including IFA.
4.1 Simulation Design

We simulated three networks mixed across fifty time units. The networks were 33×33 images where “active” pixels were in the shape of a “1”, “2 2”, or “3 3 3” with values between 0.5 and 1, and “inactive” pixels were mean zero iid normal with variance equal to 0.0001 (see Figure 2). Let $m_q$ denote the $q$th column of $M_S$. To simulate the temporal activation patterns of brain networks, we used neuRosim to convolve the canonical hemodynamic response function (HRF) with a block-design with a pair of onsets at $\{1, 20.6\}$, $\{10.8, 40.2\}$, and $\{10.8, 30.4\}$ for $m_1$, $m_2$, and $m_3$, respectively, and duration equal to 5 time units (Welvaert et al., 2011).

In the LCA scenario, noise components were generated as forty-seven independent 33×33 Gaussian random fields with FWHM=6. Temporal correlation was introduced via the mixing matrix, in which each column of $M_N$ corresponded to an AR(1) process simulated for fifty time units with AR coefficient equal to 0.47 and unit variance, where the AR coefficient was chosen based on a preliminary analysis of the fMRI data analyzed in Section 6. Additionally, noise components were scaled such that the SNR was 0.4, which approximately equals the SNR estimated in Section 6. In the noisy ICA scenario, a 33×33 Gaussian random field with FWHM=6 was simulated for $t = 1$. Then noise components were defined recursively for $t = 2, \ldots, 50$ to be equal to 0.47 times the noise at time $t - 1$ plus a realization from an independent Gaussian random field with FWHM=6.

We conducted 111 simulations with $Q^* = 2, 3$ or 4 (with fixed $Q = 3$) and initialized all algorithms from twenty random mixing matrices for each simulation and each $Q^*$. For Logis-LCA and Spline-LCA, ten of the twenty initializations were from random matrices in the principal subspace, as in Section 3.1.

4.2 Results

By inspecting the images and loadings associated with the median $PMSE(\hat{S}, S)$ for each method in the LCA scenario, we see that D-FastICA recovers a spurious component when $Q^* = 3$; PCA+Infomax and PCA+ProDenICA generally fail to unmix features; and Logis-
LCA and Spline-LCA are highly accurate (Figure 2). It is notable that estimates from PCA+Infomax and PCA+ProDenICA were sensitive to the choice of $Q^*$, as when $Q^* < Q$, an estimated latent component resembled a union of components two and three. In PCA+ProDenICA, the loadings for the estimated component were highly correlated with component three ($r = 0.75$), which mistakenly suggests components two and three are functionally connected. For $Q^* = 3$, the features in the estimated component one are faintly visible in PCA+Infomax whereas component one was not recovered by PCA+ProDenICA. In contrast, Logis-LCA and Spline-LCA clearly separated components for all $Q^*$, such that when $Q^* < 3$, the recovered components were accurate estimates of a subset of the true ($Q = 3$) components.

For the noisy-ICA scenario, the features recovered by Logis-LCA most closely resembled the truth (Figure 3). Features from component two were again faintly visible in component three for $Q^* = 2$ in both PCA+Infomax and PCA+ProDenICA, again indicating inadequate unmixing of the networks. As seen in the LCA scenario, D-FastICA recovered a spurious component for $Q^* = 3$, but accurately estimated component three for $Q^* = 4$. Spline-LCA was sensitive to the assumption on the rank of the noise, as it failed to recover component one, although it was quite accurate for components two and three. Spatial correlations in the noise appear to result in spurious disk-like features, which were detected in Spline-LCA and D-FastICA. For the simulation associated with the median error, an accurate estimate of component one was associated with a local maxima in Spline-LCA, but the spurious component had a higher likelihood. The component was recovered in some simulations, and the lower quartile of the PMSE error was accurate (Web Supplement Figure 1).

5 Data Visualization and Dimension Reduction

We used Logis-LCA and Spline-LCA for data visualization and dimension reduction in multivariate data comprising measurements from independent leaf samples (Silva et al.,
Figure 2: Network recovery from the LCA scenario with $Q = 3$ for $Q^* = 2$ (first three columns), $Q^* = 3$ (columns 4-6), or $Q^* = 4$ (columns 7-10). Images depict LCs and time-series depict the loadings ($\hat{m}_1, \ldots, \hat{m}_Q$) corresponding to the median $PMSE(\hat{S}, S)$ from 111 simulations. In the last column, “Truth” corresponds to an arbitrary noise component whereas the algorithms attempted to estimate a fourth LC.
Figure 3: Network recovery from the noisy-ICA scenario with $Q = 3$ for $Q^* = 2$ (first three columns), 3 (columns 4-6), or 4 (columns 7-10).
Fourteen variables were generated from eight to sixteen images of leaves from each of thirty species (Web Supplement Figure 2). Many of the covariates are highly correlated (Web Supplement Figure 3). We plotted the first two PCs, ICs from PCA+Infomax and PCA+ProDenICA, and LCs from Logis-LCA and Spline-LCA. Two-dimensional PCA does not reveal clear features (Figure 4). Since we are examining two dimensions, the effect of ICA is apparent as a rotation of the X- and Y-axes. Rotating the axes does not reveal any additional insight (Figure 4, Web Supplement Figure 4). In contrast, Spline-LCA clearly reveals three categories, where the green dots correspond to two plant species that have very thin leaves (species 31 and 34 in Web Supplement Figure 2), the blue category corresponds to a species with leaves that are thinner than most species but less than those comprising the green dots (species 8), and the red category corresponds to all other species. Logis-LCA also reveals structure (Web Supplement Figure 4), although the separation is less than in Spline-LCA.

PCA+ICA methods were sensitive to the number of components estimated whereas components were robustly estimated in the LCA methods. In PCA+Infomax and PCA+ProDenICA, the first two (matched) ICs for $Q^* = 5$ differed from the ICs estimated using two components, demonstrating the sensitivity of PCA+ICA methods to the number of principal components (Web Supplement Figures 4 and 5). In contrast, the first two LCs extracted from Logis-LCA and Spline-LCA when five components were estimated were nearly equivalent to the LCs estimated using two components.

6 Application to fMRI

We applied Spline-LCA to a single subject from the Social Cognition / Theory of Mind (ToM) experiment of the MGH-UCLA Human Connectome Project (HCP; additional information in Web Supplement). For details of the experimental paradigm see Barch et al. (2013). We used the minimally pre-processed data (Glasser et al., 2013) from the first session of subject
Figure 4: Data visualization and dimension reduction for the leaf dataset. The original dataset comprises 14 variables, many of which are highly correlated. The green dots correspond to *Podocarpus sp.* and *Pseudosasa japonica*; the blue dots correspond to *Neurium oleander*; the red dots correspond to all other species.

103414 from the June 5, 2014, data release. The first two volumes were removed to allow for scanner equilibration. Three-dimensional volume data were vectorized and non-brain tissue excluded using the mask provided from the HCP. This resulted in a $230,459 \times 272$ data matrix. Each voxel was treated as a replicate with $v = 1, \ldots, V$, which is analogous to ‘spatial’ ICA of fMRI (Calhoun et al., 2009). We mean centered and variance normalized each voxel’s time course prior to conducting LCA, as suggested for ICA of fMRI (Beckmann and Smith, 2004).

The application of ICA to fMRI usually assumes that voxels are iid (an exception is the approach in Lee et al. 2011, which models time-series data using the Whittle likelihood). This assumption is often not made explicitly because ICA is usually derived from the perspective of maximizing non-Gaussianity. Since the objective function maximizing non-Gaussianity can also be derived from ML theory where the non-linear function is equivalent to the log likelihood (e.g., Hyvärinen and Oja 2000), summation of the non-linear function over voxels (e.g., Equation 12 in Beckmann and Smith 2004) is mathematically equivalent to assuming the voxels are independent. Despite the violation of model assumptions, ICA recovers simulated brain networks and their loadings (Beckmann and Smith, 2004) and has proven useful in constructing models of functional connectivity that are consistent across
subjects and image acquisition centers (Biswal et al., 2010).

We used the ICA software MELODIC (FSL) to determine the number of components that would be used in an analogous ICA of this dataset, which estimated thirty components. We initiated the algorithm from fifty-six randomly generated matrices, twenty-eight of which were in the principal subspace. We selected the estimate corresponding to the largest log likelihood as our estimate of the true argmax. Depending on initialization, the algorithm took between ten minutes and 3.75 hours on a 2.666 GHz processor, where 3.75 hours represented initializations that reached the maximum number of iterations, which we conservatively chose to be equal to 300. We also completed an analogous PCA+ProDenICA with thirty components and fifty-six initializations using the R package ProDenICA (Hastie and Tibshirani, 2010).

We examined the correlation between the loadings for each component (columns of $\hat{M}_S$) to the two variables ‘mentalize’ and ‘random.’ These variables were generated by convolving each task’s onsets and durations with the canonical HRF in SPM8 (Ashburner et al., 2004). The first component was highly correlated with the mentalize and random tasks (Figure 5). This component showed activation primarily in the lateral occipital cortex. A similar component was found using PCA+ProDenICA (not depicted).

We also detected components that were estimated in Spline-LCA but not in PCA+ProDenICA. Eight out of thirty LCs had a correlation less than 0.5 with their matched IC components. In particular, component two in Spline-LCA was not correlated with any of the components in PCA+ProDenICA (max correlation among all ICs = 0.01). This component appears to correspond to an artifact due to motion and possibly other sources of noise. Its time course was correlated with three of the motion parameters from the rigid-body alignment ($r = 0.32, 0.32,$ and 0.42 for the x-transformation, x-rotation, and z-rotation parameters, respectively). Voxels were highly activated in the brainstem, which could be due to movement. Additionally, there was a positive correlation with time ($r = 0.44$), which could be related to scanner drift. Removing artifacts from fMRI detected using ICA is a popular tool that can increase
Figure 5: Selected brain networks estimated from the HCP ToM data using Spline-LCA. The first row depicts a task-activated component that was highly correlated with the mentalizing (green) and random (blue) tasks (MNI coordinates: 35,-75,8; thresholded $|s_{v1}| \geq 2$); a similar component was found using PCA+ProDenICA (not depicted). The second row appears to be an artifact (MNI: 0,-50,0; unthresholded); this component was not found by PCA+ProDenICA.

detection in subsequent mixed-modeling of voxel activation (Tohka et al., 2008). Thus, our detection of a novel artifact represents a potential benefit of LCA over current methodology.

7 Discussion

In this study, we propose a model-based method for estimating non-Gaussian latent components in the presence of Gaussian noise that has many applications including data visualization, signal processing, artifact detection, and network estimation. We presented two applications: data visualization and dimension reduction, and identifying brain networks and artifacts from neuroimagery. Our first simulation study indicates that our methods perform
well when the LCA model is true, even for low SNR. When the noisy-ICA model is true, our methods perform well in the high SNR scenario, while none of the methods perform well in the low SNR scenario. In the second simulation study, we examined performance when data contained spatiotemporal dependence and a moderately low SNR. Logis-LCA and Spline-LCA outperformed competing methods for the LCA model and Logis-LCA outperformed PCA+Infomax for the noisy-ICA model. These results suggest that Logis-LCA and Spline-LCA can be used to reveal structure for a large class of non-Gaussian observations. In our fMRI application, we simultaneously achieve dimension reduction and latent variable extraction for large image data ($T = 272$ and $V = 230,459$) and identify an artifact not identified by PCA+ICA.

The presence of local maxima in LCA can increase computational expenses, and more initializations are required for larger values of $T$. Since the set of orthogonal matrices is non-convex, local optima are also a problem in PCA+ICA (e.g., Risk et al. 2014). For fMRI data, fifty initializations appeared to be adequate when estimating thirty components with nearly three hundred time points. In general, we found that Logis-LCA was less sensitive to initialization than Spline-LCA (results not shown). It appears that the additional flexibility of Spline-LCA comes at the expense of increased detection of local maxima. We favor Spline-LCA because it can estimate sub-Gaussian densities. However, sub-Gaussian components appear to be uncommon in fMRI (sparse images are super-Gaussian). Future research should examine whether Spline-LCA offers advantages over Logis-LCA in fMRI. Additionally, developing algorithms to more efficiently address local optima is an avenue for future research.

An important advantage of LCA over existing frameworks is its robustness to misspecification of the number of estimated components. This robustness suggests LCA could be used to improve estimates of functional connectivity in fMRI studies. In contrast, estimating the correct number of components in noisy ICA is a pre-requisite to recovering valid components (Section 4, see also Allassonniere and Younes 2012). Beckmann and Smith (2004) explored
the use of probabilistic PCA to estimate the number of brain networks prior to ICA in order to avoid model over-fitting, which addresses the concern that over fitting may separate a single network into multiple networks. However, our simulations suggest that using too few components leads to inappropriately aggregated networks in PCA+ICA methods (Figures 2 and 3). In contrast, the components recovered for $Q^* \neq Q$ in Logis-LCA across model scenarios (Figures 2 and 3) and Spline-LCA for the LCA scenario (Figure 2) accurately represent functional connectivity. In the leaf data example, the first two components were nearly identical for $Q^* = 2$ and $Q^* = 5$ for LCA but differed for PCA+ICA. Although we have argued that our framework is robust to the choice of $Q^*$, we would like a rigorous method to determine the number of components. For iid data, AIC may be an effective method, but AIC and other model selection criteria are ineffective when correlation among observations is not taken into account adequately. Future research should investigate selection criteria for non-iid data.

LCA offers a computationally tractable alternative to one of the most common applications of ICA to fMRI and EEG: artifact detection. Currently, PCA+ICA is used as a pre-processing step to reveal biologically implausible loadings and/or loadings resembling physiological artifacts that can be used to de-noise data for subsequent analyses (Beckmann, 2012). In LCA, these artifacts appear as LCs since they have non-Gaussian distributions. Our detection of the artifact in component two (Figure 5) suggests LCA could be used for more powerful denoising methods over traditional PCA+ICA. Artifacts may increase and/or become more problematic when using state-of-the-art data with high-resolution, as smaller voxels are associated with smaller signals, suggesting artifact removal is increasingly important (Griffanti et al., 2014). The HCP data represent the highest resolution and fastest acquisition times currently available in fMRI, and thus LCA offers a promising alternative to ICA for artifact detection.

Code implementing Logis-LCA, Spline-LCA, and the PMSE similarity measure is available in the Web Supplement.
8 Acknowledgments

hidden

References


Automatic independent component labeling for artifact removal in fMRI. *Neuroimage*,