Uniqueness of Non-Gaussianity-Based Dimension Reduction

Fabian J. Theis, Motoaki Kawanabe, and Klaus-Robert Müller

***Abstract*—Dimension reduction is a key step in preprocessing large-scale data sets. A recently proposed method named non-Gaussian component analysis searches for a projection onto the non-Gaussian part of a given multivariate recording, which is a generalization of the deflationary projection pursuit model. In this contribution, we discuss the uniqueness of the subspaces of such a projection. We prove that a necessary and sufficient condition for uniqueness is that the non-Gaussian signal subspace is of minimal dimension. Furthermore, we propose a measure for estimating this minimal dimension and illustrate it by numerical simulations. Our result guarantees that projection algorithms uniquely recover the underlying lower dimensional data signals.**

***Index Terms*—Identifiability, independent subspace analysis, non-Gaussian component analysis, projection pursuit.**

A common and important problem in signal processing is the task of efficient dimension reduction, i.e., the search for meaningful signals within a higher dimensional data set. Classical techniques such as principal component analysis hereby define "meaningful" using second-order statistics (maximal variance), which may often be inadequate for signal detection, e.g., in the presence of strong noise. This contrasts to higher-order models including *projection pursuit* [1] or *non-Gaussian component analysis*, for short NGCA [2]–[5]. While the former classically extracts a single non-Gaussian independent component from the data set, the latter tries to detect a whole non-Gaussian subspace within the data, and no assumption of independence within the subspace is made; see Fig. 1. Applications of such non-Gaussianity-based dimension reduction techniques can for example be preprocessing in multivariate speech signals [6] and estimating speakers and their directions as alternative to more classical methods [7].

Our goal is to describe necessary and sufficient conditions in order for NGCA projections to exist and to be unique—otherwise we would not be able to speak of a well-defined "signal subspace." Of course, no basis within the signal subspace can be chosen without additional knowledge, so only the subspaces of the projection can be unique in this general model. Other indeterminacies can occur if we do not remove all independent Gaussians from the signal subspace, since in such a case different we can include different Gaussians into the signal subspace. Here we will prove (Theorem 1) that a sufficient and necessary condition for uniqueness is to search for a Gaussian subspace of maximal dimension, or correspondingly for a non-Gaussian subspace of minimal dimension.

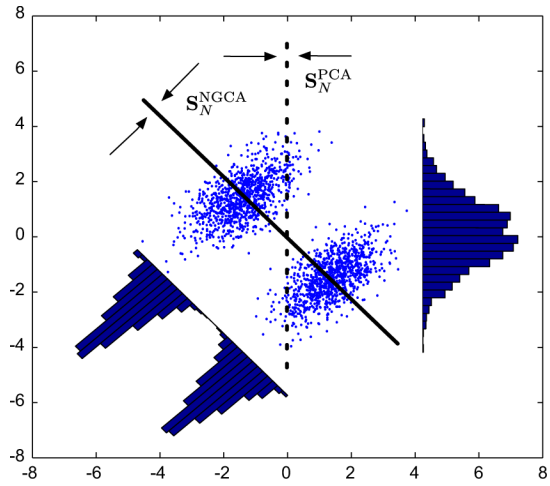

Fig. 1. Comparison of PCA- and NGCA-based dimension reduction. The signal subspace represented by the random vector $\mathbf{S}_N^{\mathrm{PCA}}$ is given by a direction of high variance, whereas in the NGCA-case, the signal subspace $\mathbf{S}_N^{\mathrm{NGCA}}$ is given by directions of non-Gaussianity. This allows for the extraction of low-power signals, see histograms of projected signals.

A practical algorithm for performing NGCA using the idea of separated characteristic functions from the uniqueness proof has been proposed already [8]. Moreover, we remark that the key idea of the uniqueness theorem without the proofs and simulations has been presented at the conference "ICA 2006" [9]. A related method for dimension reduction denoted as "colored subspace analysis" (CSA) has been introduced in [10]; instead of NGCA where we consider i.i.d. data with nontrivial higher order correlations, the CSA method deals with wide-sense stationary processes. Separability is easier to show for CSA [10], since we then only have to deal with second-order tensors, whereas for NGCA we have to take the full higher-order structure into account for the separability proof as will be shown here.

In the following, we will use the notations $\mathrm{Mat}(m \times n)$, $\mathrm{Gl}(n)$, and $\mathrm{O}(n)$ to denote the set of real $m \times n$-matrices, the group of real invertible and the group of real orthogonal $n \times n$-matrices, respectively. The transpose of a matrix $\mathbf{A}$ is denoted as $\mathbf{A}^\top$. Differentials are evaluated at $\mathbf{x}$ using the notation $\nabla f|_{\mathbf{x}}$. Matlab-notation is employed for selecting columns and rows of matrices, so $\mathbf{A}(:, 2:n)$ denotes the matrix consisting of the last $(n-1)$ columns of $\mathbf{A}$. Random variables and vectors are defined on the probability space Ω, and the notation $X \in L_2(\Omega, \mathbb{R})$ means that the random variable X is square-integrable.

I. Uniqueness of Non-Gaussian Component Analysis (NGCA)

Originally, NGCA has been formulated as finding a lower n-dimensional non-Gaussian signal within higher d-dimensional Gaussian noise, $n < d$. It can easily be seen that we can only identify the Gaussian signal orthogonal to the non-Gaussian part, see e.g., [2], [3], [6], [8]. We may therefore formulate the goal of NGCA as finding a matrix $\mathbf{W}_N \in \mathrm{Mat}(n \times d)$ such that there exists

Manuscript received July 29, 2010; revised November 22, 2010, March 03, 2011, and May 26, 2011; accepted May 27, 2011. Date of publication June 16, 2011; date of current version August 10, 2011. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Kainam Thomas Wong. F. J. Theis gratefully acknowledges partial financial supported by the Helmholtz Alliance on Systems Biology (project "CoReNe"). M. Kawanabe and K.-R. Müller acknowledge the THESEUS project (01MQ07018) funded by the German Federal Ministry of Economics and Technology. K.-R. Müller also acknowledges general support by the BMBF and DFG.

F. J. Theis is with the Institute of Bioinformatics and Systems Biology, Helmholtz Center Munich, 85764 Neuherberg, Germany, and the Department of Mathematical Science, Technical University of Munich, 85747 Garching, Germany (e-mail: fabian.theis@helmholtz-muenchen.de).

M. Kawanabe is with the Machine Learning Group, Faculty of Computer Science, Technical University of Berlin, 10587 Berlin, Germany, and also with Fraunhofer FIRST.IDA, 12489 Berlin, Germany (e-mail: motoaki.kawanabe@first.fraunhofer.de).

K.-R. Müller is with the Machine Learning Group, Faculty of Computer Science, Technical University of Berlin, 10587 Berlin, Germany (e-mail: krm@cs.tu-berlin.de).

This correspondence has supplementary downloadable multimedia material available at http://ieeexplore.ieee.org provided by the authors. This includes a simulation analysis of the uniqueness theorem, illustration on subspace estimation and a more detailed proof of Lemma 1.8. This material is 5.2 MB in size.

Color versions of one or more of the figures in this correspondence are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2011.2159600

$\mathbf{W}_G \in \mathrm{Mat}((d-n)\times d)$ with $\mathbf{W}_N\mathbf{X}$ and $\mathbf{W}_G\mathbf{X}$ being independent, and $\mathbf{W}_G\mathbf{X}$ being Gaussian. The dimension reduction problem itself can also be formulated within a generative model, leading to

$$\mathbf{X} = \mathbf{A}_N\mathbf{S}_N + \mathbf{A}_G\mathbf{S}_G \tag{1}$$

with $\mathbf{A}_N \in \mathrm{Mat}(d \times n)$ and $\mathbf{A}_G \in \mathrm{Mat}(d \times (d-n))$. Here the n-dimensional random vector $\mathbf{S}_N$ and the $(d-n)$-dimensional vector $\mathbf{S}_G$ are independent, and $\mathbf{S}_G$ Gaussian. Then $(\mathbf{A}_N, \mathbf{A}_G)^{-1} = (\mathbf{W}_N^\top, \mathbf{W}_G^\top)^\top$. This leads to the following definition of NGCA.

Definition 1.1: A factorization $\mathbf{X} = \mathbf{A}\mathbf{S}$ with $\mathbf{A} \in \mathrm{Gl}(d)$, $\mathbf{S} = (\mathbf{S}_N, \mathbf{S}_G)$ and $\mathbf{S}_N \in L_2(\Omega, \mathbb{R}^n)$ is called an *n-decomposition* of $\mathbf{X}$, $0 \le n \le d$, if $\mathbf{S}_N$ and $\mathbf{S}_G$ are stochastically independent and $\mathbf{S}_G$ is Gaussian. $\mathbf{X}$ is then said to be *n-decomposable*.

Hence, an n-decomposition of $\mathbf{X}$ corresponds to an NGCA solution. It is the higher order equivalent of an n-temporal decomposition introduced for the second-order case with temporal structure in [10]. If as before $\mathbf{A} = (\mathbf{A}_N, \mathbf{A}_G)$, then the n-dimensional vector subspace $R(\mathbf{A}_N) \subset \mathbb{R}^d$ is called the *non-Gaussian* or simply the *signal subspace*, and $R(\mathbf{A}_G)$ the *Gaussian subspace* of the decomposition; here, $R(\mathbf{A})$ denotes the range of the linear map $\mathbf{A}$ i.e., the span of its columns.

A. Indeterminacies

Clearly, the matrices $\mathbf{A}_N$ and $\mathbf{A}_G$ in the decomposition (1) cannot be unique—multiplication from the right in both subspaces using any invertible matrix leaves the model invariant. As already mentioned, an additional indeterminacy comes into play due to the fact that we do not want to fix the reduced dimension in advance. Given a realization of the model (1) with $n < d$, let $\mathbf{B}_G := \mathrm{Cov}(\mathbf{S}_G)^{\frac{1}{2}} \in \mathrm{Gl}(d-n)$. Then $\mathbf{B}_G^+\mathbf{S}_G$ is decorrelated, i.e., mutually independent because of Gaussianity. By replacing $\mathbf{A}_G$ with $\mathbf{A}_G\mathbf{B}_G$, we may therefore assume that $\mathbf{S}_G$ consists of independent components. If $\mathbf{a} := \mathbf{A}_G(:,1)$ denotes the first column of $\mathbf{A}_G$, then $\mathbf{X} = \mathbf{A}_N\mathbf{S}_N + \mathbf{A}_G\mathbf{S}_G$ can be rewritten as

$$\mathbf{X} = (\mathbf{A}_N, \mathbf{a}) \begin{pmatrix} \mathbf{S}_N \\ \mathbf{S}_G(1) \end{pmatrix} + \mathbf{A}_G(:, 2:d-n)\mathbf{S}_G(2:d-n) \tag{2}$$

and $(\mathbf{S}_N, \mathbf{S}_G(1))$ and $\mathbf{S}_G(2:d-n)$ are independent, with the second vector being Gaussian. In other words, without putting an additional condition of maximality onto the Gaussian part, different model realizations can be generated by simply moving random variables to the non-Gaussian part. This is similar to the case of temporally correlated data, where we need to remove as many white (instead of Gaussian) components to achieve uniqueness of the signal subspace [10].

B. Uniqueness Theorem

Clearly, already $\mathbf{X} = \mathbf{I}(\mathbf{X}, \emptyset)$ is (the unique) n-decomposition of $\mathbf{X}$ for $n = d$ but of little interest, so additional conditions are necessary:

Definition 1.2: $\mathbf{X}$ is denoted to be *minimally n-decomposable* if $\mathbf{X}$ is not $(n-1)$-decomposable. Then $\dim_e(\mathbf{X}) := n$ is called the *essential dimension of* $\mathbf{X}$.

For example, the essential dimension $\dim_e(\mathbf{X})$ is zero if and only if $\mathbf{X}$ is Gaussian, whereas the essential dimension of a d-dimensional mutually independent Laplacian is d. The following theorem is the main theoretical contribution of this work. It connects uniqueness of the dimension reduction model with minimality, and gives a simple characterization for it.

Theorem 1.3 (Uniqueness of NGCA): Let $0 \le n < d$. Given an n-decomposition $\mathbf{A}_N\mathbf{S}_N + \mathbf{A}_G\mathbf{S}_G$ of the random vector $\mathbf{X} \in L_2(\Omega, \mathbb{R}^d)$, the following is equivalent.

(i) The decomposition is minimal, i.e., $n = \dim_e(\mathbf{X})$.

(ii) There exists no basis $\mathbf{M} \in \mathrm{Gl}(n)$ with $(\mathbf{M}\mathbf{S}_N)(1)$ is Gaussian and independent of $(\mathbf{M}\mathbf{S}_N)(2:n)$.

(iii) The subspaces of the decomposition are unique i.e., another n-decomposition has the same non-Gaussian and Gaussian subspaces.

Note that we have to exclude $n = d$, since for any $\mathbf{X}$, the trivial d-decomposition of $\mathbf{X}$ fulfills (iii) but in general not (i,ii). Condition (ii) means that there exists no Gaussian independent component in the non-Gaussian part of the decomposition. The theorem proves that this is equivalent to the decomposition being minimal. Note that in (ii), it is *not* enough to require only that there exists no Gaussian component, i.e., $\mathbf{v} \in \mathbb{R}^n$ such that $\mathbf{v}^\top\mathbf{S}_N$ is Gaussian. We can now show uniqueness of an essential decomposition:

Corollary 1.4: The subspaces of a $\dim_e(\mathbf{X})$-decomposition of $\mathbf{X}$ are unique.

This follows from Theorem 1.3(i)⇒(iii) in the case of $\dim_e(\mathbf{X}) < d$, and holds trivially if $\dim_e(\mathbf{X}) = d$. Also note that existence of a minimal decomposition always holds.

Proof of the Uniqueness Theorem

We first remark that the theorem holds trivially for $n = 0$, because in this case $\mathbf{X}$ is Gaussian. So in the following let $0 < n < d$.

(i) ⇒ (ii) follows similarly to (2): If (ii) did not hold, then we would have a basis in which one component of $\mathbf{S}_N$ is Gaussian and independent of the rest. So we could simply move it to the Gaussian part and reduce the order of the decomposition.

Also (iii) ⇒ (i) can be easily shown: assume that the decomposition is not minimal. We may choose an $(n-1)$-decomposition $\mathbf{X} = \tilde{\mathbf{A}}_N\tilde{\mathbf{S}}_N + \tilde{\mathbf{A}}_G\tilde{\mathbf{S}}_G$, without loss of generality $\tilde{\mathbf{S}}_G \in L_2(\Omega, \mathbb{R}^{d-n+1})$ being independent (see Section I-A). The two (because $n < d$) columns $(\tilde{\mathbf{a}}_G)_1$ and $(\tilde{\mathbf{a}}_G)_2$ of $\tilde{\mathbf{A}}_G$ are linearly independent, Similar to (2), we get two n-decompositions of $\mathbf{X}$ with non-Gaussian coordinates $\left(\tilde{\mathbf{A}}_N, (\tilde{\mathbf{a}}_G)_1\right)$ and $\left(\tilde{\mathbf{A}}_N, (\tilde{\mathbf{a}}_G)_2\right)$, respectively. However, the two vectors $(\tilde{\mathbf{a}}_G)_i$ do not lie in the range of $\tilde{\mathbf{A}}_N$ and are linearly independent, so we have constructed two n-decompositions of $\mathbf{X}$ with different non-Gaussian subspaces. This contradicts (iii).

The main part of the proof now consists of showing (ii) ⇒ (iii). Assume that (ii) holds for all n-decompositions of $\mathbf{X}$. Given two such n-decompositions, then

$$\mathbf{A}_N\mathbf{S}_N + \mathbf{A}_G\mathbf{S}_G = \tilde{\mathbf{A}}_N\tilde{\mathbf{S}}_N + \tilde{\mathbf{A}}_G\tilde{\mathbf{S}}_G.$$

By applying $(\tilde{\mathbf{A}}_N, \tilde{\mathbf{A}}_G)^{-1}$, we may therefore without loss of generality assume that $\mathbf{A}_N\mathbf{S}_N + \mathbf{A}_G\mathbf{S}_G = (\tilde{\mathbf{S}}_N, \tilde{\mathbf{S}}_G)$, or in other words that $(\mathbf{X}_N, \mathbf{X}_G)$ is also an n-decomposition of $\mathbf{X}$. Our assumption is that (ii) holds for both $\mathbf{S}$ and $\mathbf{X}$.

We introduce some notation by dividing $\mathbf{A} \in \mathrm{Gl}(d)$ into

$$\mathbf{A} = \begin{array}{|c|c|} \hline \mathbf{A}_{NN} & \mathbf{A}_{NG} \\ \hline \mathbf{A}_{GN} & \mathbf{A}_{GG} \\ \hline \end{array} \tag{3}$$

with $\mathbf{A}_{NN} \in \mathrm{Mat}(n \times n)$. Altogether, we are dealing with the simple linear model $\mathbf{X} = \mathbf{A}\mathbf{S}$, where both $\mathbf{X}$ and $\mathbf{S}$ consist of an n-dimensional non-Gaussian and an independent Gaussian part. We have to show that these parts span the same subspaces, respectively, which is equivalent to showing that $\mathbf{A}_{NG} = \mathbf{A}_{GN} = 0$.

In order to state the main argument of the proof without getting lost in details, we have divided it up into a sequence of lemmata, the proofs of which are given in the Appendix.

The *characteristic function* of the random vector $\mathbf{X}$ is defined by $\Phi_{\mathbf{x}}(\mathbf{x}) := E(\exp i\mathbf{x}^\top\mathbf{X})$, and since $\mathbf{X}$ is assumed to admit second-order moments, $\Phi\mathbf{X}$ is twice continuously differentiable, i.e., $\Phi_{\mathbf{X}} \in \mathbf{C}^2(\mathbb{R}^d, \mathbb{C})$. Moreover, by definition $\Phi_{\mathbf{AS}}(\mathbf{x}) = \Phi_{\mathbf{S}}(\mathbf{A}^\top\mathbf{x})$, and the characteristic function of two mutually independent random vectors factorizes into two component characteristic functions.

Originally, we formulated the proof in terms of the logarithm of $\Phi_{\mathbf{X}}$, also known as the second characteristic function. However, a logarithm of a complex variable represents a noncanonical choice and is not defined everywhere. Hence, we reformulated the proof to only employ logarithms locally. This is done in the following lemma; here, ∇f denotes the gradient of f and $\nabla^2 f$ its Hessian.

Lemma 1.5 (Characterization of Gaussians): Let $\mathbf{X} \in L_2(\Omega, \mathbb{R}^n)$ be a random vector, and let $\mathbf{C} \in \mathrm{Gl}(n)$. Then, $\mathbf{X}$ is Gaussian with covariance $\mathbf{C}^{-1}$ if and only if its characteristic function $\Phi_{\mathbf{X}}$ satisfies

$$\Phi_{\mathbf{X}}\nabla^2\Phi_{\mathbf{X}} - \nabla\Phi_{\mathbf{X}}(\nabla\Phi_{\mathbf{X}})^\top + \mathbf{C}\Phi_{\mathbf{X}}^2 \equiv 0. \tag{4}$$

The lemma also holds in the degenerate case of an only positive semidefinite covariance; then the inverse of $\mathbf{C}$ has to be replaced by its Moore–Penrose pseudoinverse. This lemma can be easily seen to be true in the case of a one-dimensional random variable: Here, (4) obviously reduces to the differential equation $\Phi_X\Phi_X'' - \Phi_X'^2 + c\Phi_X^2 \equiv 0$,, which implies that the logarithm of Φ_X is a second-order polynomial, so indeed X is Gaussian as claimed by the lemma.

The next lemma is as modification of lemma 1.5, and can be shown similarly. Here more generally $\mathbf{e}_i \in \mathbb{R}^n$ denotes the ith unit vector.

Lemma 1.6: If a decorrelated random vector $\mathbf{X} \in L_2(\Omega, \mathbb{R}^n)$ fulfills

$$\left(\Phi_{\mathbf{X}}\nabla^2\Phi_{\mathbf{X}} - \nabla\Phi_{\mathbf{X}}(\nabla\Phi_{\mathbf{X}})^\top\right)\mathbf{e}_1 + \mathbf{c}\Phi_{\mathbf{X}}^2 \equiv 0,$$

for some constant $\mathbf{c} \in \mathbb{R}^n$, then its first component $\mathbf{X}(1)$ is Gaussian and independent of $\mathbf{X}(2:n)$.

We may now assume that the covariance of $\mathbf{S}$ (and hence also of $\mathbf{X}$) is positive definite instead of only semidefinite—otherwise, while still keeping the model, we can simply remove the subspace of deterministic components (i.e., components of variance 0, also falling into the category of sometimes called extended Gaussians), which have to be mapped onto each other by $\mathbf{A}$. Hence, we may even assume $\mathrm{Cov}(\mathbf{S}_G) = \mathbf{I}$, after whitening as described in Section I-A. This uses the fact that the basis within the Gaussian subspace is not unique. The same holds also for the non-Gaussian subspace, and the Assumption 1.3(ii) is invariant under linear transformation in the signal subspace, so we may choose any $\mathbf{B}_N \in \mathrm{Gl}(n)$ and $\mathbf{B}_G \in O(d-n)$ to get

$$\mathbf{X} = \begin{pmatrix}\mathbf{A}_{NN}\mathbf{B}_N \\ \mathbf{A}_{GN}\mathbf{B}_N\end{pmatrix}(\mathbf{B}_N^{-1}\mathbf{S}_N) + \begin{pmatrix}\mathbf{A}_{NG}\mathbf{B}_G \\ \mathbf{A}_{GG}\mathbf{B}_G\end{pmatrix}(\mathbf{B}_G^\top\mathbf{S}_G). \tag{5}$$

Here, only orthogonal matrices $\mathbf{B}_G$ are allowed in order for $\mathbf{B}_G^\top\mathbf{S}_G$ to stay white, with $\mathbf{S}_G$ being white. Removal of covariance by whitening simplifies the basis choices in the following lemmata.

Assumption (ii) from Theorem 1.3 can now be used to prove regularity of $\mathbf{A}$ in the non-Gaussian part:

Lemma 1.7: If (ii) from Theorem 1.3 holds, then the non-Gaussian transformation is invertible, i.e., $\mathbf{A}_{NN} \in \mathrm{Gl}(n)$.

The next lemma uses the generative model (1) for $\mathbf{X}$ and $\mathbf{S}$ to derive an explicit differential equation for $\Phi_{\mathbf{S}_N}$. The Gaussian part $\Phi_{\mathbf{S}_G}$ in this equation has been eliminated by applying Lemma 1.5.

Lemma 1.8: For any basis $\mathbf{B}_N \in \mathrm{Gl}(n)$, the characteristic function $\Phi_{\mathbf{S}_N} \in \mathbf{C}^2(\mathbb{R}^n, \mathbb{C})$ of the transformed non-Gaussian sources $\mathbf{S}_N = \mathbf{B}_N^{-1}\mathbf{S}_N$ fulfills the differential equation

$$\mathbf{A}_{NN}\mathbf{B}_N\left(\Phi_{\tilde{\mathbf{S}}_N}\nabla^2\Phi_{\tilde{\mathbf{S}}_N} - \nabla\Phi_{\tilde{\mathbf{S}}_N}(\nabla\Phi_{\tilde{\mathbf{S}}_N})^\top\right)\mathbf{B}_N^\top\mathbf{A}_{GN}^\top \equiv \mathbf{A}_{NG}\mathbf{A}_{GG}^\top\Phi_{\tilde{\mathbf{S}}_N}^2. \tag{6}$$

Putting these lemmata together, we can finally prove Theorem 1.3: According to Lemma 1.7, $\mathbf{A}_{NN}$ is invertible, so multiplying (6) by $\mathbf{B}_N^{-1}\mathbf{A}_{NN}^{-1}$ from the left yields

$$\left(\Phi_{\tilde{\mathbf{S}}_N}\nabla^2\Phi_{\tilde{\mathbf{S}}_N} - \nabla\Phi_{\tilde{\mathbf{S}}_N}(\nabla\Phi_{\tilde{\mathbf{S}}_N})^\top\right)\mathbf{B}_N^\top\mathbf{A}_{GN}^\top \equiv \mathbf{C}\Phi_{\tilde{\mathbf{S}}_N}^2 \tag{7}$$

for any $\mathbf{B}_N \in \mathrm{Gl}(n)$ and some fixed, real matrix $\mathbf{C} \in \mathrm{Mat}(n \times (d-n))$.

We claim that $\mathbf{A}_{GN} = 0$. If not, then there exists some $\mathbf{v} \in \mathbb{R}^{d-n}$ with $\|\mathbf{A}_{GN}^\top\mathbf{v}\| = 1$. Now choose the basis $\mathbf{B}_N$ of the non-Gaussian subspace in (5) such that $\tilde{\mathbf{S}}_N = \mathbf{B}_N^{-1}\mathbf{S}_N$ is white. This is invariant under left-multiplication by an orthogonal matrix, so we may moreover assume that $\mathbf{B}_N^\top\mathbf{A}_{GN}^\top\mathbf{v} = \mathbf{e}_1$. Multiplying (7) in turn by $\mathbf{v}$ from the right therefore shows the vector equality

$$\left(\Phi_{\tilde{\mathbf{S}}_N}\nabla^2\Phi_{\tilde{\mathbf{S}}_N} - \nabla\Phi_{\tilde{\mathbf{S}}_N}(\nabla\Phi_{\tilde{\mathbf{S}}_N})^\top\right)\mathbf{e}_1 \equiv \mathbf{c}\Phi_{\tilde{\mathbf{S}}_N}^2 \tag{8}$$

where $\mathbf{c} := \mathbf{C}\mathbf{v} \in \mathbb{R}^n$. This means that $\tilde{\mathbf{S}}_N$ fulfills the condition of Lemma 1.6, which implies that $\tilde{\mathbf{S}}_N(1)$ is Gaussian and independent of the rest. However, this contradicts Theorem 1.3(ii) for $\mathbf{S}$, hence $\mathbf{A}_{GN} = 0$. Plugging this result into (6), evaluation at $\mathbf{s}_N = 0$ shows that $\mathbf{A}_{NG}\mathbf{A}_{GG}^\top = 0$. Since $\mathbf{A}_{GN} = 0$ and $\mathbf{A} \in \mathrm{Gl}(d)$, necessarily $\mathbf{A}_{GG} \in \mathrm{Gl}(d-n)$, so $\mathbf{A}_{NG} = 0$, which completes the proof of Theorem 1.3.

II. Estimation of the Essential Dimension $\dim_e(\mathbf{X})$

The main practical consequence of the presented uniqueness theorem is that there exists a unique 'effective' dimension (i.e., the dimension of the true non-Gaussian subspace), so its estimation is possible in theory. In general however, estimating the dimension of the signal subspace is a difficult problem by itself, often ill-posed due to the fact that model assumptions only hold approximately due to sampling or systematic noise effects. Many indexes have been proposed, e.g., for PCA and ICA such as Akaike's information criterion (AIC) and the criterion for minimum description length (MDL) [11], [12]. More recently, more specific structure or additional information has been used to estimate dimensions and subspaces; see, e.g., [13] and references therein. In particular, in the application field of geosciences, there have been a lot of efforts to extend such detection criteria techniques [14]–[16]. In this section, we propose a general thresholding rule related to the original NGCA algorithm [2] for dimension estimation in NGCA and present some numerical experiments.

The key idea is that the matrix set has been generated by the NGCA algorithm in such a way that eigenvalues corresponding to eigenvectors from the non-Gaussian subspace are larger than those of the Gaussian subspace [2]. Therefore, we determine the distribution of maximal eigenvalue activity in the NGCA vectors by applying the method first to purely Gaussian data. Then according to some statistical threshold, we determine the cutoff in eigenvalues corresponding to the signal subspace. In practice, this means that we at first generate purely Gaussian data of the same total dimension d and apply the NGCA algorithm [2]. This we do multiple times to obtain samples of the eigenvalues, when non-Gaussian structures are not present at all. Then, we take the 95% percentile of the largest eigenvalues in the Monte Carlo runs as the threshold θ. If some eigenvalue of the data at hand exceeds this threshold, we interpret it to stem from non-Gaussian structure in the corresponding components. That is, the estimator Φ_n of the effective dimension is defined as $\hat{n} := \arg\max_i\{\mathrm{eigenvalue}(i) \geq \theta\}$. We apply this method to synthetic data sets (A-D) from [3], see supplementary material A. As summarized in Table I and illustrated in supplement B, this simple method works almost perfectly with the synthetic data sets. Indeed, for 100 runs and seven different mixture situations, we never over- or underestimate the dimension by more than one, and only by one in 1% of the runs. In supplement C, we

TABLE I
HISTOGRAM OF THE DIMENSION ESTIMATOR $\hat{n}$ BY THRESHOLD RULE FROM SECTION II BASED ON 100 REPETITIONS

#	data set			histogram of $\hat{n}-n$			threshold θ
	non-Gauss	d	n	-1	0	$+1$	
1	(D)	4	2	0	100	0	3.8132
2	(C)	5	2	0	100	0	3.2137
3	(C)	5	3	2	98	0	3.2137
4	(A)	10	2	0	99	1	1.5118
5	(B)	10	2	1	98	1	1.5118
6	(C)	10	2	0	98	2	1.5118
7	(D)	10	2	0	100	0	1.5118

demonstrate efficient subspace estimation given correct dimension estimation using the proposed index.

In summary, both our theoretical as well as the pragmatic dimension estimation result allows NGCA algorithms to find the unknown, unique signal space within a noisy high-dimensional data set: If we identify a non-Gaussian subspace, we know that it is the one in question given that we have removed all non-Gaussian components.

APPENDIX

We prove the lemmata necessary for the proof of the main Theorem 1.3.

Proof of Lemma 1.5: We first show that the differential (4) locally at nonzeros of $\Phi_{\mathbf{X}}$ has the solution $\exp g$, where g is a m-dimensional polynomial of degree ≤ 2. For this note that $\Phi_{\mathbf{X}}(0) = 1$ by definition, so there exists an nonempty open set U containing 0 such that a complex logarithm $\log$ is defined on $\Phi_X(U)$. Set $g := \log \Phi_X|U$. Substituting $\exp g$ for Φ_X in (4) yields equations

$$\exp(g)\left(\frac{\partial^2 g}{\partial x_i \partial x_j} + \frac{\partial g}{\partial x_i}\frac{\partial g}{\partial x_j}\right)\exp(g) - \frac{\partial g}{\partial x_i}\frac{\partial g}{\partial x_j}\exp(2g) + c_{ij}\exp(2g) \equiv 0,$$

for $i,j \in \{1,\ldots,m\}$, so $\frac{\partial^2 g}{\partial x_i \partial x_j} \equiv -c_{ij}$. Hence, g is a polynomial of degree ≤ 2, and $\Phi_{\mathbf{X}} = \exp g \neq 0$ on all of $\bar{U}$. Therefore, $\Phi_{\mathbf{X}} \neq 0$ everywhere because of continuity.

The local argument from above then shows that $\Phi_{\mathbf{X}}(\mathbf{x}) = \exp\left(-\frac{1}{2}\sum_{ij} c_{ij}x_i x_j + \sum_i b_i x_i\right)$ everywhere, where we have already used $\Phi_{\mathbf{X}}(0) = 1$. Moreover, from $\Phi_{\mathbf{X}}(-\mathbf{x}) = \overline{\Phi_{\mathbf{X}}(\mathbf{x})}$ we get $c_{ij} \in \mathbb{R}, c_{ij} = c_{ji}$ and $\mathbf{b} = i\mu$ with real $\mu \in \mathbb{R}^m$. And $|\Phi_X| \leq 1$ guarantees that $\mathbf{C} = (c_{ij})$ is positive semidefinite, actually positive definite because $\mathbf{C} \in \mathrm{Gl}(n)$. Altogether, we find that $\Phi_{\mathbf{X}}(\mathbf{x}) = \exp\left(i\mu^\top \mathbf{x} - \frac{1}{2}\mathbf{x}^\top \mathbf{C}\mathbf{x}\right)$, which means that $\mathbf{X}$ is normally distributed with mean μ and covariance $\mathbf{C}^{-1}$. The converse follows by direct calculation. ∎

Proof of Lemma 1.6: At points $\mathbf{x} \in \mathbb{R}^n$ with $g(\mathbf{x}) := \Phi_{\mathbf{X}}(\mathbf{x}) \neq 0$, we may locally choose a complex logarithm. Then, locally, $h := \log g$ can be plugged into (8), which results in

$$\mathbf{e}_i^\top\left(g(g\nabla^2 h + g\nabla h(\nabla h)^\top) - g^2\nabla h(\nabla h)^\top\right)\mathbf{e}_1 + c_i g^2 \equiv 0$$

so $\mathbf{e}_i^\top \nabla^2 h \mathbf{e}_1 \equiv -c_i$ locally, i.e., $\frac{\partial^2}{\partial x_1}\partial x_i \log g \equiv -c_i$. Hence, after integration, we get locally

$$g(\mathbf{x}) = \exp\left(-\frac{1}{2}\sum_{i=1}^n c_i x_1 x_i + d x_1\right) g_2(x_2,\ldots,x_n)$$

for some $d \in \mathbb{C}$ and function g_2 on $\mathbb{R}^{n-1}$; this equality, due to continuity, is then valid also for all $\mathbf{x} \in \mathbb{R}^n$. By assumption, $\mathbf{X}$ is white, so $c_i = 0$ for $i > 1$ and $g(\mathbf{x}) = \exp(-\frac{1}{2}c_1 x_1^2)g_2(x_2,\ldots,x_n)$, which implies that $\mathbf{X}_1$ is Gaussian ($c_1 \geq 0$ due to integrability and $d \in \mathbb{R}$ because of $\Phi_{\mathbf{X}}(-\mathbf{x}) = \overline{\Phi_{\mathbf{X}}(\mathbf{x})}$) and independent of the rest.

In order to show the next Lemma 1.7, we need the following basis choice. ∎

Lemma 2.1: Let $(\mathbf{A}_{NN}, \mathbf{A}_{NG}) \in \mathrm{Mat}(n \times (n + (d-n)))$ be an arbitrary full rank matrix. If $\mathrm{rank}\ \mathbf{A}_{NN} < n$, then we may choose coordinates $\mathbf{B}_N \in \mathrm{Gl}(n)$, $\mathbf{B}_G \in O(d-n)$ and $\mathbf{M} \in \mathrm{Gl}(n)$ such that

$$\mathbf{M}\mathbf{A}_{NN}\mathbf{B}_N = \begin{array}{|c|c|}\hline 0 & 0 \\ \hline 0 & * \\ \hline \end{array} \text{ and } \mathbf{M}\mathbf{A}_{NG}\mathbf{B}_G = \begin{array}{|c|c|}\hline 1 & 0 \\ \hline 0 & *' \\ \hline \end{array}$$

where $* \in \mathrm{Mat}((n-1)\times(n-1))$,$*' \in \mathrm{Mat}((n-1)\times(d-n-1))$ are some matrices.

Proof: We have to construct coordinates such that

$$\mathbf{e}_1^\top \mathbf{M}\mathbf{A}_{NN}\mathbf{B}_N = 0, \tag{9}$$

$$\mathbf{e}_1^\top \mathbf{M}\mathbf{A}_{NG}\mathbf{B}_G = \mathbf{e}_1^\top \tag{10}$$

$$\mathbf{M}\mathbf{A}_{NN}\mathbf{B}_N\mathbf{e}_1 = 0, \tag{11}$$

$$\mathbf{M}\mathbf{A}_{NG}\mathbf{B}_G\mathbf{e}_1 = \mathbf{e}_1 \tag{12}$$

where more generally $\mathbf{e}_i$ denotes the ith unit vector of either $\mathbb{R}^n$ or $\mathbb{R}^{d-n}$, depending on the context. Now, $\mathrm{rank}\mathbf{A}_{NN} < n$ so there exist $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n \setminus \{0\}$ with $\mathbf{A}_{NN}\mathbf{v} = \mathbf{w}^\top \mathbf{A}_{NN} = 0$. But the matrix $(\mathbf{A}_{NN}, \mathbf{A}_{NG})$ has full rank n, hence necessarily $\mathbf{w}^\top \mathbf{A}_{NG} =: \mathbf{u}^\top \neq 0$, and after possible scaling of $\mathbf{w}$ without loss of generality $\|\mathbf{u}\| = 1$.

At first choose $\mathbf{B}_N \in \mathrm{Gl}(n)$ such that $\mathbf{B}_N\mathbf{e}_1 = \mathbf{v}$, then (11) is fulfilled, independent on the choice of $\mathbf{M}$. If we now fix the first row of $\mathbf{M}$ to equal $\mathbf{w}^\top$, then also (9) holds, independent on the choice of the other rows of $\mathbf{M}$. Moreover, then $\mathbf{e}_1^\top \mathbf{M}\mathbf{A}_{NG}\mathbf{B}_G = \mathbf{u}^\top \mathbf{B}_G$, so if we choose an orthogonal $\mathbf{B}_G$ with first column $\mathbf{u}$ (possible because $\|\mathbf{u}\| = 1$), then $\mathbf{B}_G\mathbf{e}_1 = \mathbf{u}$ or $\mathbf{u}^\top \mathbf{B}_G = \mathbf{e}_1^\top$, which implies the equality (10). Finally, we also get $\mathbf{e}_1^\top \mathbf{M}\mathbf{A}_{NG}\mathbf{B}_G\mathbf{e}_1 = \mathbf{u}^\top\mathbf{u} = 1$, the first row of (12); for the other rows, we now simply choose rows 2 to n of $\mathbf{M}$ such that $\mathbf{e}_i^\top \mathbf{M}(\mathbf{A}_{NG}\mathbf{B}_G\mathbf{e}_1) = 0$ for $i > 1$ and $\mathbf{M} \in \mathrm{Gl}(n)$, which is easily possible. ∎

Proof of Lemma 1.7: First note that $\mathbf{A}_{NN} \neq 0$, otherwise $\mathbf{X}_N = \mathbf{A}_{NG}\mathbf{S}_G$ which contradicts Theorem 1.3 (ii) for $\mathbf{X}$. Indeed even $\mathbf{A}_{NN} \in \mathrm{Gl}(n)$ as we will show in the following: Assume not, then $\mathrm{rank}\mathbf{A}_{NN} < n$. According to lemma 2.1, we may therefore choose subspace coordinates $\mathbf{B}_N \in \mathrm{Gl}(n)$, $\mathbf{B}_G \in \vee(d-n)$ and coordinates $\mathbf{M} \in \mathrm{Gl}(n)$ with

$$\mathbf{M}\mathbf{A}_{NN}\mathbf{B}_N = \begin{array}{|c|c|}\hline 0 & 0 \\ \hline 0 & * \\ \hline \end{array} \text{ and } \mathbf{M}\mathbf{A}_{NG}\mathbf{B}_G = \begin{array}{|c|c|}\hline 1 & 0 \\ \hline 0 & *' \\ \hline \end{array}.$$

Moreover, due to the fact that $\mathbf{B}_G$ is orthogonal, the Gaussian random vector $\mathbf{B}_G^\top \mathbf{S}_G$ is still independent. But the non-Gaussian part of $\mathbf{X}$ equals

$$\mathbf{X}_N = \mathbf{A}_{NN}\mathbf{B}_N(\mathbf{B}_N^{-1}\mathbf{S}_N) + \mathbf{A}_{NG}\mathbf{B}_G(\mathbf{B}_G^\top \mathbf{S}_G)$$

so according to the coordinate choice, $(\mathbf{M}\mathbf{X}_N)_1 = (\mathbf{B}_G^\top \mathbf{S}_G)_1$ is Gaussian and independent of $(\mathbf{M}\mathbf{X}_N)(2 \rightharpoonup n)$, which contradicts Theorem 1.3 (ii) for $\mathbf{X}$. Hence, $\mathbf{A}_{NN}$ is invertible. ∎

Proof of Lemma 1.8: Let $\mathbf{B} \in \mathrm{Gl}(d)$ be the unit matrix except for the first $n \times n$ block equaling $\mathbf{B}_N$. Then, according to (5), $\mathbf{S} := \mathbf{B}^{-1}\mathbf{S}$ is also an n-decomposition, so $\Phi_{\tilde{\mathbf{S}}}(\mathbf{s}) = \Phi_{\tilde{\mathbf{S}}_N}(\mathbf{s}_N)\Phi_{\tilde{\mathbf{S}}_G}(\mathbf{s}_G) =: g_N(\mathbf{s}_N)g_G(\mathbf{s}_G)$ for $\mathbf{s} \in \mathbb{R}^d = \mathbb{R}^n \oplus \mathbb{R}^{d-n}$. In the following, we will use tensor notation $g_N \otimes g_G$ to indicate the factorization $(g_N \otimes g_G)(\mathbf{s}) := g_N(\mathbf{s}_N)g_G(\mathbf{s}_G)$. $\mathbf{X} = \tilde{\mathbf{A}}\tilde{\mathbf{S}}$ with $\tilde{\mathbf{A}} := \mathbf{A}\mathbf{B}$ is an n-decomposition, as well, and $f(\mathbf{x}) := \Phi_{\tilde{\mathbf{A}}\tilde{\mathbf{S}}}(\mathbf{x}) = \Phi_{\tilde{\mathbf{S}}}(\tilde{\mathbf{A}}^\top\mathbf{x})$ factorizes as above into components f_N and f_G. For indexes $i \le n < j$ from different subspaces, we therefore get $\frac{\partial}{\partial x_i}f = \left(\frac{\partial}{\partial x_i f_N}\right) \otimes f_G$ and $\frac{\partial}{\partial x_j}f = f_N \otimes \left(\frac{\partial}{\partial x_j f_G}\right)$ and hence

$$\left(f\frac{\partial^2 f}{\partial x_i \partial x_j} - \frac{\partial f}{\partial x_i}\frac{\partial f}{\partial x_j}\right)(\mathbf{x}) = 0$$

for $\mathbf{x} \in \mathbb{R}^d$. Using $f(\mathbf{x}) = (g_N \otimes g_G)(\tilde{\mathbf{A}}^\top\mathbf{x})$, some calculation shows that

$$\left(\left(g_N\frac{\partial^2 g_N}{\partial x_i \partial x_j} - \frac{\partial g_N}{\partial x_i}\frac{\partial g_N}{\partial x_j}\right) \otimes g_G^2 + g_N^2 \otimes \left(g_G\frac{\partial^2 g_G}{\partial x_i \partial x_j} - \frac{\partial g_G}{\partial x_i}\frac{\partial g_G}{\partial x_j}\right)\right)(\tilde{\mathbf{A}}^\top\mathbf{x}) = 0.$$

Note that, in this equation, we take derivatives with respect to x_i and x_j (from $\tilde{\mathbf{A}}^\top\mathbf{x}$), so chain rule applies; however, in deriving this equation, we chose to not apply it yet to save some calculations (see supplement D). We want to evaluate the term in the two inner brackets. First, we observe that chain rule yields $\left(\frac{\partial}{\partial x_i}\right)g_k(\tilde{\mathbf{A}}^\top\mathbf{x}) = \tilde{\mathbf{A}}_{Nk}^{(i)}\nabla g_k \,|_{\tilde{\mathbf{A}}^\top\mathbf{x}}$ with $\tilde{\mathbf{A}}_{Nk}^{(i)} := ((\tilde{\mathbf{A}}_{Nk})_{i,1}, \ldots, (\tilde{\mathbf{A}}_{Nk})_{i,n})$ for $k = N, G$, and similarly $\left(\frac{\partial}{\partial x_j}\right)g_k(\tilde{\mathbf{A}}^\top\mathbf{x}) = \tilde{\mathbf{A}}_{Gk}^{(j-n)}\nabla g_k \,|_{\tilde{\mathbf{A}}^\top\mathbf{x}}$. So, the term in the two inner brackets above can be rewritten into $g_k\frac{(\partial^2 g_k)}{(\partial x_i \partial x_j)} - (\frac{\partial g_k}{\partial x_i})(\frac{\partial g_k}{\partial x_j}))(\tilde{\mathbf{A}}^\top\mathbf{x}) = \tilde{\mathbf{A}}_{Nk}^{(i)}(g_k\nabla^2 g_k - \nabla g_k(\nabla g_k)^\top)\,|_{\tilde{\mathbf{A}}^\top\mathbf{x}}\,\tilde{\mathbf{A}}_{Gk}^{(j-n)\top}$, and altogether we get that

$$\left(\left(\tilde{\mathbf{A}}_{NN}^{(i)}\left(g_N\nabla^2 g_N - \nabla g_N(\nabla g_N)^\top\right)\tilde{\mathbf{A}}_{GN}^{(j-n)\top}\right) \otimes g_G^2 + g_N^2 \otimes \left(\tilde{\mathbf{A}}_{NG}^{(i)}\left(g_G\nabla^2 g_G - \nabla g_G(\nabla g_G)^\top\right)\tilde{\mathbf{A}}_{GG}^{(j-n)\top}\right)\right) \equiv 0$$

vanishes everywhere by using that the above equations hold for all $\mathbf{x}$ and that $\tilde{\mathbf{A}}^\top$ is invertible. Then,

$$\tilde{\mathbf{A}}_{NN}^{(i)}\left(g_N\nabla^2 g_N - \nabla g_N(\nabla g_N)^\top\right)\tilde{\mathbf{A}}_{GN}^{(j-n)\top} \otimes g_G^2 - g_N^2 \otimes \tilde{\mathbf{A}}_{NG}^{(i)}\mathbf{C}\tilde{\mathbf{A}}_{GG}^{(j-n)\top}g_G^2 \equiv 0. \quad (13)$$

Here, we have used that $\tilde{\mathbf{S}}_G = \mathbf{S}_G$ is Gaussian, so $g_G \equiv \Phi_{\mathbf{S}_G}$ fulfills the differential equation of Lemma 1.5, where $\mathbf{C} = \mathrm{Cov}(\mathbf{S}_G)^{-1} = \mathbf{I}$ by the whitening assumption before Lemma 1.8. Equation (13) for $i = 1, \ldots, n$ and $j = n+1, \ldots, d$ can be gathered into a matrix to read $\tilde{\mathbf{A}}_{NN}\left(g_N\nabla^2 g_N - \nabla g_N(\nabla g_N)^\top\right)\tilde{\mathbf{A}}_{GN}^\top \otimes g_G^2 \equiv g_N^2 \otimes \tilde{\mathbf{A}}_{NG}\tilde{\mathbf{A}}_{GG}^\top g_G^2$. Evaluation at $(\mathbf{x}_N, 0)$ then guarantees

$$\tilde{\mathbf{A}}_{NN}\left(g_N\nabla^2 g_N - \nabla g_N(\nabla g_N)^\top\right)\tilde{\mathbf{A}}_{GN}^\top \equiv \tilde{\mathbf{A}}_{NG}\tilde{\mathbf{A}}_{GG}^\top g_N^2 \quad (14)$$

because $g_G(0) = 1$. By construction $\tilde{\mathbf{A}}_{NN} = \mathbf{A}_{NN}\mathbf{B}_N$, $\tilde{\mathbf{A}}_{GN} = \mathbf{A}_{GN}\mathbf{B}_N$, $\tilde{\mathbf{A}}_{NG} = \mathbf{A}_{NG}$ and $\tilde{\mathbf{A}}_{GG} = \mathbf{A}_{GG}$, which finally shows that $\mathbf{A}_{NN}\mathbf{B}_N\left(g_N\nabla^2 g_N - \nabla g_N(\nabla g_N)^\top\right)\mathbf{B}_N^\top\mathbf{A}_{GN}^\top \equiv \mathbf{A}_{NG}\mathbf{A}_{GG}^\top g_N^2$, that is that (6) holds for any $\mathbf{B}_N \in \mathrm{Gl}(n)$. ∎

ACKNOWLEDGMENT

The authors would like to thanks H. Gutch for his careful proof-reading.

REFERENCES

[1] J. B. Kruskal, *Statistical Computation, Chapter Toward a Practical Method Which Helps Uncover the Structure of a Set of Observations by Finding the Line Tranformation Which Optimizes a New Index of Condensation*. New York: Academic, 1969, pp. 427–440.

[2] G. Blanchard, M. Kawanabe, M. Sugiyama, V. Spokoiny, and K. R. Müller, "In search of non-Gaussian components of a high-dimensional distribution," *J. Mach. Learn. Res.*, vol. 7, pp. 247–282, 2006.

[3] M. Kawanabe, "Linear dimension reduction based on the fourth-order cumulant tensor," in *Proc. ICANN 2005*, Warsaw, Poland, 2005, vol. 3697, pp. 151–156.

[4] M. Kawanabe and F. J. Theis, "Estimating non-Gaussian subspaces by characteristic functions," in *Proc. Int. Conf. Independent Component Anal. (ICA)*, Charleston, SC, 2006, vol. 3889, pp. 157–164.

[5] M. Kawanabe, M. Sugiyama, G. Blanchard, and K.-R. Müller, "A new algorithm of non-Gaussian component analysis with radial kernel functions," *Ann. Inst. Statist. Math.*, vol. 59, no. 1, pp. 57–75, 2007.

[6] E. Diederichs, A. Juditsky, V. Spokoiny, and C. Schütte, "Sparse non-Gaussian component analysis," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 3033–3047, 2010.

[7] H. Sayoud and S. Ouamour, "Proposal of a new confidence parameter estimating the number of speakers," *J. Inf. Hiding Multimedia Signal Process.*, vol. 1, no. 2, pp. 101–109, 2010.

[8] M. Kawanabe and F. J. Theis, "Joint low-rank approximation for extracting non-Gaussian subspaces," *Signal Process.*, vol. 87, pp. 1890–1903, 2007.

[9] F. J. Theis and M. Kawanabe, "Uniqueness of non-Gaussian subspace analysis," in *Proc. Conf. Independent Component Anal. (ICA)*, Charleston, SC, 2006, vol. 3889, pp. 917–925.

[10] F. J. Theis, "Colored subspace analysis: Dimension reduction based on a signal's autocorrelation structure," *IEEE Trans. Circuits Syst. I*, vol. 57, no. 7, pp. 1–12, 2010.

[11] H. Akaike, "A new look at statistical model identification," *IEEE Trans. Autom. Control*, vol. 19, no. 6, pp. 716–723, 1974.

[12] J. Rissanen, "A universal prior for integers and estimation by minimum description length," *Ann. Statist.*, vol. 2, pp. 416–431, 1983.

[13] A. S. Dalalyan, A. Juditsky, and V. Spokoiny, "A new algorithm for estimating the effective dimension-reduction subspace," *J. Mach. Learn. Res.*, vol. 9, pp. 1647–1678, 2008.

[14] R. E. Roger and J. F. Arnold, "Reliably estimating the noise in Aviris hyperspectral images," *Int. J. Remote Sens.*, vol. 17, no. 10, pp. 1951–1962, 1996.

[15] J. M. Bioucas-Dias and J. M. P. Nascimento, "Hyperspectral subspace identification," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 8, pp. 2435–2445, 2008.

[16] C. Chang and Q. Du, "Estimation of number of spectrally distinct signal sources in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 3, pp. 608–619, 2004.