Weierstraß-Institut
für Angewandte Analysis und Stochastik

im Forschungsverbund Berlin e.V.

Preprint ISSN 0946 – 8633

Non-Gaussian Component Analysis: New Ideas, New Proofs, New Applications

Vladimir Panov [1]

submitted: April 12, 2010

[1] Weierstrass-Institut
für Angewandte Analysis und Stochastik
Mohrenstrasse 39
10117 Berlin, Germany
E-Mail: panov@wias-berlin.de

No. 1501
Berlin 2010

2000 *Mathematics Subject Classification.* 62G05 62H99 62H30 60G35 00A73 62P20.

Key words and phrases. Dimension reduction, non-Gaussian components, NGCA, EDR subspace, classification problem, Value at Risk.

This research was partially supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

Edited by
Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS)
Mohrenstraße 39
10117 Berlin
Germany

Fax: + 49 30 2044975
E-Mail: preprint@wias-berlin.de
World Wide Web: http://www.wias-berlin.de/

Abstract

In this article, we present new ideas concerning Non-Gaussian Component Analysis (NGCA). We use the structural assumption that a high-dimensional random vector $\vec{X}$ can be represented as a sum of two components - a low-dimensional signal $\vec{S}$ and a noise component $\vec{N}$. We show that this assumption enables us for a special representation for the density function of $\vec{X}$. Similar facts are proven in original papers about NGCA ([1], [5], [13]), but our representation differs from the previous versions. The new form helps us to provide a strong theoretical support for the algorithm; moreover, it gives some ideas about new approaches in multidimensional statistical analysis. In this paper, we establish important results for the NGCA procedure using the new representation, and show benefits of our method.

Contents

Introduction

Each method for solving the dimension reduction problem has its own "point of departure". In the approach based on Independent Component Analysis, high-dimensional data are assumed to be a linear or nonlinear mixture of unknown latent non-Gaussian variables [Hyvärianen et al., 2001]. Another popular method - Principal Component Analysis - implies that the projection representing the data optimally in the least-squared sense is the best projection [Jackson, 1991].

In this paper, we discuss the new ideas concerning Non-Gaussian Component Analysis (NGCA). The method assumes that the "useful part" (which one can imagine as "a signal" or "an information") of the high - dimensional random variable belongs to some low-dimensional space, and the "rest part" has normal distribution. This assumption follows the observation that in real-world applications the "useful part" is non-Gaussian while the "rest part" has a nearly Gaussian distribution [Blanchard et al, 2006].

The aim of Non-Gaussian Component Analysis is to estimate the subspace of the "useful part", also known as an effective dimension reduction subspace (EDR subspace). The original method can be briefly explained as follows: 1) finding vectors that belong to the EDR subspace; 2) constructing a basis from vectors that were obtained on the first step. [Blanchard et al., 2006; Kawanabe et al., 2007; Dalalyan et al., 2007; Diederichs et al., 2010]

We focus on the first task - finding vectors from the EDR subspace. The proof of the original method is based on the special representation of the density function. A serious limitation is that the representation involves the covariance matrix of the noise component, which cannot be consistently estimated from the data. This fact is an obstacle in real-world applications of the method.

In this article, we describe the new representation of the density function. Our representation doesn't involve the noise covariance matrix; it only includes the covariance matrix of the observations, which can be assessed.

This paper is organized as follows. Subsection 2.1 explains the new representation. Then we formulate the main results of NGCA using this form and discuss practical issues. Afterwards we discuss advantages of the new representation. All proofs can be found in the appendices, as well as some examples of practical applications.

1 Set-up and General Scheme

This section gives a formal description of the method.

Assume that a high-dimensional random variable $\vec{X} \in \mathbb{R}^d$ can be represented as a sum of two independent components - a low-dimensional signal and a noise component. More precisely,

$$\vec{X} = \vec{S} + \vec{N}, \tag{1}$$

where

- $\vec{S}$ belongs to some low-dimensional subspace $\mathcal{S}$, $\dim \mathcal{S} = m < d$;
- $\vec{N}$ is a normal vector with zero mean and a covariance matrix Γ;
- $\vec{S}$ is independent of $\vec{N}$.

The original aim of NGCA is to recover the subspace $(\operatorname{Ker} T)^{\perp}$. Theoretical framework is presented in [1], [5], [13]. It can be split into the following steps:

(T1) A proof of the formula for the density function of X:

$$p(\vec{x}) = g(T\vec{x})\phi_{\Gamma}(\vec{x})$$

where

(i) T is a linear transformation from $\mathbb{R}^d$ to some subspace $\mathcal{S}^{\circ}$ that has dimensionality m;

(ii) g is some function from $\mathcal{S}^{\circ}$ to $\mathbb{R}$;

(iii) $\phi_{\Gamma}(\vec{x})$ is the density function for the normal distribution with zero mean and covariance matrix equals Γ.

So, the density function of $\vec{X}$ can be represented as a product of two functions. The first one is a superposition of the linear transformation T from $\mathbb{R}^d$ to some low-dimensional subspace $\mathcal{S}^{\circ}$ and some function from $\mathcal{S}^{\circ}$ to $\mathbb{R}$. The second function is the density function of the normal distribution.

(T2) A proof of the key result: if function $\psi : \mathbb{R}^d \to \mathbb{R}$ is such that

$$\mathbb{E}\left[\vec{X}\psi(\vec{X})\right] = \vec{0}, \tag{2}$$

then

$$\vec{\beta} := \mathbb{E}\left[\nabla\psi(\vec{X})\right] \in (\operatorname{Ker}\ T)^{\perp}.$$

This result gives a method for estimating vectors from $(\operatorname{Ker}\ T)^{\perp}$. It means that if one can construct a function that satisfies (2) then one can estimate a vector from the subspace $(\operatorname{Ker}\ T)^{\perp}$.

(T3) Finding a method for constructing a basis of $(\operatorname{Ker} T)^{\perp}$ from vectors that were obtained on the previous step.

One can develop a method for estimating the EDR subspace, based on (T1)-(T3):

(P1) Estimate vectors from $(\operatorname{Ker}\ \mathbf{T})^{\perp}$; the corresponding fact is proved in (T2).

(P2) Construct a basis of $\mathcal{S}$ using vectors from previous step; see item (T3).

This paper has a slightly different focus: the aim is to recover the subspace $\mathcal{S}$. Relation between the subspace $\mathcal{S}$ and the subspace $(\operatorname{Ker} T)^{\perp}$ uses an unknown matrix Γ (covariance matrix of noise). This means that it is impossible to estimate vectors from $\mathcal{S}$ using only estimates of vectors from $(\operatorname{Ker} T)^{\perp}$; in other words, original approach produces estimates vectors from $\mathcal{S}$ only in the case of white noise.

This article is organized as follows:

Subsection 2.1: a new "convenient" form of the density function of $\vec{X}$ (alternative version of the item (T1)), which doesn't use the matrix Γ; the proof is given in A.1.

Subsection 2.2: the key result for the new representation of the density function (a generalization of the original key result is given); the proof can be found in A.2.

Subsection 2.3: studying a relation between $\mathcal{S}$ and $(\operatorname{Ker} T)^{\perp}$; this relation also doesn't include matrix Γ.

Section 3: discussion about the advantages of the new form.

Appendix A: proofs of the main results.

Appendix B: some ideas about application of NGCA in multidimensional statistical analysis.

2 Main results

2.1 Density representation

The new representation for the density function is given in Theorem 1.

Theorem 1. *Let the structural assumption (1) be fulfilled. Then the density function of the random vector $\vec{X}$ can be represented in the following way:*

$$p(\vec{x}) = \mathbf{g}(\mathbf{T}\vec{x})\phi_{\Sigma}(\vec{x}), \tag{3}$$

where

- $\mathbf{T}: \mathbb{R}^d \to \mathcal{S}', \quad \mathcal{S}' := \Sigma^{-1/2}\mathcal{S}, \quad \dim \mathcal{S}' = m,$

  $$\mathbf{T}\vec{x} = \Pr_{\mathcal{S}'}\{\Sigma^{-\frac{1}{2}}\vec{x}\}, \tag{4}$$

  by Σ we denote the matrix $\mathbb{E}\left[\vec{X}\vec{X}^T\right]$.

- $\mathbf{g}: \mathcal{S}' \to \mathbb{R},$

  $$\mathbf{g}(\vec{t}) = |\Sigma^{-1/2}|\,\frac{q\left(\Pr_{\mathcal{S}'}\{\vec{t}\}\right)}{\phi_m\left(\Pr_{\mathcal{S}'}\{\vec{t}\}\right)}, \tag{5}$$

  where $q(\cdot)$ is the density function of the random variable $\Pr_{\mathcal{S}'}\{\Sigma^{-1/2}\vec{X}\}$.

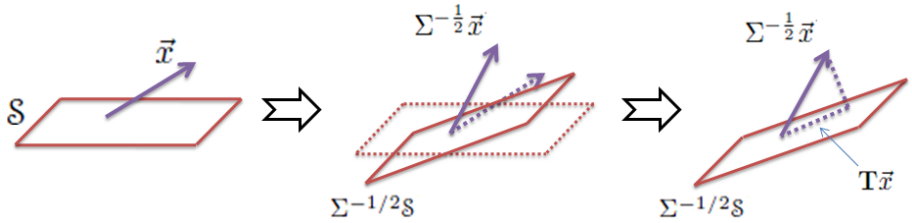

Figure 1: The action of the linear transformation $\mathbf{T}$: 1. $\vec{x}$ is transformed by $\mathcal{S}$; 2. transformed $\vec{x}$ is projected on transformed $\mathcal{S}$.

The proof is given in Appendix A.1.

Remark 1. *One can find similar results in [1], [5], and [14]. Usually such facts are stated in the following form: if assumption (1) is fulfilled, then* $\exists\, T : \mathbb{R}^d \to \widetilde{\mathcal{S}}$, $dim\, \widetilde{\mathcal{S}} = m$, $\exists g : \widetilde{\mathcal{S}} \to \mathbb{R}$ *such that*

$$p(\vec{x}) = g(T\vec{x})\phi_\Gamma(\vec{x}) \tag{6}$$

with some linear transformation T *and some function* g*. Main difference between (3) and (6) is that (6) uses the covariance matrix* Σ *while (3) relies on the covariance matrix of the Gaussian component.*

Advantages of the new form will be discussed later, see subsection 3 for details.

Remark 2. *The linear transformation* $\mathbf{T}$ *acts on* $\vec{x}$ *in the following way:*

1. *firstly,* $\mathcal{S}$ *and* $\vec{x}$ *are transformed by matrix* $\Sigma^{-1/2}$*;*
2. *secondly, transformed* $\vec{x}$ *is projected on transformed* $\mathcal{S}$*.*

Figure 1 illustrates this action.

2.2 Key result of NGCA

The key result of NGCA can be found e.g. in [14]. We state this result in a slightly different form.

Theorem 2. *Assume that the density function of a random vector* $\vec{X} \in \mathbb{R}^d$ *can be represented in the following way:*

$$p(\vec{x}) = g(T\vec{x})\phi_A(\vec{x}), \tag{7}$$

where $T : \mathbb{R}^d \to \mathcal{E}$ *is any linear transformation (*$\mathcal{E}$ *- any subspace),* $g : \mathcal{E} \to \mathbb{R}$ *- any function, and* A *- any* $d \times d$ *symmetric positive matrix.*

Let a function $\psi : \mathbb{R}^d \to \mathbb{R}$ *be such that*

$$\mathbb{E}\left[\vec{X}\psi(\vec{X})\right] = 0.$$

Then

$$\mathbb{E}\left[\nabla\psi(\vec{X})\right] \in (\mathrm{Ker}\ T)^{\perp}.$$

The proof is given in Appendix A.2.

This theorem applies to the representation from the previous subsection (see (3)) and to the former versions of theorem 1 (see (6)). In the next section, we explain why this result is useful for finding vectors from $\mathcal{S}$.

2.3 What is $(\mathrm{Ker}\,\mathbf{T})^{\perp}$?

The previous section gives us the method for finding vectors from the subspace $(\mathrm{Ker}\,\mathbf{T})^{\perp}$. Now we explain, why it is useful for our purposes (that is, finding vectors that belong to $\mathcal{S}$).

The following lemma plays the key role in practical use of NGCA.

Lemma 3. *Let* $\mathbf{T}$ *be the linear transformation defined by (4). Then*

$$(\mathrm{Ker}\,\mathbf{T})^{\perp} = \Sigma^{-1}\mathcal{S}.$$

Proof. By definition, for any x

$$\mathbf{T}\vec{x} = \mathrm{Pr}_{\Sigma^{-1/2}\mathcal{S}}\{\Sigma^{-\frac{1}{2}}\vec{x}\}.$$

It is easy to see that

$$\begin{aligned}\mathrm{Ker}\,\mathbf{T} &= \left\{\vec{x} :\ \Sigma^{-1/2}\vec{x} \perp \Sigma^{-1/2}\mathcal{S}\right\} \\ &= \left\{\vec{x} :\ \exists\vec{s} \in \mathcal{S} \mid \vec{x}^{\top}\left(\Sigma^{-1/2}\right)^{\top}\Sigma^{-1/2}\vec{s} = 0\right\} = \left\{\vec{x} :\ \exists\vec{s} \in \mathcal{S} \mid \vec{x}^{\top}\,\Sigma^{-1}\vec{s} = 0\right\} \\ &= \left\{\vec{x} :\ \vec{x} \perp \Sigma^{-1}\mathcal{S}\right\}.\end{aligned}$$

Here we use the symmetry of the matrix $\Sigma^{-1/2}$. □

From a practical point of view, the last lemma means that one can obtain estimates for vectors from the subspace S using the key result. In fact, one can estimate vectors from the space $\Sigma^{-1}\mathcal{S}$ using the key result. Denote these vectors by $\hat{\beta}_i$. Then one can estimate vectors from the space $\mathcal{S}$ by $\hat{\Sigma}\hat{\beta}_i$, where $\hat{\Sigma}$ is an estimator of the matrix Σ.

3 Discussion

This section discusses the novelty of the proposed approach relative to the original papers about NGCA ([1], [5]).

Theorem 1 stands that in our set-up the density function of $\vec{X}$ can be represented in the following way:

$$p(\vec{x}) = \mathbf{g}(\mathbf{T}\vec{x})\phi_{\Sigma}(\vec{x}),$$

where the linear transformation $\mathbf{T}$ is given by (4) and the function $\mathbf{g}$ is given by (5). In the previous papers about NGCA, another result is proved (it has been already mentioned, see remark 1):

$$p(\vec{x}) = g(T\vec{x})\phi_{\Gamma}(\vec{x}).$$

The new representation has only one principal difference - the covariance matrix of the normal component is equal to Σ while in the original version it is equal to Γ.

The new approach has some advantages.

1. One doesn't need any knowledge of Γ. In fact, the representation from Theorem 1 involves only the unknown subspace $\mathcal{S}$; on the other hand, the former form involves the subspace $\mathcal{S}$ and the matrix Γ.

2. It is clear what the linear transformation $\mathbf{T}$ and the subspace $(\operatorname{Ker}\mathbf{T})^{\perp}$ are, see figure 1, remark 2, and section 2.3 for details.

   In the contrary of the new version, action of the linear transformation T from the former version is much more involved. The explanation of this action is given below. Consider three cases.

   (i) $\Gamma = \mathbf{I}_d$ and $\mathcal{S}$ is a span of first m basis vectors. Then the linear transformation T is simply a projector onto the first m components; denote this projector by Π_m.

   (ii) The noise covariance matrix Γ is still the identity matrix $\mathbf{I}_d$; $\mathcal{S}$ is a span of some m orthogonal vectors. Denote a transaction matrix by U. Then

   $$T = \Pi_m\, U.$$

   (iii) Γ is not necessary equal to the identical matrix; $\mathcal{S}$ is any subspace with dimensionality m. In this case

   $$T = \Pi_m\, U\, \Gamma^{-1/2}. \tag{8}$$

   In the first and in the second cases and even in the third (general) case with a diagonal matrix Γ, one can show that

   $$(\operatorname{Ker} T)^{\perp} = \mathcal{S}.$$

   If Γ is not diagonal, then the formula for the subspace $(\operatorname{Ker} T)^{\perp}$ includes the matrix Γ, which cannot be estimated from the data. This fact is an obstacle in real-world applications of the method.

3. The new presentation for the density function allows us to apply NGCA to the classification problem. According to this method given in B.1, one has to test the assumption

$$\Sigma_1 = \Sigma_2,$$

where Σ_1 and Σ_2 are the covariance matrices for the first and the second groups correspondingly. This assumption is widely used and the testing problem can be solved by different methods ([6], [9]).

Possible applications of NGCA in the previous versions lead to the assumption of equality of the noise covariance matrices. Such hypothesis are more difficult to test.

A Proofs of the main results

A.1 Proof of Theorem 1

Step 1. Denote by $\vec{X}' = \Sigma^{-1/2}\vec{X}$ the standardized vector,

$$\Sigma^{-1/2}\vec{X} = \Sigma^{-1/2}\vec{S} + \Sigma^{-1/2}\vec{N}. \tag{9}$$

Introduce the notation

$$\vec{S}' = \Sigma^{-1/2}\vec{S}, \quad \vec{N}' = \Sigma^{-1/2}\vec{N}. \tag{10}$$

The first component in (9) belongs to the subspace $\mathcal{S}' := \Sigma^{-1/2}\mathcal{S}$. Denote by $\mathcal{N}'$ the subspace that is orthogonal to $\mathcal{S}'$,

$$\mathcal{N}' = \Sigma^{1/2}\mathcal{S}^{\perp}.$$

Studying a relation between the subspaces $\mathcal{S}'$ and $\mathcal{N}'$ can be found in Appendix A.3.

Vector $\mathcal{N}'$ can be decomposed into the sum of two vectors,

$$\vec{N}' = \vec{N}_{\mathcal{S}'} + \vec{N}_{\mathcal{N}'},$$

where $\vec{N}_{\mathcal{S}'} \in \mathcal{S}'$, $\vec{N}_{\mathcal{N}'} \in \mathcal{N}'$.

So, up to now the following decomposition of $\vec{X}'$ is proved:

$$\vec{X}' = \underbrace{\vec{S}' + \vec{N}_{\mathcal{S}'}}_{\in\, \mathcal{S}'} + \underbrace{\vec{N}_{\mathcal{N}'}}_{\in\, \mathcal{N}'}.$$

It is worth mentioning that the density function doesn't depend on a basis. This means that for a calculation of the density function the basis can be changed arbitrarily. Let us choose the basis such that the first m vectors $\vec{v}_1, ..., \vec{v}_m$ compose a basis of $\mathcal{S}'$ and the next $d - m$ vectors $\vec{v}_{m+1}, ..., \vec{v}_d$ compose a basis of $\mathcal{N}'$. In the following we assume that this change is already made.

Step 2. By definition, $\vec{X}'$ is a standardized vector. This step shows that the vectors $\vec{Z}' = \vec{S}' + \vec{N}_{\mathcal{S}'}$ and $\vec{N}_{\mathcal{N}'}$ are also standardized.

$$\begin{aligned}\mathbf{I}_d = \mathrm{Cov}\ \vec{X}' = \mathbb{E}\left[\vec{X}'\vec{X}'^T\right] \\ = \mathbb{E}\left[\vec{Z}'\vec{Z}'^T\right] + \mathbb{E}\left[\vec{N}_{\mathcal{N}'}\vec{N}_{\mathcal{N}'}^T\right] + \mathbb{E}\left[\vec{S}'\vec{N}_{\mathcal{N}'}^T\right] + \mathbb{E}\left[\vec{N}_{\mathcal{S}'}\vec{N}_{\mathcal{N}'}^T\right] + \mathbb{E}\left[\vec{N}_{\mathcal{N}'}\vec{S}'^T\right] + \mathbb{E}\left[\vec{N}_{\mathcal{N}'}\vec{N}_{\mathcal{S}'}^T\right]\end{aligned} \tag{11}$$

Note some facts:

(i) By the change of the basis, the last $d-m$ components of the vectors $\vec{S}'$, $\vec{Z}'$, $\vec{N}_{\mathcal{S}'}$ and the first m components of the vector $\vec{N}_{\mathcal{N}'}$ are equal to zero.

(ii) The vectors $\vec{S}' = \Sigma^{-1/2}\vec{S}$ and $\vec{N}_{\mathcal{N}'} = \Pr_{\mathcal{N}'}\{\Sigma^{-1/2}\vec{N}\}$ are independent as functions of the independent vectors $\vec{S}$ and $\vec{N}$.

(iii) $\mathbb{E}\vec{N}_{\mathcal{N}'} = \mathbb{E}\left[\Pr_{\mathcal{N}'}\{\Sigma^{-1/2}\vec{N}\}\right] = 0$, because of $\mathbb{E}\vec{N} = 0$ and (i).

Now it's easy to see that the third and the fifth summands in (11) are equal to zero. In fact,

$$\mathbb{E}\left[\vec{S}'\vec{N}_{\mathcal{N}'}^T\right] = \mathbb{E}\vec{S}'\ \mathbb{E}\mathcal{N}_{\mathcal{N}'}^T = 0.$$

So, one can rewrite (11) in the following way

$$\mathbf{I}_d = \mathbb{E}\left[\vec{Z}'\vec{Z}'^T\right] + \mathbb{E}\left[\vec{N}_{\mathcal{N}'}\vec{N}_{\mathcal{N}'}^T\right] + \mathbb{E}\left[\vec{N}_{\mathcal{S}'}\vec{N}_{\mathcal{N}'}^T\right] + \mathbb{E}\left[\vec{N}_{\mathcal{N}'}\vec{N}_{\mathcal{S}'}^T\right]. \tag{12}$$

Decompose the vectors $\vec{Z}', \vec{N}_{\mathcal{S}'}$ and $\vec{N}_{\mathcal{N}'}$ into the basis $\vec{v}_1, .., \vec{v}_d$:

$$\vec{Z}' = \sum_{i=1}^{m} z_i\vec{v}_i; \quad \vec{N}_{\mathcal{S}'} = \sum_{i=1}^{m} n_i\vec{v}_i; \quad \vec{N}_{\mathcal{N}'} = \sum_{i=m+1}^{d} n_i\vec{v}_i, \tag{13}$$

where all coefficients z_i and n_i are random values.

Equality (12) can be rewritten as follows:

$$\begin{aligned}\mathbf{I}_d = \sum_{i,i'=1}^{m} \mathbb{E}\left[z_i z_{i'}\right]\vec{v}_i\vec{v}_{i'}^{\top} + \sum_{i,i'=m+1}^{d} \mathbb{E}\left[n_i n_{i'}\right]\vec{v}_i\vec{v}_{i'}^{\top} \\ + \sum_{i=1}^{m}\sum_{i'=m+1}^{d} \mathbb{E}\left[n_i n_{i'}\right]\vec{v}_i\vec{v}_{i'}^{\top} + \sum_{i=m+1}^{d}\sum_{i'=1}^{m} \mathbb{E}\left[n_i n_{i'}\right]\vec{v}_i\vec{v}_{i'}^{\top}\end{aligned}$$

Then the second term in the right hand side is equal to $\mathbf{I}_{d-m}$, i.e.

$$\mathbb{E}\left[\vec{N}_{\mathcal{N}'}\vec{N}_{\mathcal{N}'}^T\right] = \sum_{i,i'=m+1}^{d} \mathbb{E}\left[n_i n_{i'}\right]\vec{v}_i\vec{v}_{i'}^{\top} = \mathbf{I}_{d-m}.$$

Thus, the $(d-m)$ - dimensional vector $\vec{N}_{\mathcal{N}'}$ has the standard normal distribution. Denote the density function by $\phi_{d-m}(x)$.

Step 3. Denote by $F'(\vec{x}')$ and $p'(\vec{x}')$ the distribution function and the density function of the vector $\vec{X}'$.

$$F'(\vec{x}') = \mathbb{P}\left\{\vec{X}' \leqslant \vec{x}'\right\} = \mathbb{P}\left\{\vec{Z}' + \vec{N}_{\mathcal{N}'} \leqslant \vec{x}'\right\} \tag{14}$$

Note some facts:

(i) Vectors $\vec{Z}' = \vec{S}' + \vec{N}_{\mathcal{S}'}$ and $\vec{N}_{\mathcal{N}'}$ are independent. In fact, vectors $\vec{S}' = \Sigma^{-1/2}\vec{S}$ and $\vec{N}' = \Sigma^{-1/2}\vec{N}$ are independent. Then vectors $\vec{S}'$, $\vec{N}_{\mathcal{N}'}$ and $\vec{N}_{\mathcal{S}'}$ are jointly independent (this follows from the choice of the basis). Finally, $\vec{Z}'$ and $\vec{N}_{\mathcal{N}'}$ are independent as functions of independent variables.

(ii) The basis choice (13) enables us to split the inequality

$$\vec{Z}' + \vec{N}_{\mathcal{N}'} \leqslant \vec{x}' = \sum_{i=1}^{d} x_i \vec{v}_i$$

into two:

$$\vec{Z}' \leqslant \sum_{i=1}^{m} x_i \vec{v}_i, \quad \vec{N}_{\mathcal{N}'} \leqslant \sum_{i=m+1}^{d} x_i \vec{v}_i.$$

In the sequel, the following notation is used

$$\vec{x}_{\mathcal{S}'} := \sum_{i=1}^{m} x_i \vec{v}_i, \quad \vec{x}_{\mathcal{N}'} := \sum_{i=m+1}^{d} x_i \vec{v}_i$$

The function F' can be rewritten in the following way:

$$F'(\vec{x}') = \mathbb{P}\left\{\vec{Z}' + \vec{N}_{\mathcal{N}'} \leqslant \vec{x}'\right\} = \mathbb{P}\left\{\vec{Z}' \leqslant \vec{x}_{\mathcal{S}'}, \vec{N}_{\mathcal{N}'} \leqslant \vec{x}_{\mathcal{N}'}\right\}$$
$$= \mathbb{P}\left\{\vec{Z}' \leqslant \vec{x}_{\mathcal{S}'}\right\} \mathbb{P}\left\{\vec{N}_{\mathcal{N}'} \leqslant \vec{x}_{\mathcal{N}'}\right\}.$$

Taking derivatives of the both parts of the last formula gives the representation of the density function of $\vec{X}'$.

$$p(\vec{x}') = q(\vec{x}_{\mathcal{S}'})\phi_{d-m}(\vec{x}_{\mathcal{N}'}) = \frac{q\left(\vec{x}_{\mathcal{S}'}\right)}{\phi_m\left(\vec{x}_{\mathcal{S}'}\right)}\phi_d(\vec{x}') = \frac{q(\mathrm{Pr}_{\mathcal{S}'}\{\vec{x}'\})}{\phi_m(\mathrm{Pr}_{\mathcal{S}'}\{\vec{x}'\})}\phi_d(\vec{x}'),$$

where by $q(\cdot)$ denote the density function of the random vector $\vec{Z}' = \vec{S}' + \vec{N}_{\mathcal{S}'} = \mathrm{Pr}_{\mathcal{S}'}\{\vec{X}\}$.

Step 4. The last step derives representation of the density function of the vector $\vec{X} = \Sigma^{1/2}\vec{X}$ from the density function of $\vec{X}'$. According to the well-known formula for a density transformation,

$$p(\vec{x}) = |\Sigma^{-1/2}|\, p'(\Sigma^{-1/2}\vec{x}) = |\Sigma^{-1/2}|\, \frac{q\left(\mathrm{Pr}_{\mathcal{S}'}\{\Sigma^{-1/2}\vec{x}\}\right)}{\phi_m\left(\mathrm{Pr}_{\mathcal{S}'}\{\Sigma^{-1/2}\vec{x}\}\right)}\phi_d(\Sigma^{-1/2}\vec{x}).$$

The remark $\phi_d(\Sigma^{-1/2}\vec{x}) = \phi_\Sigma(\vec{x})$ concludes the proof.

A.2 Proof of Theorem 2

Integration by parts yields

$$\mathbb{E}\,\nabla\psi(\vec{X}) = \int \nabla\left[\psi(\vec{x})\right] p(\vec{x})dx = -\int \psi(\vec{x})\nabla\left[p(\vec{x})\right] dx. \tag{15}$$

The gradient of the density function can be represented as a sum of two components:

$$\nabla p(\vec{x}) = \nabla\left[\log p(\vec{x})\right] p(\vec{x}) = \nabla\left[\log g(T\vec{x})\right] p(\vec{x}) + \nabla\left[\log \phi_A(\vec{x})\right] p(\vec{x}).$$

The summands in the right hand side can be transformed in the following way:

$$\nabla\left[\log g(T\vec{x})\right] p(\vec{x}) = \frac{\nabla g(T\vec{x})}{g(T\vec{x})} p(\vec{x})$$
$$= \nabla\left[g(T\vec{x})\right] \phi_A(\vec{x}) = T^\top \nabla_{\{T\vec{x}\}}\left[g(T\vec{x})\right] \phi_A(\vec{x})$$
$$\nabla\left[\log \phi_A(\vec{x})\right] p(\vec{x}) = -\Sigma^{-1}\vec{x}p(\vec{x}).$$

Substitution of these expressions into (15) yields:

$$\mathbb{E}\,\nabla\psi(\vec{X}) = -T^\top \int \psi(\vec{x})\nabla_{\{T\vec{x}\}}\left[g(T\vec{x})\right] \phi_A(\vec{x})p(\vec{x})dx + \Sigma^{-1}\int \psi(\vec{x})\vec{x}p(\vec{x})dx$$
$$= T^\top\vec{\Lambda} + \Sigma^{-1}\mathbb{E}\left[\vec{X}\psi(\vec{X})\right] = T^\top\vec{\Lambda}$$
$$\in Im(T^\top) = (\operatorname{Ker} T)^\perp,$$

where $\vec{\Lambda} = -\int \psi(\vec{x})\nabla_{\{T\vec{x}\}}\left[g(T\vec{x})\right] \phi_A(\vec{x})p(\vec{x})dx$. This completes the proof.

A.3 Some facts about subspaces

Denote the subspace $\Gamma\mathcal{S}^\perp$ by $\mathcal{N}$. Motivation of introducing this subspace is given in [14]: it turns out, that oblique projection onto $\mathcal{S}$ along $\mathcal{N}$ is in some sense the optimal mapping onto $\mathcal{S}$.

Lemma 4.

$$\mathcal{N} := \Sigma\mathcal{S}^\perp.$$

Proof.

$$\Sigma\mathcal{S}^\perp = \mathbb{E}\left[\vec{x}\vec{x}^\perp\right]\mathcal{S}^\perp$$
$$= \mathbb{E}\left[\vec{s}\vec{s}^\perp\right]\mathcal{S}^\perp + \mathbb{E}\left[\vec{s}\vec{n}^\perp + \vec{n}\vec{s}^\perp\right]\mathcal{S}^\perp + \mathbb{E}\left[\vec{n}\vec{n}^\perp\right]\mathcal{S}^\perp = \mathbb{E}\left[\vec{n}\vec{n}^\perp\right]\mathcal{S}^\perp,$$

because mathematical expectations in the first and in the second summands are equal to zero. □

This result means that the space $\mathcal{N}$ can be described without any knowledge of Γ.

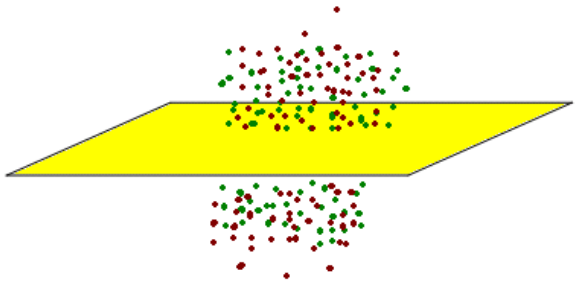

Figure 2: Non-Gaussian component analysis for the classification problem. We assume that the useful signal for both populations belongs to the same low-dimensional subspace $\mathcal{S}$; rest parts have normal distribution, but with different covariance matrices.

Lemma 5. *Let $\mathcal{S}_1$ be any subspace, A - any symmetric matrix. Then the subspace $A^{-1/2}\mathcal{S}_1$ is orthogonal to the subspace $A^{1/2}\mathcal{S}_1^{\perp}$.*

Proof. It suffices to mention that two arbitrary elements of these subspaces are orthogonal. □

Corollary 6. *The subspace $\Sigma^{-1/2}\mathcal{S}$ is perpendicular to the subspace $\Sigma^{-1/2}\mathcal{N}$.*

Proof.

$$\Sigma^{-1/2}\mathcal{N} = \Sigma^{-1/2}\Sigma\,\mathcal{S}^{\perp} = \Sigma^{1/2}\mathcal{S}^{\perp}.$$

According to lemma 5, this subspace is perpendicular to the subspace $\Sigma^{-1/2}\mathcal{S}$. □

The last fact plays the key role in the proof of Theorem 1.

B Examples of possible applications

B.1 Classification problem

In this section we discuss how one can use NGCA algorithm for solving the classification problem.

Lemma 7. *Consider two populations presented by vectors $\vec{X}_1$ and $\vec{X}_2$ from $\mathbb{R}^d$. Assume that*

$$\vec{X}_1 = \vec{S}_1 + \vec{N}_1, \qquad \vec{X}_2 = \vec{S}_2 + \vec{N}_2,$$

where

- *the useful signals $\vec{S}_1$ and $\vec{S}_2$ belong to the same low-dimensional subspace $\mathcal{S}$, $dim(\mathcal{S}) = m$;*

- *the noise components* $\vec{N}_1, \vec{N}_2$ *are normal vectors with covariance matrices* Γ_1 *and* Γ_2*;*
- $\vec{S}_1$ *is independent of* $\vec{N}_1$*,* $\vec{S}_2$ *is independent of* $\vec{N}_2$*;*
- *covariance matrix of* $\vec{X}_1$ *is equal to covariance matrix of* $\vec{X}_2$*; denote this matrix by* Σ*.*

Then the density functions of $\vec{X}_1$ *and* $\vec{X}_2$ *can be represented in the following form:*

$$\begin{aligned} p_1(\vec{x}) &= \mathbf{g}_1(\mathbf{T}\vec{x})\phi_\Sigma(\vec{x}) \\ p_2(\vec{x}) &= \mathbf{g}_2(\mathbf{T}\vec{x})\phi_\Sigma(\vec{x}) \end{aligned}$$

where

- $\mathbf{T}: \mathbb{R}^d \to \mathcal{S}', \quad \mathcal{S}' := \Sigma^{-1/2}\mathcal{S}, \quad \dim \mathcal{S}' = m,$

$$\mathbf{T}\vec{x} = \Pr_{\mathcal{S}'}\{\Sigma^{-\frac{1}{2}}\vec{x}\}$$

- $\mathbf{g}_1, \mathbf{g}_2: \mathcal{S}' \to \mathbb{R}$,

$$\begin{aligned} \mathbf{g}_1(\vec{t}) &= |\Sigma^{-1/2}| \, \frac{q_1\left(\Pr_{\mathcal{S}'}\{\vec{t}\}\right)}{\phi_m\left(\Pr_{\mathcal{S}'}\{\vec{t}\}\right)} \\ \mathbf{g}_2(\vec{t}) &= |\Sigma^{-1/2}| \, \frac{q_2\left(\Pr_{\mathcal{S}'}\{\vec{t}\}\right)}{\phi_m\left(\Pr_{\mathcal{S}'}\{\vec{t}\}\right)} \end{aligned}$$

  where $q_i(\cdot)$ *are the density functions of the random variables* $\Pr_{\mathcal{S}'}\{\Sigma^{-1/2}\vec{X}_i\}$*,* $i = 1, 2$*.*

Remark 3. *The fourth assumption is widely used in classical methods for solving the classification task, for example in Fisher discriminant analysis.*

Proof. This lemma is a straightforward corollary of Theorem 1. One has to separately apply Theorem 1 for each population. □

So, the density functions of the random vectors $\vec{X}_1$ and $\vec{X}_2$ can be represented in the following form:

$$p_1(\vec{x}) = |\Sigma^{-1/2}| \, \frac{q_1\left(\Pr_{\Sigma^{-1/2}\mathcal{S}}\{\Sigma^{-1/2}\vec{x}\}\right)}{\phi_m\left(\Pr_{\Sigma^{-1/2}\mathcal{S}}\{\Sigma^{-1/2}\vec{x}\}\right)}\phi_\Sigma(\vec{x}) \tag{16}$$

$$p_2(\vec{x}) = |\Sigma^{-1/2}| \, \frac{q_2\left(\Pr_{\Sigma^{-1/2}\mathcal{S}}\{\Sigma^{-1/2}\vec{x}\}\right)}{\phi_m\left(\Pr_{\Sigma^{-1/2}\mathcal{S}}\{\Sigma^{-1/2}\vec{x}\}\right)}\phi_\Sigma(\vec{x}) \tag{17}$$

It is worth mentioning that these representations differ only in one place: the function q_1 for the first group and the function q_2 for the second.

Lemma 7 yields that one can estimate the space $\mathcal{S}$ using the following algorithm:

1. estimating vectors from $\mathcal{S}$ using the first population (by the key result)
2. estimating vectors from $\mathcal{S}$ using the second population (by the key result)
3. constructing a basis of $\mathcal{S}$; here one can use all vectors that were obtained on the first and on the second steps

Consider a new object $\vec{x}_\circ$ and classify it into one category. Standard way for solving this problem - comparison of the density functions $p_1(\cdot)$ and $p_2(\cdot)$ at the point $\vec{x}_\circ$. So, one has to compare

$$\mathbf{g}_1\left(\mathbf{T}\vec{x}_\circ\right)\phi_\Sigma(\vec{x}_\circ) \quad \text{vs} \quad \mathbf{g}_2\left(\mathbf{T}\vec{x}_\circ\right)\phi_\Sigma(\vec{x}_\circ).$$

According to (16) and (17), it is equivalent as to compare

$$q_1\left(\Pr_{\Sigma^{-\frac{1}{2}}\mathcal{S}}\{\Sigma^{-\frac{1}{2}}\vec{x}_\circ\}\right) \quad \text{vs} \quad q_2\left(\Pr_{\Sigma^{-\frac{1}{2}}\mathcal{S}}\{\Sigma^{-\frac{1}{2}}\vec{x}_\circ\}\right). \tag{18}$$

If the EDR subspace $\mathcal{S}$ is already estimated, then the data can be projected on the low-dimensional subspace $\Sigma^{-\frac{1}{2}}\mathcal{S}$. Afterwards task (18) is a well-known problem of comparison of the two densities of some low-dimensional variables.

B.2 Portfolio Value at Risk

This subsection discusses possible applications of NGCA in the estimating of Value at Risk.

At time t an investor has some endowment W_t and an additional reserve amount R_t. An endowment can be calculated in the following way:

$$W_t = \vec{b}^{\,T}\vec{p}_t,$$

where $\vec{b}$ is a fixed allocation (a portfolio) and $\vec{p}_t$ - market prices at time t (or logarithms of market prices). The reserve amount is supposed to compensate potential changes in the market price. Investor selects this reserve amount from the following condition:

$$\mathbb{P}\left\{W_{t+h} + R_t < 0\right\} = \alpha, \tag{19}$$

where α is some some fixed constant; h is some fixed amount of time points (usually $h = 10$ days).

Value at Risk is defined as the required capital at time t; actually it is the sum of the endowment and the reserve amount at time t:

$$\mathrm{VaR}_t = W_t + R_t.$$

The economical meaning of Value at Risk can be briefly explained as the maximal loss for h days. This meaning becomes clear if one rewrites (19) as follows:

$$\mathbb{P}\left\{W_{t+h} < W_t - \mathrm{VaR}_t\right\} = \alpha.$$

On the other hand, one can rewrite the last formula as

$$\mathbb{P}\left\{W_{t+h} - W_t < -\operatorname{VaR}_t\right\} = \alpha,$$

and statistical meaning of Value at Risk becomes also clear: $(-\operatorname{VaR}_t)$ is in fact the α - quantile of the distribution of $W_{t+h} - W_t$.

Let us denote the deference $\vec{p}_{t+h} - \vec{p}_t$ by $\vec{X}_t$, the correspondent α - quantile of the random variable $\vec{b}^T \vec{X}_t$ by $q_\alpha\left(\vec{b}^T \vec{X}_t\right)$.

The aim is to estimate q_α, see [7], [8]. In this article, we follow the proposal from [2] based on Independent Component Analysis (ICA). ICA is used to represent the portfolio loss as a result of several independent non - Gaussian factors. Independence allows to estimate and study each factor independently from the others. To be more specific, let us decompose this method into two steps:

- find statistically independent components Y_t such that

$$X_t = AY_t;$$

- simulate independent components N times (we denote this simulations by $\hat{y}_t^{(k)}, k = 1..N$) and estimate VaR_t

$$\operatorname{VaR}_t = -\frac{1}{N}\sum_{k=1}^{N} \hat{q}_\alpha\left(\left\{\vec{b}^T A\hat{y}_t^{(k)}\right\}\right).$$

The main idea of this approach is to reduce a sampling from a high-dimensional to a low-dimensional variate. In the sequel, two methods for finding such variates using NGCA are discussed.

The first method based on NGCA

Assume that the random variable $\vec{X}_t$ can be represented as a sum of two components - a low-dimensional useful component and a Gaussian noise:

$$\vec{X}_t = \vec{S}_t + \vec{N}_t. \tag{20}$$

Then

$$R_t = \vec{b}^T \vec{X}_t = \vec{b}^T \vec{S}_t + \underbrace{\vec{b}^T \vec{N}_t}_{\mathcal{N}\left(0,\vec{b}^T \Gamma\vec{b}\right)}.$$

Note that the second variable has a normal distribution; the first component is the scalar product of the fixed vector and the low-dimensional random vector. This means that one can

- generate N samples $\{s_i^{(k)}\}_{i=1}^d$, $k = 1..N$ from the distribution of $\vec{S}_t$;
- generate N samples $\{n_i^{(k)}\}_{i=1}^d$, $k = 1..N$ from $\mathcal{N}\left(0, \vec{b}^T \Gamma\vec{b}\right)$;

- estimate Value at Risk

$$\mathrm{VaR}_t = -\frac{1}{N}\sum_{k=1}^{N} \hat{q}_\alpha \left(\{ \vec{b}^T s_i^{(k)} + n_i^{(k)} \} \right).$$

The second method based on NGCA

Assume that a random variable $\vec{X}$ satisfies (20). Then the density function of $\vec{X}$ can be represented in the following way:

$$p(\vec{x}) = \mathbf{g}(\mathbf{T}\vec{x})\phi_\Sigma(\vec{x}) = |\Sigma^{-1/2}| \, \frac{q\left(\mathrm{Pr}_{\Sigma^{-1/2}\mathcal{S}}\{\Sigma^{-1/2}\vec{x}\}\right)}{\phi_m\left(\mathrm{Pr}_{\Sigma^{-1/2}\mathcal{S}}\{\Sigma^{-1/2}\vec{x}\}\right)}\phi_\Sigma(\vec{x}).$$

If the subspace $\mathcal{S}$ is already estimated, than all elements in this representation are already known. This means that one can sample from this distribution and estimate Value at Risk via

$$\mathrm{VaR}_t = \frac{1}{N}\sum_{k=1}^{N} \hat{q}_\alpha \left(\left\{ \vec{b}^T x_a^{(k)} \right\} \right).$$

Acknowledgements

The Financial support from the Deutsche Forschungsgemeinschaft via SFB 649 "Economical Risks" is gratefully acknowledged.

Author would like to gratefully thank

- his science advisor professor Vladimir Spokoiny for attention to this work;
- professor Gilles Blanchard for fruitful discussions and a lot of useful advices;
- Elmar Diederichs for some implementations.

References

[1] Blanchard, G., Kawanabe, M., Sugiyama, M., Spokoiny, V., Müller, K.-R., 2006. In search of non-Gaussian components of a high-dimensional distribution. J. Mach. Learn. Res. 6, 247-282.

[2] Chen, Y., Haerdle, W., Spokoiny, V., 2007. Portfolio value at risk based on indenpendent component analysis. J. Comp. Apppl. Math. 205(1), 594-607.

[3] Dalalyan, A.S, Juditsky, A., Spokoiny, V., 2007. A new algorithm for estimating the effective dimension - reduction subspace. J. Mach. Learn. Res. 9, 1647 - 1678.

[4] Diederichs, E. Semi-parametric reduction of dimensionality, 2009. PhD thesis. Free University of Berlin.

[5] Diederichs, E., Juditsky, A., Spokoiny, V., Schütte, C., 2010. Sparse non-Gaussian component analysis. IEEE Trans. Inf. Theory. 15, 5249-5262.

[6] Duda, R.O. , Hart, P.E., Stork D.G., 2001. Patern classification. John Wiley.

[7] Gourieroux, C., Jasiak, J., 2002. Value at Risk. Manuscript, University of Toronto.

[8] Fan, J., Gu, J., 2003. Semiparametric estimation of Value-at-Risk. Econ. J. 6, 261-290.

[9] Hastie, T.J., Tibshirani, R., Friedman, J., 2001. The elements of statistical learning. Springer Series in Statistics. Springer.

[10] Hyvärinen, A.,, Oja, E., 1999. Independent Component Analysis: algorithms and applications. Neural Networks, 13: 411 - 420.

[11] Hyvärinen, A., Karhunen, J., Oja, E., 2001. Independent Component Analysis. Wiley, New York.

[12] Jackson, J.E., 1991. A user's guide to principal componenfs. Wiley, New York.

[13] Kawanabe, M., Sugiyama, M., Blanchard, G., Müller, K.-R., 2007. A new algorithm of non-Gaussian component analysis with radial kernel functions. Ann. Inst. Stat. Math. 59, 57-75.

[14] Sugiyama, M., Kawanabe, M., Blanchard, G., Müller, K.-R., 2008. Approximating the best linear unbiased estimator of non-Gaussian signals with Gaussian noise. IEICE Trans. Inform. Syst., E91-D (5), 1577-1580.