Weierstraß-Institut
für Angewandte Analysis und Stochastik

im Forschungsverbund Berlin e. V.

Preprint ISSN 0946 – 8633

Estimation of the Signal Subspace Without Estimation of the Inverse Covariance Matrix

Vladimir A. Panov[1]

submitted: October 11, 2010

[1] Weierstrass-Institute
for Applied Analysis and Stochastics
Mohrenstr. 39
10117 Berlin, Germany
E-Mail:panov@wias-berlin.de

No. 1546
Berlin 2010

2000 *Mathematics Subject Classification.* 62G05 62H99 60G35.

Key words and phrases. Dimension reduction, non-Gaussian components, NGCA .

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

Let a high-dimensional random vector $\vec{X}$ can be represented as a sum of two components - a signal $\vec{S}$, which belongs to some low-dimensional subspace $\mathscr{S}$, and a noise component $\vec{N}$. This paper presents a new approach for estimating the subspace $\mathscr{S}$ based on the ideas of the Non-Gaussian Component Analysis. Our approach avoids the technical difficulties that usually exist in similar methods - it doesn't require neither the estimation of the inverse covariance matrix of $\vec{X}$ nor the estimation of the covariance matrix of $\vec{N}$.

Contents

1 Introduction and set-up

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 (which one can imagine as än useful partör än information") and a noise component (which has a Normal distribution). More precisely,

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

where $\vec{S}$ belongs to some low-dimensional subspace $\mathscr{S}$, $\vec{N}$ is a normal vector with zero mean and unknown covariance matrix Γ, and $\vec{S}$ is independent of $\vec{N}$. Denote the dimension of $\vec{S}$ by m; up to this paper, m is fixed such that the representation (1) is unique (the existence of such m is proved by Theis and Kawanabe, 2007).

The aim of this paper is to estimate vectors from the subspace $\mathscr{S}$, which we call *the signal subspace*. A very related task, estimation of so called *the non-Gaussian subspace* $\mathscr{I}$ (the definition will be given below) is widely studied in the literature. The original method known as Non-Gaussian Component Analysis (NGCA) was proposed by Blanchard et al, 2006a, and later improved by Kawanabe et al., 2007, Dalalyan et al., 2007, Sugiyama et al., 2008, Diederichs et al., 2010.

In almost all papers mentioned above, the problem of estimation of the vectors from $\mathscr{S}$ is not considered in details; the natural estimators require the estimation of the unknown matrix Γ. The exception is an article Sugiyama et al., 2008, where one estimator is proposed. But practical usage of this method meets another problem - the estimation of the inverse covariance matrix.

Each of these tasks, estimation of the unknown matrix Γ and inverse covariance matrix, is an obstacle in real-world applications of the method. In this article, we propose a new approach, which avoids the mentioned problems.

The main theoretical fact is given in theorem 1. Together with lemma 2, this result yields a method of estimation. For proving the main result, one needs a special representation of the density of $\vec{X}$, which is given in lemma 7, discussed in section 3, and proved in section 4.

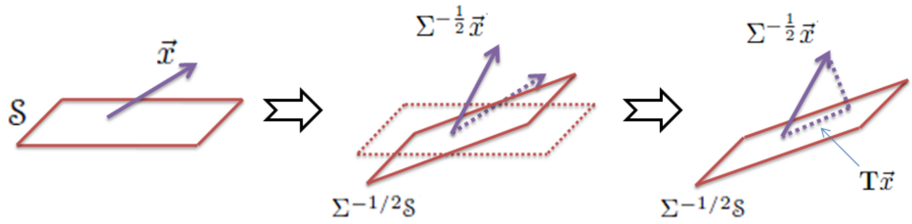

Figure 1: The action of the linear transformation $\mathbf{T}$: 1. $\vec{x}$ is transformed by $\mathscr{S}$; 2. transformed $\vec{x}$ is projected on transformed $\mathscr{S}$.

2 Estimation of the signal subspace

We begin with the main result.

Theorem 1. *Let* $\mathbf{T}: \mathbb{R}^d \to \Sigma^{-1/2}\mathscr{S}$ *be the linear transformation defined as*

$$\mathbf{T}\vec{x} := \mathrm{Pr}_{\Sigma^{-1/2}\mathscr{S}}\{\Sigma^{-1/2}\vec{x}\}, \tag{2}$$

where by Σ *we denote the covariance matrix of* $\vec{X}$*. Then*

$$\mathscr{S} = \Sigma\,(\mathrm{Ker}\,\mathbf{T})^{\perp}. \tag{3}$$

In Blanchard et al., 2006a, a transformation $\mathscr{T}$ is considered instead of $\mathbf{T}$:

$$\mathscr{T}\vec{x} := \mathrm{Pr}_{\Gamma^{-1/2}\mathscr{S}}\{\Gamma^{-1/2}\vec{x}\}.$$

In that paper, the subspace $(\mathrm{Ker}\,\mathscr{T})^{\perp}$ is called *the non-Gaussian subspace* and is in fact the main object of interest. We would like to stress here, that $\mathscr{T} \neq \mathbf{T}$, and equalities like (3) are wrong for $\mathscr{T}$.

The linear transformation $\mathbf{T}$ acts on $\vec{x}$ in the following way: firstly, $\mathscr{S}$ and $\vec{x}$ are transformed by matrix $\Sigma^{-1/2}$; secondly, the transformed $\vec{x}$ is projected on the transformed $\mathscr{S}$. Figure 1 illustrates this action. One of the main results of the NGCA approach gives the practical method for estimating vectors from $(\mathrm{Ker}\,\mathscr{T})^{\perp}$. Similar result can be formulated for the subspace $(\mathrm{Ker}\,\mathbf{T})^{\perp}$ also.

Lemma 2. *Assume that a structural assumption (1) is fulfilled. Then for any function* $\psi \in \mathscr{C}^1(\mathbb{R}^d, \mathbb{R})$ *there exists a vector* $\beta \in (\mathrm{Ker}\ \mathbf{T})^{\perp}$ *such that*

$$\mathbb{E}[\nabla\psi(\vec{X})] - \beta = \Sigma^{-1}\mathbb{E}\left[\vec{X}\psi(\vec{X})\right]. \tag{4}$$

Corollary 3. *Let a structural assumption (1) be fulfilled and let a function* $\psi \in \mathscr{C}^1(\mathbb{R}^d, \mathbb{R})$ *be such that* $\mathbb{E}\left[\vec{X}\psi(\vec{X})\right] = 0$*. Then*

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

Theorem 1 and lemma 2 yield a method for finding vectors from the subspace $\mathscr{S}$.

The first step. On the first step, one estimates vectors from the subspace $(\text{Ker } \mathbf{T})^{\perp}$ using lemma 2. Theoretically, the best way for the estimation is to find a function ψ such that $\mathbb{E}\left[\vec{X}\psi(\vec{X})\right] = 0$, and then to use the corollary. In practice, it is difficult to find such functions; usually it is more realistic to consider some ψ such that $\mathbb{E}\left[\vec{X}\psi(\vec{X})\right]$ is close to zero (but not exactly zero). In this case, according to lemma 2, the vector $\mathbb{E}\left[\nabla\psi(\vec{X})\right]$ is close to some vector from the subspace $(\text{Ker } \mathbf{T})^{\perp}$. For discussing practical issues about finding functions ψ, we refer to Diederichs, PhD dissertation, 2007.

The second step. Denote the vectors obtained on the first step by $\hat{\beta}_i$. Now one can use theorem 1 and estimate vectors from the signal subspace by $\hat{\Sigma}\hat{\beta}_i$, where $\hat{\Sigma}$ is an estimator of the matrix Σ.

Note that the inverse covariance matrix is presented in the formulae (4) but our approach doesn't require the estimation of it. In fact, lemma 2 is used only for theoretical justification of the first step; practical method described above doesn't need neither the estimation of Σ^{-1} nor the estimation of Σ. On the second step, on uses only the representation (3), which also allows to avoid the estimation of the inverse covariance matrix.

3 Density representation

The proofs of the facts formulated in the previous section lie on some special representation of the density function of $\vec{X}$. Certain representations can be also found in previous papers about NGCA. Such facts are stated in the following form: if structural assumption (1) is fulfilled than the density function of a random vector $\vec{X} \in \mathbb{R}^d$ can be represented as

$$p(\vec{x}) = g(T\vec{x})\phi_A(\vec{x}), \tag{5}$$

where $T : \mathbb{R}^d \to \mathscr{E}$ is a linear transformation ($\mathscr{E}$ - some subspace with $\dim \mathscr{E} = m$), $g : \mathscr{E} \to \mathbb{R}$ - a function, and A - a $d \times d$ symmetric positive matrix. Usually the formulae (5) is proven for $A = \Sigma$, see e.g. Kawanabe et al., 2007; rarely for $A = \Gamma$, see Sugiyama, 2008. Another way is to start with the representation (5) without giving the motivation in the spirit of (1), see e.g. Blanchard et al., 2006b.

The main result of this section can be briefly explained as follows: one can find a function g such that (5) is fulfilled with $T = \mathbf{T}$ and $A = \Gamma$. The precise formulation is given below in lemma 4.

The existence of the representation in the form (5) can be easily shown as follows. Note that the model (1) can be equivalently formulated via linear mixing model:

$$\vec{X} = A_S\vec{X}_S + A_N\vec{X}_N, \tag{6}$$

where $\vec{X}_S \in \mathbb{R}^m, \vec{X}_N \in \mathbb{R}^{d-m}$ are two random variables; $\vec{X}_N$ is a normal vector with unknown covariance matrix; $\vec{X}_S$ is independent of $\vec{X}_N$; $A_S \in \text{Matr}(d \times m), A_N \in \text{Matr}(d \times (d-m))$ are two deterministic matrices such that columns of these matrices are independent. In this formulation, the signal subspace is spanned by the columns of matrix A_S.

From (6), one can easily see that the vector X is in fact a linear transformation of the vector $\vec{X}' := (\vec{X}_S; \vec{X}_N)$ (vector $\vec{X}'$ is a concatenation of vectors $\vec{X}_S$ and $\vec{X}_N$). This yields that $p(x) \propto g(\vec{X}_S)\phi(\vec{X}_N)$, where by g we denote the density of the m-dimensional non-Gaussian component, and by ϕ - the density function of the normally distributed random variable $\vec{X}_N$. Thus, the representation (5) is proven with $T = \Gamma$.

The next theorem gives the exact representation for the density of $\vec{X}$ that is needed for our purposes.

Lemma 4. *Let the structural assumption (1) be fulfilled . Then the density function of the random vector $\vec{X}$ can be represented in the following way:*

$$p(\vec{x}) = \mathbf{g}(\mathbf{T}\vec{x})\phi_\Sigma(\vec{x}), \tag{7}$$

where

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

$$\mathbf{T}\vec{x} = \text{Pr}_{\mathscr{S}'}\{\Sigma^{-1/2}\vec{x}\}, \tag{8}$$

  by Σ we denote the covariance matrix of $\vec{X}$.

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

$$\mathbf{g}(\vec{t}) = |\Sigma^{-1/2}| \frac{q(\vec{t})}{\phi_m(\vec{t})}, \tag{9}$$

  where $q(\cdot)$ is the density function of the random variable $\mathbf{T}\vec{X}$, and $\phi_m(\cdot)$ is the density function of the m - dimensional standard normal vector.

The proof of this fact begins the next section.

4 Proofs of the main facts

Proof of the lemma 4

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

$$\Sigma^{-1/2}\vec{X} = \Sigma^{-1/2}\vec{S} + \Sigma^{-1/2}\vec{N}. \tag{10}$$

Introduce the notation

$$\vec{S}' = \Sigma^{-1/2}\vec{S}, \quad \vec{N}' = \Sigma^{-1/2}\vec{N}. \tag{11}$$

The first component in (10) belongs to the subspace $\mathscr{S}' := \Sigma^{-1/2}\mathscr{S}$. Denote by $\mathcal{N}'$ the subspace that is orthogonal to $\mathscr{S}'$. One can proof that $\mathcal{N}' = \Sigma^{1/2}\mathscr{S}^{\perp}$ (see Sugiyama et al., 2008).

Vector $\mathcal{N}'$ can be decomposed into the sum of two vectors, $\vec{N}' = \vec{N}_{\mathscr{S}'} + \vec{N}_{\mathcal{N}'}$, where $\vec{N}_{\mathscr{S}'} \in \mathscr{S}'$, $\vec{N}_{\mathcal{N}'} \in \mathcal{N}'$. So, we consider the following decomposition of $\vec{X}'$:

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

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 it such that the first m vectors $\vec{v}_1, ..., \vec{v}_m$ compose a basis of $\mathscr{S}'$ 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 $\vec{Z}' = \vec{S}' + \vec{N}_{\mathscr{S}'}$ and $\vec{N}_{\mathcal{N}'}$ are also standardized.

$$\begin{aligned}\mathbf{I}_d &= \operatorname{Cov} \vec{X}' = \mathbb{E}\left[\vec{X}'\vec{X}'^T\right] \\ &= \mathbb{E}\left[\vec{Z}'\vec{Z}'^T\right] + \mathbb{E}\left[\vec{N}_{\mathcal{N}'}\vec{N}^T_{\mathcal{N}'}\right] + \mathbb{E}\left[\vec{S}'\vec{N}^T_{\mathcal{N}'}\right] + \mathbb{E}\left[\vec{N}_{\mathscr{S}'}\vec{N}^T_{\mathcal{N}'}\right] + \mathbb{E}\left[\vec{N}_{\mathcal{N}'}\vec{S}'^T\right] + \mathbb{E}\left[\vec{N}_{\mathcal{N}'}\vec{N}^T_{\mathscr{S}'}\right]\end{aligned} \tag{12}$$

Note some facts:

(i) By the change of the basis, the last $d-m$ components of the vectors $\vec{S}'$, $\vec{Z}'$, $\vec{N}_{\mathscr{S}'}$ 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 $\vec{N}_{\mathcal{N}'} = \operatorname{Pr}_{\mathcal{N}'}\{\Sigma^{-1/2}\vec{N}\}$ are independent as functions of the independent vectors $\vec{S}$ and $\vec{N}$.

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

Now it's easy to see that the third and the fifth summands in (12) are equal to zero. In fact,

$$\mathbb{E}\left[\vec{S}'\vec{N}^T_{\mathcal{N}'}\right] = \mathbb{E}\vec{S}'\,\mathbb{E}\mathcal{N}^T_{\mathcal{N}'} = 0.$$

So, one can rewrite (12) in the following way

$$\mathbf{I}_d = \mathbb{E}\left[\vec{Z}'\vec{Z}'^T\right] + \mathbb{E}\left[\vec{N}_{\mathcal{N}'}\vec{N}^T_{\mathcal{N}'}\right] + \mathbb{E}\left[\vec{N}_{\mathscr{S}'}\vec{N}^T_{\mathcal{N}'}\right] + \mathbb{E}\left[\vec{N}_{\mathcal{N}'}\vec{N}^T_{\mathscr{S}'}\right]. \tag{13}$$

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

$$\vec{Z}' = \sum_{i=1}^{m} z_i\vec{v}_i; \quad \vec{N}_{\mathscr{S}'} = \sum_{i=1}^{m} n_i\vec{v}_i; \quad \vec{N}_{\mathcal{N}'} = \sum_{i=m+1}^{d} n_i\vec{v}_i, \tag{14}$$

where all coefficients z_i and n_i are random values.

Equality (13) can be rewritten as follows:

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

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 $\vec{X}'$.

$$F'(\vec{x}') = \mathbb{P}\left\{\vec{X}' \leqslant \vec{x}'\right\} = \mathbb{P}\left\{\vec{Z}' + \vec{N}_{\mathcal{N}'} \leqslant \vec{x}'\right\} \tag{15}$$

Note some facts:

(i) Vectors $\vec{Z}' = \vec{S}' + \vec{N}_{\mathscr{S}'}$ 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 $\vec{N}_{\mathscr{S}'}$ 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 (14) 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:

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

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

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

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

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

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

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,

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

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

Proof of the lemma 2. Here we prove a more general result:

Lemma 5. *Assume that the density function of a random vector $\vec{X} \in \mathbb{R}^d$ can be represented in the form (5), where $T : \mathbb{R}^d \to \mathscr{E}$ is any linear transformation ($\mathscr{E}$ - any subspace), $g : \mathscr{E} \to \mathbb{R}$ - any function, and A - any $d \times d$ symmetric positive matrix.*

Assume that a structural assumption (1) is fulfilled. Then for any function $\psi \in \mathscr{C}^1(\mathbb{R}^d, \mathbb{R})$ there exists a vector $\beta \in (\mathrm{Ker}\ T)^\perp$ such that

$$\mathbb{E}[\nabla\psi(\vec{X})] - \beta = \Sigma^{-1}\mathbb{E}\left[\vec{X}\psi(\vec{X})\right]. \tag{16}$$

Proof. Integration by parts yields

$$\mathbb{E}\,\nabla\psi(\vec{X}) = \int \nabla\left[\psi(\vec{x})\right] p(\vec{x})dx = -\int \psi(\vec{x})\nabla\left[p(\vec{x})\right]dx. \tag{17}$$

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:

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

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

Denote $\beta = T^\top\vec{\Lambda}$. Then

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

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. □

Proof of the theorem 1

The proof is straightforward:

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

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

Acknowledgements

Author would like to gratefully thank his science advisor professor Vladimir Spokoiny for attention to this work, as well as professor Gilles Blanchard for fruitful discussions and some useful advices.

References

Blanchard, G., Kawanabe, M., Sugiyama, M., Spokoiny, V., Müller, K.-R. (2006a), ‘In search of non-Gaussian components of a high-dimensional distribution.’, *J. Mach. Learn. Res.* **6**, 247–282.

Blanchard, G., Kawanabe, M., Sugiyama, M., Spokoiny, V., Müller, K.-R. (2006b), Non-Gaussian component analysis: a semi-parametric framework for linear dimension reduction, *in* ‘Advances in Neural Inf. Proc. Systems (NIPS 05)’, Vol. 18, MIT Press, pp. 131–138.

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.

Diederichs, E. (2007), Semi-parametric reduction of dimensionality., PhD thesis, Free University of Berlin.

Diederichs, E., Juditsky, A., Spokoiny, V., Schütte, C. (2010), ‘Sparse non-Gaussian component analysis.’, *IEEE Trans. Inf. Theory.* **15**, 5249–5262.

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.

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.

Theis, F.J. and Kawanabe, M. (2007), Uniqueness of non-Gaussian subspace analysis, *in* 'Proc. ICA 2007', Vol. 4666 of LNCS, Springer, London, pp. 917–925.