LETTER

Approximating the Best Linear Unbiased Estimator of Non-Gaussian Signals with Gaussian Noise

Masashi SUGIYAMA[†a)], *Member*, **Motoaki KAWANABE**[††b)], **Gilles BLANCHARD**[††c)], *and* **Klaus-Robert MÜLLER**[††,†††d)], *Nonmembers*

SUMMARY Obtaining the best linear unbiased estimator (BLUE) of noisy signals is a traditional but powerful approach to noise reduction. Explicitly computing the BLUE usually requires the prior knowledge of the noise covariance matrix and the subspace to which the true signal belongs. However, such prior knowledge is often unavailable in reality, which prevents us from applying the BLUE to real-world problems. To cope with this problem, we give a practical procedure for approximating the BLUE *without* such prior knowledge. Our additional assumption is that the true signal follows a non-Gaussian distribution while the noise is Gaussian.

key words: *signal denoising, best linear unbiased estimator (BLUE), non-Gaussian component analysis (NGCA), Gaussian noise*

1. Introduction and Formulation

Let $\boldsymbol{x} \in \mathbb{R}^d$ be the observed noisy signal, which is composed of an unknown true signal $\boldsymbol{s}$ and unknown noise $\boldsymbol{n}$.

$$\boldsymbol{x} = \boldsymbol{s} + \boldsymbol{n}. \tag{1}$$

We treat $\boldsymbol{s}$ and $\boldsymbol{n}$ as random variables (thus $\boldsymbol{x}$ also), and assume $\boldsymbol{s}$ and $\boldsymbol{n}$ are statistically independent. We further suppose that the true signal $\boldsymbol{s}$ lies within a subspace $\mathcal{S} \subset \mathbb{R}^d$ of *known* dimension $m = \dim(\mathcal{S})$, where $1 \leq m < d$. On the other hand, the noise $\boldsymbol{n}$ spreads out over the entire space $\mathbb{R}^d$ and is assumed to be of mean zero. Following this generative model, we are given i.i.d. observations $\{\boldsymbol{x}_i\}_{i=1}^n$. Our goal is to obtain denoised signals that are *close* to the true signals $\{\boldsymbol{s}_i\}_{i=1}^n$.

A standard approach to noise reduction in this setting is to project the noisy signal $\boldsymbol{x}$ onto the true signal subspace $\mathcal{S}$, by which the noise is reduced while the signal component $\boldsymbol{s}$ is still preserved. Here the projection does not have to be orthogonal, thus we may want to optimize the projection direction so that the maximum amount of noise can be removed.

In statistics, the linear estimator which fulfills the above requirement is called the *best linear unbiased estimator (BLUE)* [1]. The BLUE has the minimum variance among all linear unbiased estimators. More precisely, the BLUE of $\boldsymbol{s}$ denoted by $\widehat{\boldsymbol{s}}$ is defined by

$$\widehat{\boldsymbol{s}} = \boldsymbol{H}\boldsymbol{x}, \tag{2}$$

where

$$\boldsymbol{H} \equiv \operatorname*{argmin}_{\widetilde{\boldsymbol{H}} \in \mathbb{R}^{d \times d}} \mathbb{E}_{\boldsymbol{n}} \|\widetilde{\boldsymbol{H}}\boldsymbol{x} - \mathbb{E}_{\boldsymbol{n}}\widetilde{\boldsymbol{H}}\boldsymbol{x}\|^2 \quad \text{subject to } \mathbb{E}_{\boldsymbol{n}}[\widetilde{\boldsymbol{H}}\boldsymbol{x}] = \boldsymbol{s}. \tag{3}$$

In the above equation, $\mathbb{E}_{\boldsymbol{n}}$ denotes the expectation over the noise $\boldsymbol{n}$. Let $\boldsymbol{Q}$ be the noise covariance matrix:

$$\boldsymbol{Q} \equiv \mathbb{E}_{\boldsymbol{n}}[\boldsymbol{n}\boldsymbol{n}^\top], \tag{4}$$

which we assume to be non-degenerated. Let $\boldsymbol{P}$ be the orthogonal projection matrix onto the subspace $\mathcal{S}$. Then, the estimation matrix $\boldsymbol{H}$ defined by Eq. (3) is given as (see e.g., [1], [4])

$$\boldsymbol{H} = (\boldsymbol{P}\boldsymbol{Q}^{-1}\boldsymbol{P})^\dagger \boldsymbol{P}\boldsymbol{Q}^{-1}, \tag{5}$$

where † denotes the Moore-Penrose generalized inverse. Let $\mathcal{N} \equiv \boldsymbol{Q}\mathcal{S}^\perp$, where $\mathcal{S}^\perp$ is the orthogonal complement of $\mathcal{S}$. Then it is possible to show that $\boldsymbol{H}$ is an *oblique* projection onto $\mathcal{S}$ along $\mathcal{N}$ (see e.g., [4]):

$$\boldsymbol{H}\boldsymbol{x} = \begin{cases} \boldsymbol{x} & \text{if } \boldsymbol{x} \in \mathcal{S}, \\ \boldsymbol{0} & \text{if } \boldsymbol{x} \in \mathcal{N}. \end{cases} \tag{6}$$

This is illustrated in Fig. 1.

When computing the BLUE by Eqs. (2) and (5), the noise covariance matrix $\boldsymbol{Q}$ and the signal subspace $\mathcal{S}$ should be known. However, $\boldsymbol{Q}$ and $\mathcal{S}$ are often unknown in practice,

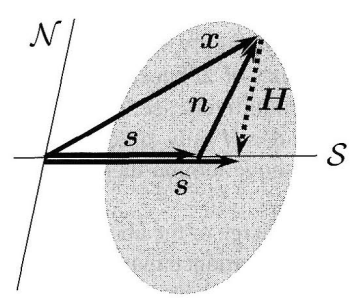


Fig. 1 Illustration of the BLUE.

Manuscript received July 20, 2007.
Manuscript revised November 8, 2007.

[†]The author is with Department of Computer Science, Tokyo Institute of Technology, Tokyo, 152–8552 Japan.
[††]The authors are with Fraunhofer FIRST.IDA, Berlin, Germany.
[†††]The author is with Department of Computer Science, Technical University of Berlin, Berlin, Germany.
a) E-mail: sugi@cs.titech.ac.jp
b) E-mail: motoaki.kawanabe@first.fraunhofer.dc
c) E-mail: gilles.blanchard@first.fhg.de
d) E-mail: klaus@first.fraunhofer.de
DOI: 10.1093/ietisy/e91–d.5.1577

and estimating them from data samples $\{\boldsymbol{x}_i\}_{i=1}^n$ is not generally a straightforward task. For this reason, the applicability of the BLUE to real-world problems has been rather limited so far.

In this paper, we therefore propose a new method which enables us to obtain an approximation of the BLUE even in the *absence* of the prior knowledge of $\boldsymbol{Q}$ and $\mathcal{S}$. Our procedure is essentially based on two facts: the noise covarince matrix $\boldsymbol{Q}$ is actually not required for computing the BLUE (see Sect. 2) and the signal subspace $\mathcal{S}$ can be estimated by a recently proposed method of dimensionality reduction called *non-Gaussian component analysis* [2], [3] (see Sect. 3). We effectively combine these two ideas and propose a practical procedure for approximating the BLUE. Our additional assumption is that the true signal follows a non-Gaussian distribution while the noise is Gaussian.

2. Computing the BLUE without Knowledge of the Noise Covariance Matrix $\boldsymbol{Q}$

In this section, we show that the noise covariance matrix $\boldsymbol{Q}$ is not needed in computing the BLUE. Let $\boldsymbol{\Sigma} \equiv \mathbb{E}_{\boldsymbol{x}}[\boldsymbol{x}\boldsymbol{x}^\top]$. Then the following lemma holds.

Lemma 1: The subspace $\mathcal{N}$ is expressed as $\mathcal{N} = \boldsymbol{\Sigma}\mathcal{S}^\perp$.

Since $\boldsymbol{s}$ and $\boldsymbol{n}$ are independent and $\mathbb{E}_{\boldsymbol{n}}[\boldsymbol{n}] = \boldsymbol{0}$, the above lemma can be immediately confirmed as

$$\begin{aligned}\boldsymbol{\Sigma}\mathcal{S}^\perp &= \mathbb{E}_{\boldsymbol{s}}[\boldsymbol{s}\boldsymbol{s}^\top\mathcal{S}^\perp] + \mathbb{E}_{\boldsymbol{x}}[\boldsymbol{s}\boldsymbol{n}^\top + \boldsymbol{n}\boldsymbol{s}^\top]\mathcal{S}^\perp + \mathbb{E}_{\boldsymbol{n}}[\boldsymbol{n}\boldsymbol{n}^\top]\mathcal{S}^\perp \\ &= \boldsymbol{Q}\mathcal{S}^\perp. \end{aligned} \tag{7}$$

Lemma 1 implies that the noise reduction subspace $\mathcal{N}$ can be characterized by $\boldsymbol{\Sigma}$, *without* using the noise covariance matrix $\boldsymbol{Q}$. Then the following lemma holds.

Lemma 2: The estimation matrix $\boldsymbol{H}$ is expressed as $\boldsymbol{H} = (\boldsymbol{P}\boldsymbol{\Sigma}^{-1}\boldsymbol{P})^\dagger\boldsymbol{P}\boldsymbol{\Sigma}^{-1}$.

Letting $\boldsymbol{P}_\perp$ be the orthogonal projection matrix onto $\mathcal{S}^\perp$, we have, for any $\boldsymbol{x} \in \mathbb{R}^d$,

$$\begin{aligned}(\boldsymbol{P}\boldsymbol{\Sigma}^{-1}\boldsymbol{P})^\dagger\boldsymbol{P}\boldsymbol{\Sigma}^{-1}(\boldsymbol{P}\boldsymbol{x}) &= (\boldsymbol{P}\boldsymbol{\Sigma}^{-1}\boldsymbol{P})^\dagger(\boldsymbol{P}\boldsymbol{\Sigma}^{-1}\boldsymbol{P})\boldsymbol{x} \\ &= \boldsymbol{P}\boldsymbol{x}, \end{aligned} \tag{8}$$

$$\begin{aligned}(\boldsymbol{P}\boldsymbol{\Sigma}^{-1}\boldsymbol{P})^\dagger\boldsymbol{P}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\Sigma}\boldsymbol{P}_\perp\boldsymbol{x}) &= (\boldsymbol{P}\boldsymbol{\Sigma}^{-1}\boldsymbol{P})^\dagger(\boldsymbol{P}\boldsymbol{P}_\perp)\boldsymbol{x} \\ &= \boldsymbol{0}, \end{aligned} \tag{9}$$

from which the above lemma can be confirmed.

Lemma 2 implies that we can obtain the BLUE using $\boldsymbol{\Sigma}$, without using $\boldsymbol{Q}$. Roughly speaking, the "$\mathcal{N}$-part" of $\boldsymbol{Q}$ should agree with that of $\boldsymbol{\Sigma}$ because the signal $\boldsymbol{s}$ lies only in $\mathcal{S}$. Therefore, it intuitively seems that we can replace $\boldsymbol{Q}$ in Eq. (5) by $\boldsymbol{\Sigma}$ because $\boldsymbol{H}$ only affects the component in $\mathcal{N}$ (see Eq. (6)). The above lemma theoretically supports this intuitive claim.

A practical advantage of the above lemma is that, while estimating the noise covariance matrix $\boldsymbol{Q}$ from the data samples $\{\boldsymbol{x}_i\}_{i=1}^n$ is not a straightforward task in general, $\boldsymbol{\Sigma}$ can be directly estimated in a consistent way as

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{i=1}^n \boldsymbol{x}_i\boldsymbol{x}_i^\top. \tag{10}$$

3. Estimating the Signal Subspace $\mathcal{S}$ by Non-Gaussian Component Analysis

The remaining issue to be discussed is how to estimate the true signal subspace $\mathcal{S}$ (or the orthogonal projection matrix $\boldsymbol{P}$). Here we additionally assume that the signal $\boldsymbol{s}$ follows a *non-Gaussian* distribution, while the noise $\boldsymbol{n}$ follows a *Gaussian* distribution.

The papers [2], [3] proposed a framework called *non-Gaussian component analysis* (NGCA). First we briefly review the key idea of NGCA and then show how we employ NGCA for finding the signal subspace $\mathcal{S}$.

3.1 Non-Gaussian Component Analysis

Suppose we have i.i.d. observations $\{\boldsymbol{x}'_i\}_{i=1}^n$, which follow a d-dimensional distribution with the following *semiparametric* probability density function:

$$p(\boldsymbol{x}') = f(\boldsymbol{T}\boldsymbol{x}')\phi_{\boldsymbol{Q}'}(\boldsymbol{x}'), \tag{11}$$

where f is a function from $\mathbb{R}^m$ to $\mathbb{R}$, $\boldsymbol{T}$ is an $m \times d$ matrix, and $\phi_{\boldsymbol{Q}'}$ is the centered Gaussian density with covariance matrix $\boldsymbol{Q}'$. Note that we know *neither* p, f, $\boldsymbol{T}$, *nor* $\boldsymbol{Q}'$, but we only know that the density $p(\boldsymbol{x}')$ is of the form of Eq. (11). Let $\mathcal{S}' = \mathcal{R}(\boldsymbol{T}^\top)$, where $\mathcal{R}(\cdot)$ is the range of a matrix. Then the following proposition holds.

Proposition 1: [2], [3] Suppose $\mathbb{E}_{\boldsymbol{x}'}[\boldsymbol{x}'\boldsymbol{x}'^\top] = \boldsymbol{I}_d$, where $\boldsymbol{I}_d$ is the d-dimensional identity matrix. For an arbitrary smooth function $h(\boldsymbol{x}')$ from $\mathbb{R}^d$ to $\mathbb{R}$,

$$\boldsymbol{\beta} \equiv \mathbb{E}_{\boldsymbol{x}'}[\boldsymbol{g}(\boldsymbol{x}')] \in \mathcal{S}', \tag{12}$$

where $\boldsymbol{g}(\boldsymbol{x}')$ is a function from $\mathbb{R}^d$ to $\mathbb{R}^d$ defined by

$$\boldsymbol{g}(\boldsymbol{x}') \equiv \boldsymbol{x}'h(\boldsymbol{x}') - \nabla h(\boldsymbol{x}'). \tag{13}$$

∇ is the differential operator.

The above proposition shows that we can construct vectors which belong to $\mathcal{S}'$. Generally, different functions $h(\boldsymbol{x}')$ produce different vectors $\boldsymbol{\beta}$. Therefore, we may construct a set of vectors which spans the entire subspace $\mathcal{S}'$. However, $\boldsymbol{\beta}$ can not be computed in practice since the unknown $p(\boldsymbol{x}')$ is included through $\mathbb{E}_{\boldsymbol{x}'}$ in Eq. (12); we employ its consistent estimator given by

$$\widehat{\boldsymbol{\beta}} = \frac{1}{n}\sum_{i=1}^n \boldsymbol{g}(\boldsymbol{x}'_i). \tag{14}$$

By applying *principal component analysis* (PCA) to the set of $\widehat{\boldsymbol{\beta}}$ obtained from a set of different $h(\boldsymbol{x}')$, we can estimate the subspace $\mathcal{S}'$.

Eqs. (14) and (13) imply that the mapping from h to $\widehat{\boldsymbol{\beta}}$ is linear. This means that we can arbitrarily change the norm

of $\widehat{\boldsymbol{\beta}}$ just by multiplying h by an arbitrary scalar. This can totally change the PCA results since vectors with larger norm have stronger impacts on the PCA solutions. Therefore, we need to normalize $\widehat{\boldsymbol{\beta}}$ (or h) in a reasonable way. A desirable normalization scheme would possess a property that resulting $\widehat{\boldsymbol{\beta}}$ has a larger norm if it is close to $\mathcal{S}'$ (i.e., the angle between $\widehat{\boldsymbol{\beta}}$ and $\mathcal{S}'$ is small). This desirable property can be approximately achieved by normalizing $\widehat{\boldsymbol{\beta}}$ by its *standard deviation* [2]. A consistent estimator of the variance of $\widehat{\boldsymbol{\beta}}$ is given by

$$\widehat{N} = \frac{1}{n}\sum_{i=1}^{n} \|\boldsymbol{g}(\boldsymbol{x}'_i)\|^2 - \|\widehat{\boldsymbol{\beta}}\|^2. \tag{15}$$

Further discussions of NGCA including sophisticated implementation, rigorous theoretical error analysis, and extensive experimental studies are given in [2], [3].

3.2 Applying NGCA for Estimating Signal Subspace $\mathcal{S}$

Here we show how NGCA is employed for approximating the BLUE.

Let us apply the *whitening* transformation to the data samples, i.e.,

$$\boldsymbol{x}' = \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{x}. \tag{16}$$

Then the following lemma holds:

Lemma 3: The probability density function of $\boldsymbol{x}'$ is expressed in the form of Eq. (11) with $\mathcal{R}(\boldsymbol{T}^\top) = \boldsymbol{\Sigma}^{-\frac{1}{2}}\mathcal{S}$.

(Proof) Let us denote $\mathcal{S}' \equiv \boldsymbol{\Sigma}^{-\frac{1}{2}}\mathcal{S}$ and $\mathcal{N}' \equiv \mathcal{S}'^{\perp} = \boldsymbol{\Sigma}^{\frac{1}{2}}\mathcal{S}^{\perp} = \boldsymbol{\Sigma}^{-\frac{1}{2}}\mathcal{N}$ (via Lemma 1). Decompose the whitened sample as follows:

$$\boldsymbol{x}' = \boldsymbol{s}' + \boldsymbol{n}'_{\mathcal{S}'} + \boldsymbol{n}'_{\mathcal{N}'}, \tag{17}$$

with $\boldsymbol{s}' = \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{s}$, and $\boldsymbol{n}'_{\mathcal{S}'}, \boldsymbol{n}'_{\mathcal{N}'}$ the orthogonal projections of $\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{n}$ on $\mathcal{S}', \mathcal{N}'$. By characterization of the whitening transform, we have $\mathbb{E}_{\boldsymbol{x}'}[\boldsymbol{x}'\boldsymbol{x}'^\top] = \boldsymbol{I}_d$. Since $\boldsymbol{n}'_{\mathcal{N}'}$ and $(\boldsymbol{s}' + \boldsymbol{n}'_{\mathcal{S}'})$ are the orthogonal projections of $\boldsymbol{x}'$ onto two orthogonal subspaces, this entails

$$\boldsymbol{O} = \mathbb{E}_{\boldsymbol{n}}[\boldsymbol{n}'_{\mathcal{N}'}(\boldsymbol{s}' + \boldsymbol{n}'_{\mathcal{S}'})^\top] = \mathbb{E}_{\boldsymbol{n}}[\boldsymbol{n}'_{\mathcal{N}'}\boldsymbol{n}'^\top_{\mathcal{S}'}], \tag{18}$$

where $\boldsymbol{O}$ is the null matrix, and the second equality holds because $\boldsymbol{s}$ and $\boldsymbol{n}$ are independent. This shows that $\boldsymbol{n}'_{\mathcal{S}'}$ and $\boldsymbol{n}'_{\mathcal{N}'}$ are independent. Let us consider an orthonormal basis $[\boldsymbol{T}^\top_{\mathcal{S}'}, \boldsymbol{T}^\top_{\mathcal{N}'}]$ such that $\mathcal{R}(\boldsymbol{T}^\top_{\mathcal{S}'}) = \mathcal{S}'$ and $\mathcal{R}(\boldsymbol{T}^\top_{\mathcal{N}'}) = \mathcal{N}'$. Let

$$z_{\mathcal{S}'} \equiv \boldsymbol{T}_{\mathcal{S}'}\boldsymbol{x}' = \boldsymbol{T}_{\mathcal{S}'}\boldsymbol{s}' + \boldsymbol{T}_{\mathcal{S}'}\boldsymbol{n}'_{\mathcal{S}}, \tag{19}$$

$$z_{\mathcal{N}'} \equiv \boldsymbol{T}_{\mathcal{N}'}\boldsymbol{x}' = \boldsymbol{T}_{\mathcal{N}'}\boldsymbol{n}'_{\mathcal{N}}. \tag{20}$$

Since $\boldsymbol{s}$ follows a non-Gaussian distribution, $z_{\mathcal{S}'}$ is also non-Gaussian. Let us denote the probability density function of $z_{\mathcal{S}'}$ by $q(z_{\mathcal{S}'})$. On the other hand, $z_{\mathcal{N}'}$ follows a centered Gaussian distribution with identity covariance matrix $\boldsymbol{I}_{d-m}$. Since $z_{\mathcal{S}'}$ and $z_{\mathcal{N}'}$ are independent, the probability density function of $\boldsymbol{x}'$ is given by

$$\begin{aligned} p(\boldsymbol{x}') &= q(z_{\mathcal{S}'})\phi_{\boldsymbol{I}_{d-m}}(z_{\mathcal{N}'}) \\ &= \frac{q(z_{\mathcal{S}'})}{\phi_{\boldsymbol{I}_m}(z_{\mathcal{S}'})}\phi_{\boldsymbol{I}_m}(z_{\mathcal{S}'})\phi_{\boldsymbol{I}_{d-m}}(z_{\mathcal{N}'}) \\ &= \frac{q(z_{\mathcal{S}'})}{\phi_{\boldsymbol{I}_{d-m}}(z_{\mathcal{S}'})}\phi_{\boldsymbol{I}_d}([z_{\mathcal{S}'}, z_{\mathcal{N}'}]^\top) \\ &= \frac{q(\boldsymbol{T}_{\mathcal{S}'}\boldsymbol{x}')}{\phi_{\boldsymbol{I}_{d-m}}(\boldsymbol{T}_{\mathcal{S}'}\boldsymbol{x}')}\phi_{\boldsymbol{I}_d}(\boldsymbol{x}'). \end{aligned} \tag{21}$$

Putting $f(z) = \frac{q(z)}{\phi_{\boldsymbol{I}_{d-m}}(z)}$, $\boldsymbol{T} = \boldsymbol{T}_{\mathcal{S}'}$, and $\boldsymbol{Q}' = \boldsymbol{I}_d$ concludes the proof. □

Lemma 3 shows that, by NGCA, we can estimate $\mathcal{S}'$ from the whitened samples $\boldsymbol{x}'$. Let $\boldsymbol{P}'$ be the orthogonal projection matrix onto the subspace $\mathcal{S}'$. We claim that $\boldsymbol{H} = \boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{P}'\boldsymbol{\Sigma}^{-\frac{1}{2}}$, which is readily checked by the characterization (6), using the fact that $\mathcal{N} = \boldsymbol{\Sigma}^{\frac{1}{2}}\mathcal{S}'^{\perp}$. Therefore, the BLUE $\widehat{\boldsymbol{s}}$ can be expressed in terms of $\boldsymbol{P}'$ and $\boldsymbol{x}'$ as

$$\widehat{\boldsymbol{s}} = \boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{P}'\boldsymbol{x}'. \tag{22}$$

Finally, replacing all the unknown quantities by their consistent estimators, we have an algorithm for approximating the BLUE (see Fig. 2).

Prepare different functions $\{h_k(\boldsymbol{x}')\}_{k=1}^L$.
$\widehat{\boldsymbol{x}}'_i = \widehat{\boldsymbol{\Sigma}}^{-\frac{1}{2}}\boldsymbol{x}_i$ with $\widehat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{i=1}^n \boldsymbol{x}_i\boldsymbol{x}_i^\top$.
$\widehat{\boldsymbol{\beta}}_k = \frac{1}{n}\sum_{i=1}^n \boldsymbol{g}_k(\boldsymbol{x}'_i)$ with $\boldsymbol{g}_k(\boldsymbol{x}') = \boldsymbol{x}'h_k(\boldsymbol{x}') - \nabla h_k(\boldsymbol{x}')$.
$\widehat{N}_k = \sum_{i=1}^n \|\boldsymbol{g}_k(\boldsymbol{x}'_i)\|^2 - \|\widehat{\boldsymbol{\beta}}_k\|^2$.
$\{\widehat{\boldsymbol{\psi}}_i\}_{i=1}^m$: m leading eigenvectors of $\sum_{k=1}^L \widehat{\boldsymbol{\beta}}_k\widehat{\boldsymbol{\beta}}_k^\top/\widehat{N}_k$.
$\widehat{\boldsymbol{P}}' = \sum_{i=1}^m \widehat{\boldsymbol{\psi}}_i\widehat{\boldsymbol{\psi}}_i^\top$.
$\widehat{\boldsymbol{s}}_i = \widehat{\boldsymbol{\Sigma}}^{\frac{1}{2}}\widehat{\boldsymbol{P}}'\widehat{\boldsymbol{x}}'_i$.

Fig. 2 Algorithm for approximating the BLUE.

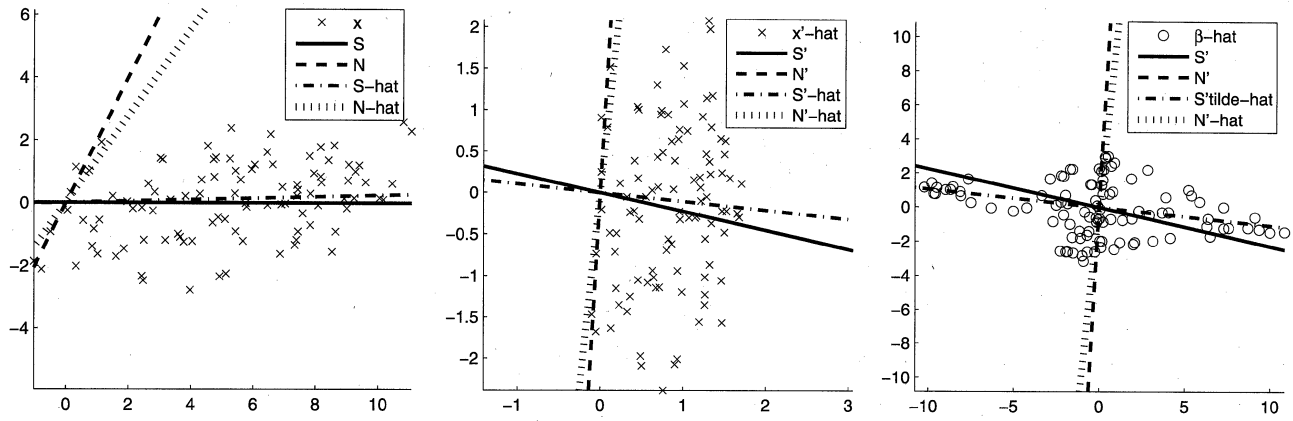
4. Numerical Example

We created two-dimensional samples, where $\mathcal{S} = (1, 0)^\top$, the first element of $\boldsymbol{s}$ follows the uniform distribution on $(0, 10)$, the second element of $\boldsymbol{s}$ is zero, and $\boldsymbol{Q} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$. We used $h_i(\boldsymbol{x}) = \sin(\boldsymbol{w}_i^\top\boldsymbol{x})$ $(i = 1, 2, \ldots, 1000)$ in NGCA, where $\boldsymbol{w}_i$ is taken from the centered normal distribution with identity covariance matrix.

The samples in the original space, the transformed samples in the whitened space, and the $\widehat{\boldsymbol{\beta}}$ functions are plotted in Fig. 3. The mean squared errors of $\{\widehat{s}_i\}_{i=1}^n$ obtained by the true BLUE (using the knowledge of $\boldsymbol{Q}$ and $\mathcal{S}$) and the proposed NGCA-based method as a function of the number of samples are depicted in Fig. 4. This illustrates the usefulness of our NGCA-based approach.

5. Conclusions

For computing the BLUE, prior knowledge of the noise covariance matrix $\boldsymbol{Q}$ and the signal subspace $\mathcal{S}$ is needed. In



(a) Samples in the original space (b) Samples in the whitened space (c) $\widehat{\beta}$ used in NGCA

Fig. 3 Illustration of simulation when $n = 100$. We used 1000 $\widehat{\beta}$ functions, but plotted only 100 points for clear visibility.

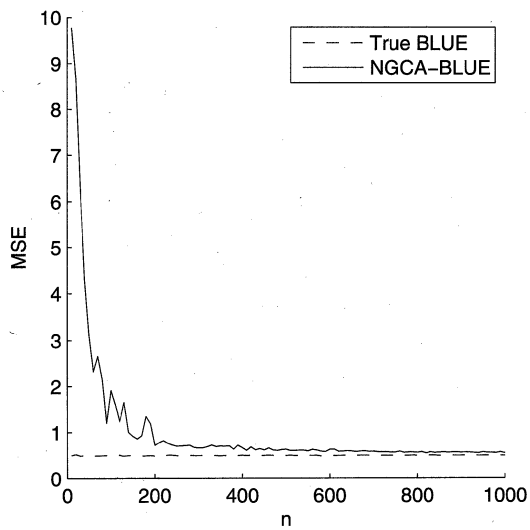


Fig. 4 Mean squared error $\sum_{i=1}^{n}(\widehat{s}_i - s_i)^2/n$ averaged over 100 runs.

this paper, we proposed an algorithm for approximating the BLUE without such prior knowledge. Our assumption is that the signal is non-Gaussian and the noise is Gaussian. We described a naive implementation of NGCA for simplicity. We may employ a more sophisticated implementation of NGCA such as *multi-index projection pursuit* (MIPP) algorithm [2] and *iterative metric adaptation for radial kernel functions* (IMAK) [3], which would further improve the accuracy.

Acknowledgments

We acknowledge partial financial support by DFG, BMBF, the EU NOE PASCAL (EU # 506778), Alexander von Humboldt Foundation, MEXT (18300057 and 17700142), and Tateishi Foundation.

References

[1] A. Albert, Regression and the Moore-Penrose Pseudoinverse, Academic Press, New York and London, 1972.
[2] G. Blanchard, M. Kawanabe, M. Sugiyama, V. Spokoiny, and K.R. Müller, "In search of non-Gaussian components of a high-dimensional distribution," Journal of Machine Learning Research, vol.7, pp.247–282, Feb. 2006.
[3] M. Kawanabe, M. Sugiyama, G. Blanchard, and K.R. Müller, "A new algorithm of non-Gaussian component analysis with radial kernel functions," Annals of the Institute of Statistical Mathematics, vol.59, no.1, pp.57–75, 2007.
[4] H. Ogawa and N.E. Berrached, "EPBOBs (extended pseudo biorthogonal bases) for signal recovery," IEICE Trans. Inf. & Syst., vol.E83-D, no.2, pp.223–232, Feb. 2000.